

THESE

*Pour obtenir le grade de
Docteur*

Spécialité : SCIENCES DE LA VIE

Option : Santé Publique *Présentée et
soutenue publiquement par Matthieu*

Hanf le 6 décembre 2011

A Cayenne

Valorisation de données libres en épidémiologie

**Intérêt des études écologiques dans le domaine des
maladies infectieuses**

Devant le jury suivant :

Bernard Carme, Professeur à l'UAG

Mathieu Nacher, Professeur à l'UAG

Jean-François Guégan, Directeur de Recherche à l'IRD

Jean-François Trape, Directeur de Recherche à l'IRD

Directeur de thèse

Co-directeur de thèse

Rapporteur

Rapporteur

Résumé

Introduction:

En épidémiologie autant les facteurs populationnels et écologiques que les facteurs individuels sont importants pour expliquer la dynamique des maladies. Les études écologiques sont maintenant considérées comme prometteuses en raison de leur capacité à intégrer dans un même modèle l'ensemble de ces facteurs. D'autre part, l'accès aux sources de données nécessaires pour mettre en place des approches multidisciplinaires en épidémiologie est devenu stratégique. Le mouvement récent de l'open data pourrait ainsi jouer un rôle important dans la pérennisation de ces démarches. Les différentes études développées dans cette thèse montrent que l'association des méthodes écologiques aux données publiques permet d'éclairer sous un angle nouveau des problématiques des maladies infectieuses.

Matériels et méthodes / résultats:

Pour étudier l'impact du climat et de ses perturbations sur la santé, il est possible de mettre en œuvre des méthodes écologiques particulières : les analyses des séries temporelles. Du fait de la grande disponibilité des données climatiques, ces méthodes sont très facilement mises en œuvre. En Guyane Française, ces dernières ont contribué à une meilleure compréhension de la dynamique de la transmission du paludisme et de la leishmaniose cutanée ainsi que des caractéristiques écologiques de leurs vecteurs locaux. Elles ont également permis de quantifier globalement l'impact du phénomène El Niño sur leur niveau épidémique. De même, la forte influence des variables climatiques sur les incidences d'une autre pathologie, l'histoplasmosse disséminée, suggère qu'une proportion importante des cas pourrait être due à une exposition récente. Ces résultats donnent de précieuses informations épidémiologiques pour les cliniciens.

Les méthodes écologiques peuvent permettre d'utiliser les données issues de la littérature scientifique pour la réalisation d'études originales. L'utilisation et l'application de ces méthodes écologiques à des données concernant la séroprévalence toxoplasmose dans les populations humaines a permis de définir des indicateurs simples permettant de juger grossièrement du risque de toxoplasmose congénitale dans une population. Dans un deuxième temps elles ont aussi permis d'identifier les principaux facteurs influençant le niveau de séroprévalence globale et donc le risque de toxoplasmose congénital associé. En extrapolant ces résultats, une première carte mondiale du risque de toxoplasmose congénitale a ainsi pu être dessinée.

Les données onusiennes sont uniques pour étudier à l'échelle populationnelle la dynamique et les déterminants importants des grandes maladies infectieuses. 3 exemples illustrant ce concept et utilisant des méthodes écologiques ont été développés. La première analyse montre que dans les pays à faible revenu et ayant un climat équatorial, une prévalence importante d'ascaridiose est associée avec une réduction par 10 de l'incidence du paludisme. Ces résultats renforcent le besoin d'une réflexion autour de la mise en place des grandes campagnes de vermifugeage. Les deux autres études démontrent l'impact significatif de la corruption sur la mortalité juvénile et l'importance des résistances aux anti-tuberculeux et soulignent elles aussi la nécessité d'une réflexion sur les mécanismes de contrôle de la corruption nécessaires à une amélioration durable de la santé des populations.

Conclusions :

Les différentes études développées dans cette thèse montrent que l'association des méthodes écologiques aux données publiques permet d'éclairer sous un angle nouveau des problématiques des maladies infectieuses. Que ce soit à un niveau local ou global, de par le grand nombre d'associations possibles entre les différents jeux de données existants, ce type d'étude fournit la flexibilité nécessaire à l'étude des interactions complexes des nombreux déterminants de la santé. Ces motifs sont à l'origine ces dernières années de son succès grandissant en épidémiologie environnementale et sociale. Cette démarche peut facilement se généraliser aux maladies chroniques non infectieuses dont la multiplicité des déterminants en font, au même titre que les maladies infectieuses, un terrain d'application idéal de ce type de méthodologie.

Mots clefs :

Épidémiologie; Complexité; Etude écologique; Données libres; Maladies infectieuses

Abstract

Introduction:

In epidemiology, populational and ecological factors are as important as individual factors to explain disease dynamics. Ecological studies are now considered promising because of their ability to integrate in the same model all these factors. On the other hand, access to the different sources of data needed to develop multidisciplinary approaches in epidemiology has become strategic. The recent open data movement could thus play an important role in the sustainability of these approaches. The various studies developed in this thesis show that the combination of ecological methods with public data could give original results in the issues of infectious diseases.

Materials and methods / results:

To study the health impacts of climate, it is possible to implement specific ecological methods called time series analysis. Because of the wide availability of climate data, these methods could be easily implemented. In French Guiana, they have contributed to a better understanding of the dynamics of transmission of malaria and cutaneous leishmaniasis and of the ecological characteristics of their local vectors. They also helped to quantify the overall impact of El Niño on their epidemic level. Similarly, the strong influence of climatic variables on the incidence of disseminated histoplasmosis, suggests that a significant proportion of cases could be due to recent exposure. These results provide valuable epidemiological information for clinicians.

Ecological methods have the flexibility to use data from the scientific literature for the realization of original studies. The application of ecological methods on data focused on toxoplasmosis seroprevalence in human populations has identified simple indicators that could be used to roughly evaluate the risk of congenital toxoplasmosis in a population. They have also identified the main factors influencing the overall level of seroprevalence and associated risk of congenital toxoplasmosis. With an extrapolation of these results, a first world map of the risk of congenital toxoplasmosis has been drawn.

The UN data are unique to study the dynamic and determinants of major infectious diseases at a population scale. Three examples of this concept, applying ecological methods, have been developed. The first analysis showed that in low-income countries with an equatorial climate, a high prevalence of Ascariasis was associated with a reduction by 10 of the incidence of malaria. These results reinforce the need for reflection on the establishment of large deworming campaigns. The other two studies demonstrate the significant impact of corruption on child mortality and on multi-resistant tuberculosis and emphasize the need for reflection on the mechanisms of corruption control necessary to significantly improve the health of populations.

Conclusions:

The various studies developed in this thesis show that the combination of ecological methods to public data sheds a new light on the issues of infectious diseases. Both at local and global level, with the large number of possible associations between the various existing datasets, this type of methodology provides the flexibility to study the complex interactions of many health determinants. These reasons are behind his growing success in environmental and social epidemiology these last years. This can easily be generalized to the field of non-infectious chronic diseases, where multiple determinants represent, as in infectious diseases, an ideal field of application of such methodology.

Key words:

Epidemiology; Complexity; Ecological study; Open data; Infectious diseases

Remerciements

*« Trois fleuves
trois fleuves coulent
trois fleuves coulent dans mes veines. »*
Léon Gontran Damas

Je remercie sincèrement les personnes qui ont encadré et évalué ce travail.

J'adresse toute ma reconnaissance à mon directeur de thèse le Pr. Bernard Carme pour m'avoir donné la chance d'effectuer ce travail, de m'avoir guidé et initié à sa passion de la parasitologie, pour m'avoir toujours laissé une précieuse indépendance dans mes prises d'initiatives, pour son souci du détail, pour sa disponibilité constante, sa gentillesse, ses encouragements, ses conseils, sa bonne humeur et son humanité. A son contact, j'ai compris la signification profonde du mot « Professeur ». Je ne saurais oublier sa patience et son aide durant ces trois années passées au CIC-EC Antilles Guyane.

Je tiens aussi à remercier vivement le Pr. Mathieu Nacher, mon co-directeur de thèse, pour avoir entrepris de me transmettre son expérience et son goût de la recherche, pour les opportunités qu'il m'a offertes, pour son enthousiasme à tout point de vue, pour ses conseils toujours pertinents et ses encouragements dans les moments difficiles, pour les discussions passionnantes que nous avons eues, et pour son intérêt permanent dans mon travail. Cette thèse n'aurait pas vu le jour sans son aide et ses conseils. Elle est le résultat d'une fréquentation quotidienne de l'école « nachérienne ».

J'adresse tous mes remerciements à mes rapporteurs, les directeurs de recherche Jean-François Guégan et Jean-François Trape pour avoir pris le temps d'évaluer ce travail et d'avoir apporté à celui-ci leurs points de vue et esprits critiques.

Je remercie également tous ceux que j'ai pu côtoyer, de près ou de loin, et qui ont contribué à la motivation ou la réflexion dans mon travail, où simplement à rendre ces 3 années en Guyane passionnantes et agréables.

Un grand merci au Pr. Pierre Couppié en tant que collaborateur pour de nombreux travaux. Je le remercie pour la disponibilité, le partage de son expérience, les remarques avisées, la gentillesse, et la confiance qu'il a déployée à travers nos différents échanges.

Je remercie aussi l'ensemble des collaborateurs du Centre Hospitalier Andrée Rosemon et de l'équipe EA 3593 avec qui il m'a été donné de collaborer. Un merci particulier à Antoine Adenis, Astrid Van Melle, Amaury Roger, Aurélia Stéfani et Célia Basurko et pour les différentes collaborations que nous avons eues ensemble.

Je remercie également le personnel du CIC-EC Antilles Guyane et de l'équipe EA 3593 de l'Université Antilles Guyane qui m'ont permis de réaliser ma thèse dans une ambiance de franche camaraderie et de bonne humeur. Merci à tous pour votre accueil et votre soutien.

Je tiens tous particulièrement à remercier mes collègues de bureau Antoine Adenis, Astrid Van Melle, Florence Fraisse, et Amaury Roger qui m'ont accompagné et supporté durant ces années. Je ne saurais comment les remercier pour leur soutien et leurs conseils, mais aussi pour leur disponibilité, leur aide précieuse, leurs oreilles attentives et leur amitié. Je les remercie aussi pour leurs commentaires et relectures avisés de ce mémoire de thèse.

Je remercie également Jean Pierre Aboulker, directeur du laboratoire SC10, où cette idée d'utiliser des données libres commença tout doucement à germer. Je le remercie pour ses encouragements et sa gentillesse. Je remercie également l'ensemble des personnes du SC10 et tout particulièrement Thomas Boulet, pour avoir continué à m'associer à ce travail, et Valérie Foubert pour sa bonne humeur.

Je remercie chaleureusement Nak(i)é Paré, amie et galérienne de thèse tout comme moi, pour les discussions statistiques et le soutien psychologique qu'elle a prodigué tout au long de cette expérience.

Enfin, il y a les proches, ceux qui voient ce travail de l'extérieur et ne comprennent pas toujours très bien ce que l'on fait ni pourquoi mais qui cependant vous encouragent et vous soutiennent quoi qu'il arrive.

Je remercie mes parents, mon frère, ma soeur et mon beau-frère pour leur soutien, leur aide sans faille et leurs encouragements. Un merci supplémentaire à ma mère pour son aide précieuse lors de l'impression de ce mémoire.

Un grand merci à mes amis guyanais, particulièrement à Pierre, Stéphane, Maélie, Estelle, Olivier, Céline, Arnaud, Julie et Bérengère pour m'avoir encouragé et pour leur indulgence face à la désocialisation qu'implique parfois la thèse.

Un autre grand merci à mes amis métropolitains pour leur amitié sans failles : Pierre, Tibo, Marco, Yann, Rémy, Nak(i)é, Loïc, et toute la troupe.

Mon dernier remerciement va à Sophie : merci pour tout.

Sommaire

RESUME	5
ABSTRACT	7
REMERCIEMENTS	9
SOMMAIRE	11
LISTE DES FIGURES.....	15
LISTE DES TABLEAUX.....	17
ABREVIATIONS	19
AVANT-PROPOS	21

PARTIE 1 : INTRODUCTION GENERALE

CHAPITRE 1 - LES DONNEES LIBRES	25
1.1 LE ROLE CROISSANT DES DONNEES NUMERIQUES DANS LE PROCESSUS DE RECHERCHE	25
1.2 LE 20 AVRIL 2010 UNE DATE HISTORIQUE	27
1.3 LE LIBRE ACCES AUX PUBLICATIONS SCIENTIFIQUES (EN ANGLAIS : OPEN ACCESS)	28
1.3.1. <i>L'Open Access c'est quoi ?</i>	28
1.3.2. <i>Historique du mouvement Open Access</i>	29
1.3.1. <i>Intérêts de l'Open Access pour les auteurs</i>	30
1.3.2 <i>Intérêt de l'Open Access pour les utilisateurs</i>	34
1.4. LE LIBRE ACCES AUX DONNEES BRUTES DES ETUDES SCIENTIFIQUES (EN ANGLAIS : OPEN DATA)	34
1.4.1. <i>Les coûts cachés des systèmes de données non libres</i>	34
1.4.2. <i>Les retombées scientifiques et socio-économiques d'une plus grande ouverture des données</i>	36
1.4.3. <i>Les retombées pour les investigateurs</i>	37
1.5. CONCLUSIONS.....	38
BIBLIOGRAPHIE	40
CHAPITRE 2 - LES METHODES ECOLOGIQUES	43
2.1. QUE VEUT DIRE « ETUDE ECOLOGIQUE » ?	43
2.2. UTILISATION DES ETUDES ECOLOGIQUES EN SANTE PUBLIQUE.....	45
2.3. LES BIAIS ECOLOGIQUES.....	47
2.4. MAIS ALORS POURQUOI FAIRE DES ETUDES ECOLOGIQUES ?	51
2.4.1. <i>Facilité, rapidité et faible coût d'une étude</i>	51
2.4.2. <i>Limitations de mesures dans les études individuelles</i>	52
2.4.3. <i>Etude de paramètres populationnels n'existant pas à l'échelle individuelle</i>	53
2.4.4. <i>Limitations méthodologiques des études individuelles</i>	54
2.5. LES BIAIS EN RELATION AVEC L'EXISTENCE DE MULTIPLES NIVEAUX D'ANALYSE : POURQUOI NE PARLONS-NOUS QUE DES BIAIS ECOLOGIQUES ?	55
2.6. CONCLUSIONS ET SYNTHÈSE SUR LES ETUDES ECOLOGIQUES.....	56
BIBLIOGRAPHIE	57
CHAPITRE 3 - LES MALADIES INFECTIEUSES	59
3.1. EPIDEMIOLOGIE DES MALADIES INFECTIEUSES	59
3.2. LES PHENOMENES D'EMERGENCE ET DE REEMERGENCE	60
3.3. CANCERS ET MALADIES INFECTIEUSES.....	60
3.4. LES COÛTS ECONOMIQUES ET SOCIAUX DES MALADIES INFECTIEUSES	61
3.5. LES CONDITIONS D'EMERGENCE ET DE CIRCULATION DES AGENTS INFECTIEUX DANS LES POPULATIONS.....	62
3.5.1. <i>L'homme dessine ses propres épidémies</i>	63
3.5.2. <i>L'environnement un élément clé pour l'émergence et la diffusion des agents infectieux</i>	66

3.5.3. <i>L'adaptation microbienne</i>	69
3.6. CONCLUSIONS SUR LES MALADIES INFECTIEUSES.....	70
BIBLIOGRAPHIE	72
CHAPITRE 4 – CAPTURER LA COMPLEXITE : LE NOUVEL ENJEU DE L'EPIDEMIOLOGIE.....	75
4.1. LA COMPLEXIFICATION DE LA RECHERCHE SCIENTIFIQUE	75
4.2. DEFINITION DE LA COMPLEXITE	76
4.3. COMPLEXITE ET EPIDEMIOLOGIE	78
4.4. IMPLICATIONS DE LA THEORIE DE LA COMPLEXITE EN EPIDEMIOLOGIE : MALADIES INFECTIEUSES, DONNEES PUBLIQUES ET METHODES ECOLOGIQUES UN TRIO GAGNANT.....	78
4.4.1. <i>La nécessaire révision du paradigme réductionniste de l'épidémiologie</i>	79
4.4.2. <i>L'appréhension des niveaux d'analyse : le nouveau défi de l'épidémiologie</i>	80
4.4.3. <i>Le rôle vital des études écologiques dans ce contexte de « complexification »</i>	82
4.4.4. <i>Les données publiques dans ce contexte de complexification : un enjeu stratégique</i>	83
4.5. DONNEES PUBLIQUES, MALADIES INFECTIEUSES, ET ETUDES ECOLOGIQUES : UN TRIO GAGNANT POUR APPREHENDER LA COMPLEXITE	84
4.6. LA GUYANE FRANÇAISE : UN TERRITOIRE COMPLEXE	84
BIBLIOGRAPHIE	86
CHAPITRE 5 - OBJECTIFS DE LA THESE	89
PARTIE 2 : TRAVAUX SCIENTIFIQUES	
CHAPITRE 6 - APPORTS DU COUPLAGE ETUDES ECOLOGIQUES/DONNEES CLIMATIQUES DANS LA COMPREHENSION DE LA DYNAMIQUE DES MALADIES INFECTIEUSES : EXEMPLE DE LA GUYANE FRANÇAISE. 93	
6.1. L'OMNIPRESENCE DU CLIMAT EN SANTE HUMAINE	93
6.2. SAISONNALITE ET PERTURBATIONS CLIMATIQUES.....	94
6.2.1 <i>La succession des saisons</i>	94
6.2.2 <i>Les perturbations climatiques : focus sur le phénomène El Niño</i>	95
6.3. LES MECANISMES D' ACTIONS DU CLIMAT SUR LA TRANSMISSION DES MALADIES INFECTIEUSES	96
6.4. ETUDES ECOLOGIQUES ET DONNEES LIBRES : DES OUTILS INDISPENSABLES EN EPIDEMIOLOGIE ENVIRONNEMENTALE	100
6.4.1 <i>Les séries temporelles : un outil de choix pour l'étude du climat</i>	100
6.4.2. <i>Les données climatiques : des données largement ouvertes à la communauté scientifique</i>	101
6.5. LA GUYANE FRANÇAISE, UN TERRAIN DE JEUX IDEAL POUR L'ETUDE DE L'INFLUENCE DU CLIMAT: EXEMPLE DU PALUDISME, DE L'HISTOPLASMOSE DISSEMINEE, ET DE LA LEISHMANIOSE CUTANEE.	103
6.5.1 <i>Le climat guyanais : un climat équatorial humide propices aux pathogènes</i>	103
6.5.2 <i>Problématiques autour du climat et du paludisme, de la leishmaniose cutanée et de l'histoplasmosse disséminée en Guyane française</i>	103
6.6. CONTEXTE ET ORIGINE DE CES ETUDES.....	107
6.7. ETUDE 1 : DETERMINATION DE LA TEMPORALITE DES REVIVISCENCES A PLASMODIUM VIVAX DANS LE VILLAGE DE CAMOPI, GUYANE FRANÇAISE	109
6.8. ETUDE 2 : INFLUENCE DU CLIMAT ET DU NIVEAU DU FLEUVE SUR L'INCIDENCE DU PALUDISME DANS LE VILLAGE DE CACAO, GUYANE FRANÇAISE	115
6.9. ETUDE 3 : LES FACTEURS DE RISQUE ENVIRONNEMENTAUX, ENTOMOLOGIQUES, SOCIO-ECONOMIQUES ET COMPORTEMENTAUX DES CRISES DE PALUDISME CHEZ LES ENFANTS AMERINDIENS DU VILLAGE DE CAMOPI, GUYANE FRANÇAISE.....	123
6.10. ETUDE 4 : LE ROLE DE L'OSCILLATION AUSTRALE EL NIÑO DANS LA VARIATION DU NOMBRE MENSUEL DE CAS DE PALUDISME A PLASMODIUM FALCIPARUM AU CENTRE HOSPITALIER DE CAYENNE, GUYANE FRANÇAISE, DURANT LA PERIODE 1996-2009.	136
6.11. ETUDE 5 : L'HISTOPLASMOSE ASSOCIEE AU VIH EN GUYANE FRANÇAISE : NOUVELLE INFECTION OU REACTIVATION ?	141
6.12. ETUDE 6 : LES VARIATIONS SAISONNIERES DE L'INCIDENCE DE L'HISTOPLASMOSE DISSEMINEE : UN ARGUMENT SUPPLEMENTAIRE EN FAVEUR DE L'HYPOTHESE DES NOUVELLES INFECTIONS ?.....	144
6.13. ETUDE 7 : CLIMAT ET LEISHMANIOSE CUTANEE EN GUYANE : ETUDE A PARTIR D'UNE SERIE DE 1302 PATIENTS VUS ENTRE 1994 ET 2010.	151
6.14. CONCLUSIONS.....	153
BIBLIOGRAPHIE	154
CHAPITRE 7 - UTILISATION DES DONNEES ISSUES DE LA LITTERATURE SCIENTIFIQUE : APPLICATION A L'ETUDE DE LA SEROPREVALENCE A TOXOPLASMA GONDII DANS LES POPULATIONS HUMAINES.....	159
7.1. LA REVOLUTION DES BASES DE DONNEES BIBLIOGRAPHIQUES	159

7.2. UTILISATION ET SYNTHESE DE LA LITTERATURE SCIENTIFIQUE EN SANTE PUBLIQUE	160
7.2.1. <i>Les revues systématiques de la littérature</i>	161
7.2.2. <i>Les méta-analyses</i>	162
7.2.3. <i>Intérêt des études écologiques dans la synthèse de la littérature scientifique</i>	163
7.3. TOXOPLASMA GONDII : UN ENJEU DE SANTE PUBLIQUE CHEZ LA FEMME ENCEINTE	164
7.4. ORIGINE ET CONTEXTE DE CES RECHERCHES	166
7.5. ETUDE 8 : CLASSIFICATION DES POPULATIONS EXPOSEES A LA TOXOPLASMOSE CONGENITALE PAR L'ANALYSE DE LEUR SEROPREVALENCE GLOBALE	168
7.6. ETUDE 9 : UTILISATION DE L'ENVIRONNEMENT GEOGRAPHIQUE ET SOCIETAL POUR PREDIRE LE NIVEAU DE SEROPREVALENCE A TOXOPLASMA GONDII ET LE RISQUE DE TOXOPLASMOSE CONGENITALE ASSOCIE	189
7.7. CONCLUSIONS	212
BIBLIOGRAPHIE	213
CHAPITRE 8 - UTILISATION DES STATISTIQUES ONUSIENNES : APPLICATIONS A L'ETUDE DES DETERMINANTS DE LA DYNAMIQUE DES RESISTANCES AUX ANTITUBERCULEUX, DU PALUDISME ET DE LA MORTALITE JUVENO-INFANTILE.....	215
8.1. LE 20 AVRIL 2010 UNE DATE HISTORIQUE	215
8.2. LES STATISTIQUES ONUSIENNES : DES INFORMATIONS STRATEGIQUES	215
8.3. AUTOUR DE LA QUALITE DES STATISTIQUES EDETEES PAR L'ONU	217
8.4. METHODES ECOLOGIQUES ET DONNEES ONUSIENNES : UN ECLAIRAGE DES EPIDEMIES A L'ECHELLE POPULATIONNELLE	218
8.5. ETUDE DES INTERACTIONS PALUDISME / GEOHELMINTHIASES A L'ECHELLE POPULATIONNELLE	218
8.6. ETUDE DU ROLE DE LA CORRUPTION SUR LA MORTALITE JUVENO-INFANTILE EN 2008.....	221
8.7. ETUDE DU ROLE DE LA CORRUPTION SUR LES MULTI-RESISTANCES AUX ANTI-TUBERCULEUX EN 2008.....	224
8.8. ORIGINE ET CONTEXTE DE CES ETUDES.....	225
8.9. ETUDE 10: UNE ANALYSE PAR ARBRE DE REGRESSION AU NIVEAU NATIONAL RETROUVE UNE ASSOCIATION SIGNIFICATIVE ENTRE DE FORTES PREVALENCES D'ASCARIS LUMBRICOIDES ET DES INCIDENCES PLUS FAIBLES DE PALUDISME	227
8.10. ETUDE 11: LA CORRUPTION TUE. ESTIMATION DE L'IMPACT GLOBAL DE LA CORRUPTION SUR LA MORTALITE JUVENO-INFANTILE	244
8.11. ETUDE 12: L'IMPACT DE LA CORRUPTION SUR LA TUBERCULOSE MULTIRESISTANTE: UNE EVALUATION QUANTITATIVE	252
8.12. CONCLUSIONS.....	263
BIBLIOGRAPHIE	264
CHAPITRE 9 - DISCUSSION GENERALE, PERSPECTIVES ET CONCLUSIONS	267
ANNEXE	
PUBLICATIONS ET COMMUNICATIONS PERSONNELLES	273
<i>Publications</i>	273
<i>Communications affichées</i>	273

Liste des figures

Figure 1 : Dépenses intérieures brutes affectées à la recherche et au développement (en milliard USD à PPA courantes), 1981-2006. Adaptée de [1].	25
Figure 2 : Un des logos du mouvement « Open Access »	28
Figure 3 : Proportions des articles scientifiques disponibles en libre accès par domaine de recherches. Adaptée de [12].	30
Figure 4 : Répartition du facteur d'impact de 8005 revues scientifiques évaluées en 2010 par Thomson Reuters	31
Figure 5 : Répartition des facteurs d'impact des journaux scientifiques libres par rapport au facteur d'impact médian des journaux évalués par l'édition 2008 (2007 en rouge) du JCR. Adaptée de [16].	32
Figure 6 : Avantage en termes de citations des articles en « Open Access » par rapport aux articles non disponibles en libre accès (tous domaines scientifiques confondus). Adaptée de [8].	33
Figure 7 : Distributions du nombre de citations en 2004-2005 de 85 essais cliniques en cancérologie en fonction du type d'accès aux données. Adaptée de [29].	38
Figure 8 : Importance des différents types d'études en santé humaine selon leur imputation causale.	43
Figure 9 : illustration des processus d'agrégation géographique, temporelle et semi-écologique.	44
Figure 10 : Taux de mortalité comparés durant trois décennies successives dans une ville avec un système d'assainissement (Beccles) et une ville sans système d'assainissement (Bungay). Adaptée de [6].	45
Figure 11 : Taux de mortalité associés au choléra par compagnies d'approvisionnement en eau potable de Londres. Adaptée de « <i>On the mode of communication of Cholera</i> » de John Snow [7].	46
Figure 12 : Illustration des sources de biais que l'on peut rencontrer dans les études écologiques	49
Figure 13 : Taux de suicide / 10^5 personnes / année en fonction de la proportion protestante de la population pour quatre groupes de provinces prussiennes, 1883-1890. Adaptée de l'ouvrage de Durkheim [15].	50
Figure 14 : Estimations des séroprévalences dans les capitales de 37 pays d'Afrique en fonction du pourcentage d'hommes circoncis. Adaptée de [17].	52
Figure 15 : (A) Série temporelle des cas de choléra observés au ICDDR, B (International Centre for Diarrhoeal Disease Research, Bangladesh, in Dhaka) compilée à partir des patients visitant le centre. (B) Série temporelle du phénomène El Niño mesurée à partir de l'index El niño 3.4. Adaptée de Pascual et al. [20].	54
Figure 16 : Graphique de la mortalité associée au cancer du sein (pour 100000 personnes) en fonction de la consommation de poissons (pourcentage de l'apport en calories) 1974-1975. Adaptée de Kaiser et al [21].	55
Figure 17 : Répartition des causes de mortalité dans le monde en 2004 [2]	60
Figure 18 : Exemples mondiaux d'émergences et de réémergences de maladies infectieuses. Les nouvelles maladies émergentes sont représentées en rouge, les maladies ré-émergentes sont représentées en bleu et les maladies intentionnellement ré-émergentes sont en noir (bioterrorisme). Adaptée à partir de [5].	61
Figure 19 : Index DALY standardisé sur l'âge en 2004. Adaptée de [7].	62
Figure 20 : Projection de l'évolution du pourcentage de la population mondiale vivant dans un environnement urbain, 1950-2050. Adaptée des données des Nations Unies [12].	64
Figure 21 : Distribution mondiale du moustique <i>Aedes albopictus</i> . Adaptée de [17].	66
Figure 22 : Distribution de la proportion de tuberculose multi résistante parmi les nouveaux cas de tuberculose, 1994-2009, adaptée de [37].	69
Figure 23 : Exemple d'œuvre obtenue par l'«art génératif», basé sur la théorie de la complexité et utilisant des algorithmes pour concevoir des œuvres se générant d'elles-mêmes.	76
Figure 24 : Schématisation des propriétés d'un système adaptatif complexe	77
Figure 25 : Schématisation de la relation entre les différentes ères de l'épidémiologie et le niveau d'analyse.	80
Figure 26 : Schématisation du principe de fonctionnement de l'immunité de communauté. Elle implique que la probabilité d'une personne de contracter une maladie infectieuse dépend en partie du niveau d'immunité de la population à laquelle elle appartient.	81
Figure 27 : Saisonnalité de la mortalité associée au choléra dans trois provinces de l'Inde (Madras, Punjab, Bengal). Adaptée de [9].	94
Figure 28 : Anomalies de températures à la surface des océans (en °C) lors de l'épisode El Niño de 1997	95
Figure 29 : (A) Fluctuation du nombre de cas de rougeoles à Glasgow sur la période 1901-1917 (B) Intensité de la transmission de la rougeole. Adaptée de [29].	97
Figure 30 : Modèle hiérarchique montrant l'influence du climat sur la transmission du choléra. Modifiée à partir de [36].	98

Figure 31 : Evolution temporelle de l'incidence de la cryptosporidiose à Brisbane (Australie) de 1996 à 2004. Adaptée de [46].	100
Figure 32 : Risque de paludisme en Guyane Française en 2010 (les zones en rouge correspondent aux zones à fort risque de paludisme, les zones en jaune foncé aux zones à faible risque). Adaptée de [58].	104
Figure 33 : Incidence annuelle par commune, calculée à partir des cas de leishmaniose recensés en dermatologie de 1994 à 2010 et pour lesquels le lieu de contamination était renseigné (1359 cas). Adaptée de [65].	105
Figure 34 : Carte mondiale de la répartition des cas décrits d'histoplasmosse américaine. Adaptée de [80].	106
Figure 35 : Evolution temporelle du nombre d'articles publiés et indexés dans PubMed [1] contenant les termes MESH : « VIH », « malaria », « toxoplasmosis », et « tuberculosis ».	160
Figure 36 : Exemple de graphique typique d'une méta-analyse. Méta-analyse de l'effet des probiotiques sur les diarrhées associées à la prise d'antibiotiques. Adaptée de [11].	162
Figure 37 : Les routes naturelles de transmission de <i>Toxoplasma gondii</i> . Adaptée de [15].	165
Figure 38 : Variabilité de la prévalence à <i>Toxoplasma gondii</i> en fonction de la zone géographique. Adaptée de [16].	166
Figure 39 : Exemple de 4 profils différents d'acquisition de l'immunité antitoxoplasmique reflétant la survenue des primo-infections au cours des ages dans une population. Adaptée de [18, 26-28].	167
Figure 40 : Page d'accueil du site Internet de la banque mondiale donnant accès à l'ensemble de ses statistiques sur 209 pays du monde et sur une période de temps remontant jusqu'à 50 ans (http://data.worldbank.org/).	216
Figure 41 : Carte en 2008 des pays exempts de paludisme et des pays endémiques en phases de contrôle, de pré-élimination, d'élimination et de prévention d'une réintroduction. Adaptée de [7].	219
Figure 42 : Répartition mondiale et statut de contrôle des géohelminthiases en 2004. Adaptée de [10].	220
Figure 43 : Tendances du taux de mortalité des moins de 5 ans (pour 1 000 naissances vivantes), par région (1990-2006). Adaptée de [1].	223
Figure 44 : Distribution de la proportion des cas de MDR-TB parmi les nouveaux cas de tuberculose, 1994-2009. Adaptée de [42].	224

Liste des tableaux

Tableau 1 : Nombre de cas déclarés, risques relatifs ajustés sur l'âge et intervalle de confiance à 95% pour les lymphomes non hodgkiniens (NHL) dans l'état du New Jersey, 1979-1987 en fonction de l'exposition aux composés organiques volatils (« TCE exposure ») et le sexe. Adapté de Cohn et collaborateurs [18].....	53
Tableau 2 : Exemple d'infection émergentes et de leurs facteurs probables d'émergence. Adaptée de [10]	63
Tableau 3 : Exemple d'agents ou de maladies infectieuses ayant un lien suspecté ou avéré avec un changement d'utilisation des terres. Adaptée de [23].	67
Tableau 4 : Exemples de facteurs climatiques influençant la transmission et la distributions de maladies vectorielles. Adaptée de [41].	99
Tableau 5 : Types de corruption dans le secteur de la santé. Adapté de [26].....	222

Abréviations

- **ADN:** Acide DésoxyriboNucléique
- **OCDE:** Organisation de Coopération et de Développement Economiques
- **PLOS:** Public Library Of Science
- **JCR:** Journal Citation Report
- **ONU:** Organisation des Nations Unies
- **COV:** Composé Organique Volatil
- **SRAS:** Syndrome Respiratoire Aigu Sévère
- **DALY:** Disability-Adjusted Life Year
- **VIH:** Virus de l'Immunodéficience Humaine
- **OMS:** Organisation Mondiale de la Santé
- **ENSO:** El Niño-Southern Oscillation
- **ARIMA:** Autoregressive Integrated Moving Average
- **SST:** Sea Surface Temperature
- **SOI:** Southern Oscillation Index
- **MEI:** Multivariate ENSO Index
- **NOAA:** National Oceanic and Atmospheric Administration
- **ZIC:** Zone Intertropicale de Convergence
- **HCC:** Histoplasma capsulatum var. capsulatum
- **CIC-EC ANTILLES GUYANE:** Centre d'Investigation Clinique Epidémiologie Clinique Antilles Guyane
- **NCBI:** National Center for Biotechnology Information
- **SDN:** Société des Nations
- **TI:** Transparency International
- **MDR-TB:** Multi-Drug-Resistant Tuberculosis

Avant-propos

« La perplexité est le début de la connaissance »

Khalil Gibran

La première question que l'on pourrait se poser à la lecture de ce mémoire est « pourquoi une thèse sur ce sujet ? ». La réponse découle directement de l'ensemble de mon travail au sein du CIC-EC Antilles Guyane et des échanges que j'ai eus avec les personnes s'investissant dans la recherche épidémiologique.

A mon arrivée en Guyane française en août 2008 au sein du CIC-EC Antilles Guyane, je me suis rapidement rendu compte que l'idée que je me faisais sur les attentes de la communauté médicale vis-à-vis du poste de biostatisticien que j'allais occuper était complètement fausse. Dès le début, j'ai été confronté à une multitude de questions plus déconcertantes les unes que les autres, d'un degré souvent plus pratique que méthodologique, et qui, de par le contexte de la Guyane, tournaient principalement autour de problématiques infectieuses. Ceux qui me posaient les questions, avant tout des médecins de formation et même des praticiens soignants en activité apportaient des éléments de réponse en rapport avec leur pratique médicale hospitalière ou de terrain. Mais le formatage en termes de recueil d'information, de plan d'analyse et d'exploitation statistique était à trouver.

Ces questions étaient par exemple les suivantes :

- « Une grande étude de cohorte portant sur une problématique particulière vient d'être lancée. Les résultats sont attendus d'ici 2 à 3 ans. Que peut-on faire avec les données existantes pour continuer à étudier cette problématique durant ce laps de temps? »
- « Il n'existe pas de données objectives sur cette pathologie en Guyane. Devons-nous lancer une grande étude? »
- « Nous possédons une base de données déjà existante, mais il manque certaines variables pour étudier une problématique particulière. Comment faire ? »
- « L'environnement influence certainement cette pathologie en Guyane. Comment étudier cette relation ? »
- « Cette base de donnée doit être valorisée. Que peut-on encore faire dessus ? »
- « La répartition dans les populations de cette pathologie obéit sûrement à ce paramètre. Comment confirmer cette relation ? »
- « On se demande à l'échelle de la population quels déterminants sont les plus importants dans la répartition géographique d'une pathologie. Comment les déterminer ? »
- On aimerait prendre en compte le contexte social pour expliquer la dynamique de cette maladie. Comment étudier ce lien ?

Première réaction face à cette multitude de questions : une profonde perplexité... Etant issu de la recherche clinique et des essais thérapeutiques où l'on n'utilise principalement que « des études prospectives randomisées multicentriques en double aveugle ayant une taille d'échantillon importante », je n'avais alors pas le bagage théorique (et surtout méthodologique) pour apporter des éléments de réponses à toutes ces questions. Par-dessus cela, en creusant plus profondément le sujet, d'autres questions d'un niveau plus méthodologique sont aussi apparues :

1. Il manque des variables dans cette base de données pour étudier une problématique particulière. Où puis-je trouver ce type de données et comment les incorporer dans une analyse ?
2. Comment intégrer dans un même modèle statistique toutes les variables recueillies par le biais des différentes sources d'information ?
3. Quels types de modèles statistiques correspondent le mieux et comment s'assurer de leur validité ?

Séances de bricolage au début, ces réflexions m'ont permis de découvrir une branche de l'épidémiologie assez dédaignée par la communauté scientifique et complètement absente dans les enseignements classiques en épidémiologie : les études écologiques. A l'époque, l'écologie n'était pour moi qu'un vague souvenir de mes études d'agronomie et comprendre le lien entre écologie et épidémiologie a alors été nécessaire. Cette recherche m'a permis petit à petit de prendre pleinement conscience du très fort lien de parenté existant entre ces deux sciences.

Par un processus identique, ces recherches m'ont aussi fait découvrir l'existence de nombreuses sources de données externes à l'hôpital, compilées par une grande variété d'organismes, et potentiellement très utiles en santé publique. La plupart étaient libres d'accès, mais paradoxalement peu valorisées en épidémiologie. L'expérience montrait pourtant qu'elles pouvaient s'avérer d'une aide précieuse pour éclairer sous un angle nouveau des problématiques de santé.

Le sujet de cette thèse n'a ainsi pas été défini clairement depuis le début mais s'est en fait surtout dessiné au fur et à mesure de l'état d'avancement de mon travail. C'est le lien logique qui unit toutes les recherches présentées dans ce mémoire.

PARTIE 1 : INTRODUCTION GENERALE

Chapitre 1 - Les données libres

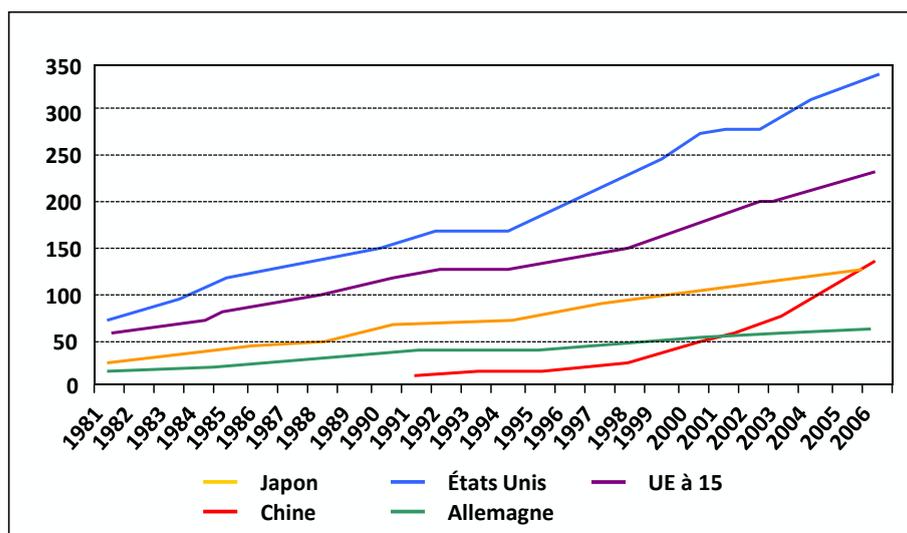
« He who receives an idea from me, receives instruction himself without lessening mine; as he who lights his taper at mine receives light without darkening me.»

Thomas Jefferson

1.1 Le rôle croissant des données numériques dans le processus de recherche

L'évolution de la recherche scientifique est actuellement caractérisée par une accélération de la croissance autant en termes d'échelle, que de portée ou de complexité. Ce développement de la recherche scientifique a parallèlement été accompagné par une hausse substantielle des coûts. Les dépenses globales de recherche et de développement dans les pays de l'OCDE ont ainsi augmenté de 163,2 à 679,8 milliards de dollars entre 1981 et 2003 (Figure 1) [1].

Figure 1 : Dépenses intérieures brutes affectées à la recherche et au développement (en milliard USD à PPA courantes), 1981-2006. Adaptée de [1].



Ces tendances ont également suscité l'implication croissante des politiques gouvernementales dans la recherche scientifique tant au niveau national et international. La mise en place de politiques de recherche a favorisé une plus grande coopération entre les chercheurs du public et du secteur privé, ainsi qu'une plus grande coopération internationale dans le domaine de la recherche publique [2]. La croissance exponentielle de l'infrastructure informatique, en particulier dans les pays de l'OCDE, a été à la fois un facilitateur et un accélérateur de ces tendances. Elle a considérablement amplifié l'ampleur, la portée et la complexité de la recherche scientifique, en permettant l'intégration des participants à la recherche et des ressources d'information de plusieurs disciplines, secteurs et pays.

Une quantité sans cesse croissante de données sur l'univers qui nous entoure est produite par des organismes gouvernementaux, institutions de recherche et de l'industrie comme une composante fondamentale de la recherche scientifique du monde entier. Pratiquement tout ce qui sert à des fins de recherche peut être décrit et enregistré dans une base de données numérique: une séquence génomique, la vitesse des particules subatomiques, une réponse à une enquête sociale, la fréquence des noms dans un corps de textes ou encore les images satellites des autres planètes,...

Comme décrit dans le colloque du National Research Council portant sur le rôle des données scientifiques et techniques et de l'information dans le domaine public en 2002 [3]:

« Les progrès rapides des technologies numériques et des réseaux au cours des deux dernières décennies ont radicalement modifié et amélioré la façon dont les données peuvent être produites, diffusées, gérées et utilisées, autant dans les sciences que dans les autres sphères de l'activité humaine. De nouveaux capteurs et instruments expérimentaux produisent actuellement une augmentation exponentielle des quantités et des types de données brutes. Cela a créé des possibilités sans précédent pour accélérer la recherche et la création de richesse se reposant sur l'exploitation de telles données [...] Il y a des secteurs entiers de la science, comme la bioinformatique, la biologie moléculaire, qui sont maintenant principalement conditionnées et dépendantes de l'utilisation de ces grandes masses de données. De nouveaux outils et logiciels aident maintenant à interpréter et à transformer les données brutes dans des configurations illimitées d'informations et de connaissances. Et l'outil de recherche le plus important et omniprésent de tous, Internet, a bouleversé l'espace et le temps dans lequel les données et informations peuvent être partagées et mises à disposition, ce qui conduit maintenant à des modes entièrement nouveaux et prometteurs de collaboration et de production en recherche scientifique. »

La production d'un jeu de données constitue donc la première étape de l'amélioration de la connaissance de certains processus de la nature ou de la société pour poursuivre la recherche et l'innovation dans ces domaines. L'offre croissante de données peut aussi fréquemment être utile à des fins autres que celles envisagées dans la collection d'origine. Beaucoup de données financées par l'état sont d'une grande valeur grâce à leur réutilisation par un large éventail de chercheurs publics et privés, dans d'autres types d'applications socio-économiques, ou par le grand public.

Les changements induits dans le processus de recherche ont été non seulement quantitatifs, mais aussi qualitatifs, et ont mené à des découvertes qui n'étaient pas possibles avant. Par exemple, des éléments de données apparemment sans lien peuvent être assemblés et donner de nouveaux résultats originaux non suspectés. La stratégie de recherche développée par Rita Colwell, ancienne directrice de la Fondation nationale américaine de recherche, dans ses études sur le choléra en est un exemple frappant [4]. En combinant de grands ensembles de données portant sur la biologie marine, les images satellitaires, l'épidémiologie, les analyses ADN, ou encore l'anthropologie sociale, elle a été en mesure d'expliquer les dynamiques de cette maladie qui, sans l'utilisation des outils informatiques et l'accès à l'ensemble des données diverses, seraient restées méconnues.

Il est clair que les données numériques jouent maintenant un rôle central dans le système scientifique mondial émergent et dans la promesse de l'e-science. Et tandis que la plupart des progrès palpables à ce jour ont eu lieu dans les pays les plus développés économiquement, le plus gros gain que l'on pourrait espérer de ce nouveau paradigme de recherche pourrait prendre place dans les pays en voie de développement. Ces changements majeurs dans la structure et la mise en place de recherches demandeuses de données provenant de l'infrastructure informatique a cependant un besoin urgent d'une meilleure organisation et planification. En effet, Un environnement plus transparent et prévisible pour l'accès et l'utilisation des ressources de données permettrait d'optimiser le système de recherches nationales et internationales.

1.2 Le 20 avril 2010 une date historique

20 avril 2010. Peu de gens connaissent la signification particulière de cette date et son coté historique, mais dans le petit monde des données libres (ou « Open Data » en anglais), cette dernière représente une petite révolution. Mais pourquoi?

Le 20 avril 2010 la banque mondiale a révélé son intention de rendre librement accessible l'ensemble des statistiques de développement qu'elle collecte quotidiennement. Cette étape attendue depuis très longtemps donna enfin accès aux étudiants, chercheurs, journalistes, économistes et organisations du monde entier à une colossale masse d'information sur l'ensemble des pays du monde. En ouvrant ses données, la banque mondiale a ainsi rejoint le mouvement grandissant de la « donnée libre » (Open Data en anglais).

Les données publiques ouvertes ont vu le jour aux États-Unis dès 1966 avec l'adoption de la loi sur le libre accès à l'information, la « Freedom of Information Act », amendée plusieurs fois notamment en 2007 par la loi « OPEN Government Act ». La libération des données s'est développée de manière formelle au niveau d'un État puis au niveau fédéral avec l'élection de Barack Obama, qui, dès son investiture a dévoilé son intention de « créer un niveau d'ouverture du gouvernement sans précédent », dans le but de « renforcer la démocratie et promouvoir son efficacité et son efficience au sein du gouvernement ». En mars 2009, le projet data.gov est lancé, puis en décembre de la même année la directive du gouvernement ouvert est publiée et pose les trois piliers de la coopération entre l'état et les citoyens qui sont la transparence du gouvernement, la participation et la collaboration.

Au Royaume Uni, un projet analogue a officiellement été lancé en janvier 2010, par Gordon Brown et sous l'impulsion de Tim Berners-Lee. Le projet oblige le gouvernement à publier, entre autres, toute dépense supérieure à 25 000 livres sterling.

En France, la tendance est apparue avec l'initiative isolée de villes comme Rennes puis Paris. La mission de l'équipe de Séverin Naudet rassemblée sous le sigle ETALAB est de créer le portail de données publiques ouvertes data.gouv.fr qui sera testé d'ici octobre 2011 [6].

Ce mouvement de « l'Open Data », porté par un nombre grandissant de chercheurs, s'est aussi développé et affirmé ces dernières années dans le monde de la recherche scientifique. La majorité des débats se sont portés dans un premier temps (et se portent encore) sur l'accès libre aux publications scientifiques

(l'«Open Access» en anglais). Actuellement, un deuxième débat, beaucoup plus radical, commence à sortir de l'ombre : l'accès libre aux données brutes de la recherche financée par des fonds publics.

1.3 Le libre accès aux publications scientifiques (en anglais : Open Access)

1.3.1. L'Open Access c'est quoi ?

Il est maintenant unanimement reconnu que les technologies de l'information et de la communication ont radicalement transformé ces 20 dernières années le monde de la recherche. Consécutivement à ce phénomène, le besoin d'adapter les pratiques scientifiques pour tirer pleinement avantage des connaissances et résultats engrangés par les chercheurs s'est fait de plus en plus sentir. Une des avancées majeures dans ce domaine a été l'émergence de l'«e-science», que l'on peut définir brièvement comme un accès facilité à travers un poste informatique et via Internet à des ressources distribuées permettant la consultation d'une très grande partie de la création et de la production intellectuelle de la communauté scientifique ainsi que des résultats associés [7].

Le libre accès (en anglais : Open Access), est la libre disponibilité en ligne de contenus numériques, qui peuvent eux-mêmes être soit libres soit sous un des régimes de propriété intellectuelle. L'Open Access est principalement utilisé pour les articles de revues ou de recherches universitaires, sélectionnés par des pairs, et publiés gratuitement.

Une des premières déclarations internationales majeures sur le libre accès, qui inclut une définition, une information de fond et une liste de signataires, est l'Open Access Initiative de Budapest de 2001. Ce rassemblement est reconnu comme le premier rassemblement historique fondateur du mouvement du libre accès. Un des logos symboliques en forme de cadenas ouvert de ce mouvement est présenté dans la Figure 2.

Il existe deux grands types de libre accès avec de nombreuses variations [8]. Dans la publication en libre accès [9], également connue comme la voie en « or » d'Open Access, les revues rendent leurs articles directement et immédiatement accessible à la consultation. Ces publications s'appellent des « revues à accès ouvert » («Open Access journals»). Un exemple de publications en Open Access est la Public Library of Science [10]. En ce qui concerne l'Open Access par auto-archivage [8], aussi appelée la voie « verte » d'Open Access, les auteurs font des copies de leurs propres articles publiés, et les rendent ouvertement accessibles. Ils le font généralement sur une page personnelle ou un dépôt institutionnel.

Figure 2 : Un des logos du mouvement « Open Access »



Le libre accès est actuellement à l'origine de beaucoup de discussions entre universitaires, bibliothécaires, administrateurs d'universités et politiques. Il existe un désaccord substantiel sur le concept de l'Open Access, avec un grand débat autour de la rémunération économique du libre accès par les universitaires.

En effet, selon les cas, il peut exister des frais de publication ; il existe un mythe selon lequel la publication en Open Access implique que l'auteur doit payer. Traditionnellement, beaucoup de revues scientifiques ont facturé des frais de mise en page, bien avant que l'Open Access soit devenu une possibilité. De récentes recherches ont démontré que les journaux en libre accès ne demandent pas de frais de publication, et sont moins taxés de frais d'auteurs que les titres traditionnels à abonnement. Quand les revues pratiquent des frais de traitement, c'est l'employeur (ou le bailleur de fonds) de l'auteur qui paye ces honoraires, et non l'auteur. De plus, des provisions sont constituées afin de couvrir les auteurs pour qui le fait de publier pourrait entraîner des difficultés financières.

1.3.2. Historique du mouvement Open Access

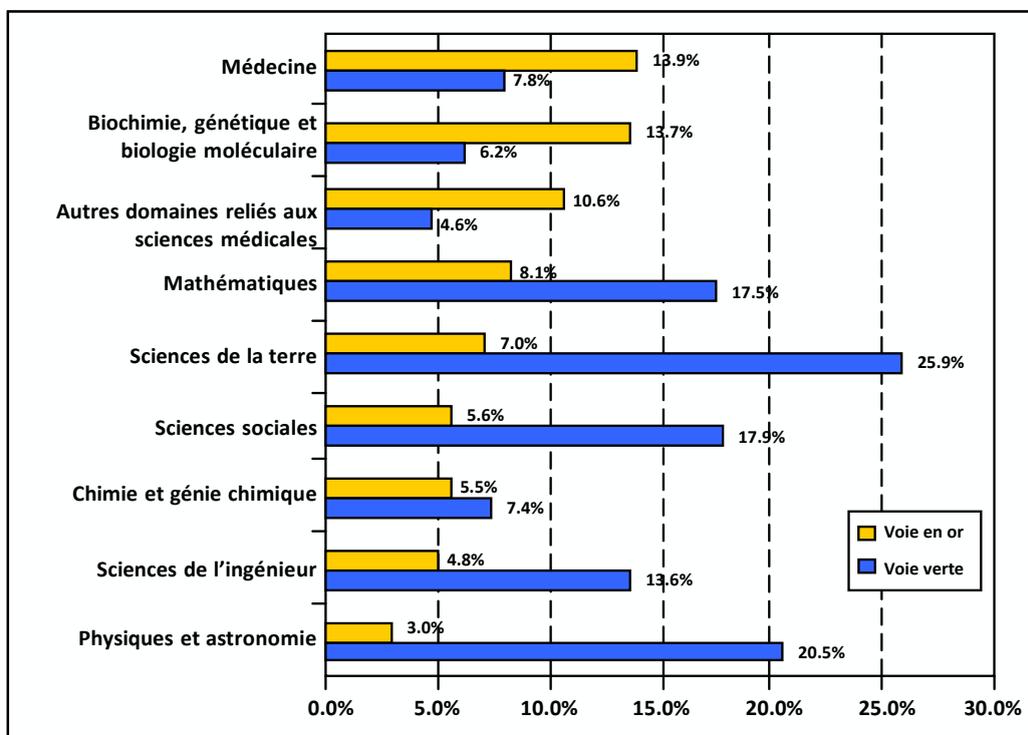
Ce mouvement moderne du libre accès aux publications (ou aussi appelé « archive ouverte ») est né avec le potentiel dégagé par l'arrivée des supports électroniques, et plus particulièrement du World Wide Web. Il est désormais possible d'éditer un article scientifique et de le rendre immédiatement accessible n'importe où dans le monde pourvu qu'il y ait des ordinateurs et des connexions internet. Le coût fixe de production de l'article peut alors être clairement séparé du coût marginal de distribution en ligne. Ces nouvelles possibilités ont émergé à un moment où le système traditionnel, basé sur l'impression papier des revues scientifiques, était en crise. Le nombre de journaux et d'articles produits a augmenté à un taux régulier ; cependant le coût moyen par journal a dépassé l'inflation alors que les budgets des bibliothèques universitaires sont restés assez statiques. Ironiquement, l'accès aux journaux scientifiques diminuait, au moment même où la technologie rendait possible, pour la première fois, un accès presque illimité. Les bibliothécaires ont joué une part importante dans le mouvement de l'archive ouverte, d'abord en alertant le corps enseignant et les administrateurs sur la crise provoquée par la croissance du coût des abonnements aux revues savantes. L'association des bibliothèques de recherches a mis en place en 1997 une coalition sur les ressources académiques et l'édition scientifique (Scholarly Publishing and Academic Resources Coalition SPARC) qui regroupait des bibliothèques universitaires et de recherches ainsi que d'autres organismes, pour faire face à la crise et pour développer des solutions de rechange, telle que l'archive ouverte.

La première archive scientifique disponible en ligne fut arXiv.org. Lancé en 1991, c'était au début un service de publication de document de travail pour des physiciens. L'existence antérieure d'une culture de publication des documents de travail en physique des particules notamment, serait une des raisons principales pour expliquer la réussite de l'arXiv [11].

Le phénomène « Open Access » représente maintenant une part non négligeable de la littérature scientifique. Une étude publiée en 2010 a montré qu'environ 20% de la production totale des articles publiés en 2008 pouvait être accessibles librement (Figure 3) [12]. 8,5% de l'ensemble de la littérature pouvaient être trouvés gratuitement sur les sites des éditeurs, dont 62% dans des revues totalement « Open Access », 14%

dans des revues payantes rendant les versions électroniques gratuites après un délai, et 24% comme articles individuellement ouverts (contre paiement) dans des revues payantes. Parallèlement, pour 11,9% des articles, des copies gratuites du texte intégral ont été trouvées ailleurs dans d'autres dépôts thématiques (43%), des dépôts institutionnels (24%) ou sur les pages d'accueil des auteurs ou de leurs départements (33%). Ces exemplaires étaient en outre classés comme copie exacte de l'article publié (38%), manuscrits acceptés pour la publication (46%) ou manuscrits soumis (15%). En discernant par domaines scientifiques, la chimie avait la plus faible part d'articles en libre accès (13%), tandis que les sciences de la Terre, avait la plus élevée (33%). En médecine et biochimie, la publication par la voie en or était plus fréquente que l'auto-archivage des manuscrits. Dans tous les autres champs l'auto-archivage dominait.

Figure 3 : Proportions des articles scientifiques disponibles en libre accès par domaine de recherches.
Adaptée de [12].



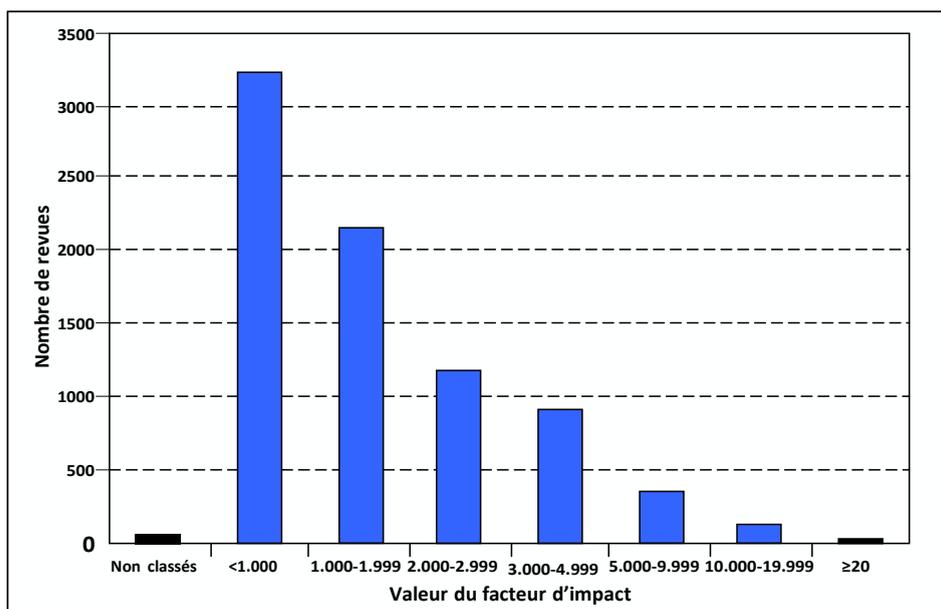
1.3.1. Intérêts de l'Open Access pour les auteurs

Pour les auteurs, la principale motivation à créer un article est l'impact de celui-ci. En effet, un article est fait pour être employé et cité. Les chercheurs se donnent traditionnellement beaucoup pour leur travail. Ils sont payés par des bailleurs de fonds de recherches et/ou leurs universités pour faire des recherches. L'article final est donc la preuve qu'ils ont effectué un travail (et non pas un gain commercial comme dans l'industrie par exemple). Plus l'article est utilisé, cité, appliqué mieux c'est pour les recherches et la carrière du chercheur. Ainsi, les chercheurs scientifiques et les bailleurs de fonds utilisent principalement deux indicateurs pour évaluer la recherche : le facteur d'impact (« impact factor » en anglais) des revues et le nombre de citations des articles.

- *La course à « l'impact factor »*

Le facteur d'impact est un indicateur qui est couramment utilisé pour évaluer la performance d'une revue scientifique. Ce facteur a été conçu par Eugène Garfield, le fondateur de l'Institut de l'information scientifique (Institute for Scientific Information) en 1960. Cet Institut a été acquis par Thomson Scientific & Healthcare en 1992. Il fait partie de Thomson Reuters à Philadelphie (USA) qui est une agence de presse canadienne et une société d'Édition professionnelle et qui édite chaque année le « Journal Citation Reports (JCR) » [13].

Figure 4 : Répartition du facteur d'impact de 8005 revues scientifiques évaluées en 2010 par Thomson Reuters



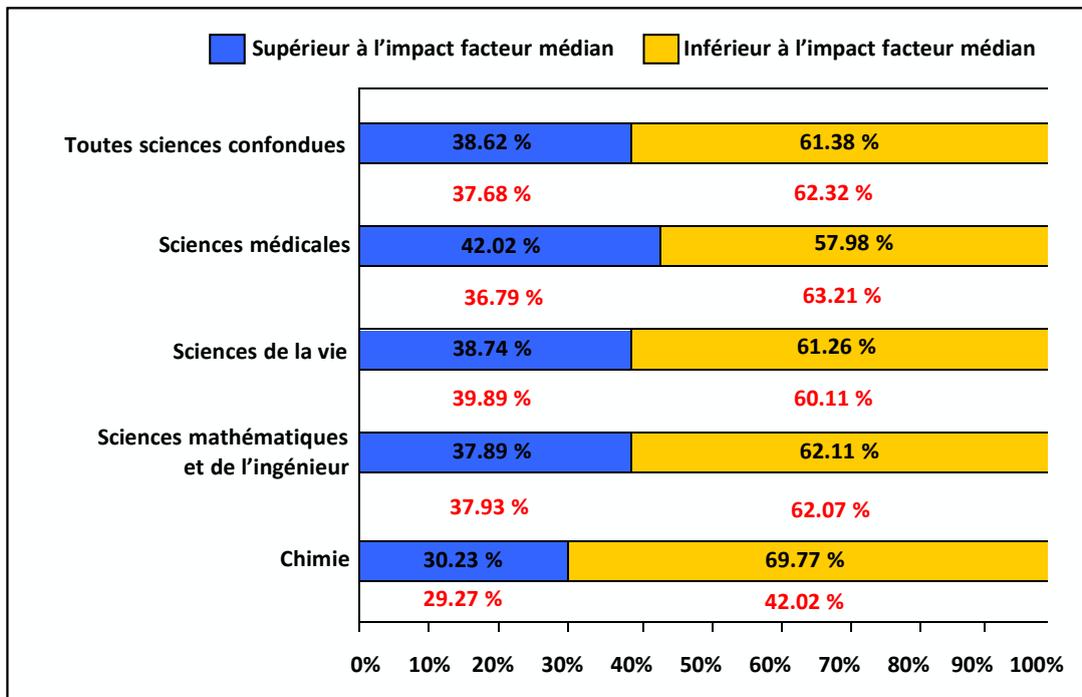
Le facteur d'impact représente, pour une année donnée, le rapport entre le nombre de citations sur le nombre d'articles publiés par un journal, sur une période de référencement de deux ans. Par exemple : pour une revue donnée le facteur d'impact en 2003 est le nombre de citations se référant à la revue donnée, trouvées dans la littérature pour l'année 2003 et concernant les articles publiés en 2001 et 2002 / nombre d'articles total publiés par la revue donnée en 2001 et 2002. La Figure 4 montre la répartition du facteur d'impact des 8005 journaux évalués par Thomson Reuters en 2010. Ainsi une publication est souvent considérée comme « d'importance » quand elle est publiée dans un journal dont le facteur d'impact est supérieur à 5.

De nombreuses critiques à l'encontre du facteur d'impact ont été formulées ces dernières années, mettant principalement en cause sa justesse à mesurer la qualité d'un journal, d'un auteur ou d'un article [14, 15]. Malgré tout il est très couramment utilisé par les sponsors et bailleurs de fonds de la recherche scientifique qui s'en servent comme critère objectif pour juger la productivité de leurs investissements. L'avantage majeur de l'utilisation du facteur d'impact pour ces derniers tient au fait qu'il se base sur une large couverture internationale de 8 400 revues publiées dans plus de 60 pays différents.

Bien que les journaux scientifiques en libre accès sont des acteurs relativement nouveaux sur le marché de l'édition, et que dans ce domaine gagner une réputation et une visibilité est un défi complexe, certains d'entre eux montrent une impressionnante tendance à l'augmentation de leur facteur d'impact depuis leurs premières années de suivi. Une étude récente a ainsi montré que l'évolution des performances des revues scientifiques en libre accès, en terme de facteur d'impact, est assez bonne [16]. Dans le domaine de la recherche médicale, ces résultats sont particulièrement prometteurs (Figure 5).

Ainsi malgré le fait que les meilleurs classements en terme de facteur d'impact sont toujours détenus par des journaux scientifiques non libres, il est de moins en moins pénalisant pour un chercheur de publier dans des journaux en libre accès. Dans certains domaines comme dans celui de la médecine, ces derniers commencent même à s'imposer comme des références n'ayant plus rien à envier à leurs équivalents non libres. La saga des journaux Plos en est un parfait exemple [10].

Figure 5 : Répartition des facteurs d'impact des journaux scientifiques libres par rapport au facteur d'impact médian des journaux évalués par l'édition 2008 (2007 en rouge) du JCR. Adaptée de [16].



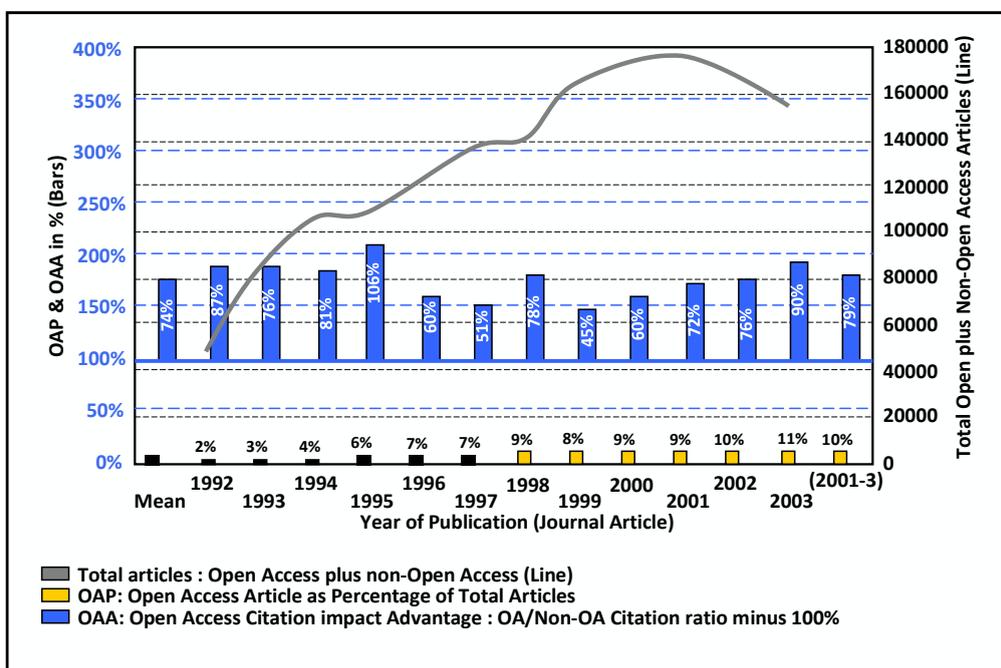
- *Le nombre de citations : l'atout majeur des publications « Open Access »*

Les chercheurs faisant des recherches et qui, une fois ces dernières finies, mettent les résultats dans un tiroir de leur bureau équivalent tout simplement à n'avoir rien fait du tout. Comme indiqué dans la maxime « publier ou mourir », les chercheurs ont la nécessité pour survivre de soumettre leurs recherches à des pairs qui les examinent et accessoirement les publient pour que d'autres puissent ensuite utiliser et appliquer leurs conclusions. Mais obtenir une évaluation positive et donc une publication des résultats ne suffit pas. Il est nécessaire que d'autres chercheurs trouvent la recherche utile et d'importance. En science, ce phénomène est principalement mesuré à travers le nombre de citations d'un article. Ainsi, le nombre de citations d'un article est la mesure la plus utilisée pour refléter l'impact scientifique dans la communauté d'une recherche. C'est un

des outils de base pour mesurer la «renommée» d’un scientifique. Plus les articles d’un chercheur sont cités, plus on estimera que leur diffusion dans la communauté scientifique aura été grande et que la réputation de ce chercheur est indéniable. En anglais cette mesure est appelée « citation impact ».

De nombreuses mesures ont été proposées, au-delà du nombre de citations simples, afin de mieux quantifier l’impact scientifique en termes de citations d’un individu. Les mesures les plus connues sont l’indice H [17] et l’indice G [18]. Une approche alternative pour mesurer l’impact scientifique d’un chercheur s’appuie sur les données d’utilisation des journaux en ligne, comme le nombre de téléchargements à partir des éditeurs par exemple [19, 20].

Figure 6 : Avantage en termes de citations des articles en « Open Access » par rapport aux articles non disponibles en libre accès (tous domaines scientifiques confondus). Adaptée de [8].



Cependant pour que la communauté scientifique soit en mesure d'utiliser et de citer les articles d'importance, elle doit d'abord être en mesure d'y accéder. C'est le problème de l'accès/impact des articles de recherche. Quelle est l'amplitude de ce phénomène ? Ces dernières années des estimations ont commencé à émerger et les résultats sont sans équivoque. Dans le domaine des sciences informatiques une étude majeure a montré que les articles dont les textes étaient disponibles gratuitement en ligne (en «Open Access») avaient en pourcentage un impact 336 fois en % supérieure en terme de citation que les articles non disponibles en libre accès [21]. D'autres études ont montré des résultats similaires dans d'autres sciences comme l'astrophysique [22, 23], les mathématiques [24] et dans le domaine des sciences en générale [8] (Figure 6).

Il semble maintenant acquis qu'un chercheur, en terme de diffusion de ces résultats et d'impact de ces travaux dans la communauté scientifique, ait tout intérêt à publier dans des journaux «Open Access» (ou tout au moins à réaliser un auto-archivage des ces articles).

1.3.2 Intérêt de l'Open Access pour les utilisateurs

La plupart du temps, les utilisateurs principaux des articles des recherches sont d'autres chercheurs. Le libre accès permet aux chercheurs, en tant que lecteurs, d'accéder à des articles de revues auxquelles leurs bibliothèques ne se sont pas abonnées. Les grands bénéficiaires de l'Open Access sont/seront les pays en voie de développement où il existe actuellement des universités sans souscription à des revues scientifiques. Cependant, tous les chercheurs en bénéficient car aucune bibliothèque ne peut s'offrir un abonnement à tous les périodiques scientifiques. La plupart d'entre elles ne sont abonnées qu'à une fraction d'entre eux [25]. Ce phénomène est résumé sous le titre de « crise de la publication périodique ».

L'Open Access étend le monde de la recherche au-delà de la communauté scientifique. Un article en libre accès peut être lu par quiconque, que ce soit un professionnel dans un domaine, un journaliste, un homme politique, un fonctionnaire ou un amateur intéressé. Ces résultats peuvent être assimilés, réutilisés par une personne tierce, ou encore combinés à d'autres pour produire des résultats originaux. Les possibilités d'assimilation, d'utilisation et d'intégration de ces résultats sont infinies.

1.4. Le libre accès aux données brutes des études scientifiques (en anglais : Open Data)

Ces dernières années le débat autour de l'accès libre aux données s'est principalement focalisé sur l'accès aux résultats digitalisés des recherches scientifiques c'est-à-dire sur l'accès aux publications des résultats scientifiques (les articles publiés dans les journaux scientifiques). Ce débat en a longtemps caché un autre : celui de l'accès aux données brutes des recherches à l'origine même des travaux de recherches et dont la création fait l'objet d'un très important investissement public chaque année.

1.4.1. Les coûts cachés des systèmes de données non libres

De plus en plus, les données numériques font leur apparition dans le monde de la recherche en tant que ressources autonomes, leur utilisation n'étant plus exclusivement liées à leurs producteurs originaux. Bien évidemment, certaines données ont peu de valeur en dehors du projet de recherche pour lesquelles elles ont été collectées ou ne sont pas utiles du fait d'un manque de qualité, d'une documentation insuffisante, ou d'autres lacunes. Malgré tout, de nombreux types de données peuvent toutefois être utilisés au-delà du champ d'application pour lequel les producteurs et utilisateurs originaux les ont engrangées. Le partage des données de la recherche publique ouvre ainsi de nouvelles possibilités pour améliorer la qualité et la productivité de la recherche. Cependant, la pleine réalisation de ce potentiel nécessite une attention particulière à la politique d'utilisation de ces données.

Parallèlement à ces nouvelles opportunités, il y a des enjeux compétitifs et d'autres raisons légitimes de restreindre l'accès aux données de la recherche publique. Les différents acteurs impliqués peuvent avoir des intérêts contradictoires lorsque l'on considère les avantages et les inconvénients d'un accès ouvert aux données. Beaucoup de chercheurs ont tendance à considérer les données qu'ils produisent à travers la

recherche publique comme une propriété individuelle ou institutionnelle, et ce point de vue est souvent renforcé en l'absence de directives claires de leurs sources de financement public.

Il y a, cependant, un certain nombre de conséquences négatives sur l'efficacité et l'efficacé du système de recherches provoquées par un accès fermé à ces données numériques.

1. Des coûts de recherche plus élevés

Un accès restreint aux données impose des inefficacités structurelles et des coûts de recherche bien plus élevés. Beaucoup de bases de données ne peuvent pas ou ne devraient pas être recréés de façon indépendante, soit parce qu'elles contiennent des observations de phénomènes uniques, des informations historiques, où coûtent très cher à générer [26]. Par ailleurs, les bases de données avec un statut de monopole qui sont maintenues sur une base exclusive auront aussi plus tendance à entraîner une hausse des prix non concurrentiels [27]. Gérer des bases de données financées sur fonds publics de manière restrictive ajoute aussi de substantiels frais administratifs supplémentaires taxant le système de recherche publique. Ceci est particulièrement vrai dans le cas de figure où des institutions publiques vendent des licences d'utilisation des données à des coûts élevés à d'autres institutions publiques.

2. Des coûts dus aux opportunités ratées

Ces coûts sont sans doute moins évidents que les précédents, mais il y a beaucoup moins de recherches intensives possibles sur les données quand ces dernières ne sont pas partagées ou disponibles en ligne facilement. Il en résulte des coûts importants liés aux occasions manquées qui sont inévitables, mais malheureusement très difficilement mesurables [28]. Une simple analogie pourrait suffire à illustrer ce problème. Tout comme il ne serait guère rentable de limiter l'accès à un télescope ou à un accélérateur de particules uniquement aux chercheurs et ingénieurs qui ont conçu l'instrument, c'est un gaspillage d'efforts et d'argent de limiter l'utilisation des données aux seuls chercheurs responsables de leur collection originale et de perdre ainsi les potentiels bénéfiques d'applications plus larges utilisant ces données.

3. Obstacles à l'innovation

La production en aval de droits d'auteur ou de biens intellectuels brevetables par les secteurs publics et privés dépend dans une large mesure de l'accès aux données et informations publiques circulant librement en amont. La surprotection ou l'indisponibilité de bases de données publiques conduisent ainsi à un lourd coût social, une taxation du système d'innovation dans chaque pays et à un ralentissement significatif des progrès scientifiques.

4. Des niveaux de coopération, d'éducation et de formation moins élevés

Ne pas rendre ses données de recherches facilement disponibles, ou ériger des barrières trop élevées à leurs accès, entraîne nécessairement une baisse de l'efficacité des coopérations interdisciplinaires, interinstitutionnelles, intersectorielles, et internationales. De la même façon, les étudiants peuvent être formés de manière moins efficace s'il ne leur est pas possible de travailler et de s'entraîner sur un large

éventail de données pour parfaire leurs connaissances. Ces obstacles sont renforcés dans les nombreux cas où les politiques de gestion n'autorisent l'accès aux données qu'à un petit nombre restreint d'investigateurs pré-approuvés et formellement associés aux projets. Dans ces cas de figure, une forte limitation à l'accès et à l'utilisation de ces données par d'autres chercheurs travaillant dans des disciplines, institutions, secteurs et pays «non approuvés» préalablement, se superposent très souvent.

5. Une qualité de données non optimale

Les données récoltées et analysées dans un environnement fermé sont souvent validées et vérifiées par une communauté nettement plus petite et moins diversifiée comparativement aux données scientifiques ouvertement disponibles. Ce phénomène accroît de manière non négligeable le risque d'une qualité des données plus faible et diminue par la même occasion la qualité des résultats de la recherche.

6. Élargissement du fossé entre les nations de l'OCDE et les pays en voie de développement

Dans l'enjeu stratégique autour de l'accès aux données pour la recherche, les pays en voie de développement sont particulièrement désavantagés par leur manque de disponibilité en données scientifiques et leurs fortes barrières à l'accès. Bien que toutes les bases de données produites dans les pays de l'OCDE ne sont pas pertinentes pour les pays moins développés, celles qui ont une large applicabilité en tant que bien public mondial sont anormalement inutilisées lorsqu'il y a un prix élevé à l'accès, et dans de nombreux cas, lorsqu'il y a un prix tout court. Des obstacles à l'accès aux données entraînent ainsi une diminution des rendements de l'investissement dans le capital social et scientifique de la recherche publique ainsi qu'une distribution inégale des retombées de ces recherches.

1.4.2. Les retombées scientifiques et socio-économiques d'une plus grande ouverture des données

Compte tenu de l'évolution et du rôle des données publiques dans la science et de la relative inefficacité du système actuel, plusieurs raisons convaincantes peuvent être avancées pour promouvoir le développement de régimes d'accès aux données plus ouverts. C'est particulièrement le cas lorsque les données sont produites au sein du gouvernement ou par des entités financées par des sources gouvernementales.

Comme la valeur des données scientifiques réside dans leur utilisation, un accès ouvert et un partage des données issues de la recherche publique offre de nombreux avantages par rapport à un système fermé et propriétaire. Ainsi ouvrir l'accès à ces données est susceptible de :

- renforcer l'investigation scientifique ouverte
- encourager la diversité d'analyses et d'opinions
- encourager les nouvelles recherches originales ainsi que les nouveaux types de recherches
- permettre la mise en place d'outils automatisés de découverte des connaissances en ligne
- permettre de vérifier les résultats précédents

- rendre possible le test d'hypothèses ou de méthodes d'analyse nouvelles ou alternatives
- appuyer les études portant sur les méthodes de collecte de données et de mesures
- faciliter l'éducation des nouveaux chercheurs
- permettre l'exploration de sujets qui n'ont pas envisagé par les investigateurs initiaux
- permettre la création de nouveaux jeux de données, informations et connaissances quand des données provenant de sources multiples sont combinées
- aider au transfert des informations factuelles et à la promotion du renforcement des capacités des pays en voie de développement
- promouvoir les collaborations interdisciplinaires, intersectorielles, interinstitutionnelles et internationales.
- permettre de maximiser le potentiel de recherche des nouvelles technologies numériques et des réseaux, et offrir ainsi un meilleur rendement de l'investissement public dans la recherche

Le libre accès aux données joue un rôle vital dans tous ces domaines. Créer une aire du jeu équitable pour les chercheurs et leurs instituts respectifs est impossible sans un accès large et efficace aux données de recherche à financement public.

1.4.3. Les retombées pour les investigateurs

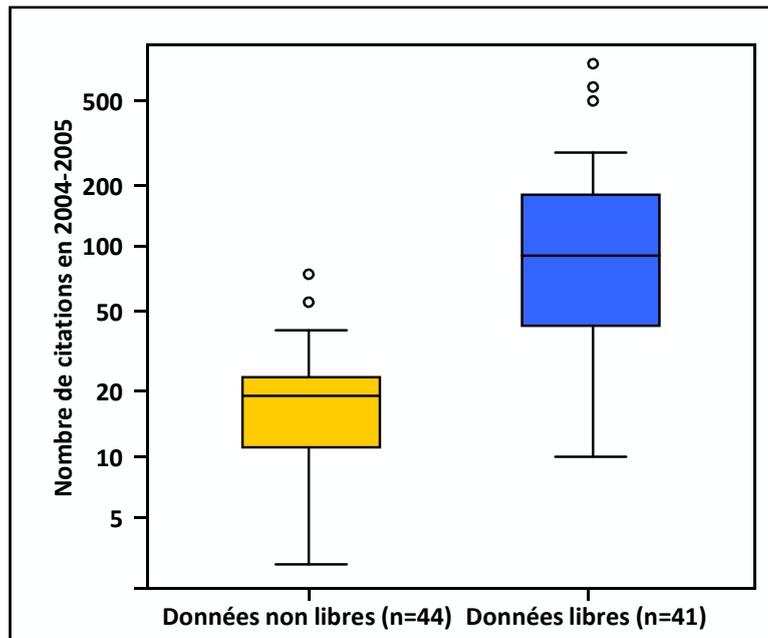
Ainsi, partager l'information facilite le processus de recherche. Le partage public des données de recherches (facteurs cliniques, résultats de patients, séquences ou puces à ADN,...) avec d'autres chercheurs permet à ces précieuses ressources de contribuer au-delà de leur analyse originale. En plus d'être utilisée pour confirmer des résultats originaux, les données brutes peuvent être utilisées pour explorer des hypothèses connexes ou nouvelles, en particulier lorsqu'elles sont combinées avec d'autres ensembles de données disponibles publiquement. Les données réelles sont indispensables pour certaines enquêtes, le développement de méthodes d'étude, de techniques d'analyse, et la création de nouveaux logiciels. L'ensemble de la communauté scientifique y retrouve également des avantages: le partage des données encourage de multiples perspectives, aide à identifier les erreurs, décourage la fraude, et est aussi utile pour la formation de nouveaux chercheurs, tout en augmentant l'utilisation efficace des fonds et des ressources en terme de population et de patients.

Cependant, alors que la communauté de chercheurs profite pleinement du partage libre de données, une grande partie du fardeau du partage retombe sur les épaules de l'investigateur de l'étude récoltant les données. Y a-t-il des avantages pour ces personnes elles même ? Il a été vu qu'un outil de valeurs pour de nombreux chercheurs était le nombre de citations de leurs publications, et que stimuler le taux de citation était une source de motivation potentiellement importante pour les auteurs des publications. Une étude récente semble ainsi montrer que les jeux de données publiquement accessibles dans le domaine de la recherche clinique en cancérologie bénéficieraient d'un taux de citation beaucoup plus élevé que leurs homologues non

libres et cela indépendamment de la date de publication, du facteur d'impact du journal et du pays d'origine de l'auteur (Figure 7) [29].

Cette corrélation entre accès libre aux données et impact de citations plus important représente ainsi un argument de force auprès des investigateurs pour promouvoir le mouvement de l'Open Data : libérer ces données représente une meilleure visibilité et reconnaissance dans la communauté scientifique.

Figure 7 : Distributions du nombre de citations en 2004-2005 de 85 essais cliniques en oncologie en fonction du type d'accès aux données. Adaptée de [29].



1.5. Conclusions

Permettre aux données qu'elles soient largement diffusées dans la communauté de chercheurs accélérerait donc le rythme des découvertes et améliorerait l'efficacité d'une recherche. Dans de nombreux champs de recherches comme, la génétique, la biologie moléculaire et les sciences sociales, le partage des données est déjà profondément ancré dans la manière de travailler des chercheurs. En génomique et génétique, la mutualisation d'études faites sur différentes populations a conduit à une explosion des connaissances sur les déterminants génétiques de la variabilité humaine en termes de santé et de maladies. Des outils bien établis permettant aux chercheurs d'accéder et d'interroger des ressources partagées de données existent maintenant et favorisent la construction de nouvelles recherches basées sur ces données publiques.

En revanche, cette culture n'a pas encore été largement adoptée par la communauté de chercheurs travaillant en santé publique. Une grande partie des infrastructures, des normes techniques, et des incitations qui sont nécessaires pour soutenir le partage des données manquent, avec par-dessus cela une réticence profonde de certains chercheurs à libérer leurs données. Trop souvent, les données sont traitées comme des biens privés des investigateurs qui cherchent à maximiser leur nombre de publications, au détriment d'une

utilisation la plus large possible. Cette situation menace de limiter à la fois les progrès possibles apportés par cette recherche ainsi que ses applications en santé publique [30, 31].

Les progrès récents dans les technologies de l'information ont révolutionné la science et tout particulièrement la santé publique et l'épidémiologie, offrant ainsi de nouvelles possibilités pour les chercheurs de partager des données et de s'appuyer sur le travail des autres. L'informatique couplé à la capacité de créer de grands jeux de données pouvant être combinés avec des informations provenant de nombreuses autres sources présente un énorme potentiel pour faire progresser le développement en matière de santé publique.

Bien que l'identification des principes et des règles au cœur du partage libre des données en santé publique soit une étape importante, les défis dans la construction d'une culture du partage, et des ressources nécessaires pour soutenir ce partage de données sont tout aussi considérables. Un renforcement des capacités et des compétences dans le milieu de la recherche pour gérer et analyser ce type de données représente aussi une étape cruciale.

D'un point de vue méthodologique, cette démocratisation des données a surtout profité pour l'instant à la modélisation mathématique en épidémiologie, qui s'applique à la description et à la prévision de la dynamique des maladies, et à l'étude de stratégies de contrôle. Il serait cependant faux de penser que cette explosion de données ne peut pas être exploitée d'un point de vue statistique, en particulier aux niveaux de l'estimation des paramètres, du test des modèles, ou encore de l'analyse de sensibilité aux hypothèses. Le véritable défi statistique des données libres repose sur la construction de modèles répondant aux spécificités du croisement de sources d'informations diverses et variées récoltées à différentes échelles géographiques et de temps. Depuis longtemps, Les études écologiques sont une des solutions statistiques utilisées pour répondre à ce type de problèmes méthodologiques.

Bibliographie

1. Organisation for Economic Co-operation and Development (OECD). **Main Science and Technology Indicators**. Paris: OECD; 2005.
2. Organisation for Economic Co-operation and Development (OECD). **The Knowledgebased Economy**. Paris: OECD; 1996.
3. Uhler P. **Discussion Framework in Julie Esanu and Paul Uhler (eds) The Role of Scientific and Technical Data and Information in the Public Domain** 2003.
4. Colwell RR. **Global climate and infectious disease: the cholera paradigm**. *Science* 1996,274:2025-2031.
5. World Bank. **The World Bank Database**; 2011.
6. ETALAB. **Le blog de la mission chargée de créer data.gouv.fr**; 2011.
7. Atkins DE, Droegemeier KK, Feldman SI, Garcia-Molina H, Klein ML, Messerschmitt DG, *et al.* **Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure**. In. Arlington, VA: National Science Foundation; 2003.
8. Harnad S, Brody T, Vallieres F, Carr L, Hitchcock S, Gingras Y, *et al.* **The Access/Impact Problem and the Green and Gold Roads to Open Access**. *Serials Review* 2004,30.
9. Directory of Open Access Journals (DOAJ). **DOAJ website**; 2011.
10. Public Library of Science (PLoS). **PLoS website**; 2011.
11. Cornell University Library. **ArXiv.org**; 2011.
12. Bjork BC, Welling P, Laakso M, Majlender P, Hedlund T, Gudnason G. **Open access to the scientific journal literature: situation 2009**. *PLoS One* 2010,5:e11273.
13. Thomson Reuters. **Journal Citation Reports (JCR)**; 2011.
14. Campbell P. **Escape from the impact factor**. *Ethics Sci Environ Polit* 2008,8:5-7.
15. Young NS, Ioannidis JP, Al-Ubaydli O. **Why current publication practices may distort science**. *PLoS Med* 2008,5:e201.
16. Giglia E. **The Impact Factor of Open Access journals: data and trends**. *International Conference on Electronic Publishing*. Helsinki (Finland) 2010.
17. Hirsch JE. **An index to quantify an individual's scientific research output**. *Proc Natl Acad Sci U S A* 2005,102:16569-16572.
18. Egghe L. **Theory and practise of the g-index**. *Scientometrics* 2006,69:131-152.
19. Bollen J, Van de Sompel H, Smith J, Luce R. **Toward alternative metrics of journal impact: A comparison of download and citation data**. *Information Processing and Management* 2005,41:1419-1440.
20. M. Couto F, Pesquita C, Grego T, Veríssimo P. **Handling self-citations using Google Scholar**. *Cybermetrics* 2009,13.
21. Lawrence S. **Online or Invisible?** *Nature* 2001:521.

22. Kurtz MJ, Eichhorn G, Accomazzi A, S. Grant C, Demleitner M, S. Murray S, *et al.* **Worldwide Use and Impact of the NASA Astrophysics Data System Digital Library.** *Journal of the American Society for Information Science and Technology* 2004,55.
23. Kurtz MJ, Eichhorn G, Accomazzi A, S. Grant C, Demleitner M, S. Murray S, *et al.* **The Bibliometric Properties of Article Readership Information.** *Journal of the American Society for Information Science and Technology* 2004,55.
24. Odlyzko AM. **The Rapid Evolution of Scholarly Communication.** *Learned Publishing* 2002,25:7-19.
25. Association of Research Libraries (ARL). **ARL statistics website**; 2011.
26. National Research Council. **A Question of Balance: Private Rights and the Public Interest in Scientific Databases**; 1999:19-20.
27. Weiss P. **Conflicting International Public Sector Information Policies and Their Effects on the Public Domain and the Economy'** In: *The Role of Scientific and Technical Data and Information in the Public Domain*: Julie Esanu and Paul Uhler; 2003:129–132.
28. National Research Council. **Bits of Power: Issues in Global Access to Scientific Data.** In; 1997:121–124.
29. Piwowar HA, Day RS, Fridsma DB. **Sharing detailed research data is associated with increased citation rate.** *PLoS One* 2007,2:e308.
30. Pisani E, Whitworth J, Zaba B, Abou-Zahr C. **Time for fair trade in research data.** *Lancet* 2010,375:703-705.
31. Walport M, Brest P. **Sharing research data to improve public health.** *Lancet* 2011,377:537-539.

Chapitre 2 - Les méthodes écologiques

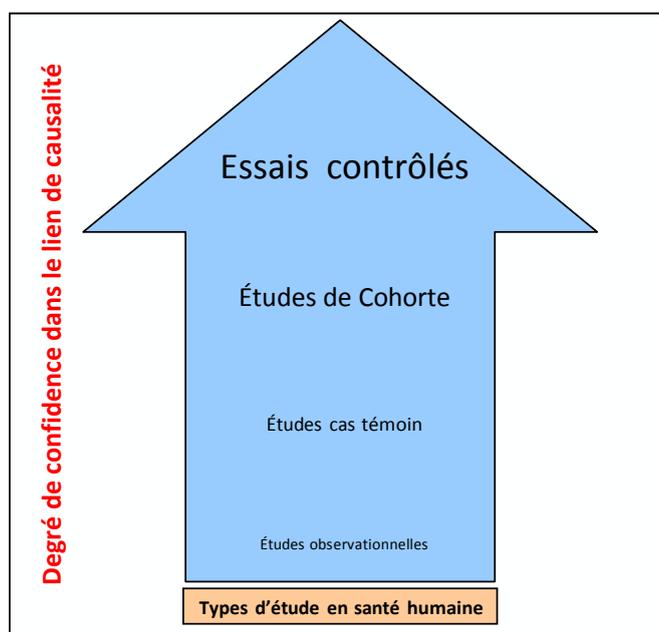
« Tous les modèles sont faux ; quelques-uns sont utiles »

George Box

2.1. Que veut dire « étude écologique » ?

L'étude de problématiques de santé s'appuie sur de nombreux types d'études, qu'il s'agisse de la mise en évidence de mécanismes biologiques, d'expérimentations animales, d'études épidémiologiques sur un nombre restreint de sujets, ou d'études à l'échelle des populations. Ces types d'études ne se prêtent pas aux mêmes interprétations quant à la nature potentielle des effets considérés, en particulier quand il s'agit de discuter leur nature causale. En se restreignant aux études sur l'homme, on peut citer quatre types d'études : l'essai contrôlé, l'étude de cohorte, l'étude cas témoin, et l'enquête épidémiologique d'observation, schémas qui, selon Sackett et al. [1], sont d'importance décroissante par rapport à leur imputation causale (Figure 8). Cependant lorsque que l'on se focalise, non pas sur le degré de causalité, mais sur la nature des données utilisées dans l'analyse, l'épidémiologie peut aussi distinguer deux grands types d'études : les études écologiques et les études individuelles. Les paragraphes suivants abordent ce type particulier d'études dites «écologiques» et qui, observées sous l'angle de la causalité, sont habituellement classées dans les études d'observations.

Figure 8 : Importance des différents types d'études en santé humaine selon leur imputation causale



Les études écologiques sont définies comme étant des études se basant sur des données agrégées à l'échelle d'un groupe ou d'une population plutôt que des données récoltées à l'échelle de l'individu. Ainsi dans les études écologiques, les unités d'analyse sont des populations ou des groupes d'individus plutôt que les

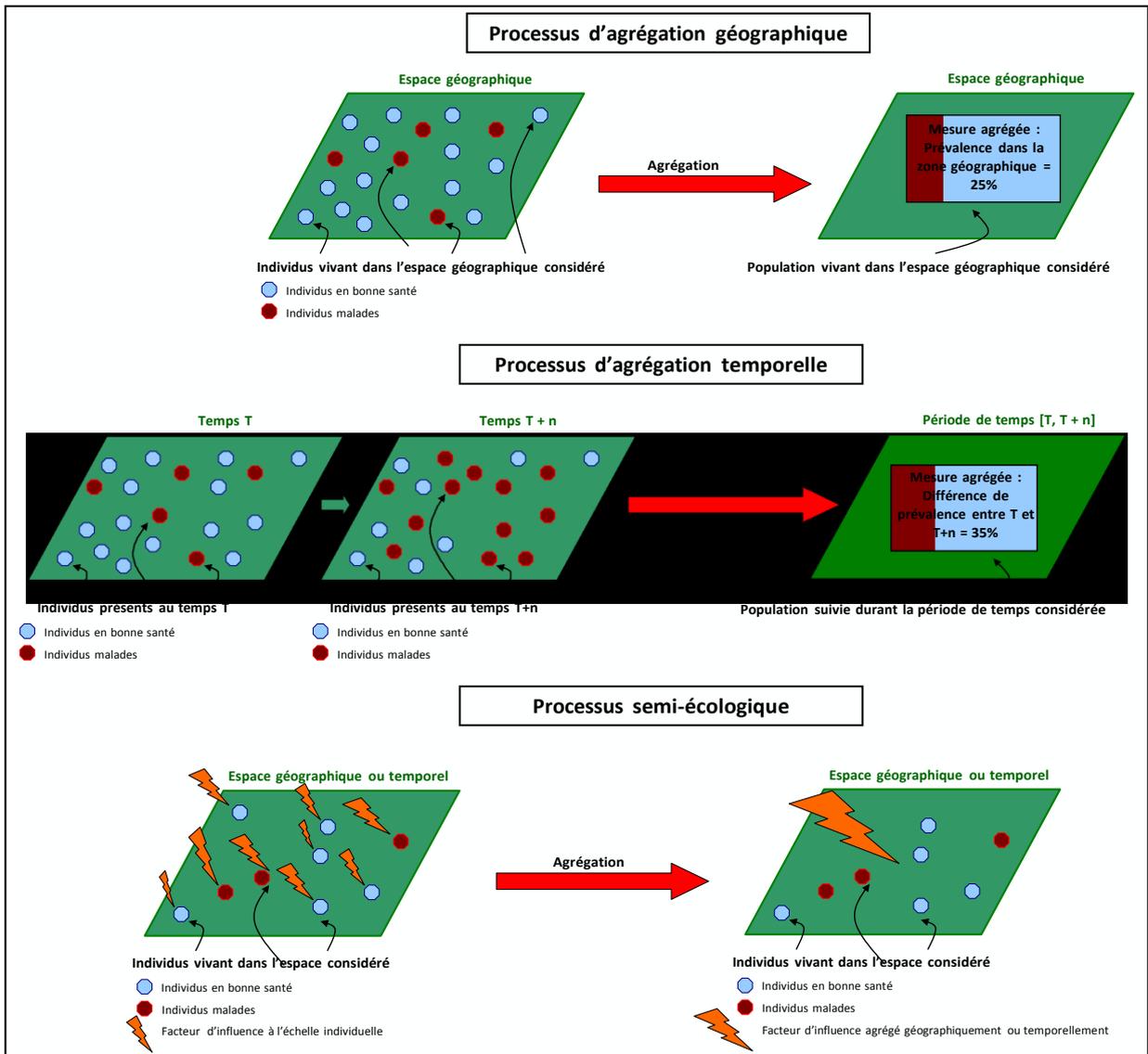
Chapitre 2 – Les méthodes écologiques

individus eux même. Dans de telles études, la variable qui a été définie à l'échelle de la population peut tout autant être une mesure agrégée résumant une caractéristique des individus d'une population (par exemple la proportion des individus de moins de 65 ans), une mesure environnementale qui décrit la localisation géographique dans laquelle la population d'étude vit ou travaille (par exemple une mesure de la pollution de l'air), ou bien encore une mesure écologique qui n'a pas « d'équivalent » à l'échelle individuelle (par exemple la densité de population ou l'existence d'une loi juridique ou d'un système de santé spécifique).

Le niveau d'agrégation peut être défini géographiquement, temporellement ou par une combinaison des deux (Figure 9). Des schémas d'étude mixtes ou à plusieurs niveaux peuvent aussi être construits. Dans ces derniers, les unités de comparaison peuvent inclure à la fois un niveau individuel et populationnel.

Des exemples d'agrégation géographique sont par exemple les études se focalisant sur l'association entre la pollution de l'air et la mortalité. La plupart du temps dans ce type d'études, la zone géographique utilisée pour l'agrégation est à l'échelle d'une ville et de ses environs [2].

Figure 9 : illustration des processus d'agrégation géographique, temporelle et semi-écologique



Partie 1 : Introduction générale

Les études qui identifient des groupes dans le temps comparent quant à elles l'évolution des incidences ou des prévalences dans le temps dans des populations bien définies géographiquement. Un exemple de ce type d'études est donné par les travaux de Salonen et al. [3]. Cette étude s'attachait à quantifier l'impact sanitaire dans le temps d'un programme communautaire promouvant des habitudes alimentaires équilibrées et un style de vie sain dans une province de l'Est de la Finlande.

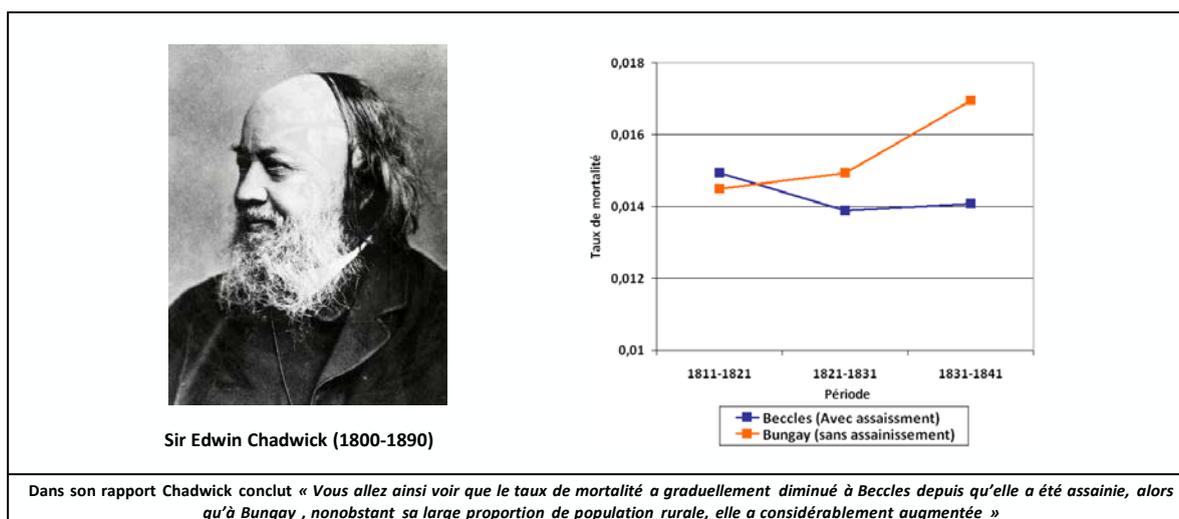
Certaines études ne peuvent être classées comme des études écologiques traditionnelles puisqu'elles mélangent une composante individuelle et une composante écologique. Ces dernières sont appelées études semi-écologiques. Un exemple de ce type d'études est le travail de Waller et al. [4] où l'influence du chlorage de l'eau potable sur les avortements spontanés était étudiée. Des données agrégées pour l'exposition étaient utilisées parallèlement aux données individuelles renseignant les avortements spontanés et les variables d'ajustement.

Pour résumer les deux points clefs qui permettent de distinguer les études écologiques des autres schémas d'étude traditionnels sont 1) un niveau d'étude à l'échelle de la population, et 2) une mesure d'exposition qui caractérise la population et non les individus. L'utilisation de ces études a une longue histoire dans beaucoup de sciences comme les sciences politiques, la géographie, la sociologie, et bien évidemment l'épidémiologie et la santé publique.

2.2. Utilisation des études écologiques en santé publique

Les études écologiques descriptives dans lesquelles les incidences des maladies ou les taux de mortalité sont comparés dans le temps ou entre plusieurs zones géographiques sont des grands classiques en épidémiologie depuis des siècles [5].

Figure 10 : Taux de mortalité comparés durant trois décennies successives dans une ville avec un système d'assainissement (Beccles) et une ville sans système d'assainissement (Bungay). Adaptée de [6].



Edwin Chadwick, réformateur social anglais du 19^{ème} Siècle investi dans l'amélioration des conditions sanitaires et de santé publique, a ainsi utilisé une approche écologique dans son célèbre rapport publié en 1842, intitulé «*Report on the sanitary condition of the labouring population of Great Britain* » [6]. L'ensemble

Chapitre 2 – Les méthodes écologiques

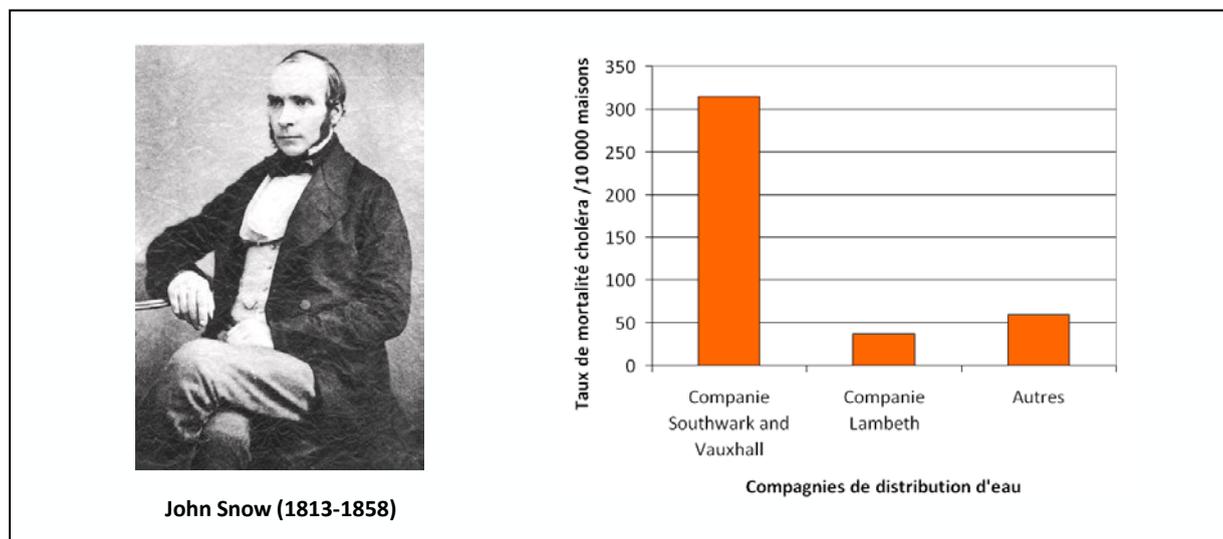
de son travail servit de justificatif à l'adoption des « lois sur les pauvres » (« poor laws »). La Figure 10 montre une partie des résultats de ce rapport dans laquelle une zone avec système d'assainissement et une sans système d'assainissement sont comparées durant 3 périodes de temps. En comparant l'évolution des taux de mortalité dans le temps des deux communautés, Chadwick fut capable d'éclaircir le lien entre assainissement et état de santé.

Très tôt dans ses recherches, un des pères de l'épidémiologie, John Snow, a utilisé une approche écologique pour comparer les taux de choléra dans les différentes maisons de Londres, et examiner si les différences observées étaient reliées aux sources d'approvisionnement en eaux [7] (Figure 11).

Au début du 20^{ème} Siècle Goldberger et al. [8] utilisèrent des schémas d'études écologiques pour démontrer que la pellagre, maladie due à la malnutrition (carences en tryptophane et en vitamine B3), n'était pas « une infection intestinale transmise à peu près de la même manière que la fièvre typhoïde », argument avancé par de nombreux médecins de l'époque. Ces résultats servirent de socle pour les études de Goldberger et Sydenstrickers qui suivirent et montrèrent que l'alimentation et les conditions socio-économiques étaient à l'origine des épidémies de pellagre sévissant dans le Sud Est des Etats-Unis [9].

Plus récemment dans le domaine des maladies chroniques, les études écologiques croisant les incidences nationales des maladies cardio-vasculaires et les prévalences de plusieurs facteurs de risque ont dessiné dans les années 80-90 les fondations de l'identification et de l'étude des causes des maladies cardio-vasculaires [10]. L'étude fondatrice de Keys et al. montra ainsi une forte corrélation entre la proportion moyenne des calories issues des acides gras saturés et la mortalité associée aux maladies coronariennes.

Figure 11 : Taux de mortalité associés au choléra par compagnies d'approvisionnement en eau potable de Londres. Adaptée de «*On the mode of communication of Cholera* » de John Snow [7].



Un autre exemple célèbre peut être donné par l'épidémiologie psychiatrique. Les études écologiques ont démontré dans ce domaine, que les incidences des psychoses transitoires aiguës variaient de manière importante en fonction du contexte socioculturel [11]. L'incidence de ces événements était jusqu'à dix fois plus importante dans les pays en voie de développement que dans les pays développés. Ces résultats ont amené

l'examen approfondi de différentes hypothèses sur les causes, en incluant les antécédents de fièvre et les comportements culturellement normés.

Ainsi, depuis l'émergence de la pratique épidémiologique, beaucoup de connaissances de santé publique ont été obtenues à l'aide des études écologiques. Malheureusement, des biais spécifiques à ces dernières existent et ont pu amener à des conclusions complètement erronées. Ces biais rassemblés sous la formule de « the ecological fallacy » sont à l'origine de la mauvaise réputation des études écologiques au sein de la communauté épidémiologique.

2.3. Les biais écologiques

Depuis leurs premières utilisations en sociologie, les méthodes écologiques ont été beaucoup critiquées du fait de leur potentialité à produire des résultats biaisés. De nombreux travaux se sont concentrés sur ces biais potentiels, spécifiques aux études écologiques, qui surgissent lors de l'agrégation des données ; un problème que Selvin [12] fut le premier à dénommer « the ecological fallacy ». Cette notion d'« ecological fallacy » correspond au raisonnement erroné qui est fait lorsque l'on extrapole au niveau individuel des résultats obtenus au niveau populationnel.

Les cas classiques de ce type de problème se retrouvent dans les situations où des données agrégées sont utilisées comme substitut à des données individuelles non disponibles. Par exemple, pour étudier dans différentes régions géographiques en l'absence de données individuelles la relation entre l'exposition à une substance X et les cancers, on peut mettre en relation l'exposition moyenne à la substance X avec l'incidence des cancers dans ces régions. Dans ce cas de figure écologique, il est impossible de savoir exactement qui est exposé à X et qui ne l'est pas. L'exposition moyenne à X est donc utilisée comme approximation grossière des expositions individuelles. Vu que l'on ne sait pas si les individus ayant développé un cancer ont été réellement exposés à X, avec ce schéma écologique on ne pourra raisonnablement pas conclure que les individus exposés à X ont un plus grand risque de contracter un cancer, même si des incidences plus fortes de cancers dans les zones où une plus grande proportion de la population exposée à X ont été retrouvées. En plus de ce problème, un manque d'information sur les autres facteurs de risque individuels ou écologiques pouvant varier d'une région à une autre peut aussi introduire un fort biais de confusion dans la relation écologique retrouvée.

Un autre exemple classique est donné par la relation entre revenu moyen et obésité quand le revenu moyen est utilisé par manque d'information au niveau individuel. Supposons que les chercheurs se focalisant sur ces problématiques trouvent qu'à l'échelle du pays des revenus moyens élevés sont associés avec une prévalence de l'obésité plus importante. Si la conclusion faite sur cette relation est que des revenus élevés sont associés avec des prévalences de l'obésité élevées, les chercheurs commettent alors une erreur écologique : la prévalence de l'obésité à l'intérieur d'un pays peut toujours être supérieure chez les individus ayant un faible revenu que chez ceux ayant un revenu élevé. De plus, les personnes vivant dans des pays à revenu moyen élevé, peuvent aussi différer grandement des personnes vivant dans des pays à revenu moyen faible au niveau des autres caractéristiques influant sur la prévalence de l'obésité.

Chapitre 2 – Les méthodes écologiques

Les sources de biais que l'on rencontre dans les études écologiques sont au nombre de 3 [13, 14]. La première se retrouve dans tous types d'études (écologiques ou pas) alors que les deux dernières sont spécifiques à ce type d'analyse :

- **Biais intra-groupe («within group bias»):** L'estimation à l'échelle individuelle de l'effet d'une variable dans une population particulière peut être biaisée par des facteurs de confusion, les méthodes de sélection utilisées ou une mauvaise classification des individus dans chacun des groupes. Si il y a un biais positif dans chacun des groupes d'étude d'une analyse écologique, il faut s'attendre à ce que les estimations écologiques résultantes le soient aussi (Figure 12).
- **Biais de spécification ou biais de confusion écologique («Confounding by group») :** Un biais de spécification peut survenir si la fréquence d'une maladie dans les populations non exposées varie d'un groupe à l'autre dans une étude écologique. De telles variations surviennent suite à une distribution hétérogène des autres facteurs de risque dans les populations d'étude. Ces autres facteurs, pouvant agir comme variables de confusion ou de modification de l'effet, ne sont pas forcément ceux mesurés à l'échelle individuelle (biais de confusion intra-groupes)
- **Biais dus à la modification de l'effet par le groupe (en anglais «Effect modification by group») :** Des biais écologiques peuvent aussi apparaître quand le groupe est un effet modifiant significatif de l'effet de l'exposition. Dans ce cas de figure, l'effet de l'exposition est susceptible de varier en fonction d'une ou plusieurs covariables. Ces covariables introduisent alors un terme d'interaction statistique important avec la mesure d'exposition et biaise ainsi les résultats de l'étude (Figure 12).

Les expressions « biais écologiques » ou « biais d'agrégation » font référence à l'ensemble des biais de spécification ainsi que des biais dus à la modification de l'effet par le groupe. En plus des sources classiques de biais, le principal problème des études écologiques est que l'hétérogénéité des niveaux d'exposition et des niveaux des covariables dans les différents groupes étudiés peut être importante. Du fait de l'agrégation, cette hétérogénéité n'est pas complètement capturée dans les études écologiques. La magnitude de ces biais peut être très grande en pratique [13, 14].

L'exemple le plus célèbre d'étude à partir de laquelle les biais écologiques ont été discutés, est celle explorant les taux de suicides des provinces de Prusse faite par Durkheim, un sociologue français du 19^{ème} siècle (Figure 13). Dans cette étude, il concluait que les taux de suicides étaient substantiellement plus importants dans les provinces où la population était en majorité protestante. Il a été montré plus tard qu'un fort biais écologique pouvait tout aussi bien expliquer ces résultats [15].

Partie 1 : Introduction générale

Figure 12 : Illustration des sources de biais que l'on peut rencontrer dans les études écologiques

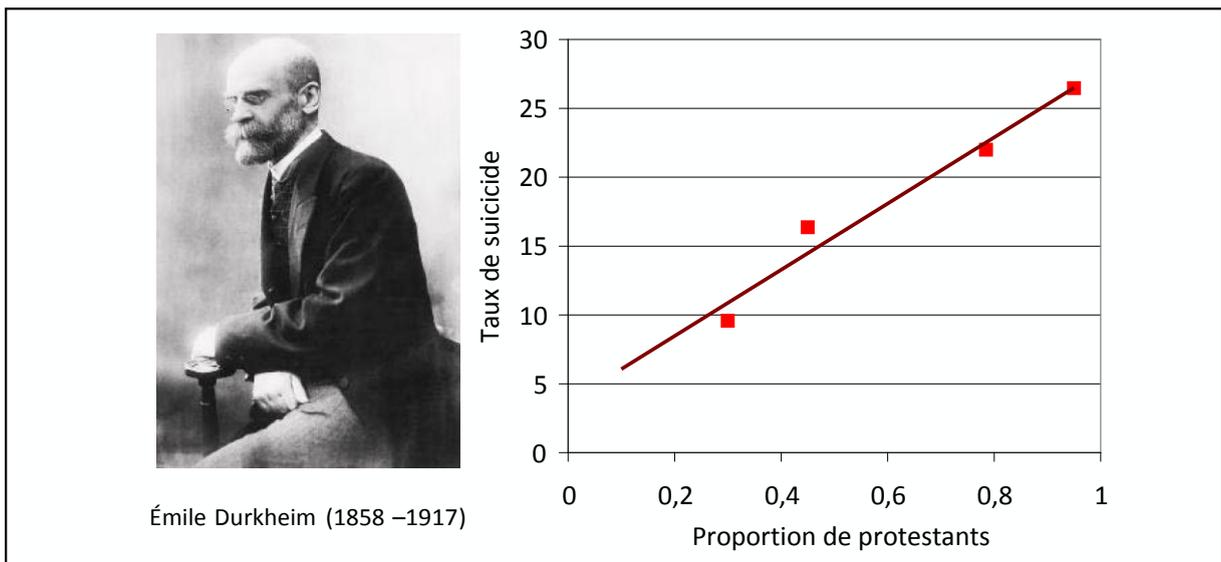
<p>A)</p> <p style="text-align: center;">Étude écologique sans biais</p> <p>Individus vivant dans les populations considérées</p> <ul style="list-style-type: none"> ● Individus en bonne santé ● Individus malades ⚡ Facteur d'influence externe <p style="text-align: center;">résultats</p> <p>Prévalence de la maladie (%)</p> <p style="text-align: center;">Intensité du facteur externe</p>	<p>SCHEMATISATION D'UNE ETUDE ECOLOGIQUE SANS BIAIS</p> <p>L'étude écologique fictive représentée ici cherche à montrer l'influence d'un facteur externe sur la prévalence d'une maladie à l'aide de 4 populations différentes.</p>
<p>B)</p> <p style="text-align: center;">Étude écologique faite à partir d'échantillonnages</p> <p style="color: red;">Biais de sélection dans chacune des populations étudiées</p> <p>Individus vivant dans les populations considérées</p> <ul style="list-style-type: none"> ● Individus en bonne santé ● Individus malades ⚡ Facteur d'influence externe <p style="text-align: center;">résultats</p> <p>Prévalence de la maladie (%)</p> <p style="text-align: center;">Intensité du facteur externe</p>	<p>SCHEMATISATION DE LA NOTION DE BIAIS INTRAGROUPE :</p> <p>Du fait d'un fort biais de sélection des individus dans chacune des populations échantillonnées, la relation trouvée par l'étude écologique basée sur l'échantillonnage des populations est inverse à celle identifiée en populations entières (schéma du haut).</p>
<p>C)</p> <p style="text-align: center;">Présence d'un facteur externe de confusion</p> <p style="color: red;">Influence du facteur 2 non prise en compte dans l'analyse</p> <p>Individus vivant dans les populations considérées</p> <ul style="list-style-type: none"> ● Individus en bonne santé ● Individus malades ⚡ Facteur d'influence externe 1 ⚡ Facteur d'influence externe 2 <p style="text-align: center;">résultats</p> <p>Prévalence de la maladie (%)</p> <p style="text-align: center;">Intensité du facteur externe 1</p>	<p>SCHEMATISATION DE LA NOTION DE BIAIS DE CONFUSION ECOLOGIQUE :</p> <p>Du fait d'un fort biais de confusion écologique du à la présence d'un 2^{ème} facteur d'influence non prise en compte dans l'analyse, la relation trouvée dans ce cas de figure est différente de celle trouvée dans le premier cas de figure (schéma du haut) où le facteur d'influence 2 est absent.</p>
<p>D)</p> <p style="text-align: center;">Présence d'un biais du à la modification de l'effet par le groupe</p> <p style="color: red;">Influence du groupe sur le facteur externe non prise en compte dans l'analyse</p> <p>Individus vivant dans les populations considérées</p> <ul style="list-style-type: none"> ● Individus en bonne santé ● Individus malades ⚡ Facteur d'influence externe ↪ Influence du groupe sur l'effet du facteur <p style="text-align: center;">Résultats</p> <p>Prévalence de la maladie (%)</p> <p style="text-align: center;">Intensité du facteur externe</p>	<p>SCHEMATISATION DE LA NOTION DE BIAIS DU A LA MODIFICATION DE L'EFFET PAR LE GROUPE :</p> <p>Du fait d'un fort biais du à la modification de l'effet par le groupe non pris en compte dans l'analyse, la relation trouvée dans ce cas de figure est différente de celle trouvée dans le premier cas de figure où ce biais est absent (schéma du haut).</p>

Chapitre 2 – Les méthodes écologiques

Cette différence pouvait aussi s'expliquer de la manière suivante : les provinces n'étant pas exclusivement protestantes ou catholiques, les taux de suicide élevés pourraient être dus aux minorités religieuses (principalement les catholiques) vivant dans ces provinces. Ces derniers auraient un taux de suicide important dans les provinces à majorité protestante. Dans une région donnée, il est tout aussi plausible que les taux de suicide dans une minorité religieuse soient plus importants que ceux de la religion majoritaire. Vivre dans un environnement majoritairement protestant aurait finalement un effet contextuel sur les taux de suicide des communautés catholiques.

De même, l'étude réalisée par Robinson en 1950 [16] est un autre exemple historique où les biais écologiques ont été discutés. Dans cette étude, Robinson s'attachait à décrire le pourcentage d'illettrisme en fonction du pourcentage d'immigrés. En calculant ces 2 indicateurs pour chacun des 48 états des Etats Unis durant l'année 1930, ce dernier trouvait que leur corrélation était de 0,53. Cette corrélation suggérait que la population immigrée avait moins tendance à être illettrée que la population native.

Figure 13 : Taux de suicide / 10^5 personnes / année en fonction de la proportion protestante de la population pour quatre groupes de provinces prussiennes, 1883-1890. Adaptée de l'ouvrage de Durkheim [15].



En réalité cette association était négative : la corrélation à l'échelle individuelle donnait -0,11 (la population immigrante était plus illettrée que celle native). Cette mauvaise inférence était due au fait que les immigrants avaient plus tendance à s'installer dans les états où les personnes natives étaient relativement plus lettrées. Du fait d'un fort biais écologique, les corrélations étaient donc totalement inversées dans cette étude.

D'autres exemples plus récents d'études avec de forts biais écologiques, sont ceux décrivant l'évolution temporelle des incidences ou des prévalences nationales de la plupart des maladies consécutives à «l'occidentalisation» des populations (comme le diabète de type 2 ou l'obésité). Ces évolutions sont très fortement corrélées aux ventes nationales de postes de télévision ainsi qu'à celles du pourcentage de la population ayant une connexion internet. Cette forte corrélation est en fait un biais écologique typique et s'explique en réalité par des variables de confusion ayant la même évolution temporelle que les ventes de

postes de télévision (dans le cas du diabète de type 2 et de l'obésité ces variables sont principalement la transformation des habitudes alimentaires ainsi que la sédentarisation).

2.4. Mais alors pourquoi faire des études écologiques ?

Couramment la perception des études écologiques par la communauté médicale est qu'elles sont des méthodes très biaisées mais faciles à mettre en place, peu onéreuses, et rapides pour étudier une problématique de santé. Ainsi avec l'avènement de l'« *evidence based medicine* » ces dernières sont souvent considérées comme une relique de la phase pré moderne de l'épidémiologie. Elles découleraient de pratiques utilisées avant que l'épidémiologie n'obtienne ses lettres de noblesse en se fondant sur un socle méthodologiquement solide se basant sur le paradigme de la théorie des essais contrôlés randomisés ainsi que sur celui des études de cohorte. Il est courant que les formations universitaires en santé publique soient souvent restreintes à enseigner la méthodologie des essais cliniques et des études de cohortes, en oubliant les autres types d'études du fait de leur plus « faible causalité ».

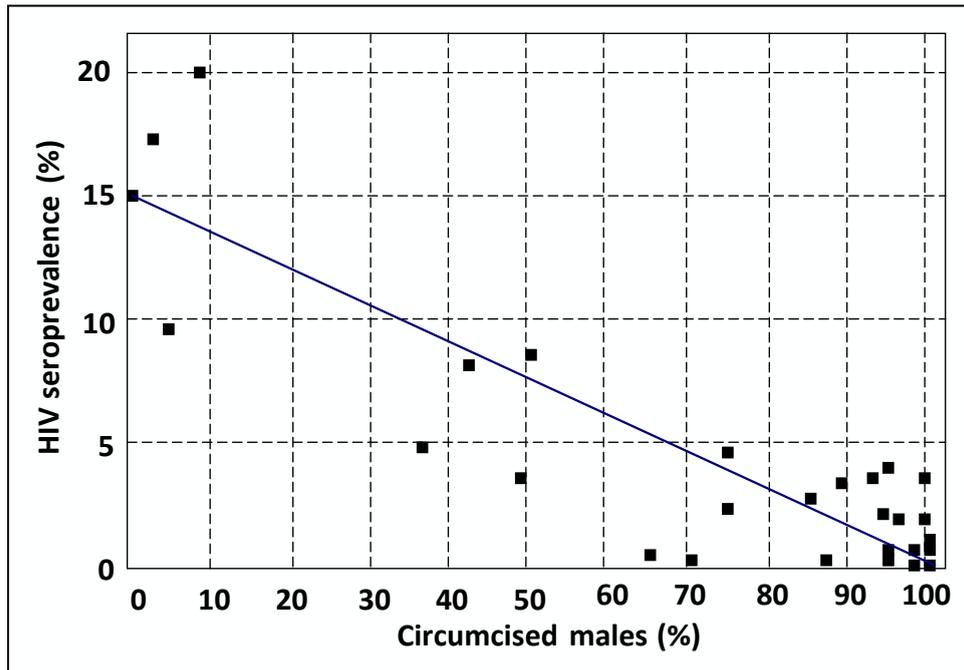
Malgré ce désintéressement de la communauté scientifique, les études écologiques ont cependant un intérêt scientifique non négligeable et des avantages très précieux dans certaines situations par rapport aux méthodes usuelles utilisées en épidémiologie. Ces avantages sont décrits dans les paragraphes suivants.

2.4.1. Facilité, rapidité et faible coût d'une étude

Les études écologiques sont peu onéreuses et nécessitent peu de temps pour leur réalisation. En effet, elles reposent sur de nombreuses sources de données secondaires, chacune impliquant différents types d'informations nécessaires pour l'analyse, et qui peuvent être aisément reliées au niveau populationnel. Par exemple, les données issues de différents recensements, enquêtes, registres nationaux, ou bases de données administratives sont assez facilement accessibles à la communauté scientifique et peuvent être reliées facilement à un niveau d'agrégation à l'échelle du pays ou des départements. Cet avantage certain des études écologiques par rapport aux autres schémas d'étude est la clé d'un de leurs rôles les plus importants en épidémiologie : celui de générer ou de tester des hypothèses de recherche nouvelles. Dans cette configuration, le rôle des études écologiques va être de soulever et d'amener des nouvelles hypothèses qui pourront dans un deuxième temps être examinées et testées plus profondément par les autres types d'études épidémiologiques.

Un exemple historique du rôle des études écologiques comme génératrices de nouveaux axes de recherche et qui déboucha sur des implications majeures en santé publique est celui du rôle protecteur de la circoncision dans la transmission du VIH des pays d'Afrique mis en évidence par Bongaarts en 1989 [17]. Dans cette étude, la comparaison nationale des taux de circoncision et des prévalences VIH mettait en évidence que les fortes prévalences de VIH étaient associées à de faibles taux de circoncision et vice versa. Les sceptiques de l'époque déclarèrent (à raison) que de nombreux autres facteurs non pris en compte dans l'analyse pouvaient faire varier les prévalence VIH nationales. Pour trancher, des études à l'échelle individuelle se montèrent. L'hypothèse était lancée ...

Figure 14 : Estimations des séroprévalences dans les capitales de 37 pays d'Afrique en fonction du pourcentage d'hommes circoncis. Adaptée de [17].



2.4.2. Limitations de mesures dans les études individuelles

Il est souvent impossible de mesurer des expositions ou des doses d'un déterminant à l'échelle de l'individu pour un grand nombre de sujets, tout du moins pas dans des échelles de temps et de budget raisonnables. Ainsi dans certains cas, les études écologiques représentent la seule alternative possible pour étudier l'effet d'une exposition sur la santé. La nature même de beaucoup d'expositions (climat, pollutions environnementales, nutrition,...) rend difficile à mettre en place tout autre schéma que des études écologiques. Pour la constitution d'une étude de cohorte sur l'exposition d'un polluant atmosphérique, il faudrait par exemple imaginer de trouver, puis de suivre pendant plusieurs dizaines d'années, de nombreux individus, contrastés sur leur exposition, pour lesquels on disposerait d'une part de mesures personnelles d'exposition grâce à des dosimètres individuels et d'autre part du relevé de l'ensemble des autres facteurs de risque potentiels. Cela n'est pas réaliste.

L'étude écologique faite en 1994 par Cohn et ses collaborateurs [18] est une bonne illustration du rôle nécessaire des études écologiques en santé environnementale et de leur utilité dans ce domaine. Les auteurs ont comparé, dans 75 municipalités de l'état du New Jersey, l'incidence des leucémies et des lymphomes non hodgkinien (déterminée à partir du registre de cancers de l'état) aux teneurs en 4 trihalomethanes et en 14 composés organiques volatils (COVs) mesurées dans toutes les ressources en eau approvisionnant les villes considérées. Ils ont ainsi trouvé que Les femmes vivant dans les agglomérations caractérisées par les plus fortes teneurs en COVs totaux présentaient, par référence aux « non-exposées » un risque relatif de lymphomes, ajusté sur l'âge, bien supérieur à celles vivant dans les agglomérations caractérisées par de faibles teneurs.

Partie 1 : Introduction générale

Tableau 1 : Nombre de cas déclarés, risques relatifs ajustés sur l'âge et intervalle de confiance à 95% pour les lymphomes non hodgkiniens (NHL) dans l'état du New Jersey, 1979-1987 en fonction de l'exposition aux composés organiques volatils (« TCE exposure ») et le sexe. Adapté de Cohn et collaborateurs [18].

TCE exposure (ppb)	Cases		RR (95% CI)	
	Males	Females	Males	Females
Total NHL				
0	487	509	1.0	1.0
0.1-5.0	235	187	1.25 (1.07-1.46)	0.95 (0.81-1.13)
>5.0	119	121	1.10 (0.90-1.35)	1.08 (0.89-1.32)
<i>Total</i>	841	817		
Intermediate grade NHL: Diffuse large cell/reticulosarcoma				
0	129	116	1.0	1.0
0.1-5.0	61	39	1.23 (0.91-1.67)	0.86 (0.60-1.24)
>5.0	26	31	0.91 (0.60-1.39)	1.21 (0.82-1.80)
<i>Total</i>	216	186		
High Grade NHL: Total				
0	23	19	1.0	1.0
0.1-5.0	9	5	1.11 (0.51-2.41)	0.71 (0.26-1.89)
>5.0	2	11	0.41 (0.09-1.76)	2.66 (1.27-5.60)
<i>Total</i>	34	35		
High Grade NHL: Non-Burkitt's				
0	15	15	1.0	1.0
0.1-5.0	7	3	1.26 (0.51-3.09)	0.53 (0.15-1.82)
>5.0	2	9	0.61 (0.14-2.65)	2.74 (1.20-6.26)
<i>Total</i>	24	27		

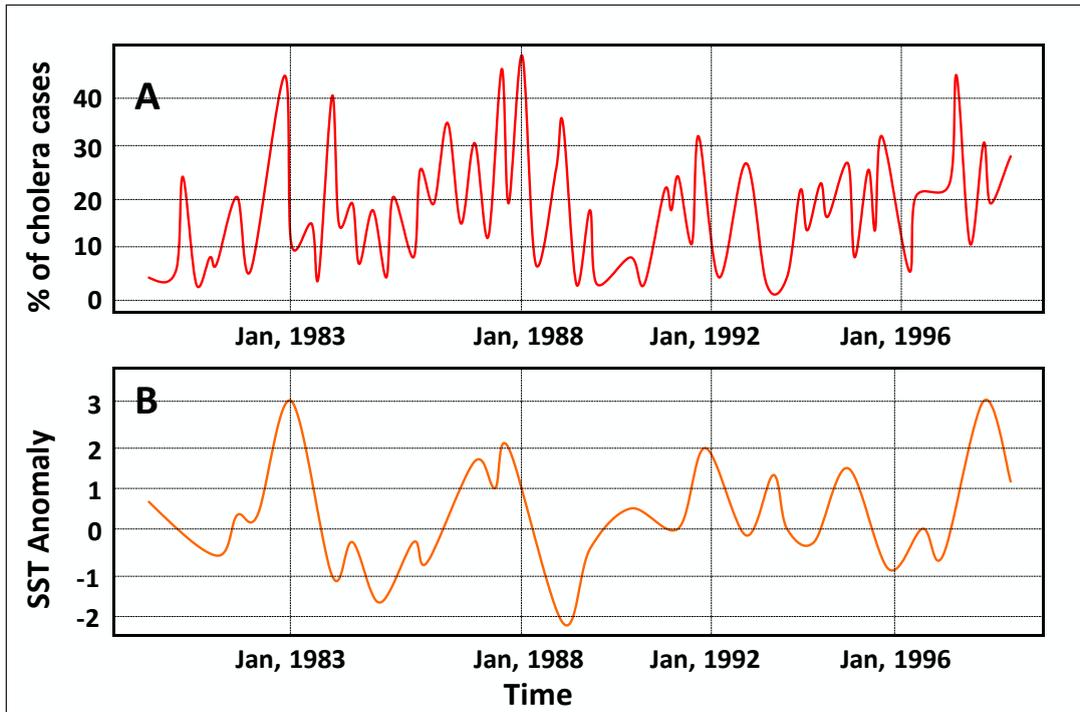
2.4.3. Etude de paramètres populationnels n'existant pas à l'échelle individuelle

Un sujet de recherche peut porter sur l'étude d'un paramètre qui n'a pas d'équivalent à l'échelle individuelle. Le niveau d'inférence est dans ce type d'études écologique et non biologique. Plus généralement, les études écologiques peuvent être utilisées pour comprendre comment le contexte influence la santé des personnes et des groupes à travers des phénomènes comme la sélection, la distribution, les interactions, l'adaptation et d'autres types de réponses qui n'ont pas d'équivalent individuel. Par exemple, ces études sont particulièrement adéquates pour évaluer l'impact d'interventions populationnelles comme les campagnes de vaccinations, la promotion d'une nouvelle loi ou celui de paramètres socioéconomiques définis au niveau de la population (comme la densité de populations par exemple). De même, dans le cadre des maladies infectieuses, elles peuvent se révéler particulièrement précieuses pour relier la maladie à des paramètres, extérieurs à l'individu, qui influencent la transmission des agents pathogènes et leur dynamique.

Le cholera est un parfait exemple pour illustrer ce concept et montrer comment des variables écologiques modèlent non seulement sa transmission mais aussi sa virulence et son épidémiologie. En vue de déterminer la source et l'hôte du vibron du choléra jusqu'alors très peu connue, d'importantes études environnementales ont été réalisées au Bangladesh entre 1987 et 1990 [19]. Des échantillons d'eaux ainsi que des différents types de planctons présents dans ces dernières ont été collectés dans la région située au niveau du delta formé par le Meghna et le Ganges. Les résultats ont montré que l'abondance de *V.cholerae* dans ces eaux augmentait avec l'abondance d'un type particulier de plancton. A partir de ces résultats, une étude écologique montra ensuite que la saisonnalité des épidémies de choléra au Bangladesh suivait une dynamique

parallèle à celle du plancton et qu'un modèle des épidémies utilisant des variables environnementales globales comme les températures de surface des mers et d'autres caractéristiques hydrologiques, pouvait être mis en place [19, 20](Figure 15).

Figure 15 : (A) Série temporelle des cas de choléra observés au ICDDR, B (International Centre for Diarrhoeal Disease Research, Bangladesh, in Dhaka) compilée à partir des patients visitant le centre. (B) Série temporelle du phénomène El Niño mesurée à partir de l'index El Niño 3.4. Adaptée de Pascual et al. [20].

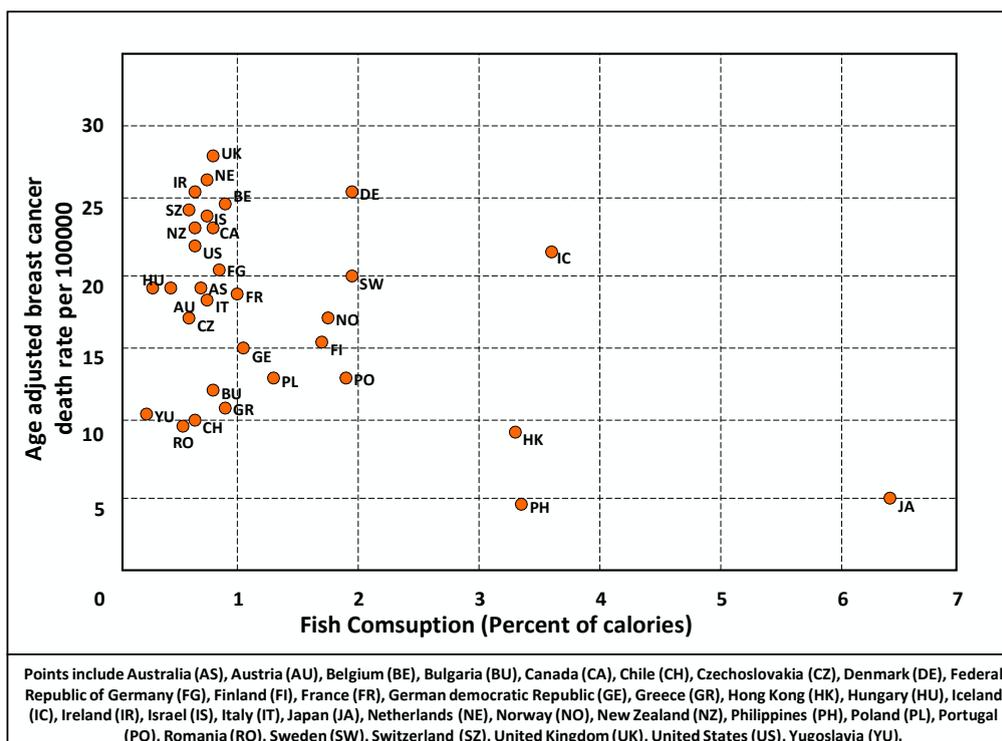


2.4.4. Limitations méthodologiques des études individuelles

Les études individuelles peuvent ne pas être adaptées pour estimer l'effet d'un paramètre sur une pathologie si ce dernier varie très peu dans la zone ou dans la population d'étude. Un schéma d'étude écologique permet de résoudre assez facilement ce type de limitations en incluant des zones géographiques plus larges ou une taille de population plus importante.

Par exemple, les habitudes alimentaires étant relativement homogènes dans certaines régions, il peut être intéressant de s'intéresser à une zone géographique plus large pour obtenir une « diversité » plus importante des pratiques alimentaires. Les études écologiques sont une réponse astucieuse à ce type de problème. En 1989, Kaiser et al [21] utilisèrent cette démarche pour étudier le rôle de la consommation de poisson sur l'incidence et la mortalité associées aux cancers du sein à l'échelle nationale. En incluant des pays ayant des profils de consommation radicalement différents, cette étude montrait une relation inverse entre le pourcentage de calories attribuables à la consommation de poissons et l'incidence et la mortalité associée au cancer du sein. Ce résultat suggérait que les acides gras de type oméga 3 contenus dans certains poissons pouvaient avoir un effet protecteur contre le cancer.

Figure 16 : Graphique de la mortalité associée au cancer du sein (pour 100000 personnes) en fonction de la consommation de poissons (pourcentage de l'apport en calories) 1974-1975. Adaptée de Kaiser et al [21].



2.5. Les biais en relation avec l'existence de multiples niveaux d'analyse : pourquoi ne parlons-nous que des biais écologiques ?

Durant la dernière décennie, la volonté d'étudier les potentiels déterminants de la santé agissant au niveau d'un groupe, ont amené les épidémiologistes à reconsidérer l'utilité et les limitations des études écologiques [22]. Dans ce contexte, le fait que des variables définies à l'échelle du groupe pouvaient fournir des informations qui ne sont pas toujours capturées au niveau individuel est apparu. Plus généralement, la reconnaissance des multiples niveaux d'organisations (par exemple du niveau moléculaire au niveau de la société) et la nécessité d'étudier les déterminants de la santé spécifiquement dans chacun de ces niveaux ont lentement émergées. La plupart des questions conceptuelles ou analytiques qui sont apparues alors dans le domaine des études écologiques dérivait en fait de l'existence de cette organisation des données sur plusieurs niveaux.

Les biais écologiques peuvent être vus comme un sous ensemble de biais dérivant de cette organisation particulière de l'information. Parce que dans les dernières années les épidémiologistes se sont principalement préoccupés de déduire des inférences sur la variabilité interindividuelle, les biais écologiques ont été exposés plus intensément sous les feux des projecteurs que leurs homologues : les biais atomistiques. L'effet d'une variable individuelle n'est pas le même sur un paramètre d'étude mesuré à l'échelle individuelle que sur son équivalent mesuré à l'échelle populationnelle. Ainsi les utilisations d'associations retrouvées à l'échelle individuelle pour faire des déductions à l'échelle populationnelle peuvent tout aussi bien amener à des conclusions erronées. Autant les biais écologiques que ceux atomistiques peuvent être finalement pensés

comme des problèmes méthodologiques inhérents à la déduction de résultats à un niveau d'échelle particulier quand les données ont été recueillies à un autre niveau d'analyse. Un exemple commun de biais atomistique peut se trouver dans les études faites en sciences politiques se focalisant sur le sondage d'opinion et le suffrage universel. Dans ce cas de figure, l'erreur la plus commune est de croire que l'opinion réelle d'une population se réduit à la simple addition des opinions individuelles.

Ainsi dans le domaine de la recherche publique, autant les variables à expliquer qu'explicatives peuvent être conceptualisées à différents niveaux d'organisation, et comprendre une variable à un niveau particulier peut requérir de prendre en compte des informations définies à un niveau d'organisation inférieur ou supérieur. Chaque système peut être vu comme niché dans un autre et dynamiquement relié à ceux au-dessus ou en dessous de lui. Dans cette configuration, chaque système possède des caractéristiques uniques qui sont différentes de la somme de ses composants. La sélection du schéma d'étude approprié doit se baser sur les questions scientifiques étudiées en incluant le niveau d'organisation sur lequel les hypothèses sont formulées ainsi que le niveau d'organisation de l'ensemble des variables utilisées dans l'analyse (cette notion sera développée plus en détails dans le chapitre 5).

2.6. Conclusions et synthèse sur les études écologiques

Malgré des biais spécifiques inhérents aux études écologiques (the « ecological fallacy »), ces dernières, en utilisant des données agrégées plus faciles à obtenir que des données individuelles, peuvent être d'une utilité précieuse en épidémiologie. Elles représentent un moyen unique d'études et de génération d'hypothèses que les épidémiologistes doivent continuer à exploiter et développer. Leur faiblesse d'imputation causale comparativement aux autres schémas d'études classiques, ne doit en rien cacher leurs autres avantages en épidémiologie. Nécessaires en épidémiologie environnementale du fait des contraintes méthodologiques, elles peuvent aussi permettre dans d'autres domaines d'étude de se focaliser sur des problématiques de santé publique à une échelle populationnelle. L'éclairage historique de l'utilisation de ces méthodes en est une illustration frappante. Plus spécifiquement, La diffusion de cette démarche écologique dans la communauté scientifique est particulièrement stratégique dans le domaine des maladies infectieuses où la dynamique des pathogènes fait intervenir un réseau complexe de déterminants agissant à des échelles diverses. C'est ce que nous allons voir dans le chapitre suivant.

Bibliographie

1. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. **Clinical Epidemiology, a Basic Science for Clinical Medicine**. Boston: Brown and Co; 1985.
2. Larrieu S, Jusot JF, Blanchard M, Prouvost H, Declercq C, Fabre P, *et al.* **Short term effects of air pollution on hospitalizations for cardiovascular diseases in eight French cities: the PSAS program**. *Sci Total Environ* 2007,387:105-112.
3. Salonen JT, Puska P, Mustaniemi H. **Changes in morbidity and mortality during comprehensive community programme to control cardiovascular diseases during 1972-7 in North Karelia**. *Br Med J* 1979,2:1178-1183.
4. Waller K, Swan SH, DeLorenze G, Hopkins B. **Trihalomethanes in drinking water and spontaneous abortion**. *Epidemiology* 1998,9:134-140.
5. Susser M. **Causal thinking in the health sciences**. New York: Oxford University Press; 1973.
6. Chadwick. **Report on the sanitary condition of the labouring population of Great Britain (originally published in 1842)**. Edimburgh: Edimburgh university Press; 1965.
7. Snow J. **On cholera (originally published in 1855)**. New York: Commonwealth Fund; 1936.
8. Goldberger J, Wheeler GA, Sydenstricker E, Tarbett RE. **A study of the relation of factors of a sanitary character to pellagra incidence in seven cotton-mill vilages of South Carolina in 1916**. *Public Health Report* 1920,35:1701-1724.
9. Terris M. **Goldberger on pellagra**: Louisiana State University Press; 1964.
10. Keys A. **Seven countries : a multivariate analysis of death and coronary heart disease**. Cambridge: Harvard University Press; 1980.
11. Susser E, Wanderling J. **Epidemiology of nonaffective acute remitting psychosis vs schizophrenia. Sex and sociocultural setting**. *Arch Gen Psychiatry* 1994,51:294-301.
12. Selvin H. **Durkheim's suicide and problems of empirical research**. *Am. J. Sociol.* 1958,63:607-619.
13. Morgenstern H. **Ecologic studies in epidemiology: concepts, principles, and methods**. *Annu Rev Public Health* 1995,16:61-81.
14. Wakefield J. **Ecologic studies revisited**. *Annu Rev Public Health* 2008,29:75-90.
15. Durkheim E. **Le suicide**. Paris: F. Alcan; 1897.
16. Robinson WS. **Ecological correlations and the behavior of individuals**. *American Sociological Review* 1950,15:351-357.
17. Bongaarts J, Reining P, Way P, Conant F. **The relationship between male circumcision and HIV infection in African populations**. *AIDS* 1989,3:373-377.
18. Cohn P, Klotz J, Bove F, Berkowitz M, Fagliano J. **Drinking Water Contamination and the Incidence of Leukemia and Non-Hodgkin's Lymphoma**. *Environ Health Perspect* 1994,102:556-561.
19. Colwell RR. **Global climate and infectious disease: the cholera paradigm**. *Science* 1996,274:2025-2031.

Chapitre 2 – Les méthodes écologiques

20. Pascual M, Rodo X, Ellner SP, Colwell R, Bouma MJ. **Cholera dynamics and El Nino-Southern Oscillation.** *Science* 2000,289:1766-1769.
21. Kaizer L, Boyd NF, Kriukov V, Tritchler D. **Fish consumption and breast cancer risk: an ecological study.** *Nutr Cancer* 1989,12:61-68.
22. Diez-Roux AV. **Bringing context back into epidemiology: variables and fallacies in multilevel analysis.** *Am J Public Health* 1998,88:216-222.

Chapitre 3 - Les maladies infectieuses

« Les maladies infectieuses : il en naîtra de nouvelles, il en disparaîtra lentement quelque unes ; celles qui subsisteront ne se montreront plus sous les formes que nous leur connaissons aujourd'hui »

Charles Nicolle

Le temps où l'on pensait que les maladies infectieuses seraient maîtrisées grâce à la généralisation des mesures d'hygiène, l'utilisation des antibiotiques et des vaccins n'est pas si éloigné. Les progrès scientifiques et technologiques laissent croire à une possible éradication de ces dernières (le cas de la variole à la fin des années 1970 avec la vaccination généralisée en étant le symbole). Cet espoir a malheureusement été déçu. On assiste à une résurgence des maladies infectieuses, des parasitoses et à l'émergence régulière de nouveaux agents infectieux [1]. Les faits sont éloquentes : permanence de l'endémie dans les pays en voie de développement, risque de pathologies d'importation dû à l'explosion des voyages intercontinentaux et à la globalisation du commerce, résistance des microbes aux antibiotiques, sida et hépatites B et C devenus endémiques, infections d'origine alimentaire, infections acquises en milieu hospitalier, menace de catastrophes économiques dues à la résurgence d'épizooties, risque croissant de transmission à l'homme de zoonoses, risque plausible du bioterrorisme, impact du réchauffement global de la planète sur les agents infectieux, leurs réservoirs et leurs vecteurs.

La mise à mal de ce grand espoir de la communauté scientifique du 20^{ème} lui aura au moins permis de prendre conscience de l'extrême complexité opérant dans la circulation des agents infectieux et des enjeux stratégiques de recherche autour de ces derniers.

3.1. Epidémiologie des maladies infectieuses

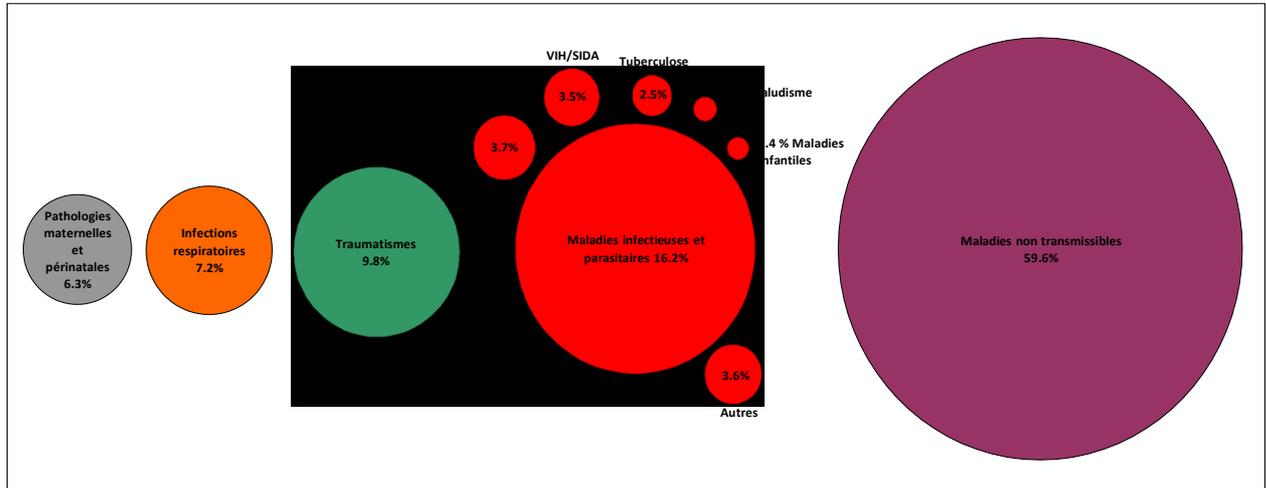
Avec une mortalité de près de 15 millions chaque année, les maladies infectieuses et parasitaires sont responsables de 23,4 % des décès causés par l'ensemble des maladies et des traumatismes survenant sur la planète [2] (Figure 17).

Les principaux types d'infections responsables de décès sont les infections respiratoires aiguës (4,3 millions par an), le sida (2,0 millions par an), les maladies diarrhéiques (2,2 millions par an), la tuberculose (1,5 million par an) et le paludisme (0,9 million par an). Plus de 90 % des maladies infectieuses humaines surviennent dans les pays en voie de développement, particulièrement chez les enfants, où l'hygiène est insuffisante et où les politiques de prévention sont inexistantes, inadaptées ou insuffisamment financées.

Le développement industriel génère aussi dans les sociétés de nouvelles conditions d'émergence infectieuse, comme les infections alimentaires par des agents prenant avantage de la chaîne du froid ou de l'industrialisation de la chaîne alimentaire, les infections nosocomiales survenant dans un environnement hospitalier de plus en plus complexe, alors que la multirésistance va croissant, les infections opportunistes chez les patients immuno-déprimés et les infections des voyageurs, qu'illustrent les 7 000 cas annuels de paludisme

observés en France. Bien que de moindre prévalence dans les pays industrialisés, les maladies infectieuses y sont encore responsables d'une mortalité non négligeable. En France par exemple, toutes causes confondues, 66 000 décès annuels peuvent être attribués à des maladies infectieuses, soit 12 % de la totalité des décès [3].

Figure 17 : Répartition des causes de mortalité dans le monde en 2004 [2]



3.2. Les phénomènes d'émergence et de réémergence

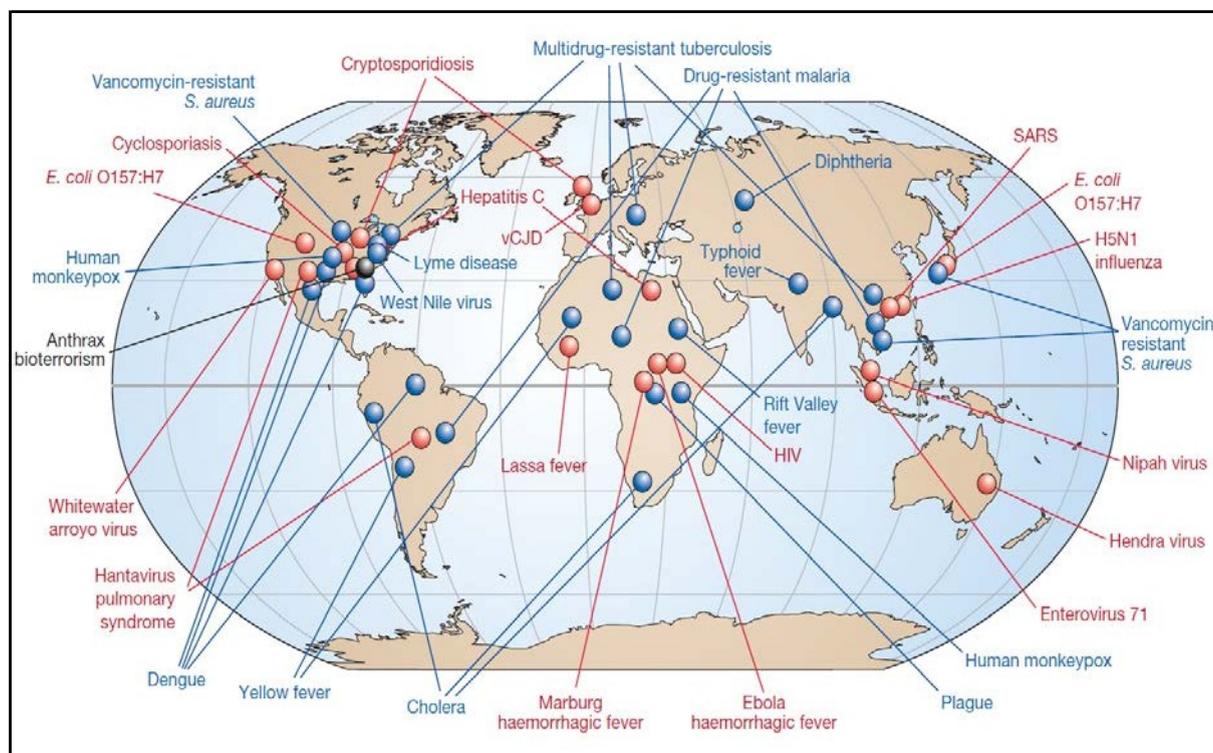
Sur cette toile de fond se superposent des phénomènes d'émergence et de réémergence. Il peut s'agir de maladies infectieuses classiques mais réapparaissant sous une forme différente ou gagnant des régions géographiques nouvelles, souvent plus sévères, du fait de la multirésistance de l'agent infectieux, comme dans le cas de la tuberculose. Il peut aussi s'agir de maladies réellement nouvelles, d'étiologie auparavant inconnue, qui explosent du fait de conditions particulières changeantes. Une série d'exemples récents illustre le rôle que joue dans l'émergence la transmission d'infections animales à l'homme suite au franchissement de la barrière d'espèces : grippe aviaire, SRAS, nouveau variant de la maladie de Creutzfeld-Jacob, virus Nipah de la chauve-souris. S'y ajoutent des virus ayant émergé précédemment, comme le virus Ebola, le virus Marburg et le VIH (Figure 18).

3.3. Cancers et maladies infectieuses

Il ne faut pas oublier qu'un pourcentage important de cancers (de 15 à 20 %) est causé par un agent infectieux, viral ou bactérien, représentant chaque année au moins 1,7 million de nouveaux cas et plus d'un million de décès. Les cancers du foie, du col utérin et de l'estomac pourraient être quasi éradiqués par la mise au point ou l'utilisation de vaccins, respectivement contre les virus des hépatites B et C, certains papillomavirus et *Helicobacter pylori* [4]. Il est par ailleurs probable que des infections sont directement ou indirectement impliquées dans l'étiologie de maladies touchant une large fraction de la population du globe, telles que l'athérosclérose, le diabète, l'allergie.

Partie 1 : Introduction générale

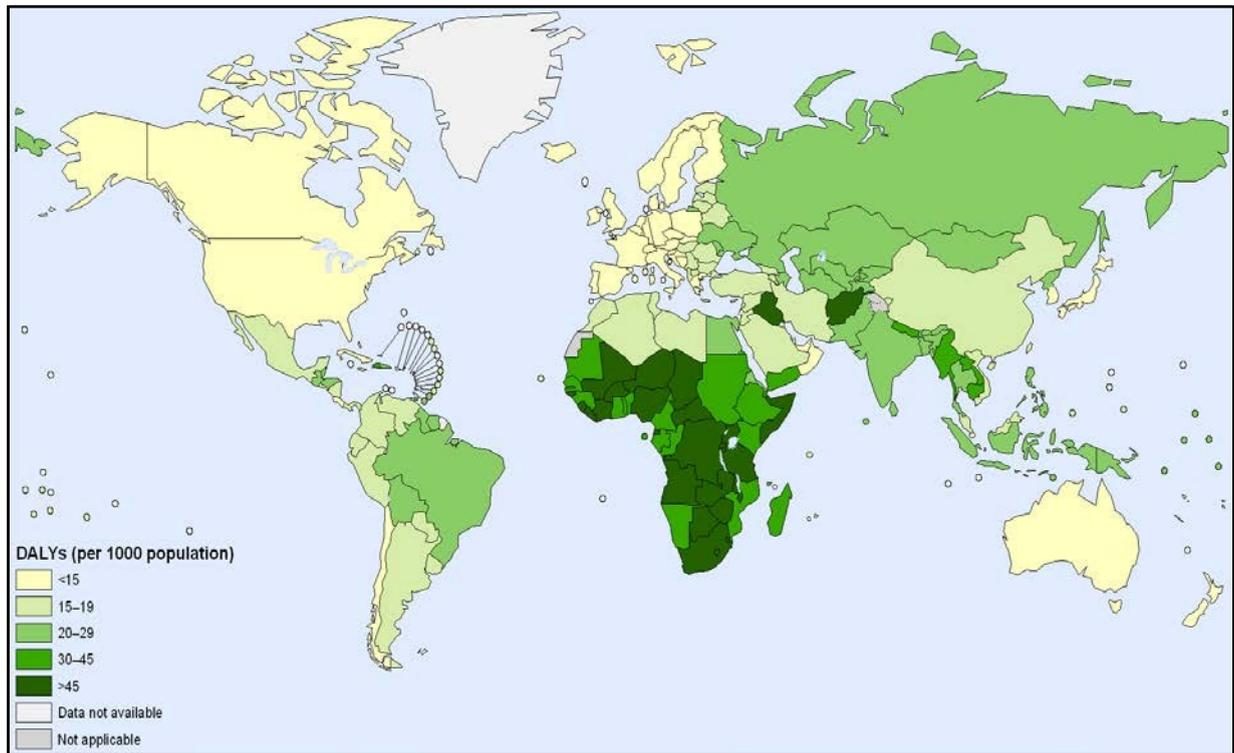
Figure 18 : Exemples mondiaux d'émergences et de réémergences de maladies infectieuses. Les nouvelles maladies émergentes sont représentées en rouge, les maladies ré-émergentes sont représentées en bleu et les maladies intentionnellement ré-émergentes sont en noir (bioterrorisme). Adaptée à partir de [5]



3.4. Les coûts économiques et sociaux des maladies infectieuses

Décès et morbidité liés aux maladies infectieuses et parasitaires humaines ont un coût économique et social considérable et un effet sur la croissance. Ces coûts peuvent être évalués globalement en incorporant les coûts directs imputables aux soins médicaux et les coûts indirects imputables à la réduction d'années d'espérance de vie et de productivité due à des morts prématurées ou à des complications chroniques. Ce poids porte essentiellement sur les populations les plus défavorisées de la planète. Grâce à l'index Daly (Disability adjusted life years) (Figure 19) qui intègre le nombre annuel de vies perdues à cause d'une maladie donnée, multiplié par un coefficient de ressources par individu, on peut calculer les pertes économiques subies. Avec un Daly de 26.2 %, les maladies infectieuses représentent la fraction la plus élevée du poids socio-économique total des maladies, largement devant les maladies neuropsychiatriques (13,1 %), la traumatologie (12,3 %), la pathologie maternelle périnatale (10,9 %), les maladies cardio-vasculaires (9,9 %) et le cancer (5,1 %) [2]. Les simulations montrent, particulièrement dans le cas de la tuberculose, du sida et du paludisme, que la prévention et le contrôle des maladies infectieuses dans les régions endémiques sont des approches socio-économiquement rentables [6]. La maîtrise des maladies infectieuses est par ailleurs un élément d'accélération de la transition démographique.

Figure 19 : Index DALY standardisé sur l'âge en 2004. Adaptée de [7].



3.5. Les conditions d'émergence et de circulation des agents infectieux dans les populations

L'émergence et la dynamique des maladies infectieuses sont régies par une diversité de facteurs pouvant soit augmenter le contact avec de nouveaux microbes (ou leurs vecteurs) soit faciliter leur dissémination. La plupart des maladies émergentes sont causées par des pathogènes, déjà présents dans l'environnement, qui à l'occasion de conditions changeantes leur conférant un avantage sélectif sont sortis de l'obscurité pour infecter de nouvelles populations hôtes (dans de rares occasions un nouveau variant peut aussi évoluer et causer une nouvelle maladie) [8, 9] (Tableau 2).

Le processus par lequel les agents infectieux peuvent se transférer des animaux aux humains ou se transférer d'un groupe isolé à une nouvelle population est appelé le « trafic microbien » [8]. Un grand nombre de paramètres influencent la dynamique de ce trafic et ont comme résultats de dessiner les émergences, la diffusion et la persistance dans les populations des maladies infectieuses.

Leur catégorisation est arbitraire mais permet tout de même d'élaborer une carte schématique des grands processus sous-jacents à l'émergence ou à la circulation des maladies infectieuses. L'émergence, la diffusion et la persistance d'une maladie infectieuse sont des phénomènes complexes qui résultent de l'interaction entre des facteurs agissant à trois niveaux différents : l'hôte c'est-à-dire la personne susceptible d'être infectée, l'agent biologique et l'environnement.

Partie 1 : Introduction générale

Tableau 2 : Exemple d'infections émergentes et de leurs facteurs probables d'émergence. Adaptée de [10]

Facteurs	Exemples de facteurs spécifiques	Exemples d'émergence de maladies
Changements écologiques (incluant ceux provoqués par le développement économique et l'utilisation des terres)	Agriculture, barrage, changements des écosystèmes hydriques	Fièvre de la vallée du rift (barrages, irrigation), fièvre hémorragique en Argentine (agriculture), fièvre hémorragique hanta en Corée (agriculture), syndrome pulmonaire à Hantavirus dans le Sud Ouest des Etats Unis
Démographie et comportement humains	Evénements sociétaux : migrations de populations (mouvement des zones rurales vers les villes), guerres, appauvrissement économique, dégradation du milieu urbain, comportements humains tels que le commerce du sexe, l'injection de drogue intraveineuse, les loisirs en plein air, l'utilisation de structures de soins intensives	Propagation du VIH et des autres maladies sexuellement transmissibles, propagation de la dengue
Voyages et commerces internationaux	Mouvements mondialisés de personnes et de marchandises	Propagation du VIH, dissémination de vecteurs tels que <i>Aedes albopictus</i> , des hantavirus véhiculés par les rats, introduction du choléra en Amérique du Sud, dissémination de la bactérie cholérique O139 par bateaux
Technologie et industrie	Productions alimentaires : globalisation de l'approvisionnement en nourritures, changements de procédés dans la production et le conditionnement Soins médicaux : nouveaux matériaux, transplantations d'organes ou de tissus, médicaments immunosuppresseurs, utilisation généralisée des antibiotiques	Procédés de fabrication de la nourriture : syndrome d'urémie hémolytique (souche d'E.coli dans des viandes contaminées et d'autres aliments), encéphalopathie spongiforme bovine, virus nipah (porc), grippe aviaire, syndromes respiratoires aigus Virus Ebola et VIH, maladies opportunistes chez les patients immunodéprimés, maladie de Creutzfeldt-Jacob par des lot d'hormone de croissance contaminés
Adaptation microbienne	Evolution microbienne et réponse à la pression de l'environnement	Dérive génétique du virus de la grippe, possibles changements génétiques des syndromes respiratoires aigus dus aux Coronavirus, développements de résistances microbiennes (VIH, résistances aux antibiotiques, tuberculose multirésistante, paludisme résistant à la chloroquine)
Interruptions des mesures de santé publiques et de contrôle	Réduction des programmes de prévention, manque d'assainissement ou de mesures de contrôle des vecteurs	Résurgence de la Tuberculose aux Etats Unis, choléra en Afrique dans les camps de réfugiés, réapparition de la diphtérie dans les pays d'Europe de l'Est dans les années 1990

3.5.1. L'homme dessine ses propres épidémies

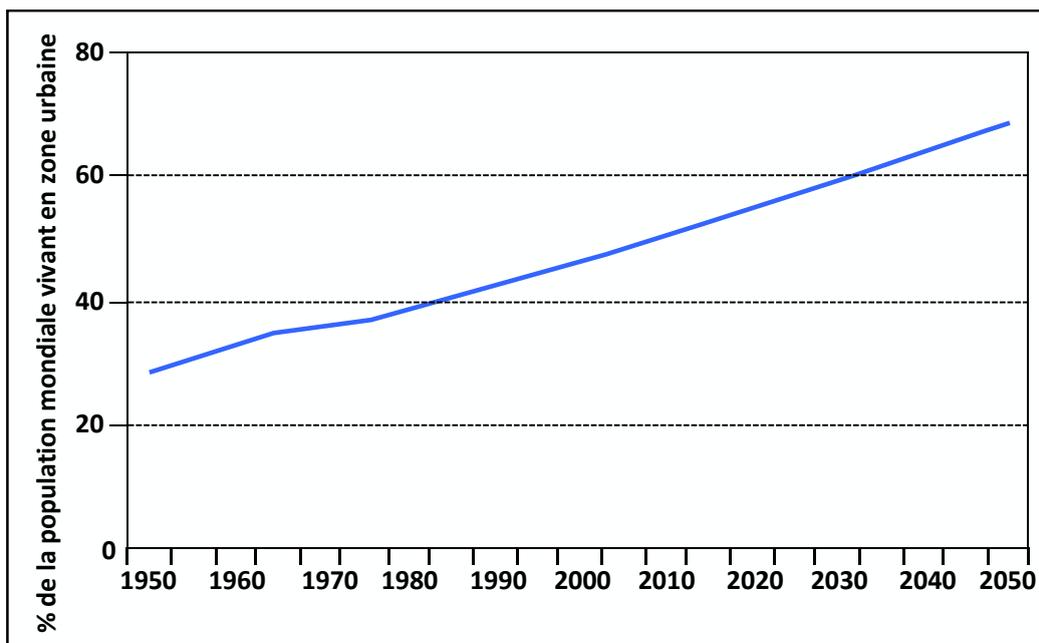
Les microbes s'appuient sur l'état sanitaire et social des populations humaines pour se développer. Les populations vulnérables et affaiblies biologiquement, tout particulièrement celles dont le tissu social est perturbé, vivant dans un contexte de privation avec un faible niveau hygiénique et une promiscuité importante sont beaucoup plus susceptibles à la colonisation microbienne [11]. Ainsi, l'ampleur de l'épidémie de peste bubonique à l'œuvre au milieu du 14^{ème} siècle en Europe reflète en réalité le faible niveau nutritionnel et la pauvreté causés par l'impact sur les récoltes des vagues de froids inhabituelles des décennies précédentes.

Chapitre 3 - Les maladies infectieuses

Les changements radicaux de la démographie humaine de ces dernières années observés à la surface du globe ont montré que les processus démographiques étaient des facteurs clés dans les phénomènes d'émergence et de transmission des maladies infectieuses. L'explosion de la taille de la population et de sa densité, l'urbanisation, la persistance de la pauvreté, l'augmentation du nombre de réfugiés de guerre, économiques, politiques, ou environnementaux sont autant de facteurs à intégrer en épidémiologie pour rendre compte de la dynamique des pathogènes.

L'environnement urbain est devenu que très récemment la forme d'habitat majoritaire de la population humaine. Dans beaucoup de régions du monde, le contexte économique encourage fortement les mouvements de masse des travailleurs ruraux vers les villes. Les Nations Unies ont estimé que d'ici à 2050, approximativement 70% de la population mondiale vivrait dans les villes (Figure 20) [12]. Ces phénomènes peuvent ainsi permettre aux infections émergeant initialement dans des milieux ruraux isolés d'atteindre de plus larges populations.

Figure 20 : Projection de l'évolution du pourcentage de la population mondiale vivant dans un environnement urbain, 1950-2050. Adaptée des données des Nations Unies [12]



Les immenses concentrations urbaines relevées dans des mégapoles dépassant parfois les 10 millions d'habitants jouent un rôle essentiel dans la dynamique des infections. Dans ces mégapoles, le risque de voir apparaître des maladies contagieuses est extrêmement important, du fait de la concentration humaine et de la promiscuité. L'impact de l'urbanisation sur la fragilisation du lien familial et social est maintenant connu : elle favorise la mobilité personnelle et les changements de réseaux sociaux. Ces phénomènes couplés aux moyens de contraception modernes, ont facilités par exemple la diffusion des maladies sexuellement transmissibles. Plus généralement, les villes ont souvent la fonction d'autoroute pour le trafic microbien. L'urbanisation rapide « booste » certaines maladies infectieuses bien établies telles que les pneumonies infantiles, les diarrhées, la

Partie 1 : Introduction générale

tuberculose ou la dengue et facilite la dissémination de maladies émergentes de manière identique à ce qui s'est passé dans le cas du SRAS à Hong Kong [13, 14].

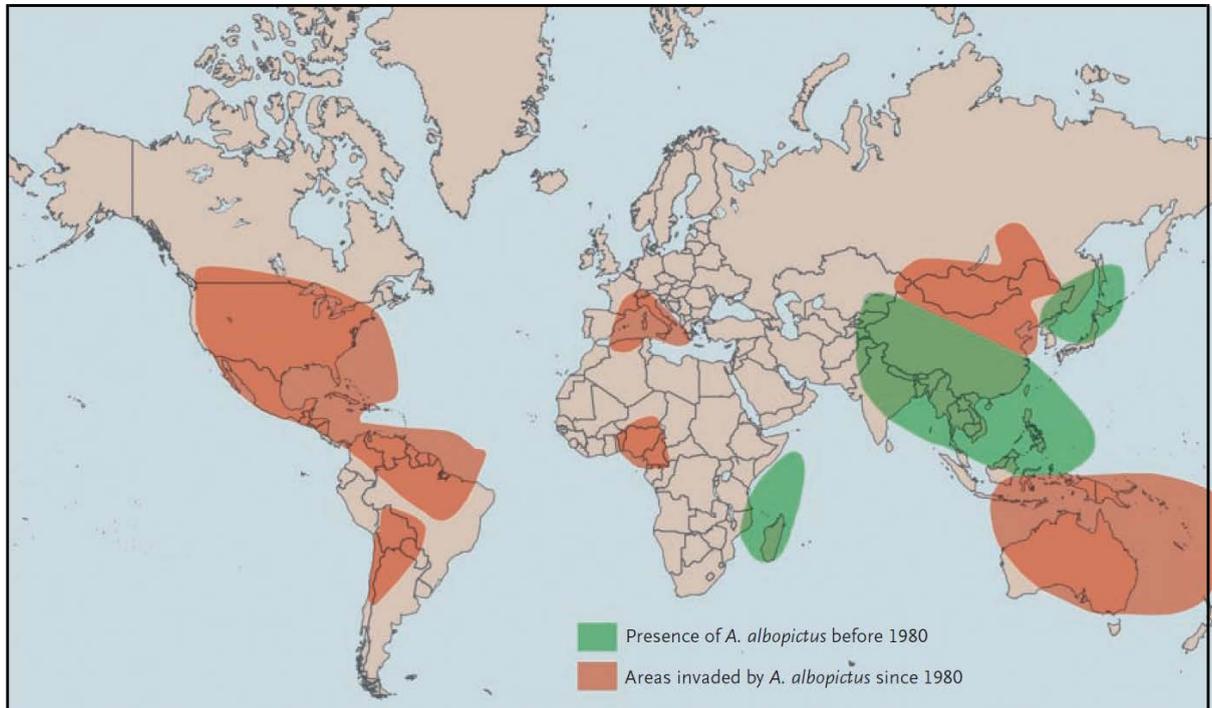
De même, les autres types de comportements humains ont aussi des effets importants sur la transmission des maladies. L'exemple le plus célèbre est donné par les maladies sexuellement transmissibles. La manière dont les comportements humains comme la sexualité et l'injection de drogues intraveineuses ont contribué à l'émergence et à la diffusion du VIH est maintenant très bien connue. D'autres facteurs responsables de changement dans la dynamique des pathogènes sont influencés par une grande variété d'actions humaines.

Dans le passé, une infection particulière touchant un peuple situé dans une zone géographique isolée pouvait être « véhiculée » dans un nouvel endroit grâce aux voyages, au commerce, et à la guerre. Le commerce entre l'Asie et l'Europe, initié avec la route de la soie et perpétué avec les croisades, a ainsi apporté en Europe les rats et la peste qu'ils véhiculaient. Par un processus identique, les bateaux transportant les esclaves de l'Afrique de l'Ouest jusqu'au nouveau monde au 16^{ème} et 17^{ème} siècle ont apporté la fièvre jaune et son vecteur, *Aedes aegypti*, dans ces nouveaux territoires. Des histoires similaires se répètent aujourd'hui, mais avec des opportunités de « passage de maladies » d'une population géographiquement distante d'une autre beaucoup plus fréquemment et intensément qu'avant. La globalisation et la politique d'ouverture des frontières associées ont causé une remarquable expansion du commerce et des mouvements humains. Ce phénomène est un déterminant majeur de la circulation actuelle des agents infectieux. Une quantité phénoménale de personnes ou de produits voyage quotidiennement autour du monde permettant la circulation d'une pathologie ou de son vecteur à travers les bateaux, les avions, containers et autres moyens [15]. On estime actuellement que plus d'un milliard de personnes prennent l'avion chaque année. Le transport international du bétail, des animaux exotiques et de compagnies participe aussi actuellement à la diffusion des maladies infectieuses [16]. Dans un monde de plus en plus interconnecté, de nouvelles maladies apparaissent à un rythme sans précédent et ont souvent la capacité de franchir les frontières et de se propager rapidement. Récemment, il a pu être mis en évidence qu'un moustique, *Aedes albopictus*, le vecteur du chikungunya, mais aussi de la dengue, s'est déplacé, probablement sur des cargos, dans des flaques d'eau (souvent, aussi, dans des pneus, dont il est difficile de vider complètement l'eau pluviale qui s'y accumule). À partir d'un foyer asiatique, ce moustique s'est répandu dans l'Océan indien, mais aussi en Europe, notamment en France, en Italie, et en Amérique (Figure 21) [17].

De plus les tendances récentes de « globalisation » des populations et de concentrations dans les milieux urbains, ont provoqué la rencontre de pathogènes avec des communautés ayant une diversité génétique et culturelle considérable les unes envers les autres. Depuis quelques années, le rôle des facteurs génétiques de l'hôte dont les caractères immunogénétiques ont été façonnés par les pressions sélectives de pathogènes différents, est mis en avant dans la dynamique des infections. La variation de réponse interindividuelle à l'infection, façonnée par l'interaction complexe des facteurs d'exposition, de la maladie, et de l'hôte modère à différentes échelles la dynamique des épidémies. La répartition mondiale de pathologies

graves et communes telles que la lèpre et la tuberculose pourraient être en fait le reflet de l'existence de gènes majeurs de susceptibilité, du moins dans certaines populations [18-21].

Figure 21 : Distribution mondiale du moustique *Aedes albopictus*. Adaptée de [17]



Enfin, depuis aussi longtemps que l'homme et les agents infectieux coexistent, les communautés humaines se sont organisées pour mettre en place des mesures de santé publiques et d'hygiène minimisant la dissémination et l'exposition aux agents pathogènes. Ces actions en réduisant le nombre de pathogènes, son réservoir, sa présence dans l'environnement, la zone géographique, ou le nombre de personnes susceptibles, sont des moyens de lutte efficace pour endiguer et freiner la circulation des agents infectieux dans les populations. Des voies d'actions classiquement utilisées sont par exemple le traitement de l'eau potable, les campagnes de vaccinations, ou encore des mesures de contrôle des populations de vecteurs. Malheureusement, ces pathogènes peuvent saisir l'opportunité de réémerger si des discontinuités dans la mise en place des mesures de prévention sont observées. Ce genre de problème a pu être observé dans les années 2000 en Amérique du Sud avec la première réapparition du choléra depuis le siècle dernier [22] qui est réapparu suite à la diminution du niveau de chlorage utilisé pour traiter les eaux dans les systèmes de distribution des eaux potables.

3.5.2. L'environnement un élément clé pour l'émergence et la diffusion des agents infectieux

Outre la diversité des paramètres intimement reliés à l'activité humaine, la dynamique des agents infectieux et de leurs vecteurs, est aussi liée à l'environnement et au système écologique dans lesquels ils évoluent. L'augmentation des incidences des maladies émergentes ou réémergentes actuelles a maintenant été reliée avec un bon nombre des changements écologiques et environnementaux majeurs se déroulant à la

Partie 1 : Introduction générale

surface de la planète et tout particulièrement avec les modifications d'utilisation des sols par les humains. La déforestation, la construction de routes ou de barrages, l'expansion urbaine, la destruction de niches écologiques par l'agriculture, l'exploitation minière, la pollution des eaux côtières ou de l'air sont autant de facteurs pouvant modifier la dynamique des anciens et nouveaux pathogènes (Tableau 3).

Tableau 3: Exemple d'agents ou de maladies infectieuses ayant un lien suspecté ou avéré avec un changement d'utilisation des terres. Adaptée de [23].

Maladies vectorielles ou zoonotiques	Sol	Eau	Humain	Autres
Paludisme	Mélioïdose	Schistosomiase	Asthme	Fièvre hémorragique
Dengue	Anthrax	Cholera	Tuberculose	Piriculariose du riz
Maladie de Lyme	Ankylostome	Shigellose	Grippe	Fièvre aphteuse
Fièvre jaune	Cocci- diomycose	Rotavirus		Trachome
Fièvre de la vallée du Rift		Salmonellose		
Encéphalite japonaise		Leptospirose		
Onchocercose		Cryptosporidiose		
Trypanosomiase				
Peste				
Filariosiose				
Méningite				
Rage				
Leishmaniose				
Fièvre de Kyasanur				
Hantavirus				
Nipah				

Les changements écologiques, qu'ils soient provoqués par l'homme ou pas, font partie des facteurs le plus souvent mis en évidence dans les phénomènes d'émergence et de circulation des maladies infectieuses. L'émergence de virus tels que Nipah, Hendra, SARS, ou celui de la grippe aviaire durant les 10 dernières années ont permis à la communauté scientifique et au grand public de se rendre compte de l'importance du lien homme-animal-écosystème dans la dynamique des maladies infectieuses [11].

Les facteurs écologiques et environnementaux favorisent les émergences en mettant en contact l'homme avec un réservoir ou un hôte d'une infection non familière mais déjà présente (la plupart du temps une zoonose ou une maladie vectorielle). Cette mise en contact s'effectue soit en créant une plus grande proximité soit par des conditions changeantes favorisant une multiplication des pathogènes, de son réservoir naturel ou de son vecteur [10].

La déforestation et la transformation des milieux forestiers en pâturage, zones agricoles, ou zone d'habitation humaine sont un exemple classique d'impact sur la dynamique des maladies infectieuses. Ces phénomènes altèrent significativement l'environnement et modifient profondément la composition des vecteurs et des pathogènes de ces zones géographiques et peuvent favoriser l'émergence ou la diffusion d'une maladie infectieuse. Des bouffées épidémiques de paludisme ont ainsi été retrouvées consécutives à la déforestation et aux changements de pratiques d'utilisation des terres en Afrique, Asie, et en Amérique latine [24]. Le sens d'influence de ce type de processus n'est pas unidirectionnel. L'émergence de la maladie de Lyme en Europe et aux Etats Unis est ainsi très largement due à la reforestation des terres qui a permis une prolifération importantes des populations de cervidés et de leurs tiques, vecteur de la maladie de Lyme [25].

Chapitre 3 - Les maladies infectieuses

La pratique de l'agriculture est un autre élément perturbateur des systèmes écologiques pouvant entraîner indirectement des émergences ou des ruptures dans la circulation des agents infectieux. C'est en effet le moyen le plus commun et le plus ancien par lequel les hommes altèrent et modifient les écosystèmes dans lesquels ils évoluent. Les hanta virus, sont connus depuis des siècles en Asie par les agriculteurs de riz. Le virus est une infection bénigne de la souris des champs *apodemus agrarius*. Ce rongeur prolifère dans les champs de riz et à leur contact, les personnes contractent souvent la maladie durant la saison des récoltes [26, 27]. Des exemples similaires peuvent être trouvés quand de nouvelles zones sont mises en cultures [23].

De même, quand l'humain colonise de nouveaux environnements pour développer une activité agricole, des nouveaux contacts entre la faune sauvage l'homme et son bétail se créent et maximisent le risque que des pathogènes passent la barrière inter espèce [23]. Un exemple de ce phénomène est celui de l'établissement de porcheries dans des zones proches de la forêt tropicale dans le nord de la Malaisie. En 1998, Dans cette région, le virus Nipah s'est transmis des chauves-souris frugivores vivant aux alentours de la ferme, au porcs de l'élevage qui à leur tour le transmirent aux fermiers de l'élevage [28].

Les ressources en eau et leur répartition sont encore d'autres facteurs influençant les émergences et la dynamique des maladies infectieuses. Les infections transmises par les moustiques ou autres arthropodes, qui incluent une bonne partie des maladies infectieuses les plus répandues géographiquement [16], sont souvent stimulées par l'expansion de zones d'eaux stagnantes, simplement parce que les larves de ces vecteurs se développent dans l'eau. Beaucoup d'exemples de tels phénomènes peuvent être trouvés dans la littérature notamment avec la création de barrages [29], les systèmes d'irrigation, ou les systèmes urbains de stockage d'eau [23].

Des facteurs environnementaux globaux comme le climat affectent aussi la dynamique de beaucoup de pathologies infectieuses et dessinent leur répartition à la surface du globe. Des modifications ou anomalies du climat peuvent avoir ainsi un impact majeur sur l'émergence et la diffusion des maladies infectieuses [30, 31]. C'est le cas par exemple pour les anomalies de type El Niño. El Niño est une anomalie prenant ses origines dans l'océan pacifique et qui perturbe durablement le climat dans de nombreuses parties du monde où il est associé soit à des pluies intenses soit à des sécheresses. Les effets d'El Niño sur les maladies infectieuses transmises par l'eau et les maladies vectorielles sont maintenant bien documentés [32]. De même, le réchauffement climatique, en affectant par exemple la répartition des vecteurs et leur compétence à transmettre le pathogène, est aussi un facteur susceptible de modifier profondément la dynamique des maladies infectieuses. Pour bon nombre de vecteurs (comme les moustiques ou les tiques) la fréquence de leur repas, leur taux de mortalité ou de reproduction sont directement corrélés à la température ambiante [33]. De plus le réchauffement climatique pourrait possiblement agrandir l'habitat de nombreux vecteurs dans des latitudes et altitudes plus grandes qu'avant [33].

D'un point plus strictement écologique, des études récentes ont montré que tout événement provoquant une diminution de la biodiversité peut avoir un impact durable sur l'émergence et la transmission des maladies infectieuses. Cette diminution de la biodiversité en modifiant l'abondance des vecteurs et des

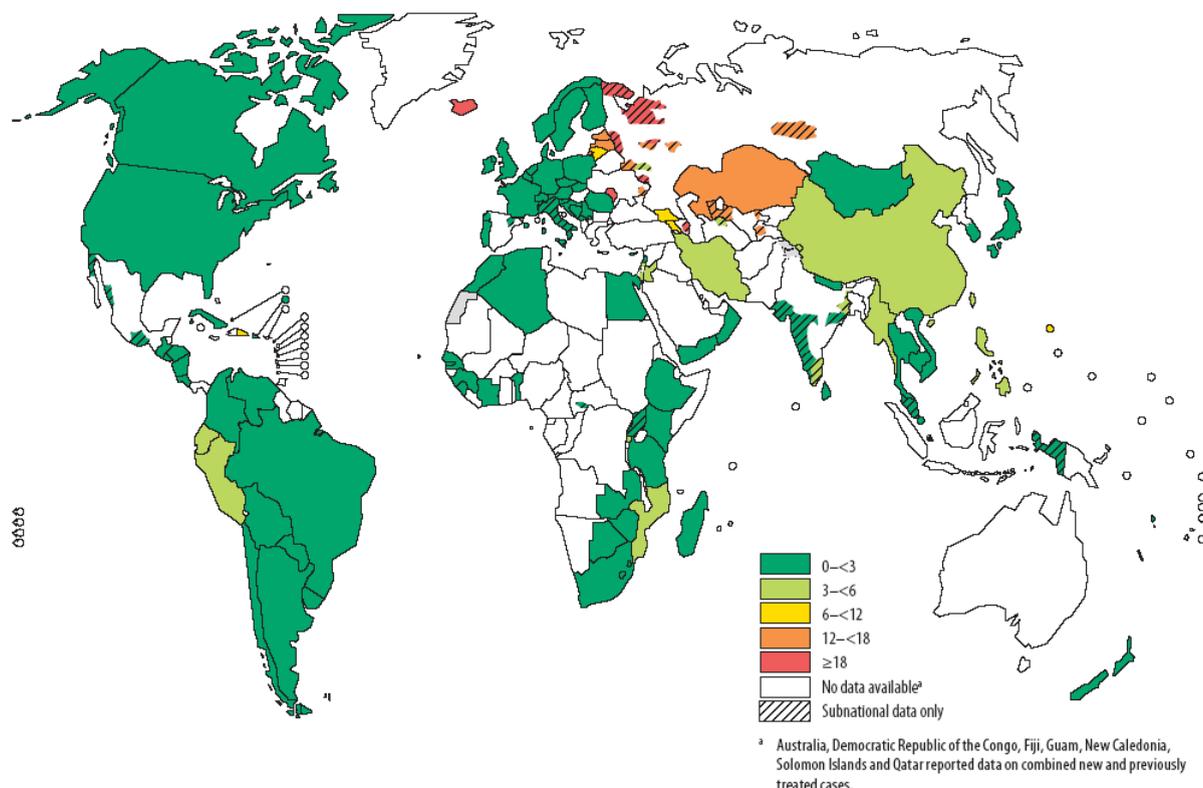
pathogènes, leurs comportements, et leurs conditions impactent très fortement la transmission des maladies infectieuses [34].

3.5.3. L'adaptation microbienne

Les microbes comme tous les autres êtres vivants s'adaptent constamment aux pressions exercées par l'environnement dans lequel ils évoluent. L'émergence de bactéries résistantes aux antibiotiques est le résultat de la généralisation et de la banalisation de l'utilisation des antibiotiques par l'homme. C'est une leçon magistrale sur l'adaptation microbienne et sur la force de la sélection naturelle. Plus surprenant encore, leurs dynamiques peuvent être dessinées par des interactions fortes entre ces derniers. L'exemple de la concomitance de la tuberculose et du VIH en est un exemple révélateur.

L'apparition de bactéries résistantes aux antibiotiques et de parasites résistants aux traitements est devenue fréquente, consécutivement à l'utilisation généralisée et parfois inappropriée de certains médicaments [35, 36]. C'est le cas par exemple ces dernières années avec l'apparition de tuberculose multi-résistante (résistante à l'isoniaside et la rifampine, deux des principaux traitements antituberculeux) et son expansion à l'ensemble du globe (Figure 22).

Figure 22 : Distribution de la proportion de tuberculose multi résistante parmi les nouveaux cas de tuberculose, 1994-2009, adaptée de [37].



Les pathogènes peuvent aussi acquérir ces gènes de résistance grâce à d'autres espèces présentes dans l'environnement, souvent non pathogènes pour l'homme mais dont l'évolution a été façonnée par la pression de sélection exercée par les antibiotiques ou d'autres médicaments. Beaucoup de virus arborent un

taux de mutations très important et peuvent rapidement évoluer en de nouveaux variants [38]. L'exemple classique ou ce type de phénomènes est retrouvé est celui de la grippe [39]. Les épidémies annuelles sont causées par une « dérive antigénique » dans une souche ayant déjà circulée précédemment. Un changement dans un site antigénique d'une protéine de surface, permet alors à ce nouveau variant de réinfecter des personnes précédemment infectées, grâce au fait que cette modification de l'antigène n'est pas immédiatement reconnue par le système immunitaire. En de rares occasions, plus fréquemment chez les pathogènes non viraux, l'évolution en un nouveau variant peut résulter dans des formes d'expressions nouvelles de la maladie. L'épidémie de fièvre purpurale observée au Brésil dans les années 1980-1990 et associée à l'émergence d'un variant clonal de *hemophilus influenza*, en est un bon exemple [40].

Par ailleurs, la diversité des pathogènes fait qu'il n'est pas du tout exceptionnel pour les personnes d'être co-infectés par différents pathogènes. Or la symptomatologie et la sévérité des co-infections ne se résument pas à la somme des effets de chaque agent pathogène. La résultante est souvent contre intuitive. Ce problème des co-infections est une constante dans l'histoire du système immunitaire. L'exemple le plus frappant est celui du VIH qui en déprimant le système immunitaire, change l'expression, la forme et l'intensité de la dynamique de nombreuses maladies infectieuses. L'immunité cellulaire est la barrière la plus importante contre les protozoaires et certaines bactéries, et sa perturbation peut être suivie d'une extrême susceptibilité aux nouvelles infections, ou favoriser des réactivations d'infections dormantes ou asymptomatiques. Il est maintenant bien connu que la remarquable réémergence de la tuberculose dans les dernières décennies est due à cet effet de potentialisation du VIH, qui a considérablement augmenté le risque de progression d'une infection latente à *Mycobacterium tuberculosis* en une maladie active, favorisant du même coup la transmission à d'autres personnes [41].

De manière identique, durant ces dix dernières années, un nombre important d'études se sont attachées à décrire les conséquences cliniques consécutives de l'interaction entre le paludisme et les nématodes gastro-intestinaux [42, 43]. Bien que certaines études n'observent aucune interactions [44, 45], un nombre important tendent à montrer une tendance à une certaine protection vis-à-vis du paludisme induite par les nématodes gastro-intestinaux, et plus particulièrement associée à *Ascaris lumbricoides* [46-51]. L'effet de cette interaction pourrait être d'une amplitude importante et avoir des conséquences de santé publiques majeures.

3.6. Conclusions sur les maladies infectieuses

L'émergence et la réémergence des maladies infectieuses et leur diffusion rapide dans le monde entier sont des défis importants pour les systèmes de santé nationaux, tout particulièrement dans les pays en voie de développement touchés par une pauvreté extrême et la dégradation de l'environnement. Les attentes dans les nouveaux vaccins et médicaments et dans la surveillance mondiale pour inverser ces tendances ont été frustrées jusqu'à ce jour par la complexité de la transition épidémiologique, en dépit des perspectives prometteuses de la recherche biomoléculaire et du génie génétique.

Partie 1 : Introduction générale

Ainsi, divers processus écologiques et sociaux, tels que la pauvreté extrême, les mouvements de population, l'urbanisation et la déforestation, favorisent l'émergence et la réémergence des syndromes infectieux et maladies et augmentent considérablement leur complexité épidémiologique. L'intensification des voyages internationaux et des migrations a également contribué à amplifier ces processus, en accélérant le mouvement des êtres humains, des animaux et des plantes et de la prolifération mondiale des virus, bactéries et champignons. La mondialisation, résultant en une incorporation rapide des nouvelles technologies et en une amplification des effets de l'urbanisation, du chômage, de l'exclusion sociale et de la pauvreté, aggrave les conséquences de ces changements sociaux et écologiques de manière inimaginable [52]. Ces conditions changeantes favorisent l'émergence et la réémergence mondiale de diverses maladies infectieuses avec des cycles et dynamiques complexes, telles que le VIH / sida, la tuberculose, la dengue, la fièvre jaune, le paludisme, la leishmaniose, la leptospirose, les syndromes pulmonaires à Hantavirus, et bien d'autres encore.

La notion de complexité est donc au cœur des questions épidémiologiques actuelles et est cruciale pour la compréhension des phénomènes liés à l'émergence d'agents pathogènes et à la dynamique des maladies. Comme un grand nombre des processus biologiques liés à l'émergence et à la dynamique des maladies sont provoquées par des interactions complexes entre l'environnement, l'individu et son milieu social, il est nécessaire de traiter de la notion même de complexité et de ces implications méthodologiques en épidémiologie et santé publique.

Bibliographie

1. Geddes A. **The future of infectious diseases.** *Antibiotiques* 2005,7:75-76.
2. World Health Organisation. **The global burden of diseases : 2004 updates** In. Geneva: World Health Organisation; 2004.
3. Péquignot F, Michel E, Le Toullec A, Jouglu E. **Mortalité par maladies infectieuses en France. Tendances évolutives et situation actuelle. Surveillance nationale des maladies infectieuses 1998-2000.** In. Saint Maurice: Institut de veille sanitaire 2002:27-34.
4. Pagano JS, Blaser M, Buendia MA, Damania B, Khalili K, Raab-Traub N, Roizman B. **Infectious agents and cancer: criteria for a causal relation.** *Semin Cancer Biol* 2004,14:453-471.
5. Fauci AS. **Infectious diseases: considerations for the 21st century.** *Clin Infect Dis* 2001,32:675-685.
6. Gordon DF, Gannon JC. **The Global Infectious Disease Threat and Its Implications for the United States.** In. Washington: National Intelligence Council; 2000.
7. World Health Organisation. **The Global Burden of Disease concept.** In. Geneva: World Health Organisation; 2004.
8. Morse SS. **Emerging viruses: defining the rules for viral traffic.** *Perspect Biol Med* 1991,34:387-409.
9. Morse SS. **Examining the origins of viruses.** In: *Emerging viruses.* new York: Oxford University Press; 1993:10-28.
10. Morse SS. **Factors in the emergence of infectious diseases.** *Emerg Infect Dis* 1995,1:7-15.
11. Weiss RA, McMichael AJ. **Social and environmental risk factors in the emergence of infectious diseases.** *Nat Med* 2004,10:S70-76.
12. Population Division of the Department of Economic and Social Affairs. **World Urbanization Prospects : The 2009 Revision.** In. New York: United Nations; 2009.
13. Alirol E, Getaz L, Stoll B, Chappuis F, Loutan L. **Urbanisation and infectious diseases in a globalised world.** *Lancet Infect Dis* 2011,11:131-141.
14. Godfrey R, Julien M. **Urbanisation and health.** *Clin Med* 2005,5:137-141.
15. Wilson ME. **Travel and the emergence of infectious diseases.** *Emerg Infect Dis* 1995,1:39-46.
16. Harrus S, Baneth G. **Drivers for the emergence and re-emergence of vector-borne protozoal and bacterial diseases.** *Int J Parasitol* 2005,35:1309-1318.
17. Charrel RN, de Lamballerie X, Raoult D. **Chikungunya outbreaks--the globalization of vectorborne diseases.** *N Engl J Med* 2007,356:769-771.
18. Altare F, Durandy A, Lammas D, Emile JF, Lamhamedi S, Le Deist F, et al. **Impairment of mycobacterial immunity in human interleukin-12 receptor deficiency.** *Science* 1998,280:1432-1435.
19. Fieschi C, Dupuis S, Catherinot E, Feinberg J, Bustamante J, Breiman A, et al. **Low penetrance, broad resistance, and favorable outcome of interleukin 12 receptor beta1 deficiency: medical and immunological implications.** *J Exp Med* 2003,197:527-535.
20. Mira MT, Alcais A, Nguyen VT, Moraes MO, Di Flumeri C, Vu HT, et al. **Susceptibility to leprosy is associated with PARK2 and PACRG.** *Nature* 2004,427:636-640.

Partie 1 : Introduction générale

21. Mira MT, Alcais A, Van Thuc N, Thai VH, Huong NT, Ba NN, *et al.* **Chromosome 6q25 is linked to susceptibility to leprosy in a Vietnamese population.** *Nat Genet* 2003,33:412-415.
22. Anderson C. **Cholera epidemic traced to risk miscalculation.** *Nature* 1991,354:255.
23. Patz JA, Daszak P, Tabor GM, Aguirre AA, Pearl M, Epstein J, *et al.* **Unhealthy landscapes: Policy recommendations on land use change and infectious disease emergence.** *Environ Health Perspect* 2004,112:1092-1098.
24. Patz JA, Graczyk TK, Geller N, Vittor AY. **Effects of environmental change on emerging parasitic diseases.** *Int J Parasitol* 2000,30:1395-1405.
25. Barbour AG, Fish D. **The biological and social phenomenon of Lyme disease.** *Science* 1993,260:1610-1616.
26. Chen HX, Qiu FX, Dong BJ, Ji SZ, Li YT, Wang Y, *et al.* **Epidemiological studies on hemorrhagic fever with renal syndrome in China.** *J Infect Dis* 1986,154:394-398.
27. Schmaljohn C, Hjelle B. **Hantaviruses: a global disease problem.** *Emerg Infect Dis* 1997,3:95-104.
28. Chua KB, Goh KJ, Wong KT, Kamarulzaman A, Tan PS, Ksiazek TG, *et al.* **Fatal encephalitis due to Nipah virus among pig-farmers in Malaysia.** *Lancet* 1999,354:1257-1259.
29. Southgate VR. **Schistosomiasis in the Senegal River Basin: before and after the construction of the dams at Diama, Senegal and Manantali, Mali and future prospects.** *J Helminthol* 1997,71:125-132.
30. Hunter PR. **Climate change and waterborne and vector-borne disease.** *J Appl Microbiol* 2003,94 Suppl:375-465.
31. Kovats RS, Haines A. **Global climate change and health: recent findings and future steps.** *CMAJ* 2005,172:501-502.
32. Kovats RS, Bouma MJ, Hajat S, Worrall E, Haines A. **El Nino and health.** *Lancet* 2003,362:1481-1489.
33. Zell R. **Global climate change and the emergence/re-emergence of infectious diseases.** *Int J Med Microbiol* 2004,293 Suppl 37:16-26.
34. Keesing F, Belden LK, Daszak P, Dobson A, Harvell CD, Holt RD, *et al.* **Impacts of biodiversity on the emergence and transmission of infectious diseases.** *Nature*,468:647-652.
35. Kunin CM. **Resistance to antimicrobial drugs--a worldwide calamity.** *Ann Intern Med* 1993,118:557-561.
36. Neu HC. **The crisis in antibiotic resistance.** *Science* 1992,257:1064-1073.
37. World Health Organisation. **Multidrug and extensively drug-resistant TB (M/XDR-TB) : 2010 global report on surveillance and response.** In. Geneva: World Health Organisation; 2010.
38. Domingo E, Holland JJ. **RNA virus mutations and fitness for survival.** *Annu Rev Microbiol* 1997,51:151-178.
39. Kilbourne ED. **The molecular epidemiology of influenza.** *J Infect Dis* 1973,127:478-487.
40. Harrison LH, da Silva GA, Pittman M, Fleming DW, Vranjac A, Broome CV. **Epidemiology and clinical spectrum of Brazilian purpuric fever. Brazilian Purpuric Fever Study Group.** *J Clin Microbiol* 1989,27:599-604.

Chapitre 3 - Les maladies infectieuses

41. Corbett EL, Watt CJ, Walker N, Maher D, Williams BG, Raviglione MC, Dye C. **The growing burden of tuberculosis: global trends and interactions with the HIV epidemic.** *Arch Intern Med* 2003,163:1009-1021.
42. Mwangi TW, Bethony JM, Brooker S. **Malaria and helminth interactions in humans: an epidemiological viewpoint.** *Ann Trop Med Parasitol* 2006,100:551-570.
43. Nacher M. **Worms and malaria: blind men feeling the elephant?** *Parasitology* 2008,135:861-868.
44. Bejon P, Mwangi TW, Lowe B, Peshu N, Hill AV, Marsh K. **Helminth infection and eosinophilia and the risk of Plasmodium falciparum malaria in 1- to 6-year-old children in a malaria endemic area.** *PLoS Negl Trop Dis* 2008,2:e164.
45. Shapiro AE, Tukahebwa EM, Kasten J, Clarke SE, Magnussen P, Olsen A, et al. **Epidemiology of helminth infections and their relationship to clinical malaria in southwest Uganda.** *Trans R Soc Trop Med Hyg* 2005,99:18-24.
46. Brutus L, Watier L, Briand V, Hanitrasoamampionona V, Razanatsoarilala H, Cot M. **Parasitic co-infections: does Ascaris lumbricoides protect against Plasmodium falciparum infection?** *Am J Trop Med Hyg* 2006,75:194-198.
47. Brutus L, Watier L, Hanitrasoamampionona V, Razanatsoarilala H, Cot M. **Confirmation of the protective effect of Ascaris lumbricoides on Plasmodium falciparum infection: results of a randomized trial in Madagascar.** *Am J Trop Med Hyg* 2007,77:1091-1095.
48. Degarege A, Animut A, Legesse M, Erko B. **Malaria severity status in patients with soil-transmitted helminth infections.** *Acta Trop* 2009,112:8-11.
49. Murray MJ, Murray AB, Murray MB, Murray CJ. **Parotid enlargement, forehead edema, and suppression of malaria as nutritional consequences of ascariasis.** *Am J Clin Nutr* 1977,30:2117-2121.
50. Nacher M, Gay F, Singhasivanon P, Krudsood S, Treeprasertsuk S, Mazier D, et al. **Ascaris lumbricoides infection is associated with protection from cerebral malaria.** *Parasite Immunol* 2000,22:107-113.
51. van Eijk AM, Lindblade KA, Odhiambo F, Peterson E, Rosen DH, Karanja D, et al. **Geohelminth Infections among pregnant women in rural western Kenya; a cross-sectional study.** *PLoS Negl Trop Dis* 2009,3:e370.
52. Possas CA. **Social ecosystem health: confronting the complexity and emergence of infectious diseases.** *Cad Saude Publica* 2001,17:31-41.

Chapitre 4 – Capturer la complexité : le nouvel enjeu de l'épidémiologie

"Aucune des cellules d'Antoine ne sait ce qui se passe quand Antoine dit son amour à Cléopâtre"
Edgar Morin

4.1. La complexification de la recherche scientifique

« Pluridisciplinarité », « Intégration », « contexte ». Ces mots sont devenus des termes stratégiques en santé publique qui sont retrouvés dans le lexique des approches multidisciplinaires récentes de la recherche biomédicale et de la médecine clinique.

Le point commun de ces nouveaux mots utilisés en recherche est qu'ils tournent tous autour de la notion de complexité et de ce qui pourrait être défini comme un changement de mentalité et une rupture vis-à-vis de l'approche réductionniste classique de la science. Depuis les approches de Descartes et Newton, le développement des connaissances a été caractérisé par le contraire de la complexité, c'est-à-dire réduire la complexité du réel à la simplicité des lois et des explications.

Cette heuristique de parcimonie a été la force motrice de l'ensemble des recherches scientifiques effectuées jusqu'au début du 20^{ème} siècle [1]. Le paradigme de la modernité, qui est enraciné dans une vision mécaniste et dans la pensée logique et mathématique, a surtout développé autour de la recherche une méthode absolue et rationnelle capable de définitivement séparer les connaissances scientifiques des pseudo-connaissances, y compris celles des disciplines sociales.

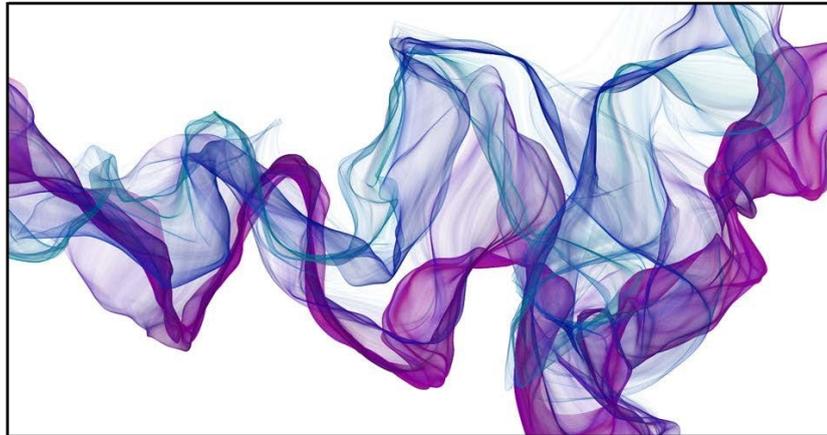
Cependant, dans la seconde moitié du 20^{ème} siècle, la modernité a subi une crise [1] : la physique quantique, la théorie d'Einstein sur la relativité, et le principe d'indétermination de Heisenberg ont complètement ébranlé ses fondements.

Ce qui a émergé est une tendance antiréductionniste, qui a ouvert la voie à un mouvement épistémologique, « la théorie de la complexité », considérée par certains comme une révolution en terme de paradigme [2]. Les signes du changement sont devenus visibles dans les dernières décennies, non seulement dans la science et la philosophie, mais aussi dans de nombreux autres champs du savoir, y compris l'architecture, les arts (Figure 23) et la littérature, les sciences sociales et politiques, et l'écologie.

Quel que soit le domaine, l'esprit de l'ère post-moderne est « complexe », du moins dans son sens le plus large: une combinaison de déterminisme et de hasard, une tolérance de l'hétérogénéité et de l'incertitude, et une attention renouvelée placée sur le contexte et l'aspect pratique.

Il est difficile, aujourd'hui, d'ouvrir un magazine de vulgarisation scientifique, ou une revue scientifique de premier plan, sans avoir lu des articles traitant de la complexité [3] La théorie de la complexité influence des domaines aussi variés que la physique [4], la cosmologie [5], la chimie [6], la géographie [7], la recherche sur le climat [8], la zoologie [9], la biologie [10], la biologie évolutive [11], la biologie cellulaire [12], les neurosciences [13], la gestion/management [14] et l'économie [15].

Figure 23 : Exemple d'œuvre obtenue par l'«art génératif», basé sur la théorie de la complexité et utilisant des algorithmes pour concevoir des œuvres se générant d'elles-mêmes.



Ces tendances se manifestent ainsi dans une variété de disciplines biomédicales [1, 16, 17]. S'agissant spécifiquement de l'épidémiologie, un débat vigoureux a surgi autour de la complexité et des méthodes utilisées ces dernières années pour la capturer.

4.2. Définition de la complexité

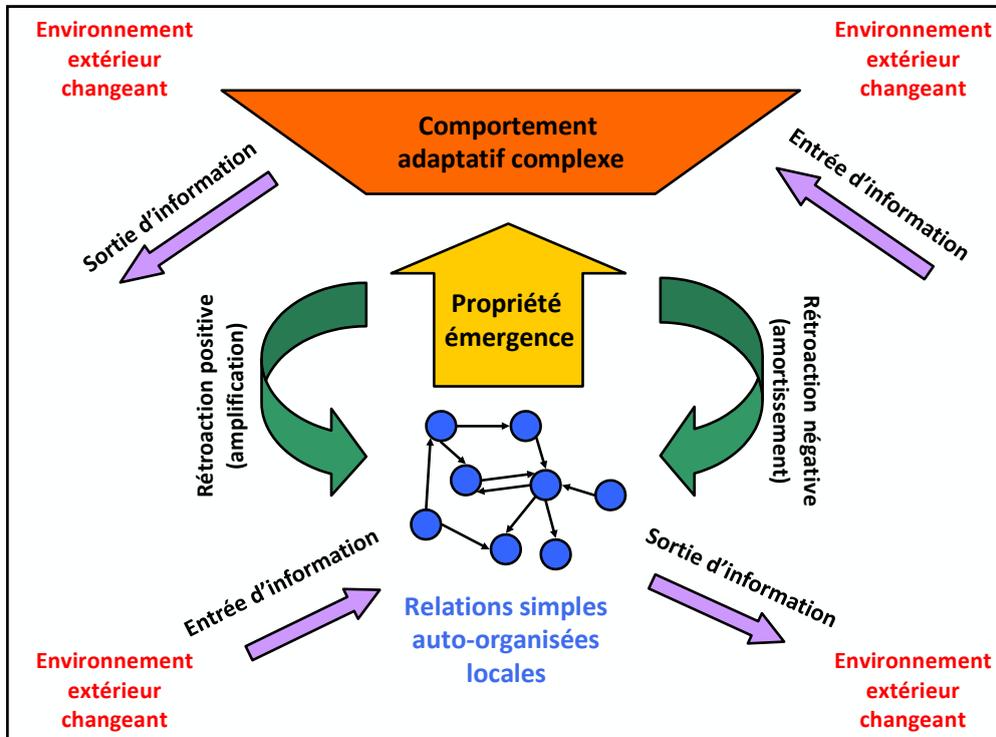
De nombreux phénomènes sont décrits comme complexes, cependant le concept de «complexité» est plus spécifique que le sens où on l'emploie couramment. La complexité est l'étude des systèmes adaptatifs complexes. Ces systèmes sont définis comme «un ensemble d'agents individuels ayant la liberté d'agir d'une manière qui n'est pas toujours totalement prévisible, et dans lesquels les actions des agents sont interconnectées entre elles de sorte que le comportement d'un agent particulier change le contexte des autres agents» [14]. Des exemples de ce type de système sont les cellules vivantes, le cerveau, le système immunitaire, les marchés financiers, les écosystèmes ou encore les populations humaines. Ils sont complexes dans le sens où il existe un grand nombre d'agents apparemment indépendants qui interagissent les uns avec les autres, mais dans lesquels la richesse des interactions permet au système dans son ensemble de tendre vers l'auto-organisation [3]. Ils sont également caractérisés par l'existence de processus non linéaires et de boucles de rétroaction dans lesquels de petits changements peuvent avoir des effets importants ne pouvant pas être étudiés en analysant simplement les composants individuels [18]. Le tout est plus que la somme de ses parties. Ces systèmes complexes peuvent exister à différents niveaux d'organisation, allant par exemple du niveau subatomique jusqu'au niveau individuel ou au niveau populationnel [19] (Figure 24).

L'exemple le plus frappant d'un système auto organisé complexe est la vie elle-même, non seulement en terme d'organismes individuels, mais aussi en terme d'évolution : les organismes s'adaptent les uns aux autres par évolution dans un écosystème finement réglé. De même, les différentes populations ont évolué à travers des modes de vie traditionnels qui doivent maintenant répondre aux changements induits par la révolution industrielle, la colonisation et la mondialisation [20].

Ainsi, une caractéristique clé de ces systèmes complexes est qu'ils sont adaptatifs. Ils ne répondent pas seulement de manière passive face à des événements, mais ils se réorganisent en un nouvel équilibre en réponse aux événements [3]. Le cerveau se réorganise pour apprendre par l'expérience, les espèces évoluent

pour atteindre un nouvel écosystème en réponse à des événements tels que des changements climatiques ou des chutes de météorites, et les populations évoluent en réaction aux changements économiques et sociaux tout en conservant souvent leur culture « traditionnelle » sous une nouvelle forme et dans un nouveau contexte.

Figure 24 : Schématisation des propriétés d'un système adaptatif complexe



Un tel équilibre dynamique n'est pas toujours atteint (des espèces, des populations et des cultures s'éteignent ou des marchés financiers s'écroulent), mais certaines nouvelles formes d'auto-organisation émergent du chaos et perdurent. Les nouvelles formes d'organisation qui peuvent survenir sont souvent imprévisibles, car de petits changements dans les conditions initiales peuvent produire de grands changements dans l'état d'équilibre finalement atteint. Cependant, bien que certains détails puissent être imprévisibles, la forme générale des nouvelles formes d'organisation est relativement prévisible et simple. Par exemple, de petits changements dans les conditions initiales de l'évolution pourraient avoir radicalement modifié son histoire, mais les « formes superficielles » de l'évolution retrouvées seraient susceptiblement les mêmes, malgré les différentes voies empruntées. Il est probable que quelque chose ressemblant à des oiseaux (avec des ailes, des plumes, etc.) aurait évolué pour remplir la même niche écologique, même si la voie suivie par l'évolution aura été radicalement différente [21].

Cela illustre un autre élément clé de la théorie de la complexité : ce qui semble chaotique et imprévisible à un niveau d'organisation (généralement bas) peut être relativement simple et stable à un autre niveau (généralement plus élevés) [19]. Personne ne serait tenté de prédire la météo à partir des mesures de molécules individuelles où tout ce que l'on peut voir est du bruit, mais les systèmes météorologiques peuvent être extrêmement simples et prévisibles lorsqu'on les observe à un niveau plus approprié [22]. Ainsi le concept de « climat » est un résumé des grandes tendances de la météo qui peut être plus prévisible, bien que le système climatique peut être sensible aux petites perturbations et peut changer au fil du temps. Ainsi, « la

nature peut produire des structures complexes, même dans des conditions simples, et peuvent obéir à des lois simples, même dans des situations complexes» [22]. Le climat peut être très complexe et difficile à prédire, sur une base journalière, mais l'hiver suit régulièrement l'été et El Niño se produit à intervalles semi-prévisibles.

Ces cinq concepts 1) auto-organisation, 2) adaptation, 3) émergence du chaos, 4) imprévisibilité des effets de petits changements dans les conditions initiales, et 5) existence de la simplicité à certains niveaux alors que le «chaos» existe à d'autres, forment les concepts fondamentaux de la complexité. Mais en quoi de tels concepts interviennent-ils en épidémiologie ?

4.3. Complexité et épidémiologie

Le lien le plus évident est que la santé d'une population peut être considérée comme un système adaptatif complexe. Une population n'est pas seulement qu'une collection d'individus, mais aussi une population unique qui a sa propre histoire, sa culture, sa structure socio-économique, et qui survit malgré d'importants changements économiques mondiaux, tout en étant affectée et façonnée en même temps par ces changements [22].

La santé d'une population façonne, et est façonnée par le contexte environnemental et socioculturel dans lequel la population vit. Ainsi, bien que l'émergence ou la dynamique d'une maladie peuvent être étudiées à différents niveaux [23] (écosystèmes, populations, individus, molécules,...), il a été avancé que le niveau populationnel est en fait le plus fondamental en épidémiologie [22].

Il y a clairement des exceptions à cette règle ; par exemple le niveau de l'écosystème est crucial si l'on veut considérer les conséquences sanitaires à long terme des changements climatiques [24] ; le niveau individuel et moléculaire, et les interactions entre différents niveaux d'analyse possibles sont également très importants [25, 26].

Cependant, le niveau populationnel est la plupart du temps le plus stratégique, car il définit les problèmes de santé publique qui devraient être abordés. Par ailleurs, il est aussi de première importance en termes scientifiques, car certains problèmes scientifiques peuvent être bien mieux compris au niveau populationnel et ne peuvent pas être réduits ou analysés uniquement au niveau moléculaire ou individuel [27].

Malgré cela il existe assez peu d'exemples de l'utilisation de la théorie de la complexité en épidémiologie, alors qu'il y a par contre de nombreux exemples de problèmes épidémiologiques (particulièrement dans le domaine des maladies infectieuses) pour laquelle la théorie de la complexité est pertinente. Bien que les études mettant l'accent sur le niveau populationnel, le contexte socioculturel et environnemental ne nécessitent pas toujours l'utilisation de la théorie de la complexité, elles rendent cependant sa valeur et son potentiel beaucoup plus intelligible.

4.4. Implications de la théorie de la complexité en épidémiologie : maladies infectieuses, données publiques et méthodes écologiques un trio gagnant

Quelles sont les implications en épidémiologie de la théorie de la complexité ?

4.4.1. La nécessaire révision du paradigme réductionniste de l'épidémiologie

La théorie de la complexité souligne ainsi les limites d'une vision réductionniste naïve. Les gens ne sont pas simplement des collections aléatoires de cellules ou de molécules, et les populations ne sont pas simplement des collections aléatoires d'individus. Les systèmes adaptatifs complexes ont une «vie» qui est plus que la somme de leurs composantes. Comprendre le fonctionnement du cerveau exige non seulement une connaissance de ses composants, mais aussi une compréhension du contexte dans lequel il fonctionne [13]. Les facteurs de risque de la maladie ne fonctionnent pas de manière isolée, mais se produisent dans un contexte démographique et environnemental particulier. Le style de vie d'un individu ne peut être compris qu'en examinant le contexte historique, culturel, social et environnemental dans lequel il survient. Ce mode de pensée a considérablement modifié les pratiques de l'épidémiologie depuis sa création. La théorie de la complexité milite pour une étude commune de l'ensemble de ces facteurs recueillis à différentes échelles et souligne que leur intégration dans une même réflexion est nécessaire, plutôt qu'une analyse réductionniste les étudiant séparément. L'histoire de l'épidémiologie est enracinée dans et a évolué à partir de ce questionnement autour de la complexité et de l'échelle d'observation adéquate. Ces questionnements successifs découlaient directement des contextes dans lesquels la science épidémiologique a évolué. Trois ères de l'épidémiologie ont ainsi été distinguées par différents auteurs [25, 26, 28] (épidémiologie traditionnelle, épidémiologie moderne et éco-épidémiologie).

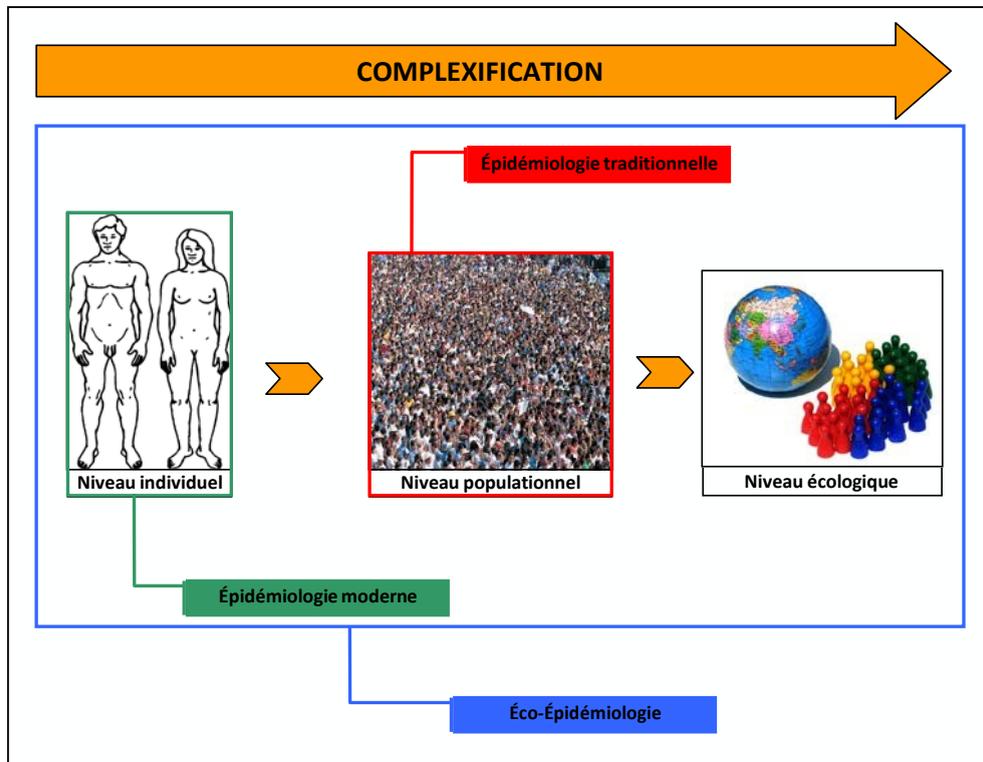
Au début des années 1800 en Angleterre, l'épidémiologie traditionnelle portait son attention sur l'étude de la population comme unité d'enquête, plutôt que l'individu. Elle était centrée sur l'importance du contexte social, et cela dans une période où le changement social était une priorité forte dans les pays industrialisés avant et au début du 20^{ème} siècle. Elle portait sur l'étude des maladies dans les populations plutôt que dans le laboratoire. L'épidémiologie traditionnelle intervenait ainsi à un niveau populationnel, communautaire ou organisationnel, tandis que l'épidémiologie moderne qui lui succéda intervenait quant à elle au niveau individuel ou moléculaire des systèmes naturels [22].

Le développement du microscope a ensuite conduit à la preuve que les micro-organismes étaient à l'origine des maladies (la théorie des germes), plutôt que des miasmes mystérieux comme suggérés avant. Ces avancées ont conduit à l'élaboration des systèmes modernes de laboratoires de santé publique, des installations de recherches, et des systèmes de surveillance coopératifs internationaux et mondiaux. Alors que l'épidémiologie traditionnelle était axée sur la santé publique des populations que ce soit au niveau de l'étude ou de celui de l'intervention, l'approche de l'épidémiologie moderne se situait au niveau de l'individu et se suffisait à elle-même. Elle s'enracinait dans les principes scientifiques tout en étant la plupart du temps déconnectée du contexte. L'épidémiologie moderne était et est toujours basée sur une perspective d'étude se concentrant sur l'observation, l'expérimentation et l'analyse, et où les résultats sont évalués statistiquement.

Au cours des 30 dernières années, et surtout au cours des 10 dernières années, un certain nombre d'auteurs ont commencé à aborder la question de changer ce paradigme de l'épidémiologie moderne [25, 26, 28]. Ils soulignaient la nécessité d'étendre l'épidémiologie à l'étude des interrelations et de la complexité des interactions intervenant dans l'émergence et la dynamique des maladies infectieuses. Les niveaux biologiques

et socio environnementaux devaient être combinés et étudiés ensemble dans la même recherche. Ils lancèrent ainsi les bases d'un nouveau type d'épidémiologie, l'éco-épidémiologie (Figure 25).

Figure 25 : Schématisation de la relation entre les différentes ères de l'épidémiologie et le niveau d'analyse



4.4.2. L'appréhension des niveaux d'analyse : le nouveau défi de l'épidémiologie

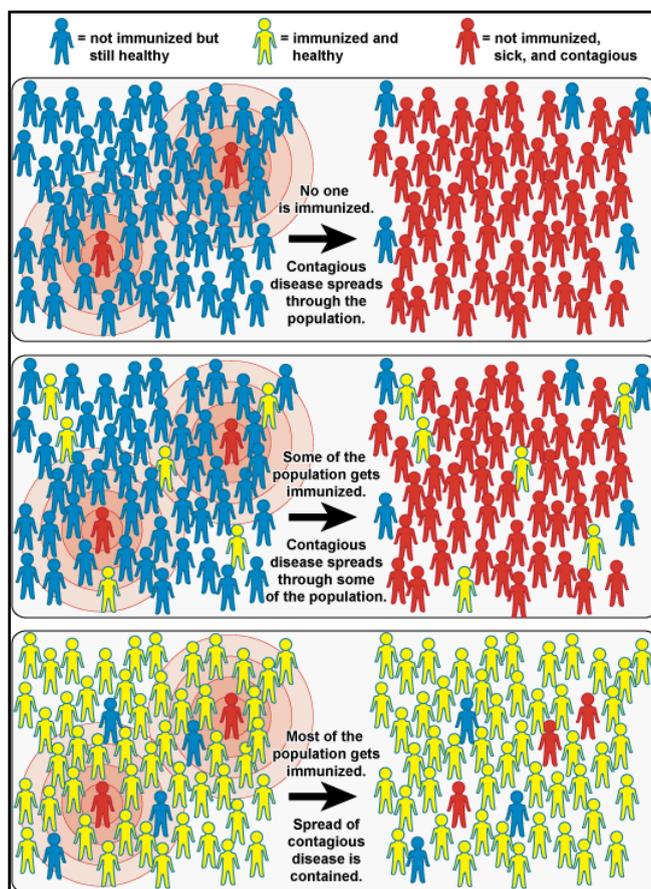
La théorie de la complexité souligne également l'importance du concept de « niveau d'analyse » [22]. Ce qui peut être chaotique à un niveau peut être simple à un autre niveau, et pour obtenir des connaissances utiles, se focaliser à un niveau d'analyse approprié est nécessaire [4]. Ainsi il n'est pas nécessaire de comprendre ce qui se passe au niveau moléculaire afin d'envoyer une fusée sur la lune, et il n'est pas nécessaire non plus de se concentrer sur le niveau moléculaire pour réaliser des améliorations en santé publique [29]. Une notion clé qui a reçu une attention particulière en épidémiologie ces dernières années a été que les déterminants d'une maladie ne peuvent pas être conceptualisés uniquement comme un attribut du niveau individuel. Il est nécessaire d'examiner les caractéristiques des groupes auxquels appartiennent les individus lorsque l'on se focalise sur les déterminants de la santé. Cette re-conceptualisation des variables définies au niveau populationnel ou écologique et étudiées au niveau individuel s'est manifestée ces dernières années, par exemple, dans l'intérêt et les débats sur les effets possibles sur la santé de variables populationnelles, telles que l'inégalité des revenus [30, 31], le capital social [32, 33], et les caractéristiques du quartier [34, 35].

L'idée que des facteurs « indépendants » des individus, définis au niveau de la population [36-39], sont importants pour la santé n'est pas nouvelle. Deux exemples bien connus sont par exemple le concept « d'immunité de communauté » pour les maladies infectieuses et de la distinction de Rose entre les causes des

Partie 1 : Introduction générale

cas et les causes des taux d'incidence des maladies chroniques. L'immunité de communauté implique que la probabilité d'une personne de contracter une maladie infectieuse dépend en partie du niveau d'immunité de la population à laquelle elle appartient [40] (Figure 26). Dans son article fondateur, " Sick Individuals and Sick Populations», Geoffrey Rose discute d'un concept lié: l'idée que les études se concentrant uniquement sur ce qui distingue les individus malades et les individus sains au sein d'une population peut manquer des déterminants importants des maladies. Les facteurs populationnels étant invariants au sein d'une population, ces derniers ne peuvent pas par conséquent être étudiés dans des études limitées à des comparaisons d'individus faites au sein d'une seule population [37]. Pour détecter ces facteurs, les chercheurs ont besoin de mettre en place des études qui comparent différentes populations (ou groupes) et recherchent des facteurs définis à l'échelle des populations.

Figure 26 : Schématisation du principe de fonctionnement de l'immunité de communauté. Elle implique que la probabilité d'une personne de contracter une maladie infectieuse dépend en partie du niveau d'immunité de la population à laquelle elle appartient.



Les discussions autour des facteurs populationnels et individuels en épidémiologie tournent souvent autour du concept que d'un côté les variables populationnelles sont importantes dans la compréhension des différences entre les populations et que d'un autre, les variables individuelles sont des facteurs importants dans la compréhension des différences entre individus. Pourtant un point crucial est qu'un ensemble de facteurs définis sur plusieurs niveaux peut être important pour comprendre les causes de la variabilité au sein

d'un unique niveau. Par exemple, à la fois des facteurs définis au niveau individuel et au niveau du groupe sont importants pour comprendre les causes des différences entre les taux de maladies de différentes populations.

De même, autant les facteurs populationnels que les facteurs individuels sont importants pour comprendre les causes de la maladie chez les individus. Par exemple, l'immunité collective, une propriété de groupe, est importante pour comprendre non seulement les raisons des différences d'incidence de la maladie entre les groupes, mais aussi pour expliquer la probabilité d'un individu de contracter la maladie.

4.4.3. Le rôle vital des études écologiques dans ce contexte de « complexification »

L'émergence de la théorie de la complexité en épidémiologie a aussi souligné la nécessité de développer de nouvelles méthodes d'analyse appropriées à cette problématique. En effet, la théorie de la complexité ne cadre pas avec la plupart des approches standards de l'épidémiologie. Une grande partie de la pensée épidémiologique moderne consistait à étudier les effets des expositions chez les individus. L'éco-épidémiologie insiste maintenant sur le fait que les populations ne sont pas seulement des collections d'individus et que le contexte de cette population ne se définit pas uniquement comme du bruit. Sa prise en compte peut être fondamentale dans la compréhension des pathologies.

Une fois ce concept reconnu, de nouvelles questions scientifiques se sont posées (se rapportant autant à l'épidémiologie, qu'à l'écologie ou à la démographie) et dont les réponses impliquaient une méthodologie très différente des techniques épidémiologiques habituelles. Ces questions ont ainsi conduit les chercheurs en santé publique à réévaluer leur conception des études écologiques et des variables écologiques. En acceptant l'existence de multiples niveaux d'analyses et de variables d'études, les chercheurs ont ainsi vu réapparaître des débats sur la nécessité et l'utilité des études écologiques pour expliquer les disparités de santé [41, 42]. En effet, ces dernières, malgré leur mauvaise réputation, sont pratiquement les seules à pouvoir répondre aux enjeux de la complexité en épidémiologie : la réalisation d'essais contrôlés randomisés pour étudier la plupart des variables contextuelles étant pratiquement impossible [41]. Les études écologiques et les techniques statistiques associées sont maintenant considérées comme particulièrement prometteuses en raison de leur capacité à intégrer dans un même modèle, à la fois les prédicteurs individuels et communautaires du niveau de santé [43].

Cependant, contrairement à d'autres domaines comme la sociologie ou l'écologie, ce changement de mentalité a encore du mal à diffuser dans le monde de la recherche épidémiologique. Cette barrière s'explique principalement par le fait que la plupart des cours ou manuels en épidémiologie passent beaucoup de temps et d'effort à détailler les aspects techniques des études de cohorte ou de cas témoins, mais donnent peu de temps ou d'espace pour les études écologiques et l'expérimentation naturelle. Des méthodes statistiques comme l'analyse de séries chronologiques ou les modèles multi-niveaux sont trop peu souvent abordées [41, 42].

4.4.4. Les données publiques dans ce contexte de complexification : un enjeu stratégique

Ainsi, l'épidémiologie tend à être de plus en plus une discipline intégrative et collaborative [44], amplifiant la nécessité d'un accès ouvert aux données. Cette discipline a mûri au cours du siècle passé, évoluant de recherches faites sur une petite échelle ou à court terme, à des expériences menées sur des individus à grande échelle, à long terme, et combinées dans des projets multidisciplinaires intégrant des données de sources variées, traitées par des approches analytiques sophistiquées. Ce type d'études nécessitent presque toujours des interactions fortes avec des disciplines voisines (par exemple, l'évolution, la génomique, la géologie, l'océanographie et la climatologie, la sociologie, l'anthropologie, l'économie, l'écologie). Ce vaste champ de compétences et connaissances génère des défis majeurs pour trouver des moyens efficaces pour découvrir, accéder, intégrer, et analyser la portée et le volume des informations intéressantes.

L'étude sur la dynamique du choléra au Bangladesh faite par Rita Colwell [45] représente un exemple convaincant de la nécessité d'un accès facilité aux données en épidémiologie et plus généralement en santé publique. Comprendre la dynamique du choléra a exigé des données portant sur le plancton, la chimie, l'océanographie, les sciences atmosphériques, la sociologie, et l'anthropologie. Elle a aussi exigé des données sur les décisions économiques, politiques et juridiques qui ont impacté les modes de gestion des épidémies de choléra et l'organisation des systèmes de santé. Malgré quelques groupes de recherche en épidémiologie bien organisés qui peuvent fournir des données libres sur les travaux qu'ils réalisent (par exemple, les enquêtes DHS « Demographic and Health Survey » [46]) la plupart des données brutes actuelles et historiques des recherches réalisées en épidémiologie sont inaccessibles ou perdues.

Par ailleurs, malgré les nombreuses études liées à une pathologie particulière, comme le VIH par exemple, seule une petite fraction des données de ces études est disponible aujourd'hui. Par conséquent, notre capacité à comprendre les effets à court terme ou chronique des pathologies humaines reste très limitée. L'accès aux données est non seulement important pour la recherche épidémiologique de base mais est également cruciale pour répondre aux préoccupations profondes des émergences, réémergences et dynamiques des maladies infectieuses que nous connaissons aujourd'hui et, inévitablement, à l'avenir. Malheureusement, seule une petite fraction des données collectées est facilement recensée, accessible, et utilisable. Dans un domaine voisin, l'écologie, il est estimé que moins de 1% des données recueillies sont accessibles après la publication des résultats associés [47]. Plutôt que de fournir un accès direct aux données, la communauté scientifique se contente « d'interprétations de données » distillées à travers des présentations et des publications. Afin de réaliser des avancées en épidémiologie majeures, possibles grâce à la synthèse des sciences se rencontrant au carrefour de l'épidémiologie, nous avons besoin de résoudre les défis technologiques et sociologiques limitant l'accès ouvert à la quantité impressionnante de données qui est engrangée quotidiennement par les activités de recherches ou de surveillance.

4.5. Données publiques, maladies infectieuses, et études écologiques : un trio gagnant pour appréhender la complexité

La notion de complexité est donc au cœur des questions épidémiologiques actuelles et est cruciale pour la compréhension des phénomènes liés à l'émergence des agents pathogènes et à la dynamique des maladies infectieuses. La prise de conscience autour de cette complexification de l'épidémiologie a ainsi amené une partie de la communauté des chercheurs en épidémiologie à repenser l'épidémiologie comme une science complexe ou la prise en compte à la fois de l'individu, de la population à laquelle il appartient et du contexte dans lequel il vit, est nécessaire pour expliquer la dynamique actuelle des maladies. Cet état des lieux est particulièrement vrai dans le cadre des maladies infectieuses.

Cette complexité souligne aussi un changement radical sur l'appréhension du niveau d'analyse à adopter en épidémiologie : autant les facteurs populationnels et écologiques que les facteurs individuels sont importants pour expliquer la dynamique des maladies chez les individus. Cette question des différents niveaux d'analyses interagissant les uns avec les autres en épidémiologie a souligné la nécessité de développer ou plutôt de sortir de l'ombre de méthodes d'analyses appropriées à cette problématique. Ainsi les études écologiques et les techniques statistiques associées, traditionnellement assez décriées en épidémiologie, sont maintenant considérées comme particulièrement prometteuses en raison de leur capacité à intégrer dans un même modèle, à la fois les prédicteurs individuels et communautaires du niveau de santé.

Pour capturer cette complexité des maladies, l'épidémiologie tend ainsi à être de plus en plus une discipline intégrative et collaborative. Un des principaux enjeux stratégiques de cette approche multidisciplinaire naissante est d'avoir accès aux nombreuses sources de données nécessaires pour mettre en place ces approches ambitieuses. Le mouvement naissant de l'open data en prônant un accès facilité aux données publiques de recherches pourrait ainsi jouer un rôle majeur dans la pérennisation de ces démarches originales et novatrices de l'épidémiologie intégrative.

L'équation « données libres + études écologiques + maladies infectieuses » semble ainsi stratégique par sa capacité à répondre aux enjeux d'une nouvelle épidémiologie plus intégrative. Son potentiel est particulièrement prometteur pour l'étude des problématiques infectieuses de zones géographiques ayant un contexte environnemental et socioculturel très riche. La Guyane Française fait partie de cette catégorie. Le sujet de cette thèse bien que généraliste est en fait un reflet des réflexions issues de l'observation des problématiques de santé spécifiques à ce territoire français d'Amérique du sud, situé dans le bassin amazonien.

4.6. La Guyane Française : un territoire complexe

Nous avons vu que l'émergence et la circulation des pathogènes sont déterminées par des facteurs très divers : socio-économiques, environnementaux et écologiques. Nombre de ces facteurs se concentrent en Guyane. L'Amazonie Française est une zone de forte biodiversité, où les phénomènes migratoires intenses, la pauvreté, et l'anthropisation de la forêt primaire (liée à la croissance démographique et à l'orpaillage clandestin) sont autant de facteurs laissant prévoir, malgré la faible densité de population, l'émergence ou la réémergence de pathologies humaines ou animales. Ceci est déjà le cas pour des pathologies telles que la

Partie 1 : Introduction générale

toxoplasmose amazonienne [48-50], et la Maladie de Chagas [51]. Ces 2 exemples récents illustrent la réalité des risques sanitaires à l'interface entre la forêt et les hommes. On peut y ajouter le paludisme [52, 53], et la résistance aux antipaludiques [54], la dengue [55], la fièvre Q [56], la leishmaniose [57-59], et peut être l'ulcère de Buruli. Aux pathogènes à transmission vectorielle, s'ajoutent les maladies à transmission hydrique (shiguelloses, et même choléra) en raison de l'absence d'eau potable pour une proportion importante de la population ; et le risque posé par les maladies à transmission respiratoire (tuberculose multirésistante favorisée par l'épidémie VIH les migrations [60] et la pauvreté).

La fréquence de la chasse en Guyane est également une activité permettant le passage de la barrière d'espèce [61]. La Guyane constitue donc une zone à risque d'émergence et de circulation particulière, nécessitant une vigilance particulière. A cette diversité de pathogènes, s'ajoute une diversité génétique et culturelle considérable des populations (origines Amérindienne, Africaine, européenne, asiatique, métissage intense) dont les caractères immunogénétiques ont été façonnés par les pressions sélectives de pathogènes différents. Ainsi, à la frontière occidentale de la Guyane essentiellement peuplée de Noir Marrons, le paludisme est essentiellement dû à *Plasmodium falciparum*, *P. vivax* étant très rare du fait de l'absence de l'antigène érythrocytaire Duffy chez les populations d'origine Africaine. Par contre, à la frontière Est, peuplée de brésiliens et d'amérindiens, *P. vivax* est plus fréquent que *P. falciparum*.

L'émergence des pathogènes peut donc s'exprimer différemment selon les populations affectées. Ceci constitue donc un terrain unique pour étudier l'importance des facteurs d'hôtes génétiques dans la physiopathologie des maladies émergentes. En outre, la diversité des cultures constitue également un champ de recherche important dans l'étude des maladies émergentes.

Ainsi, la diversité des hommes et des pathogènes, et l'intensité des facteurs qui les poussent à se croiser sont particulièrement remarquables en Guyane et en font un site d'études multidisciplinaires particulièrement intéressant, dans lequel les études écologiques et l'utilisation de données publiques trouvent leur pleine légitimité. Ces dernières vont permettre d'intégrer et d'étudier les interactions de cette masse d'information autant biologique qu'écologique et cela à différents niveaux d'étude. Par-dessus cela, de par le contexte structurel guyanais, il est souvent difficile d'implémenter les schémas d'études classiques pour étudier une problématique de santé. Cette contrainte méthodologique renforce encore la pertinence de l'utilisation de méthodes alternatives plus souples à implémenter comme les études écologiques ou l'utilisation de données libres.

Cette thèse bien que ne se limitant pas à des problématiques guyanaises trouve ainsi directement ses sources dans ce territoire et correspond à une réflexion intimement liée à la réalité et aux problématiques de santé rencontrées quotidiennement en Guyane Française.

Bibliographie

1. Materia E, Baglio G. **Health, science, and complexity.** *J Epidemiol Community Health* 2005,59:534-535.
2. Morin E. **Epistémologie. La complexité.** *Revue internationale des sciences sociales* 1974,26:607-634.
3. Waldrop MM. **Complexity: The Emerging Science at The Edge of Order and Chaos.** London: Penguin; 1992.
4. Goldenfeld N, Kadanoff LP. **Simple lessons from complexity.** *Science* 1999,284:87-89.
5. Chaisson EJ. **Cosmic Evolution: The Rise of Complexity in Nature.** Cambridge: Harvard University Press; 2001.
6. Whitesides GM, Ismagilov RF. **Complexity in chemistry.** *Science* 1999,284:89-92.
7. Werner BT. **Complexity in natural landform patterns.** *Science* 1999,284:102-104.
8. Rind D. **Complexity and climate.** *Science* 1999,284:105-107.
9. Parrish JK, Edelstein-Keshet L. **Complexity, pattern, and evolutionary trade-offs in animal aggregation.** *Science* 1999,284:99-101.
10. Oltvai ZN, Barabasi AL. **Systems biology. Life's complexity pyramid.** *Science* 2002,298:763-764.
11. Lynch M, Conery JS. **The origins of genome complexity.** *Science* 2003,302:1401-1404.
12. Weng G, Bhalla US, Iyengar R. **Complexity in biological signaling systems.** *Science* 1999,284:92-96.
13. Koch C, Laurent G. **Complexity and the nervous system.** *Science* 1999,284:96-98.
14. Plsek PE, Wilson T. **Complexity, leadership, and management in healthcare organisations.** *BMJ* 2001,323:746-749.
15. Arthur WB. **Complexity and the economy.** *Science* 1999,284:107-109.
16. Pearce N, Merletti F. **Complexity, simplicity, and epidemiology.** *Int J Epidemiol* 2006,35:515-519.
17. Wilson T, Holt T, Greenhalgh T. **Complexity science: complexity and clinical care.** *BMJ* 2001,323:685-688.
18. Goldberger AL. **Non-linear dynamics for clinicians: chaos theory, fractals, and complexity at the bedside.** *Lancet* 1996,347:1312-1314.
19. Gribbin J. **Deep Simplicity: Chaos, Complexity and the Emergence of Life.** London: Allen Lane; 2004.
20. Pearce N. **Epidemiology: populations, methods and theories.** *Eur J Epidemiol* 2004,19:729-731.
21. Cohen J, Stewart I. **The Collapse of Chaos: Discovering Simplicity in a Complex World.** London: Penguin; 1994.
22. Pearce N. **Traditional epidemiology, modern epidemiology, and public health.** *Am J Public Health* 1996,86:678-683.
23. Susser M. **Causal Thinking in the Health Sciences.** New York: Oxford University Press; 1973.
24. Pekkanen J, Pearce N. **Environmental epidemiology: challenges and opportunities.** *Environ Health Perspect* 2001,109:1-5.
25. Susser M, Susser E. **Choosing a future for epidemiology: II. From black box to Chinese boxes and eco-epidemiology.** *Am J Public Health* 1996,86:674-677.

Partie 1 : Introduction générale

26. Susser M, Susser E. **Choosing a future for epidemiology: I. Eras and paradigms.** *Am J Public Health* 1996,86:668-673.
27. Pearce N. **Epidemiology as a population science.** *Int J Epidemiol* 1999,28:S1015-1018.
28. McMichael AJ. **Prisoners of the proximate: loosening the constraints on epidemiology in an age of change.** *Am J Epidemiol* 1999,149:887-897.
29. Pearce N, McKinlay JB. **Dissent: Back to the future in epidemiology and public health: response to Dr. Gori.** *J Clin Epidemiol* 1998,51:643-646; discussion 647-649.
30. Lynch J, Harper S, Davey Smith G. **Commentary: Plugging leaks and repelling borders--where to next for the SS income inequality?** *Int J Epidemiol* 2003,32:1029-1036; discussion 1037-1040.
31. Subramanian SV, Blakely T, Kawachi I. **Income inequality as a public health concern: where do we stand? Commentary on "Is exposure to income inequality a public health concern?"** *Health Serv Res* 2003,38:153-167.
32. Kawachi I, Kennedy BP, Glass R. **Social capital and self-rated health: a contextual analysis.** *Am J Public Health* 1999,89:1187-1193.
33. Lynch J, Due P, Muntaner C, Smith GD. **Social capital--is it a good investment strategy for public health?** *J Epidemiol Community Health* 2000,54:404-408.
34. Pickett KE, Pearl M. **Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review.** *J Epidemiol Community Health* 2001,55:111-122.
35. Subramanian SV. **The relevance of multilevel statistical methods for identifying causal neighborhood effects.** *Soc Sci Med* 2004,58:1961-1967.
36. Morgenstern H. **Ecologic studies in epidemiology: concepts, principles, and methods.** *Annu Rev Public Health* 1995,16:61-81.
37. Rose G. **Sick individuals and sick populations.** *Int J Epidemiol* 1985,14:32-38.
38. Susser M. **The logic in ecological: I. The logic of analysis.** *Am J Public Health* 1994,84:825-829.
39. Von Korff M, Koepsell T, Curry S, Diehr P. **Multi-level analysis in epidemiologic research on health behaviors and outcomes.** *Am J Epidemiol* 1992,135:1077-1082.
40. Fine PE. **Herd immunity: history, theory, practice.** *Epidemiol Rev* 1993,15:265-302.
41. Ben-Shlomo Y. **Real epidemiologists don't do ecological studies?** *Int J Epidemiol* 2005,34:1181-1182.
42. Pearce N. **The ecological fallacy strikes back.** *J Epidemiol Community Health* 2000,54:326-327.
43. Diez Roux AV. **The study of group-level factors in epidemiology: rethinking variables, study designs, and analytical approaches.** *Epidemiol Rev* 2004,26:104-111.
44. Guégan JF, Choisy M. **Introduction à l'Épidémiologie Intégrative des maladies infectieuses et parasitaires.** Louvain: De Boeck Université; 2009.
45. Colwell RR. **Global climate and infectious disease: the cholera paradigm.** *Science* 1996,274:2025-2031.
46. Measure DHS. **Demographic and Health Survey website.** In. Calverton; 2011.
47. Reichman OJ, Jones MB, Schildhauer MP. **Challenges and opportunities of open data in ecology.** *Science*,331:703-705.

Chapitre 4 - Capturer la complexité : le nouvel enjeu de l'épidémiologie

48. Ajzenberg D, Banuls AL, Su C, Dumetre A, Demar M, Carme B, Darde ML. **Genetic diversity, clonality and sexuality in *Toxoplasma gondii***. *Int J Parasitol* 2004,34:1185-1196.
49. Carme B, Bissuel F, Ajzenberg D, Bouyne R, Aznar C, Demar M, *et al.* **Severe acquired toxoplasmosis in immunocompetent adult patients in French Guiana**. *J Clin Microbiol* 2002,40:4037-4044.
50. Demar M, Ajzenberg D, Maubon D, Djossou F, Panchoe D, Punwasi W, *et al.* **Fatal outbreak of human toxoplasmosis along the Maroni River: epidemiological, clinical, and parasitological aspects**. *Clin Infect Dis* 2007,45:e88-95.
51. Aznar C, La Ruche G, Laventure S, Carme B, Liegeard P, Hontebeyrie M. **Seroprevalence of *Trypanosoma cruzi* infection in French Guiana**. *Mem Inst Oswaldo Cruz* 2004,99:805-808.
52. Carme B. **Substantial increase of malaria in inland areas of eastern French Guiana**. *Trop Med Int Health* 2005,10:154-159.
53. Legrand E, Volney B, Meynard JB, Mercereau-Puijalon O, Esterre P. **In vitro monitoring of *Plasmodium falciparum* drug resistance in French Guiana: a synopsis of continuous assessment from 1994 to 2005**. *Antimicrob Agents Chemother* 2008,52:288-298.
54. Legrand E, Demar M, Volney B, Ekala MT, Quininternet M, Bouchier C, *et al.* **First case of emergence of atovaquone resistance in *Plasmodium falciparum* during second-line atovaquone-proguanil treatment in South America**. *Antimicrob Agents Chemother* 2007,51:2280-2281.
55. Dussart P, Lavergne A, Lagathu G, Lacoste V, Martial J, Morvan J, Cesaire R. **Reemergence of dengue virus type 4, French Antilles and French Guiana, 2004-2005**. *Emerg Infect Dis* 2006,12:1748-1751.
56. Gardon J, Heraud JM, Laventure S, Ladam A, Capot P, Fouquet E, *et al.* **Suburban transmission of Q fever in French Guiana: evidence of a wild reservoir**. *J Infect Dis* 2001,184:278-284.
57. Rotureau B. **Are New World leishmaniasis becoming anthroponoses?** *Med Hypotheses* 2006,67:1235-1241.
58. Rotureau B, Gaborit P, Issaly J, Carinci R, Fouque F, Carme B. **Diversity and ecology of sand flies (Diptera: Psychodidae: Phlebotominae) in coastal French Guiana**. *Am J Trop Med Hyg* 2006,75:62-69.
59. Rotureau B, Joubert M, Clyti E, Djossou F, Carme B. **Leishmaniasis among gold miners, French Guiana**. *Emerg Infect Dis* 2006,12:1169-1170.
60. Brudey K, Filliol I, Ferdinand S, Guernier V, Duval P, Maubert B, *et al.* **Long-term population-based genotyping study of *Mycobacterium tuberculosis* complex isolates in the French departments of the Americas**. *J Clin Microbiol* 2006,44:183-191.
61. Wolfe ND, Switzer WM, Carr JK, Bhullar VB, Shanmugam V, Tamoufe U, *et al.* **Naturally acquired simian retrovirus infections in central African hunters**. *Lancet* 2004,363:932-937.

Chapitre 5 - Objectifs de la thèse

Nous avons vu dans les chapitres précédents que le trio maladies infectieuses, données libres et études écologiques pouvait ainsi s'avérer très utile dans l'étude de problématiques complexes de santé publique et de surcroît nécessaire et indispensable dans d'autres situations. La partie qui suit tente d'illustrer ces notions à travers des exemples d'études portant aussi bien sur des problématiques locales (en Guyane française) que sur des problématiques plus globales abordées à l'échelle populationnelle.

Cette seconde partie est découpée arbitrairement en chapitres structurés non pas en fonction du schéma d'étude ou de la pathologie étudiée, mais plutôt en fonction de la nature des données libres utilisées.

4 chapitres ont ainsi été réalisés :

- **Chapitre 6** : Apports du couplage études écologiques/données climatiques dans la compréhension de la dynamique des maladies infectieuses : exemple de la Guyane Française.
- **Chapitre 7** : Utilisation des données issues de la littérature scientifique : Application à l'étude de la séroprévalence à *Toxoplasma gondii* dans les populations humaines.
- **Chapitre 8** : Utilisation des statistiques onusiennes : Applications à l'étude des déterminants de la prévalence/incidence des résistances aux antituberculeux, du paludisme et de la mortalité juvénile.
- **Chapitre 9** : Discussion générale, perspectives et conclusions

Le terme « données libres » dans cette deuxième partie est à prendre dans son sens le plus large : ils s'agit soit de données réellement libres (accessibles à tout le monde) comme les données onusiennes par exemple, soit de « données pseudo-libres » (librement accessibles uniquement aux les chercheurs travaillant dans une structure publique) comme les données météorologiques locales, ou soit d'un mélange de données libres et restreintes comme les données issues de la littérature scientifique par exemple. Les études abordées ici sont surtout des illustrations de situations où une ouverture totale des données serait potentiellement intéressante.

PARTIE 2: TRAVAUX SCIENTIFIQUES

Chapitre 6 - Apports du couplage études écologiques/données climatiques dans la compréhension de la dynamique des maladies infectieuses : exemple de la Guyane Française.

« Quiconque souhaitant pratiquer la médecine correctement devrait procéder ainsi: examiner en premier lieu les saisons de l'année et déterminer quels effets chacune d'elle produits »
Hippocrate

6.1. L'omniprésence du climat en santé humaine

La relation entre santé humaine et climat n'est ni un concept ni un sujet nouveau. Depuis aussi loin que l'époque d'Hippocrate, les scientifiques s'attachant à étudier la santé humaine étaient persuadés que cette dernière était étroitement liée aux cycles des saisons, aux conditions météorologiques locales, ainsi qu'à d'autres facteurs environnementaux [1]. Bien avant l'avènement de la théorie des germes de Pasteur dans le milieu des années 1800, le climat était utilisé pour décrire certaines dynamiques des maladies, et bien des voies thérapeutiques étaient fondées directement sur ces changements de l'environnement. Les partisans des miasmes, croyant que l'atmosphère empoisonnée découlant des marécages et des matières putrides était la source des maladies, remarquaient déjà que les fièvres et les affections intestinales étaient plus courantes durant les saisons chaudes et étaient souvent associées aux zones humides et mal drainées, comme les marécages par exemple [1]. Cette association motiva l'assèchement des marais dans les zones rurales et l'installation des premiers systèmes d'égouts collectifs dans les zones urbaines [1]. Bien sûr, les bienfaits thérapeutiques de ces pratiques avaient peu à voir avec l'élimination des miasmes de l'environnement, mais elles permirent de réduire la taille et le nombre des sites de reproduction des moustiques (vecteurs de maladies) et d'éliminer les déchets bruts du contact immédiat des humains (et donc les pathogènes d'origine hydrique). En dépit de l'évolution du milieu des années 1800 démontrant le rôle des microorganismes dans les maladies, la théorie des maladies transmises par les conditions météorologiques a tenu ferme en médecine jusqu'à la fin des années 1800 [1].

Durant les trois dernières décennies, le développement d'outils modernes (comme la télédétection et les images satellitaires) a conduit à des observations fascinantes suscitant un intérêt nouveau dans l'identification du rôle de l'environnement, et particulièrement des conditions météorologiques et climatiques, dans la dynamique des maladies infectieuses. Cet intérêt a été amplifié par les problèmes croissants de résistances aux antibiotiques, les émergences et les réémergences de certaines maladies infectieuses, la menace potentielle du bioterrorisme, et les débats concernant les changements climatiques. Au cours de ces dernières années les scientifiques et plusieurs organismes comme L'OMS ont publié des articles et rapports

Chapitre 6 - Apports du couplage études écologiques/données climatiques dans la compréhension de la dynamique des maladies infectieuses : exemple de la Guyane Française.

s'attardant sur le lien entre climat et santé [2-4]. Des progrès importants ont été réalisés aboutissant à l'incorporation des facteurs climatiques dans les modèles prédictifs de certaines maladies, comme notamment le choléra [5].

6.2. Saisonnalité et perturbations climatiques

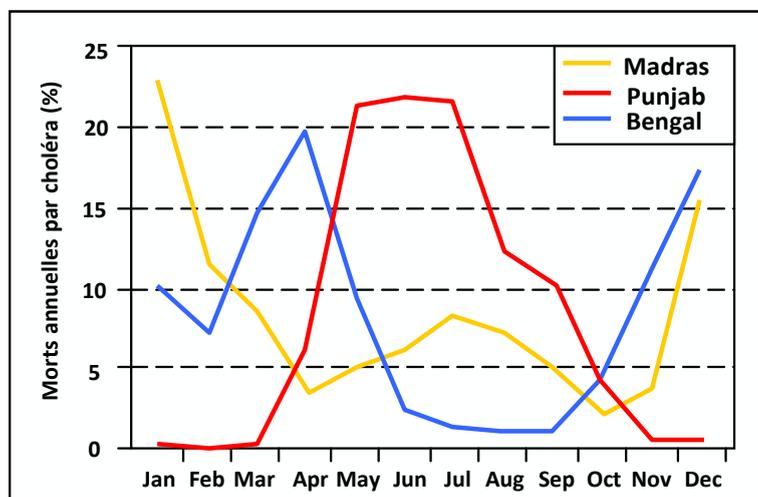
Le climat, de par ses inconstances, influe sur la dynamique des maladies infectieuses. Son action résulte souvent en une propagation complexe et non constante des agents infectieux dans les populations humaines. La variabilité du climat s'explique principalement par l'existence de deux types de phénomènes climatiques : la saisonnalité et les perturbations climatiques.

6.2.1 La succession des saisons

L'observation de changements cycliques dans les conditions climatiques a conduit les communautés humaines à diviser les années calendaires en saisons. Les saisons résultent de la révolution annuelle de la Terre autour du Soleil et de l'inclinaison relative de l'axe de la Terre par rapport au plan de révolution. En fonction de l'endroit où l'on se situe sur le globe, les caractéristiques et le découpage de ces saisons varient.

Dans les régions tempérées et polaires, les saisons sont marquées par des changements dans l'intensité de la lumière solaire qui atteint la surface de la Terre et on se base donc sur la température pour identifier les saisons. On y distingue généralement quatre saisons par année calendaire : le printemps, l'été, l'automne et l'hiver. Dans certaines régions tropicales et subtropicales, du fait de caractéristiques climatiques particulières, il est plus courant de distinguer la saison des pluies (ou humide, ou mousson) par rapport à la saison sèche, car les quantités de précipitations peuvent varier de façon plus spectaculaire que la température moyenne. Dans d'autres régions tropicales une division à trois voies (saison chaude, pluvieuse et fraîche) est utilisée. L'impact de cette saisonnalité sur la dynamique des maladies infectieuses peut être colossale en pratique. De nombreux exemples de pathologies affectées par la saisonnalité du climat peuvent être trouvés dans la littérature (Figure 27)[6-8].

Figure 27 : Saisonnalité de la mortalité associée au choléra dans trois provinces de l'Inde (Madras, Punjab, Bengal). Adaptée de [9].



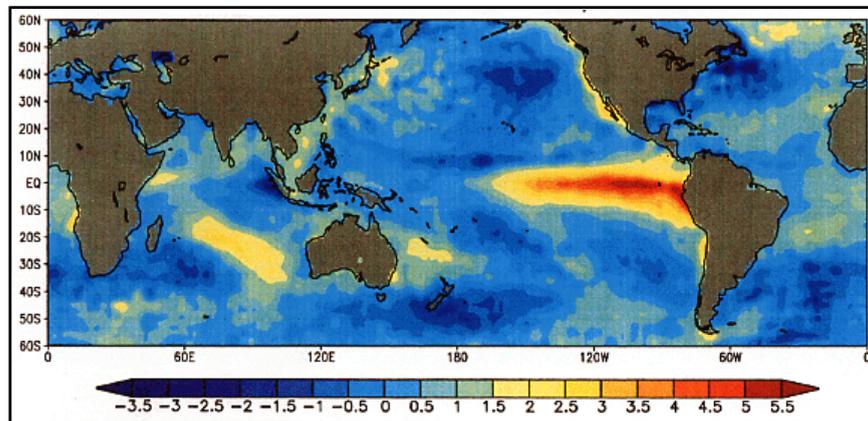
6.2.2 Les perturbations climatiques : focus sur le phénomène El Niño

Des phénomènes climatiques plus globaux influencent eux même ce climat local et viennent perturber les cycles climatiques habituels. Le réchauffement climatique est un exemple célèbre. Le phénomène El Niño, aussi très médiatisé, est également une de ces perturbations. Il est considéré comme le plus important cycle climatique contribuant à la variabilité interannuelle du climat et comme l'origine d'événements climatiques extrêmes tels que des pluies abondantes, des sécheresses et des tempêtes.

El Niño, littéralement « courant de l'Enfant Jésus » en espagnol, est nommé ainsi parce qu'il apparaît peu de temps après Noël. Il désigne, à l'origine, un courant côtier saisonnier au large du Pérou et de l'Équateur [10]. Ce terme est maintenant utilisé pour désigner une variation quasi-périodique du climat, différente du climat usuel, qui se produit dans l'Océan Pacifique. Cette variation apparaît à des intervalles de temps irréguliers de 2-7 ans et persiste entre neuf mois et deux ans.

El Niño est caractérisé par une élévation anormale de la température de l'océan dans la partie Est de l'océan Pacifique Sud, représentant une extension vers le Sud du courant chaud péruvien. L'épisode contraire est appelé par opposition La Niña. Il a été relié à un cycle de variation de la pression atmosphérique globale entre l'Est et l'Ouest du Pacifique que l'on nomme l'Oscillation Australe et l'on unit souvent les deux sous le titre de ENSO [11]. E.N.S.O est l'acronyme d'El Niño Southern Oscillation, c'est à dire "l'Oscillation Australe El Niño", nom donné par les climatologues et océanographes au 20^{ème} siècle aux variations interannuelles de la circulation océanique et atmosphérique observées à l'échelle de l'ensemble du Bassin Pacifique équatorial et tropical (Figure 28).

Figure 28 : Anomalies de températures à la surface des océans (en °C) lors de l'épisode El Niño de 1997



Le phénomène El Niño perturbe la configuration habituelle des précipitations ainsi que la circulation atmosphérique aux latitudes tropicales et peut avoir des répercussions à grande échelle sur le climat dans de nombreuses régions du monde. Son apparition déplace les zones de précipitations vers l'Est dans l'Océan Pacifique et empêche la remontée d'eau froide le long de la côte de l'Amérique du Sud. Sans que l'on soit encore capable d'expliquer toutes les relations physiques, El Niño fait partie à travers le globe d'anomalies dans la circulation qui peuvent dérouter les cyclones tropicaux de leurs routes habituelles, déplacer les zones de précipitation et de sécheresse et changer localement le niveau de la mer par le changement de la pression moyenne [12].

Chapitre 6 - Apports du couplage études écologiques/données climatiques dans la compréhension de la dynamique des maladies infectieuses : exemple de la Guyane Française.

El Niño tient sa renommée des catastrophes climatiques qu'il engendre et qui ont déjà causé des milliers de morts [13, 14]. Pour l'heure, son impact sur les maladies infectieuses semblent concerner essentiellement les éruptions violentes de maladies telles que le choléra [15, 16], les hantavirus [17], les diarrhées déshydratantes [18], ainsi que le développement d'épidémies lié à des pathogènes ayant une forte composante vectorielle telles que la dengue [19-21], le paludisme [22-24] et la fièvre de la vallée du Rift [25].

6.3. Les mécanismes d'actions du climat sur la transmission des maladies infectieuses

De nombreuses maladies infectieuses (en particulier celles à transmission vectorielle) sont géographiquement et temporellement limitées par les variations de variables environnementales locales comme la température, l'humidité et les précipitations. L'impact direct du climat sur les maladies infectieuses peut se produire par trois voies principales: les effets sur le comportement humain, les effets sur l'agent pathogène de la maladie; et les effets sur le vecteur de la maladie.

6.3.1. Influence du climat sur l'homme et ses comportements : implications pour les maladies infectieuses

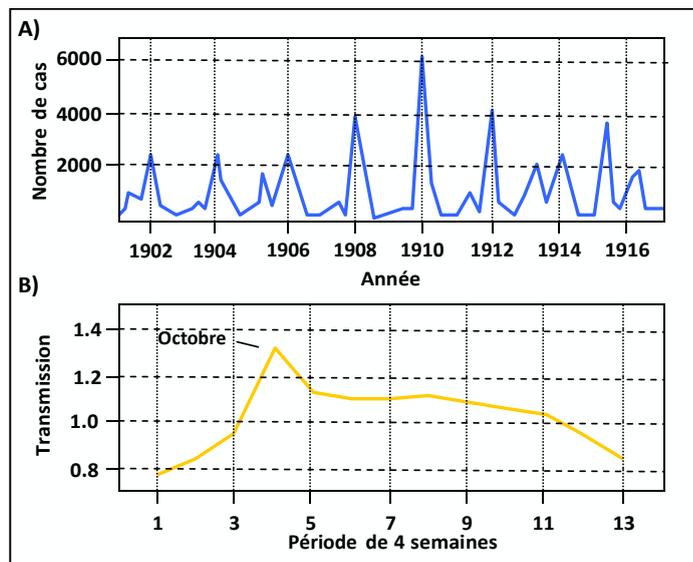
La variabilité du climat a une influence directe sur le comportement humain (par exemple les occupations, migrations, et modes de vie saisonniers), qui à leur tour peuvent avoir un impact significatif sur les modes de transmission des maladies. Les variations annuelles des interactions sociales dans les communautés humaines (dessinées indirectement par le climat) peuvent modifier l'intensité de la transmission de nombreuses infections contagieuses transmises par voie directe. Chez les humains, une série d'événements calendaires régulent les contacts par lesquelles les maladies contagieuses telles que les oreillons, la varicelle et la grippe se propagent [26-28]. La forte saisonnalité des infections de grippe en Europe, par exemple, peut être pensée comme une conséquence de la tendance accrue chez les humains à consacrer une plus grande part de leur temps à des activités d'intérieur pendant les mois d'hiver. Inversement, le pic d'incidence de la gastro-entérite dans les régions tempérées des pays développés pendant les mois d'été peut être lié à des changements du comportement humain (par exemple, plus de pique-niques et autres plats cuisinés à l'extérieur) associée à des températures plus chaudes.

L'exemple le plus frappant de ce type de phénomènes est donné par la rougeole, une maladie infantile contagieuse. Dans l'une des premières études de la saisonnalité des dynamiques des maladies infectieuses, Soper en se concentrant sur les notifications de rougeole à Glasgow de 1901 à 1917, a montré un pic évident de la transmission en octobre, qui coïncidait avec le début du calendrier scolaire [29] (Figure 29). Depuis lors, un certain nombre de chercheurs [27, 28], en utilisant une méthodologie plus raffinée et en examinant des données supplémentaires, ont conclu que les transmissions de maladies infantiles sont dans une grande proportion un reflet de l'augmentation régulière des taux de contact durant l'ouverture de la période scolaire et de sa diminution pendant les vacances.

Partie 2 : Travaux scientifiques

D'autre part, ces dernières années des travaux semblent montrer qu'il existe chez l'homme des rythmes annuels qui influencent les fonctions immunitaires [30, 31]. Des études récentes sur les humains et d'autres mammifères suggèrent que les systèmes immunitaires sont affaiblis pendant l'hiver, et par des conditions météorologiques difficiles et cela d'une manière qui pourrait affecter leur capacité à se défendre contre les maladies infectieuses. Plusieurs mécanismes pourraient entraîner des variations saisonnières dans les défenses de l'hôte. Ces changements dans l'immunité peuvent influencer la propagation du parasite par des probabilités de transmission plus importantes ou des taux de récupération plus faibles (prolongeant la durée de la période infectieuse).

Figure 29 : (A) Fluctuation du nombre de cas de rougeoles à Glasgow sur la période 1901-1917 (B) Intensité de la transmission de la rougeole. Adaptée de [29].



6.3.2. Influence du climat sur les pathogènes

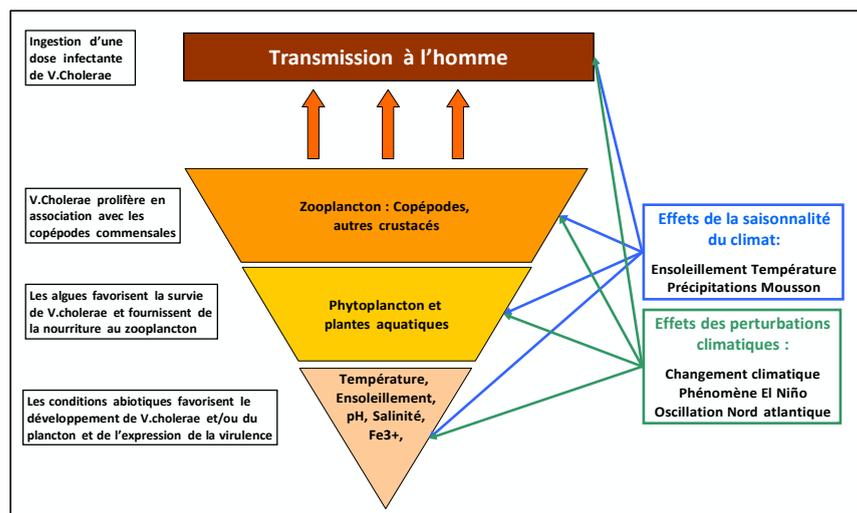
Pour de nombreux parasites intestinaux, les stades infectieux relâchés dans l'environnement sont vulnérables aux variations de températures, d'humidité et de pluviométrie avant leur rencontre avec un nouvel hôte [32, 33]. Chez les humains, de fortes variations saisonnières des prévalences des nématodes gastro-intestinaux sont régulièrement retrouvées. Pour les parasites dont les œufs sont retrouvés dans les selles, la température peut influencer à la fois le taux de développement du parasite et la survie des formes externes. Pour le développement des stades larvaires, la température doit en premier lieu excéder un certain seuil. L'augmentation de la température ambiante au-dessus de ce seuil peut ensuite raccourcir le temps nécessaire pour le développement de l'agent pathogène. La présence d'humidité, est également très importante pour le développement, la survie, et la mobilité des larves helminthiques. Des conditions de sécheresse peuvent limiter la transmission et provoquer la migration des larves dans le sol, bien que des pluies excessives peuvent « lessiver » les œufs et les stades infectieux [34].

Cependant l'influence du climat sur les maladies dont les formes infectieuses sont relâchées dans l'environnement est souvent beaucoup plus complexe que les mécanismes décrits plus haut. Le choléra est un bon exemple pour illustrer cette complexité. La dynamique du choléra dans les régions endémiques montrent

Chapitre 6 - Apports du couplage études écologiques/données climatiques dans la compréhension de la dynamique des maladies infectieuses : exemple de la Guyane Française.

des cycles saisonniers marqués et réguliers [9]. Le défi majeur dans l'explication du rôle du climat dans les épidémies de choléra, est que presque toutes les variables météorologiques et environnementales, notamment la pluviométrie et la température, sont sous l'influence omniprésente de la mousson. Ainsi, l'humidité semble être importante pour la transmission, en particulier dans les régions plus sèches, mais les inondations saisonnières peuvent également réduire les concentrations de bactéries dans l'environnement et réduire la salinité à un niveau défavorable [9]. Les changements saisonniers de température peuvent de plus influencer sur le taux de croissance du pathogène et la prolifération de plancton dans les habitats aquatiques d'une manière qui pourrait augmenter la survie du pathogène à travers son attachement aux copépodes et aux cyanobactéries [35]. Cette complexité de l'action du climat sur la dynamique du choléra réside dans son action simultanée sur les deux principales voies de transmission du choléra : la transmission primaire (des réservoirs aquatiques à la population humaine), et la transmission secondaire (d'humain à humain via la transmission féco-orale) (Figure 30).

Figure 30: Modèle hiérarchique montrant l'influence du climat sur la transmission du choléra. Modifiée à partir de [36].



6.3.3. Influence du climat sur les vecteurs

Les maladies à transmission vectorielle sont celles qui exigent un hôte (le plus souvent un arthropode) pour transmettre l'agent pathogène. La variabilité du climat et de la météorologie peut affecter à la fois les populations de vecteurs et à la fois l'amplification de l'agent pathogène dans les différents hôtes. La répartition géographique et la dynamique des populations de vecteurs sont ainsi étroitement liées à des conditions spécifiques de température, de précipitations et d'humidité. Les maladies vectorielles à virus, protozoaires et nématodes filaires sont particulièrement susceptibles de covarier avec les conditions environnementales [37, 38] (Tableau 4).

Dans de nombreuses régions, les variations saisonnières de la température limitent l'abondance des moustiques, des tiques et des autres arthropodes vecteurs qui meurent, ne parviennent pas à se développer ou deviennent moins actifs du fait de températures inférieures à celles communément rencontrées durant certaines périodes [39]. Par exemple en zones tempérées, dans le cas de la maladie de Lyme et des infections

Partie 2 : Travaux scientifiques

transmises par les tiques, des hivers trop froids peuvent réduire le nombre de nymphes de l'hôte porteur présents durant les mois d'été et d'automne et freiner ainsi la transmission de ces maladies durant cette même période [40].

Tableau 4 : Exemples de facteurs climatiques influençant la transmission et la distributions de maladies vectorielles. Adaptée de [41].

Maladie (agent causatif)	Vecteurs	Facteurs climatiques d'influence retrouvés dans la littérature	Effet de la variabilité climatique
Maladies parasitaires vectorielles			
Paludisme (<i>P. vivax</i> , <i>P. falciparum</i>)	Moustiques	Température, précipitations, humidité, El Niño, Temperature de surface des mers	Distribution de la maladie, développement du pathogène dans le vecteur, développement, reproduction, activité, distribution et abondance des vecteurs, intensité et schéma de la transmission, épidémies
Leishmaniose (<i>Leishmania spp.</i>)	phlébotomes	Température, précipitations, El Niño	Incidence de la maladie, épidémie, abondance, comportement et distribution des vecteurs
Maladie de Chagas (<i>Trypanosoma cruzi</i>)	triatomes	Température, précipitations, humidité, El Niño	Distribution des vecteurs, augmentation de l'infestation des maisons par les vecteurs
Onchocercose (<i>Onchocerca volvulus</i>)	Simulies	Température	Intensité de la transmission
Arboviroses			
Dengue (<i>Virus de la dengue</i>)	Moustiques	Température, précipitations	Epidémies, reproduction du moustique, abondance, intensité de la transmission (période extrinsèque de transmission)
Fièvre jaune (<i>Virus de la fièvre jaune</i>)	Moustiques	Température, précipitations	Epidémies, reproduction du moustique, abondance, intensité de la transmission (période extrinsèque de transmission)
Chikungunya (<i>Virus du chikungunya</i>)	Moustiques	Température, précipitations	Epidémies, reproduction du moustique, abondance, intensité de la transmission (période extrinsèque de transmission)
Fièvre à virus West Nile (<i>Virus West Nile</i>)	Moustiques	Température, précipitations	taux de transmission, développement du pathogène dans le vecteur, distribution de la maladie et du vecteur
Fièvre de la vallée du Rift (<i>Virus de la fièvre de la vallée du Rift</i>)	Moustiques	Précipitations, Temperature de surface des mers	Epidémies, reproduction du moustique, abondance, intensité de la transmission (période extrinsèque de transmission)
Fièvre de la Ross River (<i>Virus de la fièvre de la Ross River</i>)	Moustiques	Température, précipitations, Temperature de surface des mers	Epidémies, reproduction du moustique, abondance, intensité de la transmission (période extrinsèque de transmission)
Encéphalite à tiques (<i>Virus de l'encéphalite à tiques</i>)	Tiques	Température, précipitations, humidité	Distribution du vecteur, phénologie de la recherche d'hôte par le vecteur
Maladies bactériennes et rickettsiennes			
Borréliose de Lyme (<i>B. burgdorferi</i> , <i>B. garinii</i> , <i>B. afzelii</i> , ou autres <i>Borrelia</i>)	Tiques	Température, précipitations, humidité	Fréquence des cas, phénologie de la recherche d'hôte par le vecteur, distribution du vecteur
Tularémie (<i>Francisella tularensis</i>)	Tiques	Température, précipitations	Fréquence des cas, et apparition
Anaplasmose granulocytaire humaine (<i>Anaplasma phagocytophilum</i>)	Tiques	Température, précipitations	Distribution du vecteur, phénologie de la recherche d'hôte par le vecteur
Ehrlichiose monocytique humaine (<i>Ehrlichia chaffeensis</i>)	Tiques	Température, précipitations	Phénologie de la recherche d'hôte par le vecteur
Peste (<i>Yersinia pestis</i>)	Tiques	Température, précipitations, humidité, El Niño	Developpement et maintenance du pathogène dans le vecteur, survie et reproduction des vecteurs et hôtes, distributions de la maladie, ocurrences de pandémies historiques et d'épidémies régionales

Dans les régions chaudes, les précipitations saisonnières peuvent augmenter l'abondance des moustiques et des autres vecteurs ayant des stades larvaires aquatiques, et dont la reproduction est étroitement liée à la disponibilité en sites de reproduction [42, 43]. Dans d'autres cas, l'excès de pluie peut avoir des effets catastrophiques sur les populations de vecteurs locaux en cas d'inondation durant lesquelles un « lessivage » des sites de reproduction s'opèrent. Par ailleurs, les moustiques atteignent leur maturité sexuelle plus tôt et s'alimentent plus fréquemment sous des températures plus chaudes, ce qui peut potentiellement augmenter le taux de transmission du pathogène. En plus des effets sur l'abondance des vecteurs et des comportements, la température influe sur le taux de développement du parasite dans les vecteurs. Par exemple, le taux de répllication du virus de la dengue et des parasites du paludisme transmis par les moustiques augmente à des températures plus chaudes [44, 45]. Ceci est important puisque, avec des conditions plus fraîches en automne ou au printemps, les pathogènes risquent de ne pas mûrir assez rapidement pour être transmis avant la mort des vecteurs adultes, d'où une réduction potentielle de la proportion de moustiques infectieux et donc du risque d'infection.

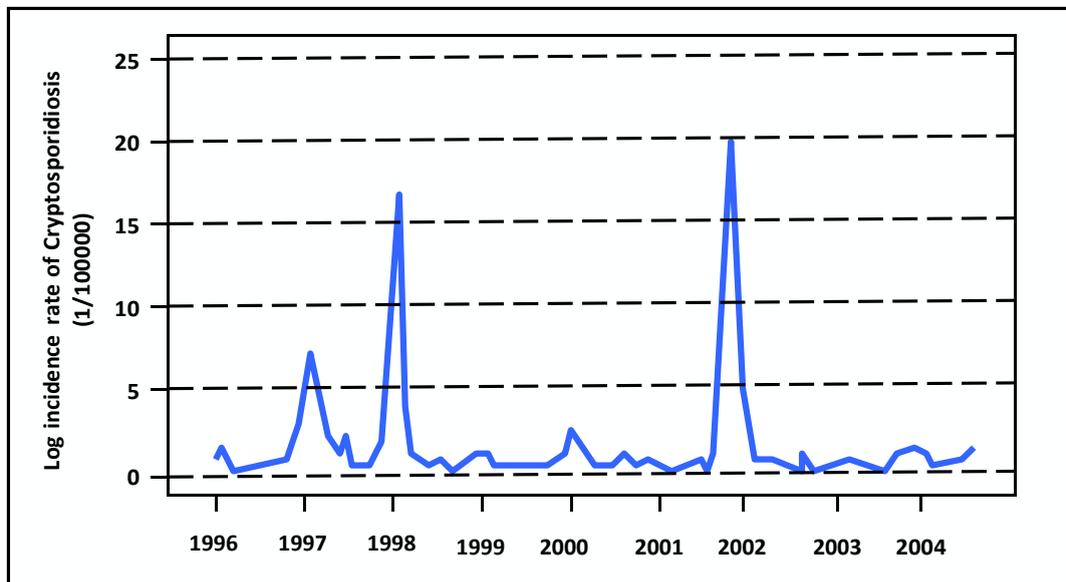
6.4. Etudes écologiques et données libres : des outils indispensables en épidémiologie environnementale

6.4.1 Les séries temporelles : un outil de choix pour l'étude du climat

Pour étudier l'impact du climat sur l'état de santé d'une population, outre les schémas classiques de l'épidémiologie observationnelle, c'est à dire les études de type cas-témoins ou de cohortes (et révéler une liaison statistique entre un paramètre climatique particulier et un état sanitaire) il est possible de mettre en œuvre des méthodes différentes, basées sur l'enregistrement en continu d'indicateurs pertinents de l'état de la santé d'une population. Ces indicateurs étant définis à une échelle populationnelle, leurs analyses nécessitent des méthodes statistiques particulières appartenant aux méthodes écologiques : les analyses des séries temporelles.

En statistiques, ou dans les sciences du traitement du signal, de l'économétrie et de la finance mathématique, une série temporelle (ou chronologique) est une séquence de points, mesurée à des temps successifs espacés par un pas de temps réguliers. Des exemples de séries temporelles sont les valeurs de clôture quotidiennes de l'indice Dow Jones, le volume du débit annuel du Nil à Assouan ou l'évolution de l'incidence de pathologies particulières comme la cryptosporidiose (Figure 31).

Figure 31 : Evolution temporelle de l'incidence de la cryptosporidiose à Brisbane (Australie) de 1996 à 2004.
Adaptée de [46].



Les méthodes statistiques appliquées aux séries chronologiques ont été initialement développées en économétrie, puis utilisées dans de nombreux autres domaines, comme l'écologie, la physique ou les sciences de l'ingénieur. Les attentes initiales des chercheurs résidaient dans la prédiction, et l'objectif principal était ainsi de produire une estimation précise des mesures à venir, étant donnée une série observée. Les approches statistiques standards adoptées à cette fin s'appuyaient généralement sur les processus autorégressifs à moyenne mobile (ARIMA) et ses modèles dérivés [47]. Dans les 20 dernières années, ce type d'analyses a

Partie 2 : Travaux scientifiques

commencé à être particulièrement utilisé en épidémiologie en raison de la grande disponibilité de séries de données administratives ou médicales systématiquement collectées, comme les données de comptage associées à la mortalité ou à la morbidité d'une pathologie particulière. En épidémiologie, les méthodes d'analyse des séries chronologiques ont ainsi été soumises à un développement méthodologique intense durant les 15 dernières années. En plus de l'intérêt initial de prédiction, un autre objectif très important de l'analyse des séries temporelles dans le domaine biomédical est d'évaluer l'association entre un résultat et une ou plusieurs séries de prédicteurs : l'accent est mis sur l'estimation plutôt que sur la prédiction.

Les tendances temporelles d'indicateurs tels que les incidences ou prévalences de maladies, ou de comportements comme le tabagisme sont maintenant souvent utilisées par les professionnels de santé publique pour l'aide à l'évaluation des besoins de la santé, la planification des services, et l'élaboration des politiques. L'examen de ce type de données permet aussi de prédire l'avenir et les fréquences des taux d'occurrence. Les études sur les tendances temporelles peuvent se concentrer sur les éléments suivants:

- Les changements temporels d'un indicateur. Ce type d'études peut être utilisé par exemple si l'on veut savoir si l'utilisation d'un service a augmenté ou diminué au fil du temps, et avec quelle cinétique l'augmentation ou la diminution s'est produite.
- L'étude de l'impact d'un facteur temporel sur un autre. Les exemples les plus célèbres de ce type d'études sont celles cherchant à quantifier l'impact de la pollution de l'air sur la fréquentation des services hospitaliers ou les incidences des maladies respiratoires et cardio-vasculaires.
- La comparaison d'une période à une autre période de temps. Par exemple, on peut chercher à évaluer l'impact d'un programme de désaccoutumance au tabac, en comparant les taux de tabagisme avant et après l'événement.
- La comparaison d'une zone géographique ou d'une population à une autre.
- La réalisation de projections futures pour aider par exemple à la planification des services de santé en estimant les besoins et les ressources susceptibles.

Le grand avantage des études de séries temporelles est d'utiliser des données généralement facilement accessibles car mesurées en routine (données de mortalité dans une population, données d'hospitalisation, etc.). D'autre part, les analyses de séries temporelles, bénéficiant souvent de longues périodes de données, voient leur puissance statistique être tout à fait honorable. Dans le domaine infectieux, beaucoup de maladies sont très bien documentées temporellement autant d'un point de vue des occurrences que de leurs impacts. Leur utilisation dans ce domaine s'est donc généralisée ces dernières années [18, 46, 48-50]. Dans l'étude de l'impact du climat et de ses perturbations, ces méthodes ont un avantage supplémentaire: elles reposent sur la grande disponibilité des données climatiques.

6.4.2. Les données climatiques : des données largement ouvertes à la communauté scientifique

Les données climatiques sont des données très facilement accessibles à la communauté de chercheurs. Ces dernières connaissent en effet un niveau d'ouverture très important.

Chapitre 6 - Apports du couplage études écologiques/données climatiques dans la compréhension de la dynamique des maladies infectieuses : exemple de la Guyane Française.

- **Les mesures climatiques locales**

En France, l'organisme français de météorologie, Météo-France, est un établissement public administratif [51]. Il est chargé de la prévision et de l'étude des phénomènes météorologiques. L'observation du climat est le premier maillon de la chaîne de l'activité de ce type d'établissement. Les organismes nationaux chargés de la surveillance du climat disposent maintenant d'une batterie d'appareils de mesures (satellites, réseaux de stations de mesures au sol, bouées, radiosondages, capteurs embarqués sur les avions de ligne,...) pour réaliser leurs tâches : surveiller et prédire le climat. Dans ce but, des données comme les précipitations, les températures, l'humidité sont enregistrées quotidiennement dans des stations disséminées sur l'ensemble du territoire pour rendre compte localement de la variabilité du climat. En France, une forte collaboration entre Météo-France et la recherche publique existe depuis longtemps. L'accès aux données météorologiques est grandement facilité pour les chercheurs de la recherche publique : Les données climatiques passées et présentes ont ainsi un statut proche du libre : sous réserve de payer les frais de port, les données brutes sont gracieusement fournies et expédiées par Météo-France si ces dernières sont destinées à des activités de recherches. Ce principe est largement répandu dans d'autres pays du monde.

- **Les mesures des perturbations climatiques : l'exemple d'El Niño**

Pour surveiller et tenter de prédire le phénomène El Niño, des indices de mesures quantitatifs précis et à une échelle de temps assez fine ont ainsi été créés. Il en existe plusieurs :

- Les températures de surface du Pacifique (en anglais Sea Surface Temperatures : SST). 4 régions ont été définies dans le Pacifique (respectivement nommées NIÑO 1 à NIÑO 4). Chaque région donne différents types d'informations sur le phénomène El Niño. A partir de ces dernières, des anomalies de température de surface sont alors définies comme les écarts dans une région donnée par rapport à la moyenne à long terme des températures durant la période 1961-1990. Chacun de ces indices est disponible gratuitement sur le site Internet de la NOAA [6].

- L'index de l'Oscillation Australe (en anglais Southern Oscillation Index : SOI). Ce dernier est un index défini comme la différence normalisée de pression atmosphérique entre Tahiti en Polynésie Française et Darwin en Australie. Des différences minimales peuvent exister dans les calculs de différents centres. Les événements El Niño sont associés avec des valeurs négatives importantes et les événements La Niña avec des valeurs positives importantes. Bien que la corrélation entre les années El Niño et les valeurs du SOI ne soit pas parfaite elle est cependant très élevée. Ici aussi ces valeurs sont accessibles publiquement sur le site Internet de beaucoup d'organismes spécialisés [6, 7].

- L'index multivarié ENSO (en anglais Multivariate ENSO Index : MEI). Ce dernier a été développé dans les années 90 pour quantifier finement les variations de l'oscillation australe El Niño [6]. Cet index est dérivé des valeurs climatiques habituelles (température, vent, pression à la surface de l'eau, température de l'air et ensoleillement total). Il est également disponible gratuitement sur le site Internet de la NOAA [6].

6.5. La Guyane française, un terrain de jeux idéal pour l'étude de l'influence du climat: exemple du paludisme, de l'histoplasmosse disséminée, et de la leishmaniose cutanée.

6.5.1 Le climat guyanais : un climat équatorial humide propices aux pathogènes

La Guyane bénéficie d'un climat de type équatorial humide, tempéré par les alizés. Elle se situe dans une zone de circulation atmosphérique Est/Ouest induite par les deux ceintures anticycloniques subtropicales (Açores au Nord et Ste Hélène au Sud) qui s'affrontent au niveau de la Zone Intertropicale de Convergence (ZIC). La ZIC balaie la Guyane deux fois par an au rythme d'une migration Nord-Sud et détermine un cycle saisonnier caractérisé par quatre saisons de durées inégales: une petite saison des pluies de mi-novembre à fin janvier; une petite saison sèche ou «petit été de mars» se produisant en général entre début février et mi-mars; une grande saison des pluies de fin mars à fin juillet ; une grande saison sèche de fin juillet à mi-novembre. La pluviométrie annuelle est de 3 mètres en moyenne sur la bande côtière et de 2,5 mètres dans les régions de l'intérieur. Les variations peuvent être importantes d'une année à l'autre pour une même station. Cette variabilité interannuelle est partiellement liée au phénomène de circulation atmosphérique global El Niño naissant dans le Pacifique Sud. Durant un épisode El Niño, il est attendu que le climat guyanais soit plus chaud et plus sec, alors qu'un épisode La Niña devrait plutôt entraîner une augmentation des précipitations accompagnée de températures plus fraîches [52, 53]. Les températures moyennes sont quasi-constantes tout au long de l'année, autour de 26,5 °C. Toutefois en saison des pluies les températures sont plus faibles et les écarts diurnes moins importants qu'en saison sèche. L'humidité relative est toujours supérieure à 75% et peut atteindre 100%. L'ensoleillement, modulé par la nébulosité, est en moyenne de 2200 heures par an, et apporte une énergie moyenne de 6 kWh/m²/jour. Les vents sont faibles à modérés et relativement constants tout au long de l'année.

La distribution de la biodiversité se répartit généralement selon un gradient latitudinal. Le nombre d'espèces animales et végétales augmente au fur et à mesure que l'on se rapproche de l'équateur [54]. Ceci s'observe pour de nombreux groupes taxonomiques. Bien que la majorité des études sur la biodiversité se soit penchée sur des organismes libres, de récents travaux [55] montrent que cette loi s'applique également aux pathogènes humains. De par sa proximité de l'équateur, la Guyane Française possède donc un climat et un terrain particulièrement favorable à la survie et au développement de nombreux pathogènes. Le paludisme, la leishmaniose cutanée et l'histoplasmosse disséminée font partie des maladies infectieuses rencontrées quotidiennement sous ces latitudes.

6.5.2 Problématiques autour du climat et du paludisme, de la leishmaniose cutanée et de l'histoplasmosse disséminée en Guyane française

- Climat et paludisme en Guyane française

Chapitre 6 - Apports du couplage études écologiques/données climatiques dans la compréhension de la dynamique des maladies infectieuses : exemple de la Guyane Française.

Le paludisme est un important problème de santé publique en Guyane française avec au cours de la dernière décennie un nombre de cas annuel moyen de l'ordre de 4.000, bien que non endémique sur la région du littoral là où se concentre la grande majorité de la population. En effet, les régions d'endémie palustre ont les plus forts taux d'incidence d'Amérique du Sud [56] même si une diminution se dessine depuis 2008 (moins de 3.000 cas en 2010). Ces foyers sont situés le long des grands fleuves frontaliers: Surinam à l'Ouest (foyer du Maroni), Brésil à l'

Figure 32 : Risque fort risque de palu

ident aux zones à 8].



Trois espèces sont présentes. *P.vivax* et *P.falciparum* prédominent largement. *P.malariae* est beaucoup plus rare. Pendant longtemps *P.falciparum* prédominait. A partir de 2002, *P.vivax* est devenu progressivement majoritaire avec un maximum en 2008 (2/3 des cas). Cette évolution fut la conséquence d'une diminution importante du paludisme dans la région ouest de la Guyane où *P.falciparum* dominait largement associée à une forte recrudescence de *P.vivax* à l'est du département [57]. Ce phénomène était lié avant aux activités d'orpaillage menées le plus souvent hors de tout contrôle par des immigrants brésiliens. Toutefois, le nombre de cas répertorié a fortement diminué depuis 2 ans y compris au sein des populations autochtones de la région. Les moyens disponibles dans ce département français (moyens financiers et équipements sanitaires) contribuent à la relative rareté des formes sévères et une faible mortalité, et ceci malgré des caractéristiques épidémiologiques défavorables (régions enclavées d'accès difficile, importance des déplacements de certaines populations exposées, souvent en situation irrégulière, souches de *P.falciparum* chimiorésistant).

Anopheles darlingi constitue depuis toujours le principal vecteur. Malgré tout, les caractéristiques écologiques des différents vecteurs du paludisme en Guyane française ne sont pas complètement élucidées [59, 60]. En l'absence de données entomologiques, étudier les facteurs environnementaux et particulièrement

Partie 2 : Travaux scientifiques

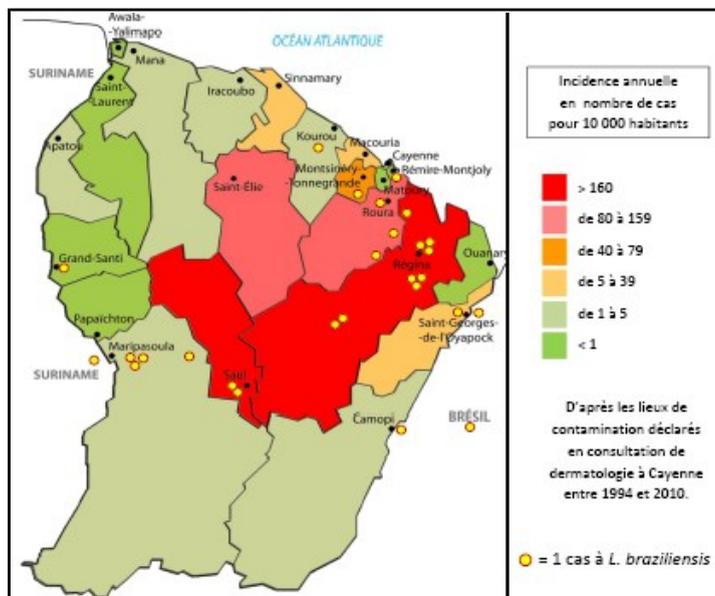
le climat conditionnant la rencontre entre les humains et les parasites, en se basant l'évolution temporelle des incidences pourrait être une stratégie intéressante pour palier à ce problème et fournir des pistes de réflexions pour la mise en place de stratégies de contrôle et de prévention efficaces.

D'autre part, dans les zones géographiques où le climat est influencé par l'Oscillation Australe El Niño, une très forte corrélation entre les épidémies de paludisme et les épisodes El Niño a souvent été observée [61] et cela particulièrement dans plusieurs pays d'Amérique du Sud [24, 62, 63]. A un niveau régional, la Guyane Française est située dans une zone potentiellement à risque de sécheresse en période El Niño et ou un risque épidémique vis-à-vis du paludisme n'est pas à écarter [61]. Une seule étude, faite à l'échelle populationnelle et portant sur plusieurs pays d'Amérique du Sud, a jusque-là été réalisée pour quantifier l'impact d'El Niño sur la circulation du paludisme en Guyane Française [24]. Réalisée avec des données partielles et sur une période d'étude trop courte, les résultats de cette dernière ne permettent pas de conclure sur le rôle exact d'El Niño dans la circulation du paludisme. Certaines études suggèrent qu'il est possible d'utiliser les relations statistiques qui existent entre El Niño et l'apparition d'épidémies pour prédire les éruptions de maladies [63]. Etudier la relation entre paludisme et El Niño en Guyane Française pourrait ainsi déboucher sur une meilleure anticipation des épidémies de paludisme en Guyane Française.

- **Climat et leishmaniose cutanée en Guyane française**

Le même raisonnement peut s'appliquer à une autre pathologie à forte composante vectorielle endémique en Guyane : la leishmaniose cutanée [65] (Figure 33).

Figure 33 : Incidence annuelle par commune, calculée à partir des cas de leishmaniose recensés en dermatologie de 1994 à 2010 et pour lesquels le lieu de contamination était renseigné (1359 cas). Adaptée de [65].



L'incidence annuelle y est estimée à 2 cas pour 1000 habitants [66]. Depuis les premières observations de *Leishmania (Vianna) guyanensis* par Floch dans les années 1950, quatre autres espèces de leishmanies autochtones ont été décrites chez l'homme en Guyane : *L.(V) braziliensis*, *L.(L) amazonensis*, *L.(V) lainsoni* et

Chapitre 6 - Apports du couplage études écologiques/données climatiques dans la compréhension de la dynamique des maladies infectieuses : exemple de la Guyane Française.

L.(V) naiffi. Ces espèces ont toutes un tropisme cutané. *L.(V) guyanensis* est l'espèce prédominante [67] ; le plus souvent responsable de formes localisées, elle peut parfois entraîner des formes diffuses. Les vecteurs exclusifs de la leishmaniose sont les phlébotomes [68]. Ce sont des diptères hématophages mesurant de 2 à 5 mm. Le cycle de *L. guyanensis* se déroule dans la canopée avec le phlébotome arboricole *Lu. umbratilis* comme vecteur et au moins un mammifère réservoir, le paresseux didactyle [69]. Comme pour d'autres maladies vectorielles, la variation saisonnière du nombre de cas suggère un rôle du climat dans la dynamique de transmission de la leishmaniose cutanée. Au Venezuela et en Argentine, des corrélations de la densité vectorielle et de variables environnementales diverses ont été décrites [70, 71] ainsi que la corrélation de la densité vectorielle et du nombre de cas de leishmaniose [70, 72]. L'influence des variations climatiques interannuelles liées au phénomène El Niño sur les niveaux épidémiques de la leishmaniose viscérale a été décrite au Brésil [48]. En Guyane, il a été décrit que l'ensoleillement et le rayonnement UV pouvaient influencer la présentation clinique de la leishmaniose cutanée et l'efficacité du traitement par pentamidine [73, 74]. En suivant l'hypothèse d'une corrélation entre variations climatiques et incidence de la leishmaniose, investiguer la corrélation du climat et des cas de leishmaniose recensés au centre hospitalier de Cayenne pourrait permettre d'identifier des paramètres climatiques servant d'indicateur des années et des périodes de l'année à risque de leishmaniose.

- **Climat et histoplasmosse disséminée chez les patients VIH+ en Guyane Française**

Avec une incidence des cas de sida déclarés dix fois supérieure à l'Île de France et un taux de prévalence de l'infection par le VIH chez les parturientes régulièrement supérieur à 1%, la Guyane est le département français où la situation de l'épidémie à VIH est la plus préoccupante [75, 76]. En Guyane, avec une incidence de 2,9 pour 100 personnes années chez les personnes infectées par le VIH, l'histoplasmosse est la première pathologie inaugurale au stade sida et, la première cause de décès au cours de l'infection par le VIH [78, 79]. Les histoplasmoses sont des infections fongiques, dues à des champignons dimorphiques du genre *Histoplasma*. La seule variante présente en Guyane est *Histoplasma capsulatum* var. *capsulatum* (Hcc), agent de l'histoplasmosse américaine ou histoplasmosse à petites formes ou « maladie de Darling » (Figure 34).

Figure 34 : Carte mondiale de la répartition des cas décrits d'histoplasmosse américaine. Adaptée de [80].



La transmission à l'homme se fait essentiellement par voie respiratoire, lors de l'inhalation de poussières riches en spores d'Hcc ventilées dans l'environnement. Les formes mycéliennes, produisant ces

Partie 2 : Travaux scientifiques

spores se retrouvent essentiellement dans les sols contaminés par les déjections d'oiseaux grégaires ou de chauve-souris (poulaillers, bâtiments à l'abandon, grottes, chantiers de démolition, etc.). En Guyane française, les cliniciens assument en général que l'histoplasmosse chez les patients VIH+ est une réactivation d'une infection passée survenant suite à une immunodépression sévère. Mais la question fait débat [81, 82]. Du fait de la forte saisonnalité des activités humaines et de la forte composante environnementale du cycle de l'histoplasmosse, l'étude de l'impact temporel du climat sur les incidences d'histoplasmosse disséminée pourrait fournir des éléments de réponse à ce questionnement. Elles pourraient aussi permettre la mise en place de voies thérapeutiques adaptées à cette pathologie en Guyane Française.

6.6. Contexte et origine de ces études

L'impact du climat sur les maladies infectieuses en Guyane constitue depuis longtemps un sujet de réflexion et d'intérêt pour le Pr. Carme (Chef de service du laboratoire de parasitologie du CHAR) ainsi que pour le Pr. Couppié (chef de service du service de dermatologie du CHAR) et le Pr. Nacher (président du COREVIH Guyane).

A travers la réalisation de différentes thèses de médecine et de sciences orientées sur la parasitologie (paludisme) associée à des problématiques spécifiques au service de dermatologie (leishmaniose) et du VIH (histoplasmosse), des collaborations se sont établies avec les différents étudiants en thèse au sein du CIC-EC Antilles Guyane et de l'équipe EA3593 pour étudier plus profondément le rôle du climat sur ces différentes pathologies.

Durant ces études, en se penchant sur le lien entre climat et santé en Guyane Française nous avons découvert par hasard l'importante littérature se focalisant sur le lien entre épisodes El Niño et santé. Une recherche bibliographique plus poussée sur l'action d'El Niño en Amérique du Sud nous a convaincu de la pertinence de s'attarder plus longtemps sur son impact potentiel en Guyane Française.

Les études spécifiques au paludisme se sont faites principalement à partir de trois jeux de données :

- 1) Des données locales reflétant le niveau de paludisme à l'échelle d'un village que sont les données des villages de Camopi et de Cacao.
 - A Camopi, un village amérindien, le Pr. Carme a constitué depuis 10 ans une cohorte exhaustive des enfants âgés de moins de 5 ans [83]. L'ensemble des événements médicaux liés au paludisme pour l'ensemble des enfants de ce village est enregistré dans une base de données. Après discussion avec le Docteur Aurélia Stéphanie, travaillant activement sur cette cohorte dans le cadre de sa thèse de sciences, une collaboration fut mise en place et initia ce travail.
 - A Cacao, un village Hmong, où du fait d'une épidémie à *P.vivax* durant les années 2000, une cohorte des personnes résidentes avait été mise en place. Après discussion avec le Docteur Célia Basurko, alors en thèse de médecine de santé publique s'intéressant au lien entre paludisme et environnement sur cette cohorte, une collaboration fut également mise en place.

Chapitre 6 - Apports du couplage études écologiques/données climatiques dans la compréhension de la dynamique des maladies infectieuses : exemple de la Guyane Française.

- 2) Des données renseignant sur le niveau de paludisme régional que sont les données de diagnostics confirmés de paludisme faits dans le laboratoire de parasitologie du CHAR (le centre de référence en Guyane française)

Pour pouvoir étudier le rôle du climat dans la transmission du paludisme, du fait que *P.vivax* est l'espèce dominante depuis les années 2000 en Guyane, ainsi que du relatif échec de l'épidémiologie moléculaire à différencier les nouvelles infections des reviviscences [84], un travail préliminaire a dû être réalisé en amont pour mettre au point un critère statistique de substitution permettant de différencier efficacement ces deux types d'accès. C'est le sujet de la première étude présentée dans ce chapitre.

Durant l'année 2010 le Docteur Antoine Adenis, a réalisé au sein du CIC-EC Antilles Guyane son travail de thèse de médecine générale portant sur le thème « Etude comparative de la tuberculose et de l'histoplasmosse chez les patients infectés par le VIH ». Au cours de nombreuses discussions communes avec le Pr. Mathieu Nacher, le médecin délégué du CIC-EC Antilles Guyane, portant sur nos deux sujets de conversations principaux de l'époque, l'histoplasmosse disséminée chez les patients VIH+ d'un côté et l'impact du climat de l'autre, la zone d'ombre autour de l'impact potentiel du climat sur cette pathologie nous est apparue. A partir de ce constat, l'idée d'une étude spécifique à l'histoplasmosse fut évoquée et initia ce travail collaboratif utilisant les données des patients VIH suivis au CHAR de Cayenne.

Durant l'année 2011 le Docteur Amaury Roger, a réalisé au sein du CIC-EC Antilles Guyane son travail de thèse de médecine générale portant sur le thème « La leishmaniose cutanée en Guyane : Etudes épidémiologique, clinique et thérapeutique des cas recensés au centre hospitalier de Cayenne de 1994 à 2010 ». Le Pr. Couppié l'ayant chargé de confirmer le rôle prépondérant du climat sur les occurrences des cas de leishmaniose cutanée retrouvé dans une étude précédente, une collaboration se mis en place pour appliquer la même méthodologie que celle utilisée dans les études sur le paludisme et l'histoplasmosse disséminée.

6.7. Etude 1 : Détermination de la temporalité des reviviscences à *Plasmodium vivax* dans le village de Camopi, Guyane française

Introduction : Le paludisme est un problème de santé publique majeur en Guyane française, où *P.vivax* est devenu l'espèce de paludisme dominante dans les années 2000. Comme dans d'autres zones endémiques, il est important d'identifier la temporalité des reviviscences à *P.vivax* et d'essayer de distinguer efficacement les nouvelles infections des reviviscences.

Matériels/Méthodes : Cette étude a été menée chez les enfants nés entre le 1^{er} Janvier 2001 et le 31 Décembre 2008 à Camopi, un village amérindien situé dans la forêt amazonienne (n = 325), à partir des données d'une étude prospective de cohorte ouverte. Les taux d'attaque primaire et secondaire à *P.vivax* ont été calculés en utilisant une analyse de survie. A partir de la différence entre les taux d'attaques primaires et secondaires, cette étude visait à estimer indirectement le taux de reviviscence à *P.vivax* et d'en évaluer son évolution dans le temps.

Résultats : Sur les 1042 attaques de paludisme enregistrés, 689 (66%) étaient dues à *P.vivax* (sans infection mixte). 151 enfants ont eu leur première attaque à *P.vivax* dont 106 ayant eu aussi leurs deuxième attaque à *P.vivax*. En l'absence d'un traitement par primaquine, il a été démontré que les reviviscences à *P. vivax* survenaient essentiellement au cours des trois premiers mois après la première attaque. Globalement il a été estimé que 30% des enfants n'ont jamais eu de reviviscence, 42% ont eu une reviviscence avant le premier mois après l'attaque primaire, 59% avant le deuxième mois et 63% avant le troisième mois.

Conclusions : Cette étude a confirmé que le type des reviviscences à Camopi est compatible avec le schéma décrit pour la souche Chesson de *P.vivax* (tropicale). En outre, en raison de l'évolution temporelle du taux de reviviscence, une règle arbitraire simple de classification a pu être construite : avant 90 jours après la première attaque, l'attaque secondaire est une reviviscence, après 90 jours, c'est une nouvelle infection. Une gestion adaptée des cas de paludisme basée sur ces résultats peut être imaginée.

Publiée dans *Malaria Journal* le : 4 Décembre 2009

Research

Open Access

Determination of the *Plasmodium vivax* relapse pattern in Camopi, French Guiana

Matthieu Hanf¹, Aurélia Stéphani², Célia Basurko¹, Mathieu Nacher¹ and Bernard Carme^{*1,2}

Address: ¹Centre d'Investigation Clinique - Epidémiologie Clinique Antilles Guyane CIC-EC Inserm 802, Centre Hospitalier de Cayenne, Guyane Française and ²Laboratoire Hospitalo-Universitaire de Parasitologie et Mycologie Médicale, Equipe EA3593, UFR Médecine - Université des Antilles et de la Guyane, Cayenne, Guyane Française

Email: Matthieu Hanf - matthieu.hanf@ch-cayenne.fr; Aurélia Stéphani - aurelia.stefani@gmail.com; Célia Basurko - c.basurko@free.fr; Mathieu Nacher - m_nacher@lycos.com; Bernard Carme* - carme.bernard@wanadoo.fr

* Corresponding author

Published: 4 December 2009

Received: 1 August 2009

Malaria Journal 2009, **8**:278 doi:10.1186/1475-2875-8-278

Accepted: 4 December 2009

This article is available from: <http://www.malariajournal.com/content/8/1/278>

© 2009 Hanf et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Malaria is a major public health problem in French Guiana, where *Plasmodium vivax* has become the dominant malaria species since 2000. As in others endemic areas, it is important to specify the pattern of vivax malaria relapses and to try to discriminate efficiently re-infections from relapses.

Methods: This study was conducted in children born between January 1, 2001 and December 31, 2008 in Camopi, an Amerindian village located in the Amazon forest (n = 325), using an open cohort design. Primary and secondary attack rates of *P. vivax* were calculated using survival analysis. With the difference between the primary and secondary rates, this study aimed to estimate indirectly *P. vivax* relapse rate and evaluate its time evolution.

Results: Of the 1042 malaria attacks recorded, 689 (66%) were due to *P. vivax* (without mixed infection). One hundred and fifty one children had their primary attack with *P. vivax* and 106 had their two first attacks with *P. vivax*. In the absence of primaquine treatment, it was shown that *P. vivax* relapses mainly occurred during the first three months after the first attack. Thirty percent of children never had a relapse, 42% had a relapse before the first month after primary attack, 59% before the second month and 63% before the third month.

Conclusion: This study confirmed that the relapse pattern in Camopi was compatible with the pattern described for the *P. vivax* Chesson (tropical) strain. In addition, due to the relapse rate time evolution, a simple arbitrary classification rule could be constructed: before 90 days after the primary attack, the secondary attack is a relapse; after 90 days, it is a re-infection. Adapted management of malaria cases based on these results could be devised.

Background

Traditionally, *Plasmodium vivax* was considered as a relatively benign malaria parasite, which could be easily

treated. For the past 10 years, this has been questioned [1]. It has been shown that malaria could be involved in severe disease [2]. Furthermore, *P. vivax* has recently re-

emerged [3] and drug resistances, mainly chloroquino-resistance have developed in several areas[4]. Primaquine treatment, the only therapeutic option against relapse, might also be failing [5]. It requires a 14-day course and G6PD deficiency screening, which is difficult and costly in some areas.

In French Guiana, *P. vivax* has become the dominant malaria species since 2000, with more than 60% of malaria attacks [6]. Knowledge about its relapse pattern is limited. *Plasmodium falciparum* remains predominant on the Maroni River, where the maroon populations are mostly Duffy negative and, therefore, resistant to *P. vivax*. *Plasmodium malariae* has been observed in 2% of malaria cases.

Differentiation of re-infection from relapse is very difficult; most articles use a combined risk or rate of re-infection and relapse [7,8], or estimated the relapse risk based on the frequency of *P. vivax* during the non-transmission season [9,10]. Although molecular epidemiology introduced new methods to discriminate re-infection from relapse, these methods are very expensive and, in many cases, they could not address the question accurately [11-13]. In French Guiana, there is a widespread belief that *P. vivax* was of the 'tropical' type, which is characterized by an early primary attack, followed by a short latent period before appearance of frequent relapses during the next year, but no published evidence from French Guiana or elsewhere in South America supports this belief.

The study objectives were to determine the relapse pattern of *P. vivax* using an alternative method and to propose a simple classification to easily differentiate re-infections from relapses.

Methods

This research was carried out in Camopi, an Amerindian village (Wayampi and Emerillon) located in the Amazon forest along the Oyapock River, which marks the eastern border with Brazil. The incidence of malaria there is around 40% [14]. A retrospective study using an open cohort design was carried out with all Camopi children born between January 1, 2001 and December 31, 2008.

The children came to the health center approximately once a month, generally at the time of illness or for systematic visits (i.e. vaccination). Every six months, there is a verification that all newborn children are included in the cohort, that all children are present in the village, and that all the malaria data is collected. Data for children for whom follow-up by the health center had been interrupted was right censored at the interruption date. It was assumed that all malaria attacks were recorded at the local health centre [15]. The Camopi health center ensures free

early diagnosis and treatment of malaria cases. Because of the extreme isolation, no other health facilities (public or private) are available in this part of the Amazonian forest.

The list of all clinical malaria episodes, their date of occurrence, Plasmodium species identification were made in the Camopi health center by two physicians, two nurses, and one Amerindian assistant. All thin blood smears were first analysed in Camopi by a trained nurse and then checked by the parasitological unit of Cayenne General Hospital, the referent malaria center in French Guiana.

Malaria was defined as temperature $> 38^{\circ}\text{C}$ at the time of consultation or fever within the past 48 hours associated with Plasmodium asexual forms on a thin blood smear (detection threshold: 50 *Plasmodium*/ μl). During the study period, all confirmed cases of *P. vivax* received a three-day treatment of chloroquine (total 25 mg/kg). Cases with mixed infections and therapeutic treatment failure (attack in the two weeks following the first attack treated correctly) were excluded from the study. In Camopi, treatment with a combination of primaquine (14 days) and chloroquine (3 days) was initiated in 2005, but only for *P. vivax* cases following a first relapse, which did not affect analysis in this study. This prescription of primaquine implies respecting French prescription rules: systematic screening for G6PD deficiency (counter-indication if G6PD deficiency) and delivery as a nominal temporary use authorization.

All parents were given an explanation and written consent was obtained for the study. The study was reviewed and approved by the Ethical Committee of Antilles-Guyane. Two groups of patients were analysed. The first one included the whole population of Camopi children born between January 1, 2001 and December 31, 2008. It was assumed that the first *P. vivax* attack after birth in this group was due to a new infection. The second one included only those who had had a primary attack due to *P. vivax*. In this group, the secondary *P. vivax* attack occurring more than two weeks after a correctly treated primary attack was due to either re-infection or relapse. Therefore, it was assumed that the difference between primary and secondary attack rate of *P. vivax* could be an estimate of the relapse rate. According to the above reasoning, the yearly incidences and the monthly rates of attacks in these groups were estimated using survival analysis methods. Data were right censored by time of child departure from village or death, or time to attack if attack was due to another Plasmodium species. Monthly attack rates were calculated by derivation of the survival curves. Monthly relapse rates were then calculated by difference between the secondary and primary attack rates. All statistical analyses were conducted using R 2.7.1 and the package Epicalc [16].

Results

Between January 2001 and December 2008, 1,042 malaria attacks were recorded in Camopi in 325 children born during this period, of which 319 (31%) were due to *P. falciparum*, 689 (66%) to *P. vivax*, 21 (2%) to mixed infections and 13 (1%) to *P. malariae* or *Plasmodium sp.* One hundred and fifty one persons had their primary attack with *P. vivax* and 106 had their two first attacks with *P. vivax*. Annual incidence and survival time in the two groups were synthesised in table 1. Under the working hypothesis 0.12 relapses occurred each year per person who had had a first attack caused by *P. vivax*. Using incidences of each group, the proportion of relapses in the whole secondary *P. vivax* attacks was estimated at 28%.

Figure 1 compares the monthly rate of first *P. vivax* attack since birth and the monthly rate of secondary attack since first *P. vivax* attack classified by the time to attack. The pattern of changes in both groups was clearly distinct. The rate of primary attack seemed to be constant and uniform along the time and was approximately 0.03% per month and 100 persons at risk. On the contrary, the rate of secondary attacks had a sharp peak during the three first months following the first attack and reached a uniform and constant level after four months comparable to the rate of primary attacks. When relapse rate temporal trends were computed according to the working hypothesis, *P. vivax* relapses were shown to mainly occur during the first three months after the primary attack. A survival curve for relapses was computed. Thirty percent of children never had a relapse, 42% had a relapse before the first month after the primary attack, 59% before the second month and 63% before the third month. After eight months the probability to relapse was nil.

Discussion

This study showed that, in Camopi, the peak of relapses of *P. vivax* occurred in the first three months after the first attack. After that, most secondary attacks were mainly due to a new *P. vivax* re-infection. These observations confirmed the assumption of a pure tropical pattern with a

short latent period for *P. vivax* relapses in French Guiana (Chesson strain [17]). In India, the existence of both tropical and temperate zone types of *P. vivax* characterized by distinct incubation period for relapse were clearly suggested [9].

The current biological methods to differentiate re-infections from relapses have their limits. This study suggests that this simple alternative can estimate the relapse rate in *P. vivax* infection indirectly. However, this method has its own limitations. First, this method to estimate the relapse rate cannot be used in areas where primaquine is associated to chloroquine to treat the first and the second malaria attacks.

Second, it cannot differentiate the relapses from re-infections in individual patients; it estimates the overall relapse rate in the population. It seems that this limitation can be partially resolved by a simple arbitrary classification rule inferred from the results of this study. At an individual scale, before 90 days there is a strong probability that the secondary attack is a relapse, after 90 days it is more probable that the secondary attack is a re-infection. By positing that the primary rate of infection is the same than the secondary rate of re-infection, it can be assumed that immunity against *P. vivax* after the primary attack is negligible. As shown elsewhere [18,19], this does not seem to be an unreasonable hypothesis.

Third, similarly, because this study was based on the follow-up of children from birth, the re-infection rate based on the primary attack rate was under-estimated because of the maternal protective immunity during the first 12 months after birth. However, this effect plays in the opposite way of the previous bias and it didn't seem to influence significantly study results.

Conclusion

This study confirmed that the relapse pattern in Camopi was compatible with the pattern observed with the *P. vivax* Chesson (tropical) strain. In addition, due to the relapse

Table 1: Annual incidence rate^{§*} and survival time of the primary and secondary *P. vivax* attack in the Camopi children population.

	STUDIED POPULATION	NUMBER OF SUBJECTS	TIME AT RISK	INCIDENCE RATE	SURVIVAL TIME (YEAR)			RELAPSE INCIDENCE RATE
					25%	50%	75%	
PRIMARY <i>P. VIVAX</i> ATTACKS	The Camopi Children population	325	571.76	0.31	1.23	2.33	4	0
SECONDARY <i>P. VIVAX</i> ATTACKS	Children having had a primary <i>P. Vivax</i> attack	151	100.42	0.43	0.76	2.14	3.58	0.12

§: Rate unit: attack/persons at risk/per year

*: Right censoring was used in survival analysis

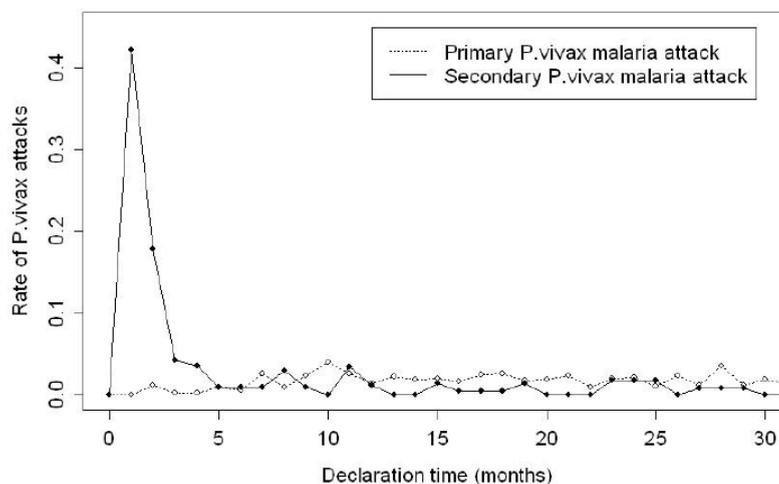


Figure 1
Monthly attack rates[§] since birth for the primary *P. vivax* malaria attacks and monthly attack rates[§] since the first attack for secondary *P. vivax* attacks. [§] Rate unit: attack/100 persons at risk/per month.

rate time evolution, a simple arbitrary classification rule could be constructed: before 90 days after the primary attack, the secondary attack is a relapse; after 90 days it is a re-infection. Adapted management of malaria cases based on these results could be devised: systematic prescription of primaquine for all relapses defined by this three-month rule, despite the obligations of screening for G6PD deficiency and administrative formalities involved on French territories. These results could be used in French Guiana and elsewhere in South America to discriminate *P. vivax* relapses from re-infections.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MH participated in the research design, performed data analysis and interpretation, and prepared the manuscript. AS was responsible for data collection and participated in manuscript revision. CB participated in data analysis and manuscript revision as well. MN provided guidance on data analyses and was involved in the interpretation of data and manuscript revision. BC has designed the cohort and initiated the study, and was involved in the interpretation of data and manuscript revision. All authors read and approved the final manuscript

Acknowledgements

The authors thank the staff of the Health Center of Camopi, particularly Alain Maniglier, Anne Viaud, and Christelle Monnerville for their assistance.

The authors gratefully acknowledge the financial support from EREMIBA project conducted by the UMR 2724 Genetique et Evolution des Maladies Infectieuses (GEMI, IRD) and the Vi-Ve-Pro-Guyane project conducted by Unité de Recherche en biologie parasitaire (URBEP, IMTSSA).

References

- Mendis K, Sina BJ, Marchesini P, Carter R: **The neglected burden of *Plasmodium vivax* malaria.** *Am J Trop Med Hyg* 2001, **64**:97-106.
- Kochar DK, Saxena V, Singh N, Kochar SK, Kumar SV, Das A: ***Plasmodium vivax* malaria.** *Emerg Infect Dis* 2005, **11**:132-134.
- Yeom JS, Ryu SH, Oh S, Lee WJ, Kim TS, Kim KH, Kim YA, Ahn SY, Cha JE, Park JW: **Status of *Plasmodium vivax* malaria in the Republic of Korea during 2001-2003.** *Am J Trop Med Hyg* 2005, **73**:604-608.
- Sumawinata IW, Bernadeta, Leksana B, Sutamihardja A, Purnomo, Subianto B, Sekartuti, Fryauff DJ, Baird JK: **Very high risk of therapeutic failure with chloroquine for uncomplicated *Plasmodium falciparum* and *P. vivax* malaria in Indonesian Papua.** *Am J Trop Med Hyg* 2003, **68**:416-420.
- Baird JK: **Neglect of *Plasmodium vivax* malaria.** *Trends in Parasitology* 2007, **23**:533-539.
- Carme B: **Substantial increase of malaria in inland areas of eastern French Guiana.** *Trop Med Int Health* 2005, **10**:154-159.
- Abeku TA, van Oortmarssen GJ, Borsboom G, de Vlas SJ, Habbema JD: **Spatial and temporal variations of malaria epidemic risk in Ethiopia: factors involved and implications.** *Acta Trop* 2003, **87**:331-340.
- Tian L, Bi Y, Ho SC, Liu W, Liang S, Goggins WB, Chan EY, Zhou S, Sung JJ: **One-year delayed effect of fog on malaria transmission: a time-series analysis in the rain forest area of Mengla County, south-west China.** *Malar J* 2008, **7**:110.
- Adak T, Sharma VP, Orlov VS: **Studies on the *Plasmodium vivax* relapse pattern in Delhi, India.** *Am J Trop Med Hyg* 1998, **59**:175-179.
- Prasad RN, Virk KJ, Sharma VP: **Relapse/reinfection patterns of *Plasmodium vivax* infection: a four year study.** *Southeast Asian J Trop Med Public Health* 1991, **22**:499-503.
- Craig AA, Kain KC: **Molecular analysis of strains of *Plasmodium vivax* from paired primary and relapse infections.** *J Infect Dis* 1996, **174**:373-379.

Chapitre 6 - Apports du couplage études écologiques/données climatiques dans la compréhension de la dynamique des maladies infectieuses : exemple de la Guyane Française.

Malaria Journal 2009, 8:278

<http://www.malariajournal.com/content/8/1/278>

12. Kirchgatter K, del Portillo HA: **Molecular analysis of *Plasmodium vivax* relapses using the MSP1 molecule as a genetic marker.** *J Infect Dis* 1998, **177**:511-515.
13. Veron V, Legrand E, Yrinesi J, Volney B, Simon S, Carme B: **Genetic diversity of msp3alpha and msp1_b5 markers of *Plasmodium vivax* in French Guiana.** *Malar J* 2009, **8**:40.
14. Carme B, Lecat J, Lefebvre P: **Le Paludisme en Guyane dans le foyer de l'Oyapock. Incidence des accès palustres chez les amérindiens de Camopi.** *Med Trop (Marseille)* 2005, **65**:149-154.
15. Hustache S, Nacher M, Djossou F, Carme B: **Malaria risk factors in amerindians children in French Guiana.** *Am J Trop Med Hyg* 2007, **76**:619-625.
16. Chongsuvivatwong V: **Analysis of epidemiological data using R and Epicalc.** Edited by: McNeil. Special Programme for Research & Training in Tropical Diseases (TDR); 2007.
17. Ehrman FC, Ellis JM, Young MD: ***Plasmodium vivax* Chesson strain.** *Science* 1945, **101**:377.
18. Molineaux L, Storey J, Cohen JE, Thomas A: **A longitudinal study of human malaria in the West African Savanna in the absence of control measures: relationships between different *Plasmodium* species, in particular *P. falciparum* and *P. malariae*.** *Am J Trop Med Hyg* 1980, **29**:725-737.
19. Mason J: **Patterns of *Plasmodium vivax* recurrence in a high-incidence coastal area of El Salvador, C. A.** *Am J Trop Med Hyg* 1975, **24**:581-585.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp



6.8. Etude 2 : Influence du climat et du niveau du fleuve sur l'incidence du paludisme dans le village de Cacao, Guyane française

Introduction : Les profils épidémiologiques des maladies à transmission vectorielle comme le paludisme, sont fortement associés aux conditions environnementales. La compréhension de l'effet du climat sur la survenue des cas de paludisme peut donner des éclaircissements sur les caractéristiques des populations des vecteurs anophéliens endémiques à une région particulière. L'association temporelle entre les facteurs météorologiques et hydrographiques et la survenue des cas de paludisme a été étudiée dans le village de Cacao en Guyane française durant une période d'épidémie de paludisme causée essentiellement par *P.vivax*.

Matériels/Méthodes : Une cohorte de cas confirmés de paludisme à *P.vivax* survenus entre 2002 et 2007 a été étudiée pour rechercher les associations temporelles potentielles entre le nombre de nouveaux épisodes d'infection survenant chaque mois, et la moyenne, le minimum, et le maximum des températures mensuelles, le cumul des précipitations mensuelles et la moyenne mensuelle de la hauteur de la rivière qui borde le village. Ces analyses ont été réalisées à l'aide de séries chronologiques (modèles ARIMA).

Résultats : Les facteurs climatiques favorisant l'ampleur de l'épidémie ont été identifiés à court terme (température minimale basse au cours du mois), à moyen terme (température maximale basse 2 mois avant) et à long terme (température maximale basse 9 mois avant et niveau de la rivière bas 12 mois avant). L'analyse par corrélations croisées montrait que les effets de ces facteurs avaient un impact plus important au début de la petite saison des pluies.

Conclusions : L'association entre le niveau du fleuve et le nombre cas de paludisme fournit des indices pour mieux comprendre la dynamique de la transmission du paludisme ainsi que les caractéristiques écologiques des vecteurs dans la région.

Publiée dans *Malaria Journal* le : 4 Février 2011



RESEARCH

Open Access

Influence of climate and river level on the incidence of malaria in Cacao, French Guiana

Célia Basurko^{1,2}, Matthieu Hanf¹, René Han-Sze³, Stéphanie Rogier¹, Philippe Héritier⁴, Claire Grenier³, Michel Joubert³, Mathieu Nacher^{1,2}, Bernard Carne^{1,2,5*}

Abstract

Background: The epidemiological profiles of vector-borne diseases, such as malaria, are strongly associated with environmental conditions. An understanding of the effect of the climate on the occurrence of malaria may provide indirect insight into the anopheles mosquito vectors endemic to a particular region. The association between meteorological and hydrographical factors and the occurrence of malaria was studied in a village in French Guiana during an epidemic caused essentially by *Plasmodium vivax*.

Methods: A cohort of confirmed cases of *P. vivax* malaria occurring between 2002 and 2007 was studied to search for an association between the number of new infection episodes occurring each month, mean, maximum and minimum monthly temperatures, cumulative rainfall for the month and the mean monthly height of the river bordering the village, with the aid of time series. Cross-correlation analysis revealed that these meteorological factors had large effects on the number of episodes, over a study period of 12 months.

Results: Climatic factors supporting the continuance of the epidemic were identified in the short-term (low minimum temperatures during the month), medium-term (low maximum temperatures two months before) and long-term (low maximum temperatures nine months before and high lowest level of the river 12 months before). Cross-correlation analysis showed that the effects of these factors were greatest at the beginning of the short rainy season.

Conclusion: The association between the river level and the number of malaria attacks provides clues to better understand the environment of malaria transmission and the ecological characteristics of the vectors in the region.

Background

Malaria may be seen as the combination of a *Plasmodium*, an *Anopheles* mosquito and a human host. Factors modifying the presence, survival or abundance of any one of these actors are likely to favour the maintenance or breaking of the parasite cycle. Environmental conditions are among the factors likely to affect this parasite cycle. The temperature of the water collecting in hollows, the presence of vegetation or predators and altitude may all favour *Anopheles* mosquitoes or limit their development within a given geographic zone [1,2].

The ecological characteristics of the vectors of malaria in French Guiana have not been completely elucidated [3,4]. In the absence of entomological data, the goal of

this study was to investigate the environmental factors conditioning the encounter between humans and parasites, focusing on the incidence of malaria episode in the village of Cacao in French Guiana. Over a period of 14 years, there were no cases of malaria in this village. Then, between 2002 and 2007, an epidemic caused essentially by *Plasmodium vivax* occurred, affecting almost 43% of the inhabitants of the village.

To better understand the determinants of this epidemic, this manuscript presents results studying the association between meteorological and hydrographical factors and the occurrence of malaria among the inhabitants of Cacao.

Methods

A cohort of patients from the Cacao health centre presenting between January 1st 2002 and December 31st 2007 with an episode of *vivax* malaria confirmed by

* Correspondence: carne.bernard@wanadoo.fr

¹Centre d'Investigation Clinique Epidémiologie Clinique Antilles Guyane CIC-EC CIE 802, Cayenne General Hospital, Cayenne, French Guiana
Full list of author information is available at the end of the article



© 2011 Basurko et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

biological tests was studied. The only health facility within a radius of 80 kms was located in the village. A malaria case corresponding to an attack of malaria was defined as an episode of fever (fever recorded or history of fever in the previous 48 hours, associated to the finding of asexual forms of *Plasmodium* in the blood using Giemsa-stained thick and thin blood smear. in the absence of any other clear aetiology. The parasitological results were validated by a biologist from Cayenne University Hospital Parasitology and Mycology Laboratory (LHUPM). Only patients residing permanently in the village during the study period were included. To ensure data quality, the population that was living temporarily in Cacao (seasonal workers, visitors, tourists...) was not included in the analysis. Therefore, the sample considered in the analysis, remained stable. Patients were included in the cohort from the day the positive sample for *Plasmodium vivax* was obtained. The primary endpoint was the occurrence of a new episode of *P. vivax* malaria (detected by the health centre in Cacao and subsequently biologically confirmed, excluding cases of treatment failure). Uncomplicated malaria cases were treated with chloroquine. None of the patients received 8-aminoquinoline. Relapses were excluded by applying the rule of three months as shown recently in French Guiana for local strains of *P. vivax*. Indeed, in a study carried out in Camopi, a village located in an endemic area, the secondary *P. vivax* attack was considered as a relapse when it occurred during the first three months after the first attack [5].

The association between climatic factors and the monthly incidence of *P. vivax* malaria was investigated by a time-series model of the autoregressive integrated moving average (ARIMA) type. The series was non-stationary and the variance grew exponentially. First, logarithmic transformation ensured the normality and homogeneity of the variance of the residuals. Then, to make the series stationary, a first-order differentiation was performed [6]. The parameters of the ARIMA and the seasonal parameters were determined with the aid of Akaike's information criterion (AIC), the autocorrelation function (ACF) and the partial autocorrelation function (PACF). The model chosen was ARIMA(1,1,1)(0,1,1)₁₂. The ARIMA residuals were not (auto-) correlated and were normally distributed according to the residuals plot and the autocorrelogram of residuals. The Ljung-Box Q test was also applied to confirm that the residual series were white noise.

The meteorological factors studied (measured at the Météo France weather station at the village) were selected on the basis of the specific climatic features in the study zone. This zone has a humid tropical environment and is covered by a dense forest with large amounts of aerial biomass. Humidity is about 90% all

year round. The village of Cacao is situated about 80 km east of Cayenne, the main city of French Guiana (Figure 1). The mountain tops are low, at an altitude of only 30 to 400 m. The village of Cacao is bordered to the North West by a river, the Comté. The region has four seasons: i) a short rainy season (November to January); ii) a "short summer" (February and March); iii) a long rainy season (mid-March to mid-July), and iv) the dry season (August to November).

The following factors were studied: i) mean minimum temperature for each month; ii) mean maximum temperature for each month; iii) cumulative rainfall for each month; iv) the lowest level of the Comté for each month and v) the highest level of the Comté for each month. These climatic data were measured daily and introduced into the model in the form of the mean monthly value for a period of up to 12 months. These factors were tested in the univariate analysis, with all factors identified as significant at the 10% level being introduced into the multivariate model. A stepwise descending approach was used to obtain the final model. The residuals of the models were visualised graphically (by autocorrelation and non systematic tendency). Spearman's rank correlation analysis was carried out to assess the association between the occurrence of malaria attacks and meteorological factors.

Statistical significance was set at $p < 0.05$. The data was analyzed using STATA 10.0 (STATA Corporation, College Station, Texas).

Results

The village of Cacao is inhabited by 839 people, 359 of whom presented at least one attack of *P. vivax* malaria between 2002 and 2007. The Cacao health centre detected 817 attacks of *P. vivax* malaria and 40 attacks of *Plasmodium falciparum* malaria among the inhabitants of this village during this epidemic. The incidence of attacks due to new infections (616 episodes) followed a seasonal pattern over the 12 months of the year, with a peak in April (Figure 2).

Between 2001 and 2007, the months of March, April, May and June had the highest levels of rainfall during the year (cumulative rainfall levels greater than 400 mm per month) and, in these months, the height of the river exceeded 4 m. Minimum and maximum temperatures remained fairly stable, with a slight increase in maximum temperatures observed between August and November.

The mean minimum temperature fell below 21°C between June and September 2003, reaching a trough of 19.4°C in July 2003. The mean maximum temperature exceeded 35°C during the following periods: September and October 2002, from August to November in 2003 and September and October in 2004 and 2005, with a

Chapitre 6 - Apports du couplage études écologiques/données climatiques dans la compréhension de la dynamique des maladies infectieuses : exemple de la Guyane Française.



Figure 1 Localization of Cacao in French Guiana.

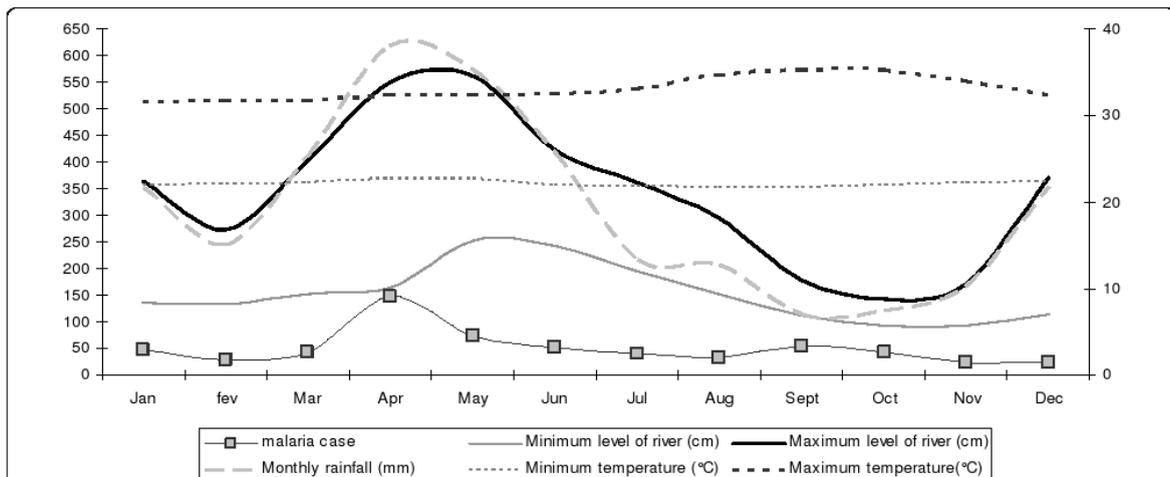
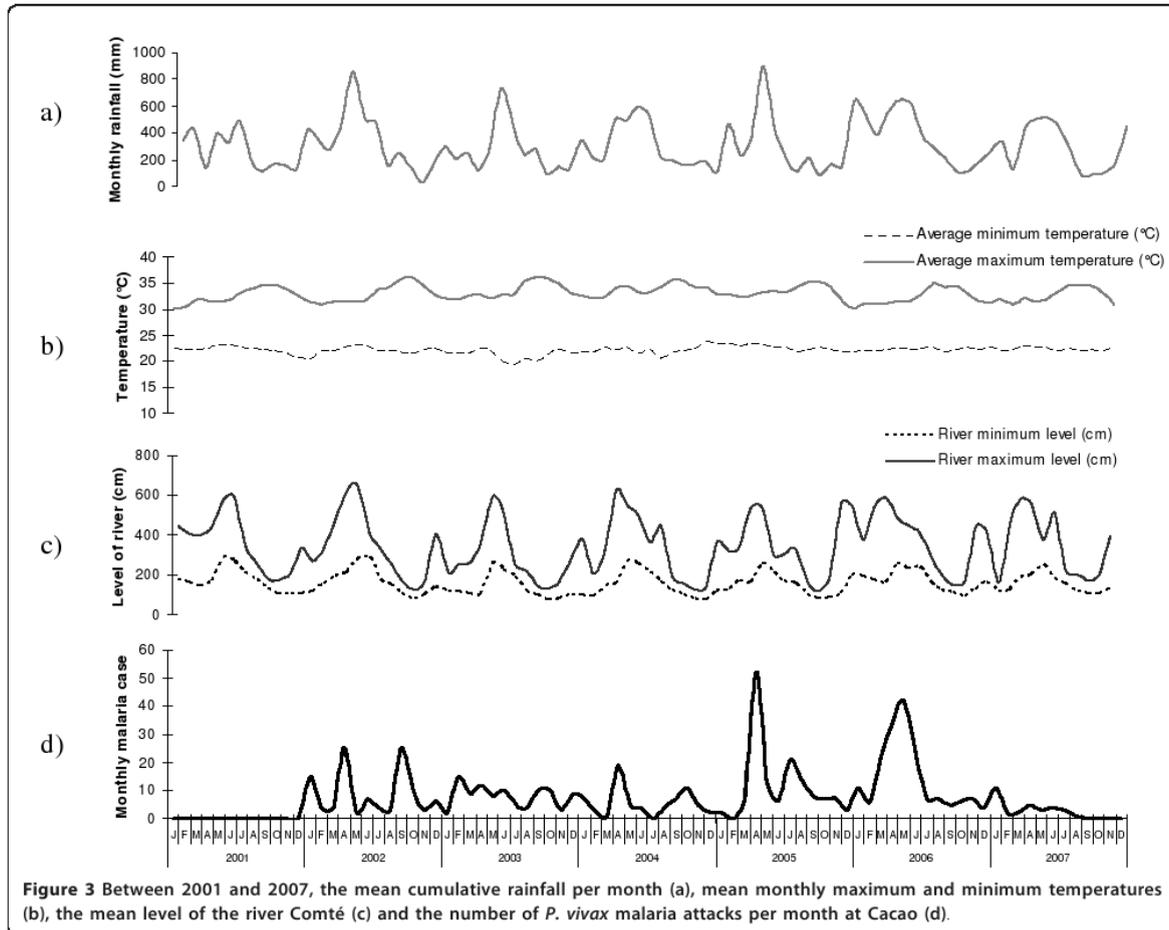


Figure 2 The incidence of malaria was highest in 2005 and 2006 (figure 2a); the Comté reached its highest level between 2002 and 2007 in May 2002, when it reached a height of 6.4 m, and its lowest level in December 2004, at 75 cm (figure 2b).



peak at 36.3°C during August 2003 (Figure 3b). Precipitation levels followed a strongly seasonal pattern (Figure 3a) during the study period. Rainfall levels were highest in April 2002, May 2003 and April 2005 (cumulative precipitation of 893.2 mm for this month).

The introduction of meteorological factors into the univariate ARIMA model (1,1,1)(0,1,1)₁₂ identified nine relevant factors: i) mean minimum temperatures at time t and t-12 months (factor 1 and 2); ii) mean maximum temperature at t-1, t-2 and t-9 months (factor 3,4,5); iii) the minimum level of the Comté at t-11 and t-12 months (factor 6 and 7); and iv) the maximum level of the Comté at t-8 and t-12 months (factor 8 and 9).

The multivariate ARIMA model obtained with the inclusion of these nine variables, applying a stepwise descending procedure, is summarized in Table 1. The incidence of a malaria episode at time t was positively correlated with the minimum level of the Comté at t-12 months and inversely correlated with the minimum

temperature at time t and with maximum temperature at times t-2 and t-9 months.

The relationship between the incidence of malaria and the minimum temperature (Figure 4a) was strongest in December ($\rho = -0.7143$; $p = 0.1108$). Correlations

Table 1 ARIMA regression model of the logarithm of monthly malaria cases (2001-2007) against meteorological factors in Cacao, French Guiana

	MODEL (AIC = 145.23)		
	Coefficient	SEM	p-value
Minimum temperature	-0.272	0.15	0.007
Maximum temperature (lag 2)	-0.307	0.13	0.002
Maximum temperature (lag 9)	-0.312	0.098	0.001
Minimum level of river (lag 12)	0.0091	0.0032	0.005
AR	-0.084	0.176	0.635
MA	-0.719	0.098	0.000
Seasonal MA	-0.505	0.13	0.000
Constant	0.037	0.015	0.012

AR: autoregressive term MA: moving average term.

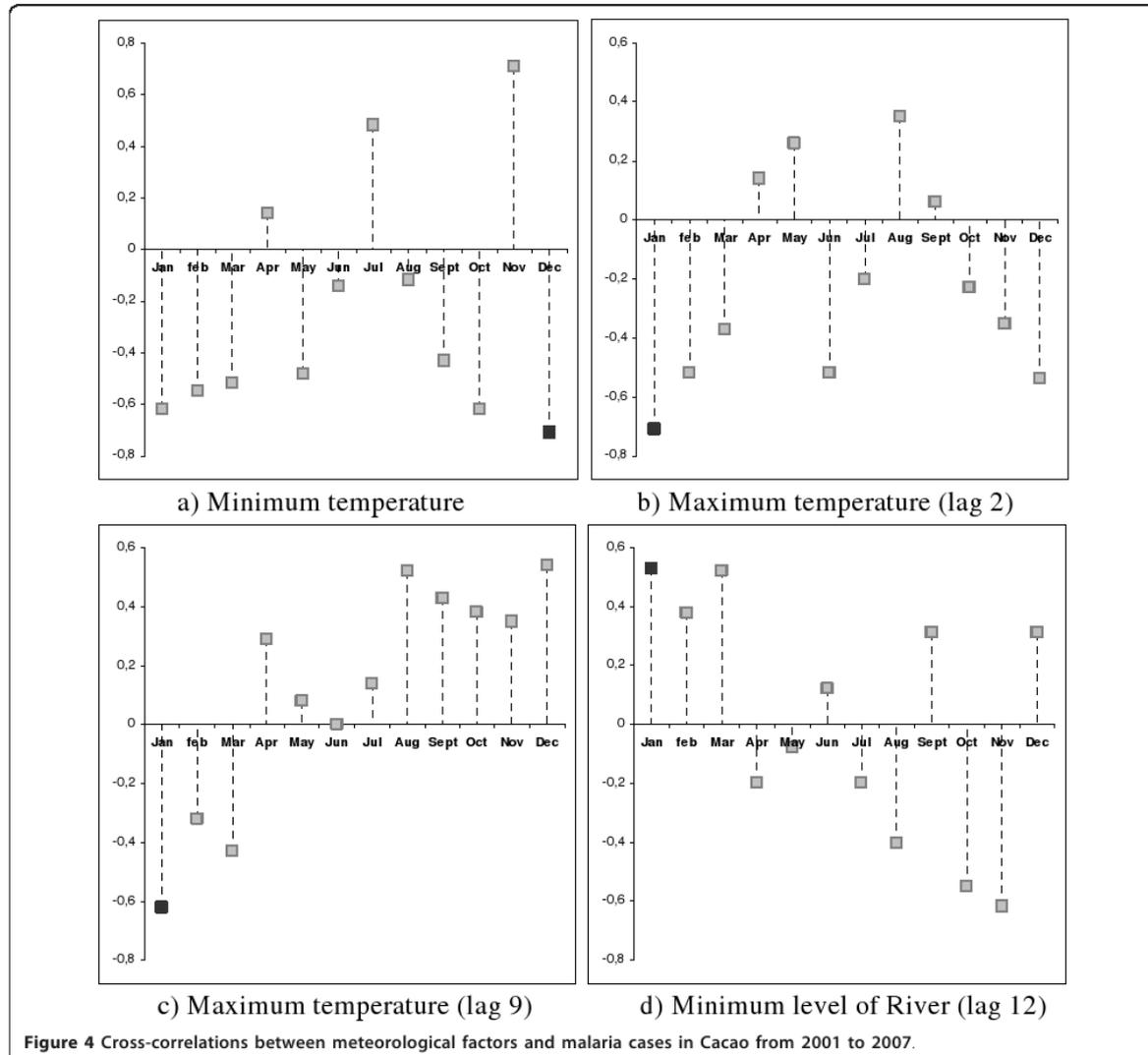


Figure 4 Cross-correlations between meteorological factors and malaria cases in Cacao from 2001 to 2007.

between the number of attacks and maximum temperatures two months before (Figure 4b) and nine months before (Figure 4c) were strongest in January ($\rho = -0.71$; $p = 0.117$ and $\rho = -0.62$; $p = 0.1911$, respectively). The same was true for the minimum level of the Comté river (Figure 4d; $\rho = 0.53$; $p = 0.27$)

Discussion

This study was carried out on a small cohort, limiting the power of statistical analyses, including cross-correlations in particular. However, it should be noted that the medical data obtained are exhaustive for the village, homogeneous (malaria caused by *P. vivax* only) and high quality (identical diagnostic and therapeutic management over a period of five years).

Very little entomological knowledge is currently available for the vector of malaria at Cacao [3]. Vector control programmes are, therefore, based on data from the other countries of South America, for *Anopheles darlingi* [7-11]. The lack of local specificity of the knowledge regarding the vector results in intervention programmes needing to be massive, with poor cost/efficacy ratios.

This study could indirectly help to improve this ratio, by providing clues to better understand the transmission of the parasite between vector and host in this region. Indeed, the results of this study showed certain climatic factors could participate in the maintenance of the epidemic in the short-, medium- or long-term.

First, the positive association between the river level and the incidence of malaria could be explained by the

location of potential larval habitats. The banks of the Comté river are encumbered with voluminous tree roots and are, therefore, not straight. During periods of high water, these features favour the creation of pockets of water with weak currents, bordered by vegetation and ideal for the development of anopheles mosquito larvae [11-13]. The persistence of a minimum level of water in the river ensures the maintenance of these larval habitats and adequate development of the surrounding vegetation. Without entomological evidence, this hypothesis could not be confirmed. However, others authors in America have also proposed this. In the north of Peru, a number of climatic factors (temperature and rainfall figures) have been studied, but only the level of the river Nanay was found to be significantly associated with the incidence of malaria due to *P. vivax* [14]. Similarly, a study in Suriname identified associations between mosquito density and the level of the river [15]. In Brazil, flooding of the fields bordering the river Branco has also been suggested as a reason for increases in the number of *An. darlingi* bites per person during the rainy season, close to the river [16]. In Venezuela, the abundance of *An. darlingi* was positively correlated with the maximum level of the Orinoco river because of the creation of lagoons during the river overflow [17].

The time lag between the occurrence of malaria and the associated meteorological factors, such as maximum temperature nine months previously and the level of the Comté 12 months previously, would suggest that meteorological conditions in a given year may affect malaria in the following year. With persistent larval habitats in all seasons due to sufficient river levels and low temperatures, the transmission of malaria is maintained into the next year. These environmental factors could create the appropriate conditions to preserve mosquito larvae with sufficient humidity in the dry season. Another hypothesis would be the need for a number of parasitic amplification cycles to make malaria transmission patent. The long-term impact of climate on malaria has already been reported in the literature. Indeed, a seven-month delayed effect of fog was associated with malaria incidence in a Chinese study. The authors explained this delay with the same biological hypothesis [6]. Similarly, significant increases of malaria mortality and morbidity have been showed in years following recognized El Niño events [18].

The intertropical convergence zone (ITCZ) or meteorological equator determines the seasons in French Guiana. It descends the northern hemisphere, crossing French Guiana in mid-December and bringing with it cumulonimbus clouds, which generate rainfall. Cross-correlation analysis showed that the time at which temperatures and the level of the Comté river seemed to

influence the occurrence of cases coincided with the period during which the ITCZ passes through French Guiana for the first time. The resulting increase in precipitation, with high humidity, favours an increase in the number of mosquito breeding sites and the development of the aquatic stages of the mosquito [19].

Finally, an increase in temperature (both minimum and maximum) has often been identified as a factor accelerating the gonotrophic cycle of anopheles mosquitoes, decreasing the interval between egg-laying episodes and increasing vector abundance [20-23]. The development cycle of the parasite is also accelerated by high temperatures. However, if the temperature is too high, it may also jeopardise the survival of adult mosquitoes and dry up breeding sites more rapidly, particularly during the dry season. Indeed, in this analysis, the number of malaria cases in 2003 and 2004 fell sharply when temperatures were particularly high.

Conclusion

A better understanding of the ecology of the anopheles mosquitoes endemic to the environment should make it possible to concentrate human and material resources more effectively in vector control programmes. In the absence of entomological and behavioural data, analyses of the incidence of malaria over time has provided information about the transmission of *P. vivax* in the region, and this information may be useful for the prevention and control of future epidemics.

Acknowledgements

We thank the staff of the Health Centre of Cacao and the staff of CIC-EC for their assistance.

Author details

¹Centre d'Investigation Clinique Epidémiologie Clinique Antilles Guyane CIC-EC CIE 802, Cayenne General Hospital, Cayenne, French Guiana. ²EA3593, UFR Médecine - Université des Antilles et de la Guyane, Cayenne, French Guiana. ³Département des centres de santé, Cayenne General Hospital, Cayenne, French Guiana. ⁴Météo-France Cayenne French Guiana. ⁵Laboratoire Hospitalo-Universitaire de Parasitologie et Mycologie Médicale, Centre Hospitalier de Cayenne, rue des Flamboyants, BP 6006, F- 97306 Cayenne (French Guiana).

Authors' contributions

CB participated in the research design, performed data analysis and interpretation, and prepared the manuscript. MH participated in the statistical analysis. RHS, FG, CG and MJ participated in interpretation of data and manuscript revision as well. SG participated in research method. MN provided guidance on data analyses and was involved in the interpretation of data and manuscript revision. BC was responsible for data collection, initiated the study, and was involved in the interpretation of data and manuscript revision. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 30 October 2010 Accepted: 4 February 2011
Published: 4 February 2011

Chapitre 6 - Apports du couplage études écologiques/données climatiques dans la compréhension de la dynamique des maladies infectieuses : exemple de la Guyane Française.

Basurko et al. *Malaria Journal* 2011, **10**:26
<http://www.malariajournal.com/content/10/1/26>

Page 7 of 7

References

1. Tangad MC, Ngundu WI, Judith N, Mbuh Tendongfor N, Simard F, Wanji S: **Climate change and altitudinal structuring of malaria vectors in south-western Cameroon: their relation to malaria transmission.** *Trans R Soc Trop Med Hyg* 2010, **104**:453-460.
2. Sinka ME, Rubio-Palis Y, Manguin S, Patil AP, Temperley WH, Gething PW, Van Boeckel T, Kabaria CW, Harbach RE, Hay SI: **The dominant Anopheles vectors of human malaria in the Americas: occurrence data, distribution maps and bionomic précis.** *Parasite Vectors* 2010, **3**:72.
3. Girod R, Gaborit P, Issaly J, Carinci R, Basurko C: **Anophèles et transmission du paludisme dans le village et les environs de Cacao (Guyane).** *Bull Soc Pathol Exot* 2009, **102**:199-205.
4. Dusfour I, Carinci R, Gaborit P, Issaly J, Girod R: **Evaluation of four methods for collecting malaria vectors in French Guiana.** *J Econ Entomol* 2010, **103**:973-976.
5. Hanf M, Stéphani A, Basurko C, Nacher M, Carme B: **Determination of the Plasmodium vivax relapse pattern in Camopi French Guiana.** *Malar J* 2009, **8**:278.
6. Tian L, Bi Y, Ho S, Liu W, Liang S, Gobbins WB, Chan EY, Zhou S, Sung JJ: **One-year delayed effect of fog on malaria transmission: a time-series analysis in the rain forest area of Mengla County, south-west China.** *Malar J* 2008, **7**:110.
7. Roper M, Torres R, Goicochea C, Andersen E, Guarda J, Calampa C, Hightower A, Magill A: **The epidemiology of malaria in an epidemic area of the Peruvian Amazon.** *Am J Trop Med Hyg* 2000, **62**:247.
8. Zeilhofer P, Santos E, Ribeiro A, Miyazaki R, Santos M: **Habitat suitability mapping of Anopheles darlingi in the surroundings of the Manso hydropower plant reservoir, Mato Grosso, Central Brazil.** *Int J Health Geogr* 2007, **6**:7.
9. Charlwood JD: **Biological variation in Anopheles darlingi Root.** *Memórias do Instituto Oswaldo Cruz* 1996, **91**:391-398.
10. Giglioli G: **Nation-wide malaria eradication projects in the Americas. III Eradication of Anopheles darlingi from the inhabited areas of British Guiana by DDT residual spraying.** *J Natl Mal Soc* 1951, **10**:142-161.
11. Vittor AY, Pan W, Gilman RH, Tielsch J, Glass G, Shields T, Sánchez-Lozano W, Pinedo VV, Salas-Cobos E, Flores S, Patz JA: **Linking deforestation to malaria in the Amazon: characterization of the breeding habitat of the principal malaria vector, Anopheles darlingi.** *Am J Trop Med Hyg* 2009, **81**:5-12.
12. Grieco JP, Rejmánková E, Achee NL, Klein CN, Andre R, Roberts D: **Habitat suitability for three species of Anopheles mosquitoes: larval growth and survival in reciprocal placement experiments.** *J Vector Ecol* 2007, **32**:176-87.
13. Manguin S, Roberts DR, Andre RG, Rejmankova E, Hakre S: **Characterization of Anopheles darlingi (Diptera: Culicidae) larval habitats in Belize, Central America.** *J Med Entomol* 1996, **33**:205-211.
14. Bautista C, Chan A, Ryan J, Calampa C, Roper M, Hightower A, Magill A: **Epidemiology and spatial analysis of malaria in the northern Peruvian Amazon.** *Am J Trop Hyg* 2006, **75**:1216-1222.
15. Rozendaal J: **Relations between Anopheles darlingi breeding habitats, rainfall, river level and malaria transmission rates in the rain forest of Suriname.** *Med Vet Entomol* 1992, **6**:16-22.
16. Barros F, Honório N: **Man biting rate seasonal variation of malaria vectors in Roraima, Brazil.** *Memórias do Instituto Oswaldo Cruz* 2007, **102**:299-302.
17. Magris M, Rubio-Palis Y, Menares C, Villegas L: **Vector bionomics and malaria transmission in the Upper Orinoco River, Southern Venezuela.** *Mem Inst Oswaldo Cruz* 2007, **102**:303-311.
18. Bouma MJ, Dye C: **Cycles of malaria associated with El Niño in Venezuela.** *JAMA* 1997, **278**:1772-1774.
19. Gomez-Elipe A, Otero A, Van Herp M, Aguirre-Jaime A: **Forecasting malaria incidence based on monthly case reports and environmental factors in Karuzi, Burundi, 1997 - 2003.** *Malar J* 2007, **6**:129.
20. Haghdoost A, Alexander N, Cox J: **Modelling of malaria temporal variations in Iran.** *Trop Med Int Health* 2008, **13**:1501-1508.
21. Mabaso M, Vounatsou P, Midzi S, Da Silva J, Smith T: **Spatio-temporal analysis of the role of climate in inter-annual variation of malaria incidence in Zimbabwe.** *Int J Health Geogr* 2006, **5**:20.
22. Clements A, Barnett A, Cheng Z, Snow R, Zhou H: **Space-time variation of malaria incidence in Yunnan province, China.** *Malar J* 2009, **8**:180.
23. Kristan M, Abeku TA, Beard J, Okia M, Rapuoda B, Sang J, Cox J: **Variations in entomological indices in relation to weather patterns and malaria**

incidence in East African highlands: implications for epidemic prevention and control. *Malar J* 2008, **7**:231.

doi:10.1186/1475-2875-10-26

Cite this article as: Basurko et al.: Influence of climate and river level on the incidence of malaria in Cacao, French Guiana. *Malaria Journal* 2011 **10**:26.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



6.9. Etude 3 : Les facteurs de risque environnementaux, entomologiques, socio-économiques et comportementaux des crises de paludisme chez les enfants amérindiens du village de Camopi, Guyane française

Introduction : Le paludisme est un problème majeur de santé publique en Guyane française. Les communautés amérindiennes restent les plus touchées. Une étude précédente a souligné le rôle prépondérant des facteurs environnementaux dans la survenue du paludisme à Camopi. Cependant, tous les paramètres impliqués dans la transmission n'ont pas été clairement identifiés. Une nouvelle enquête a été menée afin de mieux clarifier les facteurs de risque de la survenue des crises de paludisme à Camopi.

Matériels/Méthodes : Une cohorte ouverte d'enfants de moins de 7 ans a été mise en place sur la base des cas de paludisme confirmés biologiquement durant la période 2001-2009. Les données épidémiologiques et environnementales ont été recueillies à l'aide de deux questionnaires structurés. Les données ont été analysées avec un modèle multivarié de Cox à événements multiples. Les influences du climat et du niveau du fleuve sur l'incidence du paludisme ont été évaluées à l'aide d'analyse de séries chronologiques (ARIMA). Des corrélations de Spearman ont été utilisées pour quantifier la relation entre les taux d'attaque d'*Anopheles darlingi* et les taux d'incidence du paludisme.

Résultats : L'incidence annuelle sur la période d'étude était de 238 pour 1000 pour *P.falciparum*, de 514 pour 1000 pour *P.vivax* et de 21 pour mille pour les infections mixtes. L'analyse de survie multivariée associée une incidence plus élevée du paludisme avec un lieu de vie sur le bord de la rivière Camopi plutôt que sur le bord de la rivière Oyapock, situé loin du centre du hameau de Camopi, dans les maisons avec de nombreux occupants et avec le fait d'aller se coucher tard. Au contraire, vivre dans une maison débarrassée de toute végétation à moins de 50 m et située à une distance élevée de la forêt ont été associés à un risque plus faible. Les caractéristiques météorologiques et hydrologiques apparaissaient comme très corrélées avec l'incidence du paludisme et cela à plusieurs décalages de temps différents. Le taux d'attaque d'*Anopheles darlingi* a également été positivement corrélé à l'incidence du paludisme chez les enfants un mois plus tard.

Conclusions : L'incidence du paludisme demeure élevée chez les jeunes enfants malgré l'apparition d'une immunité chez les enfants autour de 3 ans. L'environnement proche mais aussi les paramètres météorologiques jouent un rôle important dans la transmission du paludisme chez les enfants de moins de 7 ans à Camopi.

Publiée dans *Malaria Journal* le : 23 Août 2011



RESEARCH

Open Access

Environmental, entomological, socioeconomic and behavioural risk factors for malaria attacks in Amerindian children of Camopi, French Guiana

Aurélia Stefani¹, Matthieu Hanf^{1,2}, Mathieu Nacher^{1,2}, Romain Girod³ and Bernard Carne^{1,2,4*}

Abstract

Background: Malaria is a major health issue in French Guiana. Amerindian communities remain the most affected. A previous study in Camopi highlighted the predominant role of environmental factors in the occurrence of malaria. However, all parameters involved in the transmission were not clearly identified. A new survey was conducted in order to clarify the risk factors for the presence of malaria cases in Camopi.

Methods: An open cohort of children under seven years of age was set up on the basis of biologically confirmed malaria cases for the period 2001-2009. Epidemiological and observational environmental data were collected using two structured questionnaires. Data were analysed with a multiple failures multivariate Cox model. The influence of climate and the river level on malaria incidence was evaluated by time-series analysis. Relationships between *Anopheles darlingi* human biting rates and malaria incidence rates were estimated using Spearman's rank correlation.

Results: The global annual incidence over the nine-year period was 238 per 1,000 for *Plasmodium falciparum*, 514 per 1,000 for *Plasmodium visa* and 21 per 1,000 for mixed infections. The multivariate survival analysis associated higher malaria incidence with living on the Camopi riverside vs. the Oyapock riverside, far from the centre of the Camopi hamlet, in a home with numerous occupants and going to sleep late. On the contrary, living in a house cleared of all vegetation within 50 m and at high distance of the forest were associated with a lower risk. Meteorological and hydrological characteristics appeared to be correlated with malaria incidence with different lags. *Anopheles darlingi* human biting rate was also positively correlated to incident malaria in children one month later.

Conclusions: Malaria incidence in children remains high in young children despite the appearance of immunity in children around three years of age. The closeness environment but also the meteorological parameters play an important role in malaria transmission among children under seven years of age in Camopi.

Background

Malaria is a major public health problem in French Guiana, a French overseas region located in South America and separated from Brazil and Suriname by the Oyapock and Maroni Rivers, respectively. The Amazon forest covers 94% of the territory. This area is among the most affected by malaria in South America [1]. To date, French Guiana counts around 3,800 malaria cases every year. The overall involvement of *Plasmodium falciparum* and *Plasmodium vivax* appears to be equal, but these two species present an

uneven distribution on the territory. Only 2.6% of cases are attributed to *Plasmodium malariae*. Most transmission occurs inland along the rivers, whereas the seashore, where 75% of the inhabitants live, is almost free of transmission [2].

A study in Camopi, an Amerindian village on the Eastern border of French Guiana, reported a mean annual malaria incidence of 486 per 1,000 over the 2000-2002 period and thus labelled the region as a hotspot for malaria [3]. A more recent study using a retrospective cohort design at the same site reported a global incidence of 935 per 1,000 for children under five years of age, 70% of the cases being caused by *P. vivax*. This investigation highlighted the predominant role of environmental factors

* Correspondence: carne.bernard@wanadoo.fr

¹EPaT Team (EA3593), UFR de Médecine - Université des Antilles et de la Guyane, Cayenne, French Guiana

Full list of author information is available at the end of the article



© 2011 Stefani et al.; licensee BioMed Central Ltd. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

surrounding households (*i.e.* clearing of vegetation; distances to the river and to the forest). In addition, ethnicity was also identified as an independent risk factor for malaria [4].

Anopheles darlingi, a common and efficient vector in the Americas, has been considered as the primary vector in French Guiana for 50 years [5,6]. This vector has the characteristics of being anthropophilic and endo-exophagic. The nycthemeral activity of *An. darlingi* in the coastal area of French Guiana presents a bimodal rhythm with a peak at dusk and one at dawn, superimposed on a nocturnal activity reaching its peak in the middle of the night [7]. This mosquito species is widely distributed in French Guiana and is responsible for malaria transmission in Amerindian villages of the Upper-Maroni on the western border with Suriname [8,9]. In Camopi, longitudinal surveys have been conducted from January 2003 to December 2006. Most of the anopheline species collected were *An. darlingi*, with a mean human biting rate around one bite/person-night that showed marked seasonality with a peak of abundance from April to July [10].

In the present study, the main objective was to determine risk factors for malaria in children under seven, including additional variables from a previous study [4]. Thus biological, ecological, meteorological, hydrological, entomological, socioeconomic and behavioural variables were studied in relation to malaria.

Methods

Study area

The study was conducted in Camopi, a village along the Oyapock River which represents the border between France and Brazil. The village consists of a main central hamlet and 28 hamlets within 15 km² located along the Oyapock and Camopi Rivers. The 1,200 inhabitants are mainly Amerindians of Wayampi and Emerillon ethnic groups, which are respectively located on the Oyapock and Camopi riversides. They live a traditional life and their main occupations are a subsistence slash and burn agriculture [11], hunting, fishing and picking. Crop culture areas are increasing gradually as these Amerindian populations, former nomads, have settled with the creation of structures, such as the health centre and the school. The houses are wood huts, locally called "carbets", which have a roof of palm leaves, steel sheet or tarpaulin. Nevertheless, modern concrete houses are progressively replacing the traditional ones, particularly in the principal hamlet. Camopi is located in the Amazon rainforest. The climate is equatorial with an annual average temperature of 27°C and a high humidity $\geq 80\%$. In French Guiana, four seasons are identified: the long rainy season from April to June, the long dry season from July to November, the short rainy season from December to February and the short dry season in March. The cumulated precipitation is

around 2,700 mm annually. The village is isolated from the inhabited seashore and the journey in dugout canoes to the first main town, Saint-Georges, takes between 4 and 8 hours, depending of the river level and the season. Tourism is not permitted in Camopi and specific authorization is required for all non-residents wishing to reach the village. The health centre, which is located in the main hamlet of Camopi, ensures free early diagnosis and treatment.

Study design

An open cohort study of children under seven years of age followed from January 1, 2001 to December 31, 2009 was carried out. The exposed population consisted of the totality of the children of Camopi born between January 1, 1994 and December 31, 2008. Malaria data were not censored after the first malaria infection and the children were followed until the day they completed seven years of age. The list of the children was based on the health centre registry which is assumed to be thorough. The children come to the health centre approximately once a month, generally for an illness or for systematic visits (*i.e.* vaccination). Every six months, there is a verification that all newborn children are included in the cohort, that all malaria data is collected and whether children have moved in another hamlet of Camopi or have moved out of Camopi. Data related to children for whom follow-up had been interrupted was right censored at the interruption date. It was assumed that all malaria attacks were recorded at the local health centre due to the isolation of the population and its limited mobility.

Clinical and parasitological diagnosis

Malaria was defined as fever (temperature $\geq 38^\circ\text{C}$ at the time of consultation or during the previous 48 hours) associated with a positive thin blood smear for *Plasmodium* asexual forms. Blood smears were first examined in Camopi by nurses trained in microscopy and then checked at the parasitological laboratory of Cayenne Hospital. When blood smear examinations were not feasible, rapid diagnostic tests (OptiMAL[®] test) were performed. The list of all clinical malaria episodes, their date of diagnosis, *Plasmodium* species and parasitaemia (when available) was established.

Collecting socio-economical and behavioural data

Data were provided by a Knowledge, Attitudes, Practices and Behaviour (KAPB) questionnaire administered to every child's mother. A previous study provided data for children born between January 1, 1996 and December 31, 2005 [4]. In the present study, the cohort was extended to children born between January 1, 1994 and December 31, 2008. Questions about the knowledge of malaria and its transmission, behavioural habits, use of prophylactic measures (use of mosquito nets, domestic insecticides

and topical repellents), socioeconomic status (knowledge of French language, education, physical possessions) and slash and burn culture grounds frequentation were asked during the visits.

Collecting data on the direct household environment

Environmental exposure was evaluated by direct observation of the patients' houses. Every household was geo-localised with a global positioning system (GPS) Magellan® eXplorist™600 and was described (structure of walls and roof). For each *carbet*, distances to the nearest household, to the forest and to the river were provided by GPS or direct measurement for distances < 50 m. The composition of the environment (percentage of water, bare soil, shrubs, slashed vegetation and forest) within the 50 m around them was visually estimated by the interviewer. The presence of a creek in the vicinity and the inundability were assessed. The frequency of intervention of the county mosquito control service (SDD) was also reported on the questionnaire.

Environmental data: rainfall, temperature and river levels

Rainfall and temperature records were provided by the French meteorological services in French Guiana, which collects data daily from an automatic station in the village of Camopi. Similarly, daily data on the Oyapock River levels were provided by the national department of environmental services and collected by the hydrometric station of Saut Maripa, 90 km downstream from Camopi. Although Saut Maripa lies at a considerable distance from Camopi, river levels at these two locations have been measured simultaneously in the past and these measurements allowed the river levels at Camopi for the period 2001-2009 to be reliably estimated from those recorded at Saut Maripa. The records of daily rainfall, minimum and maximum temperatures, and minimum and maximum river levels were converted into monthly cumulated rainfall, monthly mean of minimum and maximum temperature, and monthly minimum and maximum river levels. Monthly number of days without rain was extracted from daily rainfall records.

Entomological data

Dataset was provided by the Medical Entomology Unit of the Institut Pasteur of French Guiana. Field mosquito collections were made in the main hamlet and in some hamlets along the Camopi and Oyapock Rivers from January 2003 to December 2006. Collecting methodology has been previously described [10]. For the need of the study, collection sites were clustered into four groups represented Figure 1. Human biting rates (HBR), representing the number of female anopheline bites/person-night, were estimated for all four groups of collecting sites.

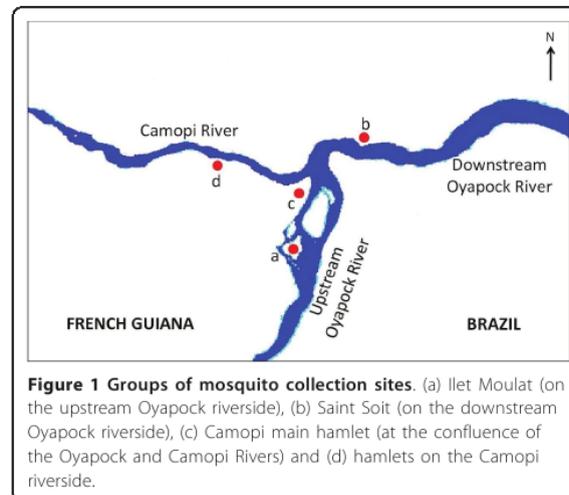


Figure 1 Groups of mosquito collection sites. (a) Ilet Moulat (on the upstream Oyapock riverside), (b) Saint Soit (on the downstream Oyapock riverside), (c) Camopi main hamlet (at the confluence of the Oyapock and Camopi Rivers) and (d) hamlets on the Camopi riverside.

Statistical analysis

A survival analysis was used to explore the malaria risk factors. The entry date into the study was January 1, 2001 or the birth date for children born after this date. Exit time was December 31, 2009 or the seventh birthday of the child if he or she reached seven before this date. Time at risk was defined by the entire period from entry to exit. Fifty-seven variables were tested. The influence of each variable on malaria incidence was assessed using the Kaplan-Meier method and Cox modelling. The proportional hazards assumption was tested graphically and with tests based on Schoenfeld residuals. All variables with statistical significance at log-rank test in univariate analysis (p -value < 0.2) were introduced in the multivariate analysis. All models were stratified by year of birth, accounting for age (which violated the proportional hazards assumption) and the period of birth (due to malaria incidence variations according the year). The most parsimonious model was obtained using the log-likelihood ratio test to remove variables from the saturated model. Convergence of this model with both forward and backward automatic stepwise elimination was checked. Survival analysis was made with Stata8® software (StataCorp, College Station, TX).

The associations between malaria incidence and cumulated rainfall, number of days without rain, temperatures and river water levels was investigated by an autoregressive integrated moving average (ARIMA) model. First, a log-transformation was made to stabilise the variance. The optimal combination of ARIMA and seasonal parameters was identified by using the Akaike's information criterion (AIC) [12]. The best-fit model was ARIMA (1,0,0)(0,1,1)₁₂. Meteorological factors were tested in univariate analysis, with all factors identified at the 5% level being introduced in the multivariate model. A backwards

stepwise approach was used to obtain the final model. Spearman's rank correlation analysis was carried out to assess the association between malaria incidence and meteorological factors identified by the multivariate analysis. In order to examine whether the association between weather and malaria remains constant or whether it is particularly strong in certain months, the response and predictor time-series were also modeled separately and Spearman's rank order correlations between their residuals series at the appropriate lag were examined subsequently. Time-series analysis was realised using the R[®] software.

Associations between human biting rate and malaria incidence during the same month and the following month were also investigated by calculating Spearman's rank order correlation coefficients for the period 2003-2006, depending on the availability of entomological data. A p-value of < 0.05 was considered indicative of a statistically significant correlation.

Ethical considerations

The protocol was approved by the CCTIRS (Committee on information processing in research in the field of health) and the CNIL (National commission for computing and liberties). Informed consent was provided by one of the parents of every child before admission into the study. A written consent was signed by the investigator, the respondent parent and the interpreter before the administration of the questionnaire. All malaria cases were treated when diagnosed.

Results

Study population characteristics

Five hundred and forty-one children, 288 boys (53%) and 253 girls (47%), were included into the cohort. The ethnic distribution of the children was as follows: 237 Wayampi (44%), 187 Emerillon (34%) and 117 persons of mixed ethnic groups (22%). Regarding spatial distribution, 23% of the children lived in the main hamlet and 37% and 40% of the children lived in the hamlets situated on the Oyapock and Camopi riversides, respectively. The total time at risk was 2279 person-years. During the follow-up, 1,773 malaria cases were observed: 542 caused by *P. falciparum*, 1,171 by *P. vivax*, 48 by mixed infections and 12 by unidentified species. There were 10 deaths during the follow-up but none of them was due to malaria. Children had on average 0.7 malaria episodes per year and the median age for the first malaria attack was 2.3 years. During the study period, 23 children emigrated out of the study area and 25 moved from one hamlet to another. The former were censored at the date of their departure. The latter were censored at the moving date and, at the same date, re-introduced in the study as new individuals, in the

new residential hamlet. Thus, the analysis took into account $541 + 25 = 566$ individuals.

Most children (75%) spent all their nights under mosquito nets. Around 80% of the interviewed mothers considered malaria as a severe and urgent disease and 57% of them said they consulted the shaman in addition to the health centre's staff. Many families used domestic insecticides and topical repellents (66% and 63%, respectively). Forty seven percent of the mothers mentioned mosquitoes as malaria vectors. Only 32% could name at least one anti-malarial treatment whereas 93% knew at least one malarial symptom (fever was mentioned in 77% of the responses). Considering socio-economic characteristics, 68% of the families owned a dugout canoe and 53% owned an out-board motor. Over 85% of homes had electricity (generator or solar panels) and more than 60% had a television set.

Inter-annual and intra-annual variations in malaria incidence

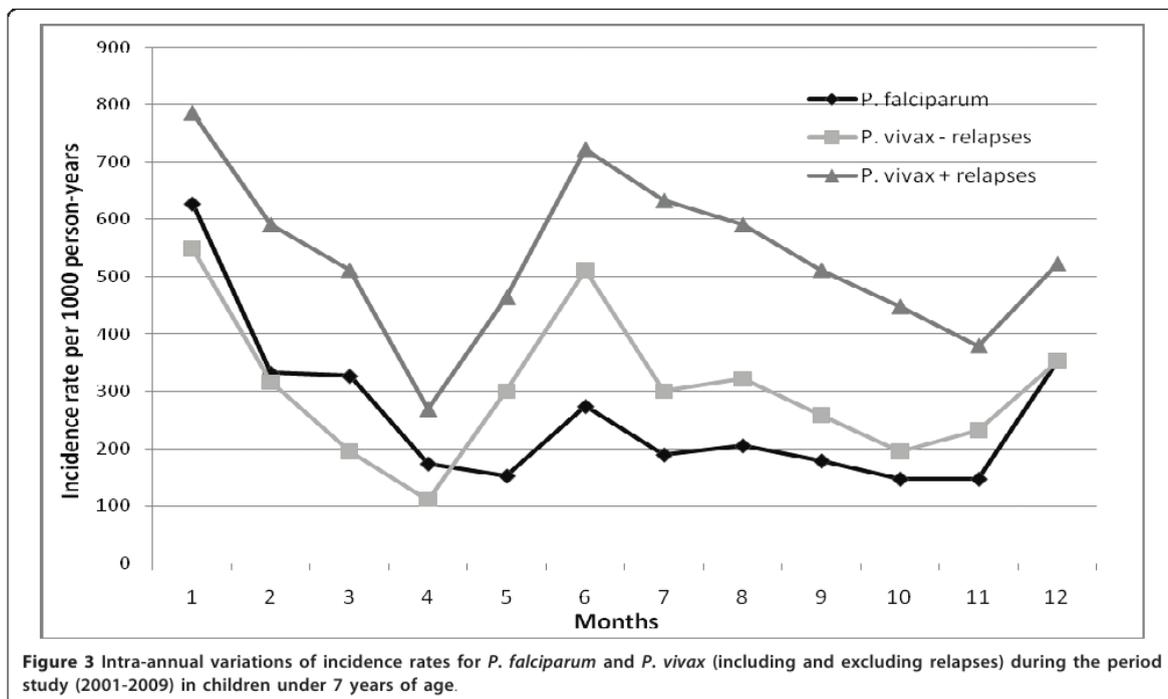
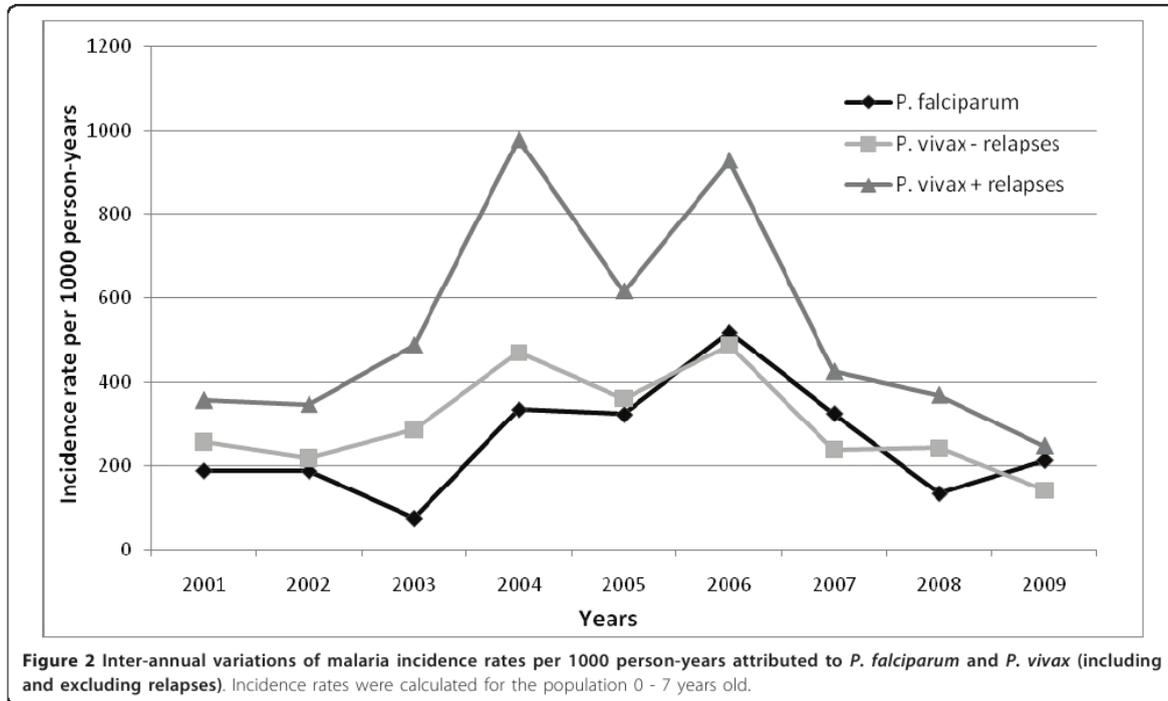
The overall annual incidence during the study period was 773 per 1,000 person-years. Regarding *Plasmodium* species, the incidences were 238, 514 and 21 per 1,000 person-years for *P. falciparum*, *P. vivax* and mixed infections, respectively. *Plasmodium vivax* relapses in French Guiana have a pure tropical pattern with a short latent period (Chesson strain [13]) [14]. Therefore, for a given child, each *P. vivax* episode occurring within 90 days after another *P. vivax* episode was considered as a relapse. By removing relapses from the database, we focussed on malaria transmission and not on the overall incidence, eliminating the noise. Thus, if relapses due to *P. vivax*, which represent 43% of *P. vivax* attacks, are excluded, the annual transmission incidence rates became 248, 292 and 11 per 1,000 person-years for *P. falciparum*, *P. vivax* and mixed infections, respectively. This varied greatly from year to year, with two peaks of incidence in 2004 and 2006 (Figure 2).

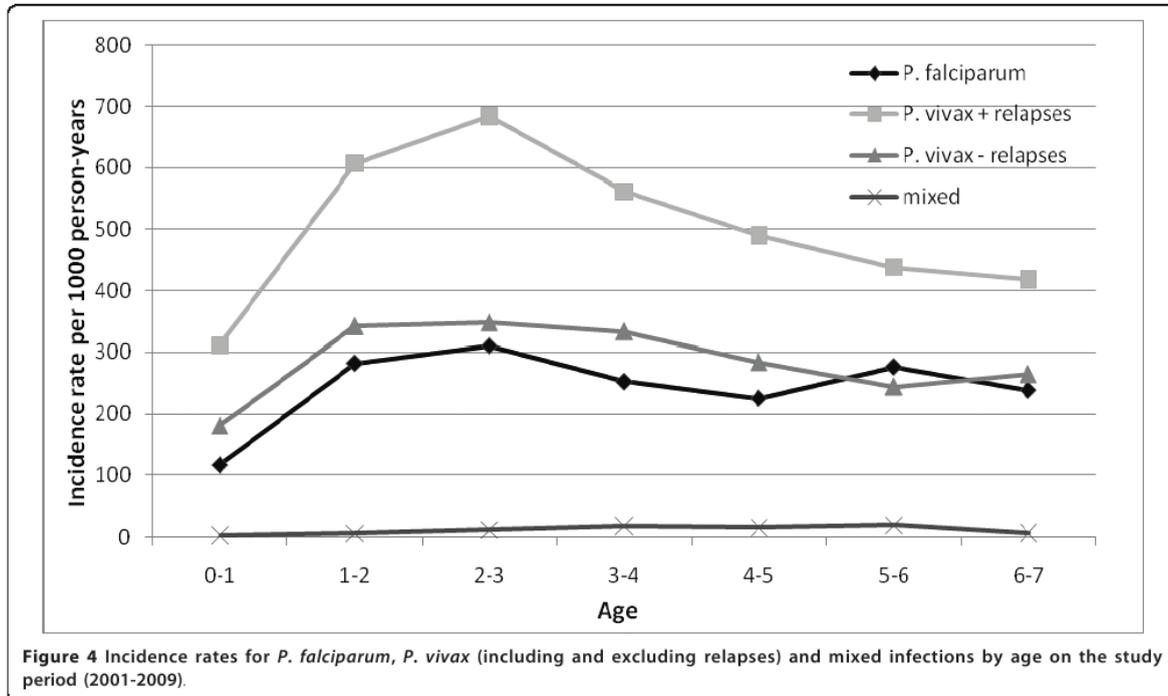
When studying the intra-annual variability of the incidence rate, two peaks were observed for *P. vivax*: one in January and another in June, whereas only one high peak was observed in January for *P. falciparum*. (Figure 3).

Malaria incidence by age

From birth to one year of age, the incidence was relatively low (117 per 1,000 and 180 per 1,000 person-years for *P. falciparum* and *P. vivax*, respectively) (Figure 4). The incidence strongly increased up to 2-3 years of age, where the incidence remained the highest (310 per 1,000 and 348 per 1,000 person-years for *P. falciparum* and *P. vivax*, respectively). After this age, the malaria incidence began to decrease until five years of age and remained more or less stable until seven

Chapitre 6 - Apports du couplage études écologiques/données climatiques dans la compréhension de la dynamique des maladies infectieuses : exemple de la Guyane Française.





years of age, around 250 per 1,000 person-years for both species.

Survival analysis

All malaria attacks were analysed regardless of the *Plasmodium* species involved. Among the 57 variables tested, 30 were significantly associated with malaria incidence in university analysis. The multivariate analysis identified six variables associated with occurrence of malaria transmission (Table 1). With regard to the living place, living on the downstream Oyapock riversides or on the Camopi riversides vs. on the upstream Oyapock riversides, as living in an isolated hamlet (more than 2 km from the main hamlet) and in a house with more than seven occupants, were associated with higher risk for malaria transmission. Regarding the environmental characteristics of the peridomiliary space, living in houses cleared of surrounding vegetation within 50 m and living far from the forest were associated with a lower risk of malaria. With regard to behavioural data, going to bed after 7:00 pm. was associated with an increased risk of malaria transmission. Any significant difference of malaria incidence between girls and boys was observed.

Intra-annual variations of climate parameters

In Camopi, the seasonality is determined by the rainfall, leading to a rainy season from December to June and a dry season from July to November (Figure 5). There is a

peak of malaria incidence at the beginning of the rainy season and another one at the end of it. The highest temperatures coincide with the driest months of the year. The lowest temperature (16.4°C) and the lowest mean minimum temperature (20.9°C) were recorded in June 2007. The highest temperature (38,0°C) and the highest mean maximum temperature (36.1°C) were recorded in November 2008. Malaria incidence was significantly higher during the rainy period than during the dry period (p-value < 0.001).

Association between climatic and hydrologic data and malaria incidence

Time-series analysis revealed nine relevant factors: minimum temperature at time t-3, t-9 and t-11 months; mean minimum temperature at t-1 and t-12 months; mean temperature at t-2 months, the number of days without rain at t-9 months and the maximum river level at t and t-1 month. The multivariate ARIMA model obtained with the inclusion of these nine variables, applying a stepwise descending procedure, is summarized in Table 2.

Spearman's correlations between malaria incidence and meteorological factors for which there is at least one statistically significant correlation between residual series are presented in Figure 6. The relationship between malaria incidence and the maximum river level at lag 0 was strongest in June (negative) and in October, November and December (positive). The correlation between incidence

Chapitre 6 - Apports du couplage études écologiques/données climatiques dans la compréhension de la dynamique des maladies infectieuses : exemple de la Guyane Française.

Table 1 Risk factors for malaria attacks

Variable	Categories	No. of children	Incidence rate	Hazard ratio	95% CI	p value
River	Upstream Oyapock	262	0.407	1.00		
	Downstream Oyapock	76	0.551	1.35	1.08-1.69	0.008**
	Camopi	228	0.729	1.37	1.13-1.66	0.002**
Percentage cleared around home within 50 metres	<50	150	0.686	1.00		
	50-75	279	0.542	0.81	0.71-0.93	0.002**
	>75	96	0.432	0.84	0.67-1.05	0.119
Distance to Camopi, metres	≤500	137	0.407	1.00		
	501-2000	208	0.451	1.13	0.93-1.38	0.208
	>2000	221	0.754	1.61	1.33-1.94	<0.001**
Distance to the forest, metres	≤50	169	0.737	1.00		
	51-150	202	0.538	0.79	0.68-0.91	0.001**
	>150	156	0.393	0.73	0.60-0.88	0.001**
Number of occupants	<7	223	0.512	1.00		
	7-10	195	0.594	1.17	1.01-1.35	0.037*
	≥11	107	0.591	1.26	1.07-1.48	0.005**
Bedtime, hours	<7:00 pm	103	0.489	1.00		
	≥7:00 pm	354	0.584	1.18	1.01-1.38	0.034*

Multivariate analysis of the risk of malaria transmission using a multiple failure Cox model, stratified for year of birth. The saturated model included 32 variables, which were significant at $p < 0.20$ in univariate analysis. Analysis time = 2279 person-years, CI = confidence interval. Symbols * and ** correspond to statistical significances at 0.05 and 0.01 first-order risks, respectively.

and mean minimum temperature one month before was strongest in April. The correlation between incidence and mean temperature two months before was strongest in January and February. The association between incidence and minimum temperature three months earlier was strongest in May (negative) and October (positive) whereas the relationship between incidence and minimum temperature eleven month earlier was strongest in November. The correlation between incidence and number of days without rain nine months before was highest in March.

Relationship between entomological data and malaria incidence

Human biting rates of the different collected anopheline species according to the four groups of hamlets are

presented Figure 7. Although *An. darlingi* was the main overall collected species, the *Anopheles nuneztovari* HBR was higher in the hamlets located on the Camopi River. A peak of HBR was observed in May for *An. darlingi* and a lower peak was observed in April-May for *An. nuneztovari*. No other peak of HBR was observed during the year for these two species. Regarding *An. darlingi* characteristics, a significant correlation was observed between malaria incidence and human biting rate (HBR) recorded one month earlier (p value = 0.03).

Discussion

During the nine years of the study, the global incidence rate in children of Camopi remained very high with two epidemics in 2004 and 2006, reaching 800 and 1,000 per 1,000 person-years respectively. During the first year of

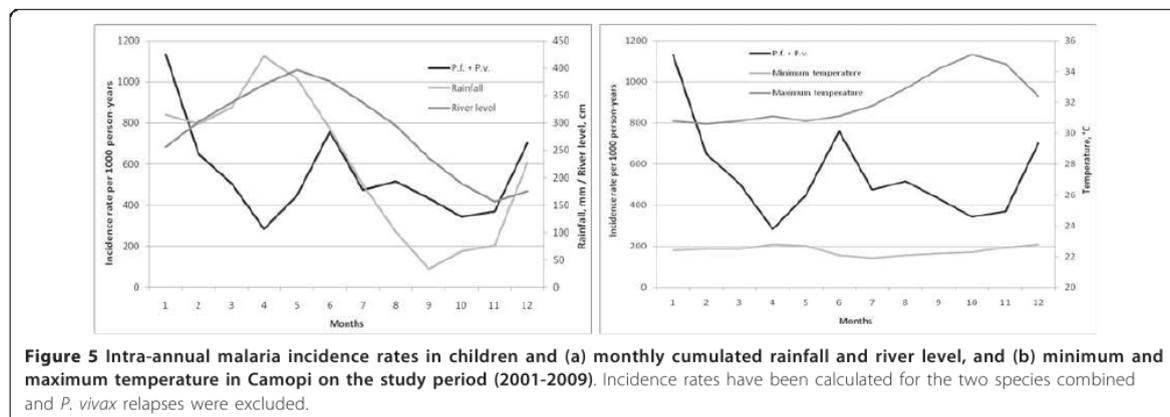


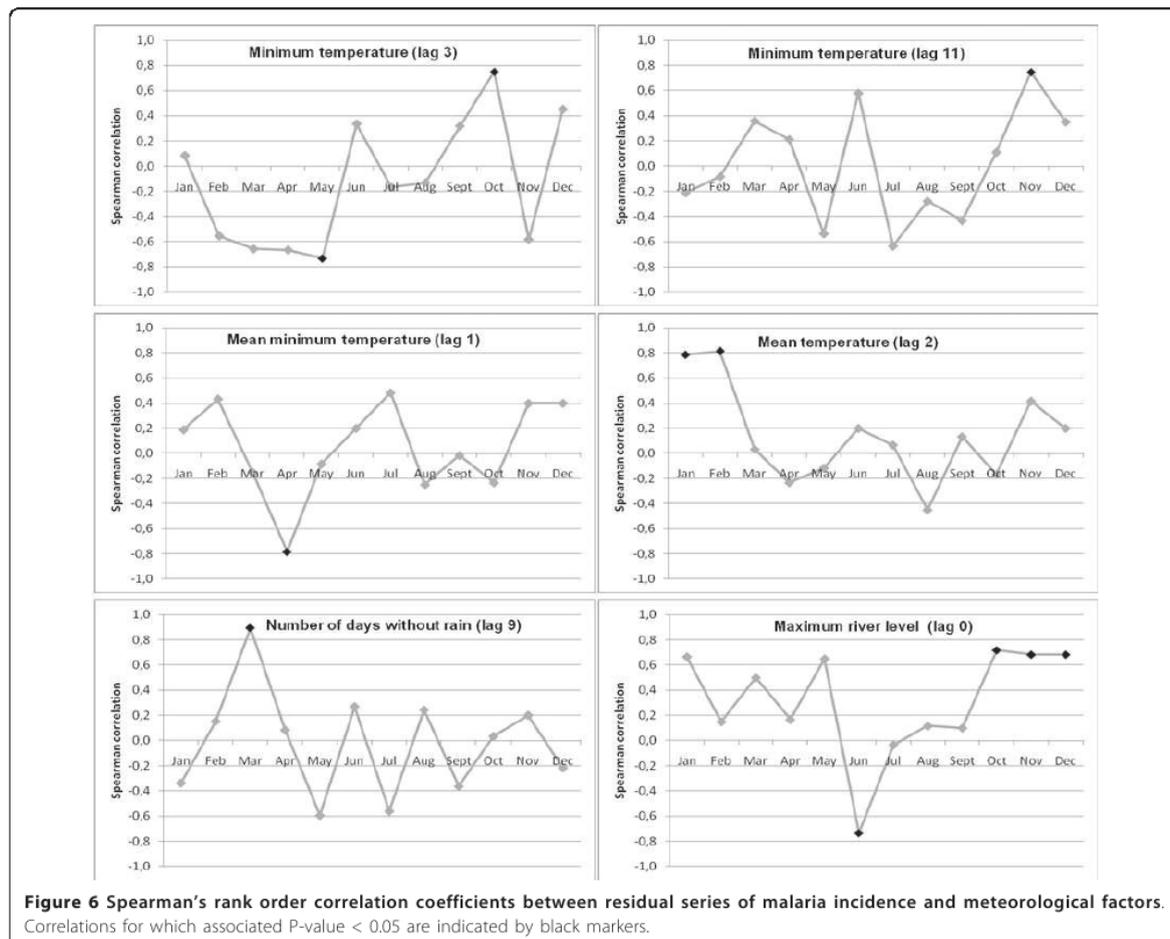
Table 2 ARIMA regression of the monthly malaria incidence in children (2001-2009) against meteorological and hydrological factors in Camopi, French Guiana

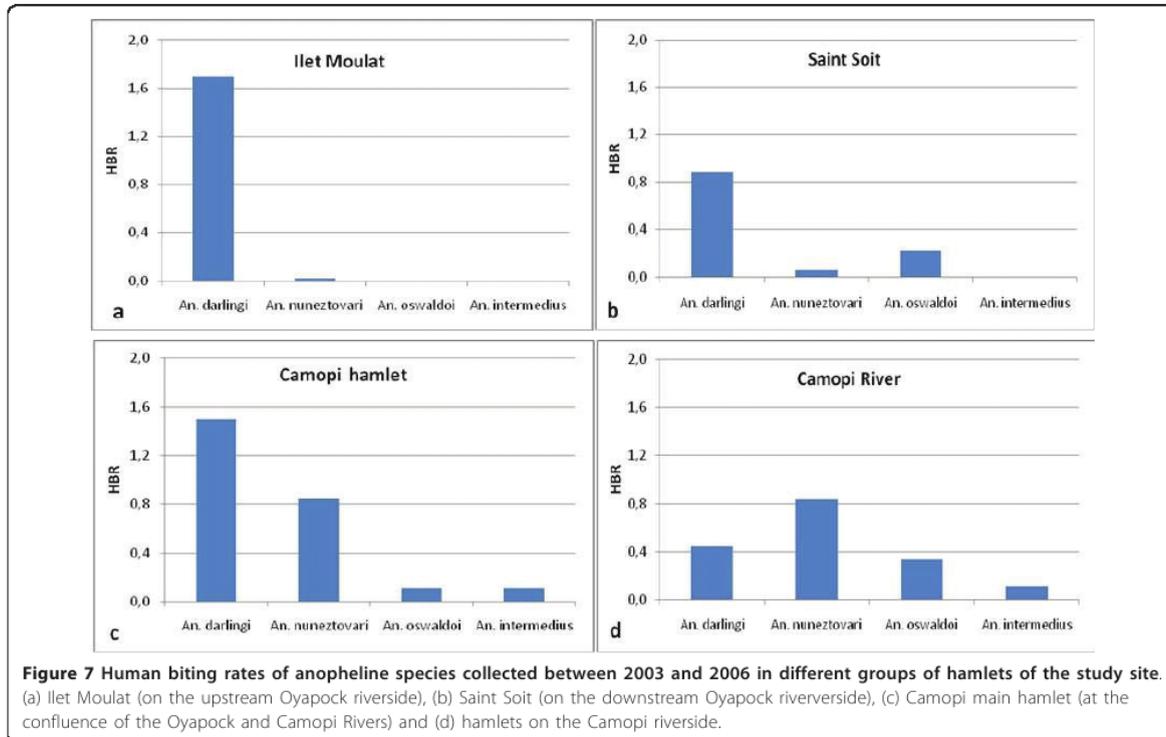
	Coefficient	Standard Error	P value
AR*	0.293	0.110	0.004
SMA**	-1.000	0.275	<0.001
Maximum river level (lag 0)	0.003	0.002	0.020
Maximum river level (lag 1)	0.004	0.002	0.018
Mean of minimum temperature (lag 1)	0.278	0.140	0.023
Mean temperature (lag 2)	0.498	0.171	0.002
Minimum temperature (lag 3)	0.128	0.076	0.046
Minimum temperature (lag 9)	0.160	0.076	0.018
Number of days without rain (lag 9)	0.035	0.015	0.011
Minimum temperature (lag 11)	0.144	0.086	0.046
Mean of minimum temperature (lag 12)	0.355	0.153	0.010

*AR: autoregressive parameter; ** SMA: seasonal moving average parameter; AIC of the model = 232.58. Lag scale = months.

life, the incidence was quite low presumably because the presence of maternal antibodies provides partial immunity. At one year of age, the incidence began to increase sharply to reach a maximum between two and three

years old (almost 1,000 per 1,000 person-years). Immunity seems to begin to develop at this age, for both species, but appeared to be mainly protective against *P.vivax* relapses. Amerindian children have different





activities as they grow and gain independence in their movements in the village. Moreover, around seven or eight years of age, boys begin to go to the forest for hunting and to the river for fishing whereas girls accompany their mother to slash. Hence, around five years of age, a different exposure may be responsible for the small increase of *P. falciparum*.

The survival analysis showed a strong association between environmental exposure and malaria transmission in children under seven years old. Indeed, the analysis permitted to identify several risk factors linked to the environmental characteristics of the surroundings of the house. Malaria transmission increased with distance from the central main hamlet of Camopi. Therefore, children living in isolated hamlets had an increased risk of malaria. However, a higher number of people in the same home (more than seven occupants) were independently associated with a higher risk of malaria. Human aggregation is likely to increase the probability for vectors near homes to be infected [4,15].

Regarding other environmental factors, the proportion of cleared vegetation within 50 m around the houses was a protective factor for malaria, as previously described. Indeed, this characteristic of land cover is not favourable for the rest of adults and the maintenance of breeding sites. Proximity to the forest was associated with a higher

risk of malaria. According to other authors, when houses are located not far from the forest, *An. darlingi* returns to the forest after feeding [16-19]. However, non-environmental factors may be partly responsible for the relation between increasing incidence and the distance from the main hamlet of Camopi. There was a significantly different risk of malaria according to the ethnic group in univariate analysis. A previous study in Camopi found a strong association of ethnicity with first malaria attack, even after adjusting for behavioural and environmental factors [4]. These results could suggest that Emerillon children have higher genetic susceptibility than Wayampi children. However, this phenomenon was not visible in our multivariate analysis, which did not take into account *P. vivax* relapses and thus was then closer to the transmission phenomenon.

Furthermore, another variable is likely to better explain the malaria incidence while being correlated to ethnicity. This is the case of the variable "river" divided in three groups: the upstream Oyapock riverside where the Wayampi live, the downstream Oyapock where the mixed ethnic groups live and the Camopi riverside where the Emerillon live. The behaviour of the residents and the protection measures used also play a role in the incidence of the disease. Indeed, the use of topical repellents and domestic insecticides and interventions of the county

mosquito control service were significantly associated with a lower risk of transmission in univariate analysis. In addition, children who used to go to sleep after 7:00 pm had a higher risk of transmission. It is likely that children going to sleep earlier are protected by mosquito net at dusk, when *An. darlingi* reaches its first peak of activity [7].

Climatic and hydrologic variations appeared to have an impact on malaria incidence at relatively short-term (lag 0 to lag 3) and at longer term (lag 9 to lag 12). Considering the short-term effect, some plausible explanations can be put forward. The mean minimum temperature was globally positively associated with malaria incidence one month later and the minimum temperature was positively associated with malaria incidence three months later. Moreover, the mean temperature was positively associated with a higher incidence that occurred two months later, and especially at the beginning of the rainy season. These observations could be explained by the fact that an increase temperature shortens the interval between egg-laying episodes and enhances the larval development. Moreover, a higher temperature is also likely to accelerate the sporogonic cycle of the *Plasmodium*. Thus, a high temperature may allow better survival of vector populations and therefore a higher transmission that could be responsible for an increasing incidence in the following months. Conversely, lower minimum temperatures may be responsible for a decreased incidence, slowing the sporogonic cycle of the parasite and decreasing vector survival. Regarding hydrological factors, incidence rates were positively associated with the maximum river level at the same month and one month earlier. Thus, particularly high water may create flooding on the river bed, leading to the creation of suitable larval breeding sites, particularly at the end of the year. This phenomenon has been observed along the Maroni River and in other countries of Central and South America [20-24]. Thus, an increase in the larval anopheline abundance may increase the malaria transmission related to the adult stage. This has been previously observed at a weekly temporal resolution in Camopi where vector abundance was positively correlated with the river level a few weeks earlier [10].

Regarding the long-term impact of climate on malaria incidence, it is difficult to grasp the meaning of these correlations and the statistical results have to be considered with caution. The significant results could be due to unidentified confounding factors or residual effects of seasonal factors that are not taken into account at a short term. This long-term hypothetical effect has been previously observed in Cacao, French Guiana, where meteorological conditions in a given year may affect malaria in the following year [25]. Nevertheless, the biological impact of meteorological factors on vector populations over a long

period can only have hypothetical explanations. Overall, annual climate seasonality was linked to malaria seasonality as observed by others [26,27]. In French Guiana, a global and durable climate anomaly such as El Niño episode is likely to increase malaria [28].

Anopheles darlingi human biting rate was correlated to the malaria incidence rate in children one month later. This is consistent with what was previously found in correlating entomological data with malaria incidence in the general population [10]. Nevertheless, none of the 148 specimens of this species collected from January 2003 to December 2006 in Camopi was found naturally infected with *Plasmodium* [10]. This study focused on young children with the hypothesis of a nightly transmission due to the characteristics of *An. darlingi* [5]. Other authors mentioned that *An. darlingi* has a 24 hours activity and can be found outside during the day in French Guiana [7]. Other anopheline species could play a role in the transmission of malaria including during the morning [6,29,30] or in a sylvatic environment around the hamlets [10]. Indeed, *An. nuneztovari* could play a role in the transmission, along the Camopi River where the incidence remains higher. Indeed, this species could be a secondary vector when present in sufficient numbers. This species, exophilic and aggressive on humans, may be collected on humans in large numbers in some Amazonian areas [20,31,32].

Given that it can be assumed that *An. darlingi* transmitted *P. vivax* and *P. falciparum*, it was intriguing to observe that *P. vivax* had a high transmission in May and June whereas *P. falciparum* had a much lower transmission during the last period. A tentative explanation is that the age composition of the *An. darlingi* population may depend on the seasons and the environment [9,33]. Thus, long-lived females that would be good vectors for the two plasmodial species might be responsible for malaria transmission in December and January. On the other hand, females with a lower life expectancy would be poor vectors for *P. falciparum* which has a longer extrinsic cycle than *P. vivax*, and should be responsible for the transmission of *P. vivax* in May and June only. Unfortunately, the comparison of survival rates of females in these two periods could not be performed due to the low numbers of mosquitoes collected. Another explanation is that *P. vivax* relapses in the months following the first peak provide gametocytes to the emerging vector population thus amplifying the *vivax* transmission.

Even if the distribution of malaria is determined by climatic and other geographic factors which affect mosquito and *Plasmodium* reproduction at a given time, malaria is also influenced by environmental changes [34]. Therefore, the impact of deforestation on malaria transmission that has been previously described [18,34-36] should be investigated in this area where the settlement of Amerindian

populations coupled with gold-mining activities constantly cause forest openings.

Conclusions

In Camopi, malaria incidence in children remained very high over the 2001 - 2009 period despite the appearance of immunity around three years of age. *Plasmodium falciparum* and *P. vivax* incidences presented two different intra-annual patterns. The occurrence of malaria is mainly due to the environmental features around the homes, and not so much to individual behaviours or population habits. Although malaria incidence is associated with a complex array of variables, the present data gives a clear picture of where households should not be built, and how the peridomestic space should be kept to minimize the risk of malaria. Moreover, meteorological and hydrological variations such as minimum temperatures and maximum river water level seem to have a relevant impact on malaria incidence in the short term (from zero to three months-lags). Nevertheless, the results of the present study should be carefully extrapolated to the general population because transmission in adolescents and adults is partly different due to a different exposure and a stronger immunity.

Acknowledgements

The study was supported by the Centre National d'Etudes Spatiales (CNES) and the Fonds Social Européen (FSE). Climate data were provided by MétéoFrance Guyane Française. We warmly thank all the inhabitants of Camopi and the sanitary staff of Camopi, particularly Christelle Suitman, Anne Viaud and Alain Maniglier. We also thank the interpreters and guide in Camopi: Dimitri Suitman, François Miso and Merlin Sakeu. We gratefully thank Dr. Isabelle Dusfour (Unité d'entomologie médicale de l'Institut Pasteur de la Guyane) for her review of the entomological aspects of the manuscript and Dr. Claire Grenier and Dr. Michel Joubert from the Département des centres de santé of the Cayenne Hospital.

Author details

¹EPaT Team (EA3593), UFR de Médecine - Université des Antilles et de la Guyane, Cayenne, French Guiana. ²Centre d'Investigation Clinique - Epidémiologie Clinique Antilles Guyane (CIE 802 INSERM), Cayenne General Hospital, French Guiana. ³Unité d'Entomologie Médicale, Institut Pasteur de la Guyane, Cayenne, French Guiana. ⁴Laboratoire Hospitalo-Universitaire de Parasitologie Mycologie, Cayenne General Hospital, French Guiana.

Authors' contributions

AS participated in the research design, data collection, analysis and interpretation, and prepared the manuscript. MH participated in the data analysis and interpretation, and revised the manuscript. MN provided guidance in data analysis and contributed to data interpretation and manuscript revision. RG was involved in the entomological data interpretation and manuscript revision. BC was responsible for overall scientific management, designed the cohort and was involved in the interpretation of data and manuscript revision. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 15 May 2011 Accepted: 23 August 2011
Published: 23 August 2011

References

1. Carme B, Venturin C: [Malaria in the Americas](in French). *Med Trop (Mars)* 1999, **59**:298-302.
2. Carme B, Ardillon V, Girod R, Grenier C, Joubert M, Djossou F, Ravachol F: [Update on the epidemiology of malaria in French Guiana] (in French). *Med Trop (Mars)* 2009, **69**:19-25.
3. Carme B, Lecat J, Lefebvre P: [Malaria in an outbreak zone in Oyapock (French Guiana): incidence of malaria attacks in the American Indian population of Camopi] (in French). *Med Trop (Mars)* 2005, **65**:149-154.
4. Hustache S, Nacher M, Djossou F, Carme B: Malaria risk factors in Amerindian children in French Guiana. *Am J Trop Med Hyg* 2007, **76**:619-625.
5. Floch H: La lutte antipaludique en Guyane française. L'anophélisme. *Riv Malariol* 1955, **24**:57-65.
6. Mouchet J, Nadire-Galliot M, Poman JP, Claustre J, Bellony S: Le paludisme en Guyane. Les caractéristiques des différents foyers et la lutte antipaludique. *Bull Soc Pathol Exot* 1989, **82**:393-405.
7. Pajot F-X, Le Pont F, Molez J-F, Degallier N: Agressivité d'*Anopheles (Nyssorhynchus) darlingi* Root, 1926 (Diptera, Culicidae) en Guyane française. *Cahiers de l'ORSTOM, Série entomologie médicale et parasitologie* 1977, **15**:15-22.
8. Girod R, Gaborit P, Carinci R, Issaly J, Fouque F: *Anopheles darlingi* bionomics and transmission of *Plasmodium falciparum*, *Plasmodium vivax* and *Plasmodium malariae* in Amerindian villages of the Upper-Maroni Amazonian forest, French Guiana. *Mem Inst Oswaldo Cruz* 2008, **103**:702-710.
9. Fouque F, Gaborit P, Carinci R, Issaly J, Girod R: Annual variations in the number of malaria cases related to two different patterns of *Anopheles darlingi* transmission potential in the Maroni area of French Guiana. *Malar J* 2010, **9**:80.
10. Girod R, Roux E, Berger F, Stefani A, Gaborit P, Carinci R, Issaly J, Carme B, Dusfour I: Unravelling the relationships between *Anopheles darlingi* (Diptera: Culicidae) densities, environmental factors and malaria incidence: understanding the variable patterns of malarial transmission in French Guiana (South America). *Ann Trop Med Parasitol* 2011, **105**:107-122.
11. Demaze MT, Manusset S: L'agriculture itinérante sur brûlis en Guyane française: la fin des durabilités écologique et socioculturelle ? *Cahiers Outre Mer* 2008, **61**:31-48.
12. Box G, Jenkins G: *Time series analysis: Forecasting and control* San Francisco: Holden-Day; 1970.
13. Ehrman FC, Ellis JM, Young MD: *Plasmodium vivax* Chesson strain. *Science* 1945, **101**:377.
14. Hanf M, Stefani A, Basurko C, Nacher M, Carme B: Determination of the *Plasmodium vivax* relapse pattern in Camopi, French Guiana. *Malar J* 2009, **8**:278-278.
15. Barrera R, Grillet ME, Rangel Y, Berti J, Aché A: Temporal and spatial patterns of malaria reinfection in northeastern Venezuela. *Am J Trop Med Hyg* 1999, **61**:784-790.
16. Rozendaal JA, Van Hoof JP, Voorham J, Oostburg BF: Behavioral responses of *Anopheles darlingi* in Suriname to DDT residues on house walls. *J Am Mosq Control Assoc* 1989, **5**:339-350.
17. Charlwood JD: Biological variation in *Anopheles darlingi* Root. *Mem Inst Oswaldo Cruz* 1996, **91**:391-398.
18. Tadei WP, Thatcher BD, Santos JM, Scarpassa VM, Rodrigues IB, Rafael MS: Ecologic observations on anopheline vectors of malaria in the Brazilian Amazon. *Am J Trop Med Hyg* 1998, **59**:325-335.
19. Tadei WP, Dutary Thatcher B: Malaria vectors in the Brazilian amazon: *Anopheles* of the subgenus *Nyssorhynchus*. *Rev Inst Med Trop Sao Paulo* 2000, **42**:87-94.
20. Rozendaal JA: Observations on the biology and behaviour of Anophelines in the Suriname rainforest with special reference to *Anopheles darlingi* Root. *Cahiers de l'ORSTOM, Série entomologie médicale et parasitologie* 1987, **25**:33-43.
21. Rozendaal JA: Relations between *Anopheles darlingi* breeding habitats, rainfall, river level and malaria transmission rates in the rain forest of Suriname. *Med Vet Entomol* 1992, **6**:16-22.
22. Manguin S, Roberts DR, Andre RG, Rejmankova E, Hakre S: Characterization of *Anopheles darlingi* (Diptera: Culicidae) larval habitats in Belize, Central America. *J Med Entomol* 1996, **33**:205-211.

23. Camargo LM, dal Colletto GM, Ferreira MU, Gurgel S de M, Escobar AL, Marques A, Krieger H, Camargo EP, da Silva LH: **Hypoendemic malaria in Rondonia (Brazil, western Amazon region): seasonal variation and risk groups in an urban locality.** *Am J Trop Med Hyg* 1996, **55**:32-38.
24. Gil LHS, Tada MS, Katsuragawa TH, Ribolla PEM, da Silva LHP: **Urban and suburban malaria in Rondônia (Brazilian Western Amazon) II: perennial transmissions with high anopheline densities are associated with human environmental changes.** *Mem Inst Oswaldo Cruz* 2007, **102**:271-276.
25. Basurko C, Hanf M, Han-Sze R, Rogier S, Héritier P, Grenier C, Joubert M, Nacher M, Carme B: **Influence of climate and river level on the incidence of malaria in Cacao, French Guiana.** *Malar J* 2011, **10**:26.
26. Olson SH, Gangnon R, Elguero E, Durieux L, Guégan JF, Foley JA, Patz JA: **Links between climate, malaria, and wetlands in the Amazon Basin.** *Emerg Infect Dis* 2009, **15**:659-662.
27. Galardo AKR, Zimmerman RH, Lounibos LP, Young LJ, Galardo CD, Arruda M, D'Almeida Couto AAR: **Seasonal abundance of anopheline mosquitoes and their association with rainfall and malaria along the Matapi River, Amapá, [corrected] Brazil.** *Med Vet Entomol* 2009, **23**:335-349.
28. Hanf M, Adenis A, Nacher M, Carme B: **The role of El Niño Southern Oscillation (ENSO) on variations of monthly *Plasmodium falciparum* malaria cases at the Cayenne General Hospital, 1996-2009, French Guiana.** *Malar J* 2011, **10**:100.
29. Pajot F-X, Molez J-F, Le Pont F: **Anophèles et paludisme sur le Haut-Oyapock (Guyane française).** *Cahiers de l'ORSTOM, Série entomologie médicale et parasitologie* 1978, **16**:105-111.
30. Raccurt CP: **[Malaria, anopheles, the anti-malaria campaign in French Guyana: between dogmatism and judgment](in French).** *Med Trop (Mars)* 1997, **57**:401-406.
31. Panday RS: ***Anopheles nuneztovari* and malaria transmission in Suriname.** *Mosq news* 1977, **37**:728-737.
32. Rojas JE, Milano MS, Avila IG: **[Studies on preadult and adult forms of *Anopheles nuñeztovari* (Diptera: Culicidae) Gabaldon 1940 in an originally malarial area in Mérida state, Venezuela](in Spanish).** *Rev Cubana Med Trop* 2002, **54**:127-133.
33. Hiwat H, Issaly J, Gaborit P, Somai A, Samjhawan A, Sardjoe P, Soekhoe T, Girod R: **Behavioral heterogeneity of *Anopheles darlingi* (Diptera: Culicidae) and malaria transmission dynamics along the Maroni River, Suriname, French Guiana.** *Trans R Soc Trop Med Hyg* 2010, **104**:207-213.
34. Patz JA, Graczyk TK, Geller N, Vittor AY: **Effects of environmental change on emerging parasitic diseases.** *Int J Parasitol* 2000, **30**:1395-1405.
35. Vittor AY, Gilman RH, Tielsch J, Glass G, Shields T, Lozano WS, Pinedo-Cancino V, Patz JA: **The effect of deforestation on the human-biting rate of *Anopheles darlingi*, the primary vector of *Falciparum* malaria in the Peruvian Amazon.** *Am J Trop Med Hyg* 2006, **74**:3-11.
36. Vittor AY, Pan W, Gilman RH, Tielsch J, Glass G, Shields T, Sánchez-Lozano W, Pinedo VV, Salas-Cobos E, Flores S, Patz JA: **Linking deforestation to malaria in the Amazon: characterization of the breeding habitat of the principal malaria vector, *Anopheles darlingi*.** *Am J Trop Med Hyg* 2009, **81**:5-12.

doi:10.1186/1475-2875-10-246

Cite this article as: Stefani et al.: Environmental, entomological, socioeconomic and behavioural risk factors for malaria attacks in Amerindian children of Camopi, French Guiana. *Malaria Journal* 2011 **10**:246.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



6.10. Etude 4 : Le rôle de l'Oscillation Australe El Niño dans la variation du nombre mensuel de cas de paludisme à *Plasmodium falciparum* au Centre Hospitalier de Cayenne, Guyane Française, durant la période 1996-2009

Introduction : Depuis longtemps, le paludisme est un problème de santé publique en Guyane Française. Cette dernière est située en Amérique du Sud dans une zone potentiellement à risque de sécheresse lors de la manifestation d'un événement El Niño et où une aggravation des épidémies de paludisme pourrait survenir consécutivement aux perturbations climatiques engendrées par l'Oscillation Australe El Niño. Cette étude se propose d'étudier le lien existant entre niveau d'épidémie de paludisme et manifestations El Niño en Guyane Française.

Matériels/Méthodes : Une analyse par série temporelle utilisant une méthodologie ARIMA a été développée pour étudier les corrélations temporelles entre le nombre de cas mensuel de paludisme à *P.falciparum* au centre hospitalier de Cayenne et l'Oscillation Australe El Niño mesurée à l'aide d'un index quantifiant l'oscillation australe (SOI). La période d'étude s'étendait entre les années 1996 et 2009.

Résultats : Une influence significative positive à un décalage de 3 mois d'El Niño sur le nombre mensuels de cas à *P.falciparum* a été retrouvée par l'analyse ($p < 0.001$). L'incorporation des données quantifiant le phénomène El Niño dans le modèle permettait une réduction de 4% de l'AIC du modèle.

Conclusions : Bien que l'étude ait identifié un lien statistique significatif, la valeur prédictive de l'Oscillation Australe El Niño pour le paludisme à *P.falciparum* semble marginale. Ce résultat est une information importante pour le développement de futurs systèmes de surveillance capables de détecter précocement les épidémies de paludisme en Guyane Française. Des travaux supplémentaires sont aussi nécessaires pour renseigner la dépendance locale du paludisme vis-à-vis d'El Niño.

Publiée dans *Malaria Journal* le : 22 Avril 2011

RESEARCH

Open Access

The role of El Niño southern oscillation (ENSO) on variations of monthly *Plasmodium falciparum* malaria cases at the cayenne general hospital, 1996-2009, French Guiana

Matthieu Hanf^{1*}, Antoine Adenis¹, Mathieu Nacher¹ and Bernard Carne^{1,2}

Abstract

Background: Malaria remains a serious problem in French Guiana, which is at potential risk for drought linked with the El Niño Event and where there could be a risk of malaria epidemic after the onset of an El Niño event.

Methods: A time series analysis using ARIMA was developed to investigate temporal correlations between the monthly *Plasmodium falciparum* case numbers and El Niño Southern Oscillation (ENSO) as measured by the Southern Oscillation Index (SOI) at the Cayenne General Hospital between 1996 and 2009.

Results: The data showed a positive influence of El Niño at a lag of three months on *P. falciparum* cases ($p < 0.001$). The incorporation of SOI data in the ARIMA model reduced the AIC by 4%.

Conclusions: Although there is a statistical link, the predictive value of ENSO to modulate prevention intervention seems marginal in French Guiana. However, additional work should refine the regional dependence of malaria on the ENSO state.

Background

In French Guiana, malaria remains a serious problem with a number of malaria cases observed every year ranging from 3,500 to 4,500. Both *Plasmodium falciparum* and *Plasmodium vivax* are prevalent. Since 2000, 60% of the cases are due to *P. vivax* [1].

The ENSO phenomenon refers to the cyclic warming and cooling of the equatorial Pacific Ocean coupled with changes in the atmospheric pressure across the Pacific. This is the most important climatic cycle contributing to worldwide interannual variability in climate and the likelihood of climatic anomalies. The two extremes of ENSO are El Niño (a warm event) and La Niña (a cold event), which can create rainfall and temperature fluctuations [2].

The El Niño-Southern Oscillation (ENSO) is the best-known example of quasi-periodic natural climate

variability on the interannual time scale. It comprises changes in sea temperature in the Pacific Ocean (El Niño) and changes in atmospheric pressure across the Pacific Basin (the Southern Oscillation), together with resultant effects on world weather. El Niño events occur at intervals of 2-7 years. In certain countries around the Pacific and beyond, El Niño is associated with extreme weather conditions that can cause floods and drought. The other extreme (La Niña) (cooling) is the reverse pattern of climate anomalies.

There is strong evidence that El Niño Southern Oscillation (ENSO) is associated with increased risk of malaria in regions of the world where climate is linked to the ENSO cycle [3]. These include, among others, countries in South America [4-8]. At a regional level, French Guiana is located in a potential risk area for drought linked with the El Niño Event and where there could be a risk of malaria epidemic after the onset of an El Niño event [3].

The purpose of the investigations reported here was to test and quantify the influence of ENSO on malaria

* Correspondence: matthieu@hanf.fr

¹Centre d'Investigation Clinique Epidémiologie Clinique Antilles Guyane CIC-EC CIE 802, Cayenne General Hospital, EHPAD, Avenue des Flamboyants, BP 6006 97300 Cayenne, French Guiana, France

Full list of author information is available at the end of the article

transmission in French Guiana using monthly data of malaria cases at the Cayenne General Hospital. Because of the difficulty to discern *P. vivax* malaria relapses from re-infections, this study focused only on *P. falciparum* malaria transmission.

Methods

ENSO data

The Southern Oscillation Index (SOI) is designed to measure the strength and phase of the Southern Oscillation. SOI is a measure based on the differences in the atmospheric pressure between Tahiti in the eastern equatorial pacific and Darwin in Australia (West Pacific), expressed as a standard deviation from the norm. SOI is used to quantify the strength of an ENSO event. It is generally negative during an El Niño event and positive during a La Niña event. SOI was taken from the Australian bureau of meteorology [9].

Malaria data

In spite of some encouraging results to statistically differentiate *P. vivax* malaria relapses from re-infections at individual scale [10], due to the difficulties to discern *P. vivax* malaria relapses from re-infection in endemic areas as French Guiana [11], this study focused only on *P. falciparum* transmission.

All clinical malaria episodes diagnosed for patients consulting the emergency service of Cayenne hospital were retrospectively collected. Malaria was defined as temperature $> 38^{\circ}\text{C}$ at the time of consultation or fever within the past 48 hours associated with Plasmodium asexual forms on a thin blood smear (screening sensitivity: 50 Plasmodium per μl)

Data were extracted from the Parasitological Laboratory files using designated software in accordance with legal procedures concerning confidentiality. Malaria data from 1996 to 2009 were recorded as numbers of cases of *P. falciparum* malaria per month.

Data analysis

ARIMA models were used to evaluate the relationship between SOI and monthly malaria cases. Before, the ARIMA methodology application, a log-transformation of the series was made to stabilize variance.

Modelling with ARIMA involves the estimation of a series of parameters to account for the inherent dynamics in the time series, including the trends and autoregressive and moving average processes. The general model introduced by Box and Jenkins [12] includes autoregressive and moving average parameters, and explicitly includes differencing in the formulation of the model. An ARIMA (p, d, q) model comprises three types of parameters: the autoregressive parameters (p), number of differencing passes (d), and moving average parameters (q). The

multiplicative seasonal ARIMA (p, d, q) (P, D, Q)s model is an extension of the ARIMA method to time series in which a pattern repeats seasonally over time. Analogous to the simple ARIMA parameters, the seasonal parameters are: seasonal autoregressive (P), seasonal differencing (D), and seasonal moving average parameters (Q). The length of the seasonal period is represented by s.

Various permutations of the order (p, d, q) (P, D, Q)s were computed, and the optimal combination of parameters using the « Akaike's information criterion (AIC) » was chosen. The smallest values of Akaike's information criterion (AIC) were set as the standard to identify the best-fit model [12]. The correlogram and partial correlogram graphs were also used to help in fixing the value of orders to include in the model. The likelihood ratio test was used to determine if inclusion of other covariates helped to improve the fit of the model.

In the model, SOI, at lags of 0 to 12 months, respectively, was then fitted into the ARIMA model to screen for potential predictors of the *P. falciparum* case numbers. Lag series significantly associated with malaria, (p-value < 0.05), were singled out to fit the best multivariate ARIMA model. The Ljung-Box Q test was applied to ascertain whether the residual series were white noise.

Results

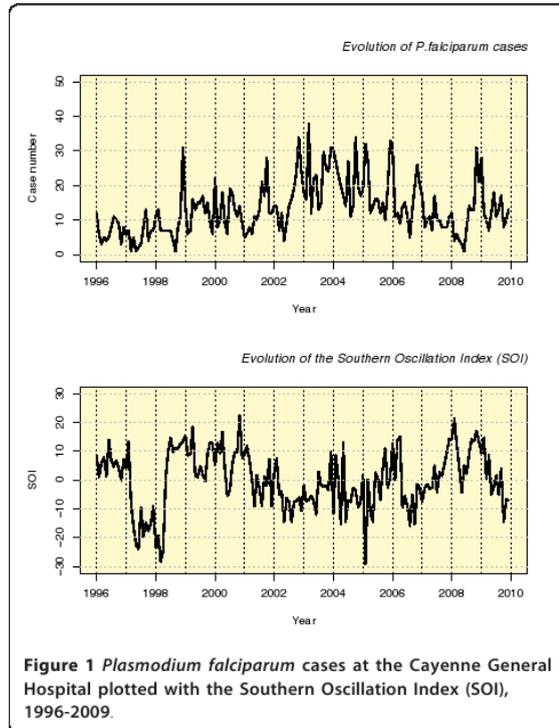
The monthly observed cases of *P. falciparum* malaria, and SOI, were shown in Figure 1. Of all the models tested, the ARIMA (1, 1, 1) model fitted the data best according to AIC. The best fitting model showed that no significant seasonal pattern explained the monthly cases of *P. falciparum* malaria alone. The results of ARIMA analysis reveal that the logarithm of the malaria cases was significantly influenced by SOI (Table 1).

In this model, the local moving average parameter is -0.8961 ($p = < 0.001$), and the auto-regression term is 0.3813 ($p = < 0.001$). With SOI data in the model, logarithm of *P. falciparum* cases number was negatively associated with SOI at a lag of three months ($\beta = -0.0121$, $p = 0.006$). The incorporation of SOI data in the ARIMA model reduced the AIC by 4%.

Discussion

There were two main limitations in the present study. The first one is that this study has used aggregated data coming from the whole heterogenic territory of French Guiana and has to be completed by additional studies at a local scale to determine whether a greater proportion of the variance is predicted, and thus giving a more practical value to the observed correlations in terms of prevention.

The second one is that, although 60% of malaria cases are caused by *P. vivax* in French Guiana, *P. vivax* cases were not included in this analysis due to impossibility to



differentiate relapses from re-infections due to missing detailed patient follow-up data.

Nevertheless, this study showed a consistent temporal correlation of ENSO with *P. falciparum* cases in Cayenne General Hospital with a three-month time lag from 1996 through 2009. Although ENSO influence on malaria transmission was suspected in French Guiana, this study is the first to identify a statistically significant correlation between El Niño and malaria cases in this area. A decreasing in SOI three months before was found to act as a positive factor on *P. falciparum* transmission.

The identification of a three-month lag allowed to quantify the delay between ENSO perturbation and consequences on malaria epidemics in French Guiana.

Table 1 Estimated coefficients by ARIMA Model between SOI and logarithmic transformation of malaria cases in Cayenne General Hospital, 1996-2009*

	Coef.	Std. Err.	P.value
AR	0.3813	0.0941	< 0.001
MA	-0.8961	0.0479	< 0.001
SOI lag 3	-0.0121	0.0049	< 0.001

* AR: Autoregressive parameter
 MA: moving average parameter
 AIC of the model: 261.12

Possible explanations for the identified association between ENSO and malaria in French Guiana include the effect of climate on the population dynamics of vectors (mainly *Anopheles darlingi* in French Guiana) through changes in population densities or survival rates, through availability of adequate breeding sites, but also through parasite associated parameters and man-vectors contacts.

The data indicated that ENSO explain 4% of the variation in malaria case numbers in the studied region. At the study scale, the remaining 96% of the variation possibly involved non-climatic causes, such as population immunity and socio-environmental factors that influence the breeding and ecology of mosquito vectors. This result suggested that changes in French Guiana are more likely to be a reflection of the efficiency of local control measures, population movements, therapeutic change, and others social factors [13]. In French Guiana, the role of recent immigration for gold-mining activity is well known as a main drivers of malaria epidemics [14].

To conclude, although there is a statistical link, the predictive value of ENSO to modulate prevention intervention seems marginal. However, the weak predictable consequence of ENSO for malaria transmission is an important piece of information for the development of future early warning systems able to detect efficiently malaria epidemics in the area.

Author details

¹Centre d'Investigation Clinique Epidémiologie Clinique Antilles Guyane CIC-EC CIE 802, Cayenne General Hospital, EHPAD, Avenue des Flamboyants, BP 6006 97300 Cayenne, French Guiana, France. ²Laboratoire Hospitalo-Universitaire de Parasitologie et Mycologie Médicale, Equipe EA3593, UFR Médecine - Université des Antilles et de la Guyane, Cayenne, Guyane Française, France.

Authors' contributions

MH participated in the research design, performed data analysis and interpretation, and prepared the manuscript. AA participated in the interpretation of data and manuscript revision as well. MN provided guidance on data analyses and was involved in the interpretation of data and manuscript revision. BC was responsible for data collection, initiated the study, and was involved in the interpretation of data and manuscript revision. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 29 September 2010 Accepted: 22 April 2011

Published: 22 April 2011

References

1. Carne B, Ardillon V, Girod R, Grenier C, Joubert M, Djossou F, Ravachol F: [Update on the epidemiology of malaria in French Guiana]. *Med Trop (Mars)* 2009, **69**:19-25.
2. Trenberth K: **General characteristics of El Niño-Southern Oscillation. Teleconnections: Linking Worldwide Climate Anomalies** Cambridge, UK: Cambridge University Press; 1991, 13-42.
3. Kovats RS, Bouma MJ, Hajat S, Worrall E, Haines A: El Niño and health. *Lancet* 2003, **362**:1481-1489.

Chapitre 6 - Apports du couplage études écologiques/données climatiques dans la compréhension de la dynamique des maladies infectieuses : exemple de la Guyane Française.

Hanf *et al. Malaria Journal* 2011, **10**:100
<http://www.malariajournal.com/content/10/1/100>

Page 4 of 4

4. Bouma MJ, Dye C: **Cycles of malaria associated with El Nino in Venezuela.** *JAMA* 1997, **278**:1772-1774.
5. Bouma MJ, Poveda G, Rojas W, Chavasse D, Quinones M, Cox J, Patz J: **Predicting high-risk years for malaria in Colombia using parameters of El Nino Southern Oscillation.** *Trop Med Int Health* 1997, **2**:1122-1127.
6. Bouma MJ, Sondorp HE, van der Kaay HJ: **Climate change and periodic epidemic malaria.** *Lancet* 1994, **343**:1440.
7. Gagnon AS, Smoyer-Tomic KE, Bush AB: **The El Nino southern oscillation and malaria epidemics in South America.** *Int J Biometeorol* 2002, **46**:81-89.
8. Poveda G, Rojas W, Quinones ML, Velez ID, Mantilla RJ, Ruiz D, Zuluaga JS, Rua GL: **Coupling between annual and ENSO timescales in the malaria-climate association in Colombia.** *Environ Health Perspect* 2001, **109**:489-493.
9. S.O.I. (Southern Oscillation Index) Archives. [<http://www.bom.gov.au/climate/current/soihtm1.shtml>], Access date: 04/02/2010.
10. Hanf M, Stephani A, Basurko C, Nacher M, Carme B: **Determination of the *Plasmodium vivax* relapse pattern in Camopi, French Guiana.** *Malar J* 2009, **8**:278.
11. Veron V, Legrand E, Yrinesi J, Volney B, Simon S, Carme B: **Genetic diversity of msp3alpha and msp1_b5 markers of *Plasmodium vivax* in French Guiana.** *Malar J* 2009, **8**:40.
12. Box G, Jenkins G: **Time series analysis: Forecasting and control.** San Francisco: Holden-Day; 1970.
13. Hustache S, Nacher M, Djossou F, Carme B: **Malaria risk factors in Amerindian children in French Guiana.** *Am J Trop Med Hyg* 2007, **76**(4):619-625.
14. Carme B: **Substantial increase of malaria in inland areas of eastern French Guiana.** *Trop Med Int Health* 2005, **10**:154-159.

doi:10.1186/1475-2875-10-100

Cite this article as: Hanf *et al.*: The role of El Niño southern oscillation (ENSO) on variations of monthly *Plasmodium falciparum* malaria cases at the cayenne general hospital, 1996-2009, French Guiana. *Malaria Journal* 2011 **10**:100.

Submit your next manuscript to BioMed Central
and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



6.11. Etude 5 : L'histoplasmosse associée au VIH en Guyane Française : nouvelle infection ou réactivation ?

Introduction : *Histoplasma capsulatum var. capsulatum* est l'infection opportuniste à VIH la plus fréquente en Guyane française, conjointement à la tuberculose, et est la première cause de mortalité liée au sida. Elle survient habituellement chez des patients présentant une immunodépression sévère. En zones non endémiques elle est présumée comme étant essentiellement une réactivation d'une infection passée. Pour les patients résidant dans des zones d'endémicité, il n'est pas possible de vérifier si l'histoplasmosse disséminée est due à une nouvelle infection ou à la réactivation d'une ancienne infection.

Matériels/Méthodes : Des modèles de séries temporelles intégrant les totaux mensuels des premiers épisodes d'histoplasmosse disséminée ainsi que des variables climatiques ont été définis afin d'identifier et de quantifier les relations éventuelles entre ces variables. Les incidences mensuelles d'histoplasmosse chez les patients VIH suivis à l'Hôpital Général de Cayenne entre le 1^{er} Janvier 1999 et le 31 Décembre 2007 ont été obtenues à partir de la base de données française pour le VIH (FHDH).

Résultats : Dans le meilleur modèle en termes d'ajustement, l'incidence de l'histoplasmosse était positivement associée à des précipitations avec un décalage de 7 mois, la température minimale à un décalage de 5 mois, l'humidité minimum à un décalage de 9 mois et la moyenne mensuelle d'ensoleillement par jour. Elle était négativement associée à la température minimale du mois et l'humidité maximale à un décalage de 6 mois.

Conclusions : Cette forte influence des variables climatiques sur les incidences d'histoplasmosse suggère qu'une proportion importante des cas est pourrait être due à une exposition récente, et pose la question de l'utilisation d'une prophylaxie primaire en Guyane française.

Publiée dans *AIDS* le 10 mars 2010

Research Letter

AIDS 2010, **24**:000–000

HIV-associated histoplasmosis in French Guiana: recent infection or reactivation?

Matthieu Hanf^a, Antoine Adenis^a, Pierre Couppie^{b,c}, Bernard Carme^{a,c} and Mathieu Nacher^{a,c,d}

In order to determine whether HIV-associated disseminated histoplasmosis was a recent infection or a reactivation, time series of first episodes of disseminated histoplasmosis were analyzed. Climatic variables were associated with histoplasmosis incidence. This suggested an important proportion of cases were due to recent exposure, and therefore primary prophylaxis may be warranted in French Guiana.

French Guiana is the French overseas territory where the HIV epidemic is most preoccupying. HIV prevalence among pregnant women is 1.5%. AIDS incidence is 10 times greater than in metropolitan France. *Histoplasma capsulatum* var. *capsulatum* infection seems to be a major AIDS-defining illness in endemic areas of South and Central America. It is the most frequent opportunistic infection due to HIV in French Guiana, along with tuberculosis, and the first cause of AIDS-related death [1]. It usually occurs in patients with severe immunosuppression and therefore is presumed to be mostly a recurrence of a past infection in nonendemic areas [2]. For patients residing in areas of endemicity, it is not possible to ascertain whether disseminated histoplasmosis is due to a new infection or to the reactivation of an old infection [3,4]. In this perspective, the main objective was to quantify the relation between environmental factors and disseminated histoplasmosis using an ARIMA model with exogenous variables [5]. Time series data with monthly totals of first episodes of disseminated histoplasmosis were modeled with climatic variables in order to identify and quantify an eventual relation between the two.

Monthly Histoplasmosis incidences in HIV patients followed at the Cayenne General Hospital between 1 January 1999 and 31 December 2007 were obtained from the French Hospital Database for HIV (FHDH). Clinical and therapeutic data are entered by trained technicians. The FHDH is approved by the commission nationale informatique et libertés. Climatic data including monthly rainfall, minimum temperature, maximum temperature, minimum humidity, maximum humidity and relative humidity were obtained from the Cayenne station of the French meteorological institute.

ARIMA models are the most general class of statistical models for analyzing and forecasting a time series.

Multiplicative seasonal ARIMA models were used to evaluate the relationship between climatic factors and monthly histoplasmosis incidence rates. Modeling with seasonal ARIMA involves the estimation of a series of parameters to account for the inherent dynamics in the time series, including the trend, and seasonal and nonseasonal autoregressive and moving average processes. The series was natural log-transformed to stabilize variance. The smallest values of Akaike's information criterion (AIC) were set as the standard to identify the best-fit model.

In the model, each of the climatic input series, at lags of 0–12 months, respectively, was then fitted into the ARIMA model to screen for potential climatic predictors of histoplasmosis incidence. Those input series significantly associated with histoplasmosis incidence, with a *P* value of less than 0.05, were singled out to fit the best multivariate ARIMA model. A manual descendent variable selection was used to eliminate the nonsignificant variables in the final multivariate model. The Ljung-Box Q-test was applied to ascertain whether the residual series was white noise.

Of all the models tested, the ARIMA (0, 0, 0) model fitted the data best according to AIC, showing that no particular pattern explained histoplasmosis incidence.

In the best fitting model with climatic data, histoplasmosis incidence was positively associated with rainfall at a lag of 7 months ($\beta = 0.001$, $P = 0.001$), minimum temperature at a lag of 5 months ($\beta = 0.091$, $P = 0.026$), minimum humidity at a lag of 9 months ($\beta = 0.011$, $P = 0.004$) and monthly mean daily insolation ($\beta = 0.001$, $P = 0.009$). Negatively associated climatic variables with histoplasmosis incidence were minimum temperature of the month ($\beta = -0.114$, $P = 0.008$) and maximum humidity at a lag of 6 months ($\beta = -0.107$, $P < 0.001$).

The incorporation of climatic data in the ARIMA model reduced the AIC by 70%. The monthly observed and predicted histoplasmosis incidences are shown in Fig. 1.

Although clinicians in the region usually assume that disseminated histoplasmosis is a relapse of a past infection consecutive to the decline of cellular immunity, we here show evidence that a significant proportion of cases is in fact influenced by environmental factors.

This suggested that climatic factors favored the development and the dispersion of molds, thereby increasing the

DOI:10.1097/QAD.0b013e32833999c9

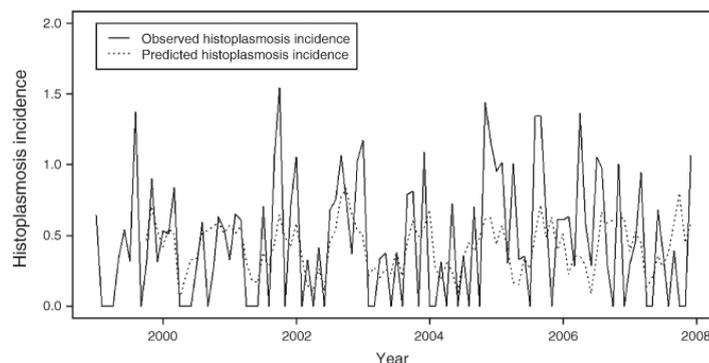


Fig. 1. Observed versus predicted histoplasmosis incidence in HIV patients, Cayenne General Hospital, 1999–2007. Predicted histoplasmosis incidences were estimated using a multiplicative seasonal ARIMA with exogenous variables.

risk of exogenous exposure and explaining the fluctuations of histoplasmosis incidence (70% in term of AIC). An alternative hypothesis, however, could be the influence of climate on the level of immunodepression.

This study leads to important conclusions. First, further studies have to be done to confirm and explain this trend of environmental exposure. Second, primary prophylaxis should be discussed in French Guiana. Given that different communities do not have the same risk of disseminated histoplasmosis, the indication for prophylaxis could be adapted to the patients' environment as suggested by some authors [6].

^aCentre d'Investigation Clinique Epidémiologie Clinique Antilles Guyane CIC-EC CIE 802; ^bDepartment of Dermatology; ^cDepartment of Parasitology and Mycology – Team EA 3593, Université des Antilles et de la Guyane; and ^dCOREVIH, Cayenne General Hospital, Cayenne, French Guiana, France.

AQ1

Received: 27 January 2010; revised: 1 March 2010; accepted: 10 March 2010.

References

- Couppie P, Sobesky M, Aznar C, Bichat S, Clyti E, Bissuel F, *et al.* **Histoplasmosis and acquired immunodeficiency syndrome: a study of prognostic factors.** *Clin Infect Dis* 2004; **38**:134–138.
- Wheat LJ. **Histoplasmosis: a review for clinicians from nonendemic areas.** *Mycoses* 2006; **49**:274–282.
- Kauffman CA. **Histoplasmosis: a clinical and laboratory update.** *Clin Microbiol Rev* 2007; **20**:115–132.
- McKinsey DS, Spiegel RA, Hutwagner L, Stanford J, Driks MR, Brewer J, *et al.* **Prospective study of histoplasmosis in patients infected with human immunodeficiency virus: incidence, risk factors, and pathophysiology.** *Clin Infect Dis* 1997; **24**:1195–1203.
- Box G, Jenkins G. *Time series analysis: forecasting and control.* San Francisco: Holden-Day; 1970.
- Wheat LJ, Freifeld AG, Kleiman MB, Baddley JW, McKinsey DS, Loyd JE, Kauffman CA. **Clinical practice guidelines for the management of patients with histoplasmosis: 2007 update by the Infectious Diseases Society of America.** *Clin Infect Dis* 2007; **45**:807–825.

6.12. Etude 6 : Les variations saisonnières de l'incidence de l'histoplasmosse disséminée : un argument supplémentaire en faveur de l'hypothèse des nouvelles infections ?

Introduction : L'histoplasmosse disséminée (HD) est la première cause de décès liés au sida en Guyane française. Sans traitement, cette maladie est rapidement fatale. En Guyane française, les cliniciens assument en général que L'HD est une réactivation d'une infection passée. Une étude récente a cependant montré qu'une grande part des fluctuations temporelles de l'incidence de l'histoplasmosse peut être expliquée par des facteurs climatiques et a donc postulé que les cas de DH pourrait être dans une grande proportion provoqués par de nouvelles infections. La description de la saisonnalité de la DH pourrait donner des informations précieuses pour tester l'hypothèse des nouvelles contaminations.

Matériels/Méthodes : Les cas d'histoplasmosse chez les patients VIH suivis à l'Hôpital Général de Cayenne entre le 1er Janvier 1999 et le 31 Décembre 2007 ont été obtenus à partir de la base de données française pour le VIH (FHDH). Un modèle multivarié de Cox à événements simples a été utilisé pour évaluer la relation brute et ajustée sur le taux de CD4 et le traitement antirétroviral entre le premier événement de DH et les deux grandes saisons existant en Guyane Française.

Résultats : Après ajustement sur le taux de CD4 et la prise d'un traitement antirétroviral, l'incidence de la DH était significativement plus élevée pendant la petite saison des pluies/longue saison sèche que pendant la petite saison sèche/longue saison des pluies (hazard ratio de 1,7 [1.1 à 2.5], P = 0,01).

Conclusions : Ces résultats donnent à la fois de précieuses informations épidémiologiques pour les cliniciens et a la fois un argument supplémentaire en faveur de l'hypothèse soutenant qu'une proportion importante des cas serait probablement due à une exposition récente. Par conséquent, l'utilisation d'une prophylaxie primaire doit donc être discutée dans les zones endémiques.

Soumise à *Transactions of The Royal Society of Tropical Medicine and Hygiene* le 8 juillet 2011

Manuscript

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Disseminated Histoplasmosis seasonal incidence variations: a supplementary argument for recent infection?

Matthieu HANF*, Antoine ADENIS*, Bernard CARME* °, Pierre COUPPIE† °, Mathieu NACHER*‡ °

* Centre d'Investigation Clinique Epidémiologie Clinique Antilles Guyane CIC-EC INSERM CIE 802, Cayenne General Hospital, Cayenne, French Guiana.

† Department of Dermatology, Cayenne General Hospital, Cayenne, French Guiana.

° Department of Parasitology and Mycology – Team EA 3593, Université des Antilles et de la Guyane, Cayenne General Hospital, Cayenne, French Guiana

‡ COREVIH, Cayenne General Hospital, Cayenne, French Guiana.

Corresponding Author : Matthieu HANF, Centre d'Investigation Clinique Epidémiologie Clinique Antilles Guyane CIC-EC CIE 802, Cayenne General Hospital, EHPAD, Avenue des Flamboyants, BP 6006 97300 Cayenne, French Guiana, France. E-mail address: matthieu@hanf.fr

ABSTRACT

A retrospective cohort study was conducted in order to determine seasonal variations of the incidence of first cases of disseminated histoplasmosis in HIV persons in Cayenne, French Guiana. After adjusting for CD4 counts and antiretroviral treatment, the incidence of disseminated histoplasmosis was significantly higher during the Short Wet Season–Long Dry Season than during the Short Dry Season–Long Wet Season (Adjusted Hazard ratio 1.7 (1.1-2.5), P= 0.01). This result gives both valuable epidemiologic information to clinicians and a supplementary argument in favour of the hypothesis that an important proportion of cases were due to recent exposure. Therefore, the use of a primary prophylaxis has to be discussed in endemic areas.

TEXT

1
2
3
4 French Guiana is the French overseas territory where the HIV epidemic is most
5 preoccupying. HIV prevalence among pregnant women is 1.5%. AIDS incidence is 10 times
6 greater than in metropolitan France. Disseminated histoplasmosis (DH) is one of the most
7 frequent opportunistic infections due to HIV in French Guiana, and is the first AIDS defining
8 illness with an incidence rate of 1.5 per 100 person-years [1]. Furthermore, DH is the first
9 causes of AIDS-related death in French Guiana [2].

10
11
12 It generally occurs when a patient's CD4+ T lymphocyte count is inferior to 100
13 cells/mm³. Without treatment, this disseminated disease has a rapidly fatal course.

14
15
16 In French Guiana, clinicians in the region usually assumed that DH was a relapse of a
17 past infection consecutive to the decline of cellular immunity [3]. However, a recent study has
18 shown that a major part of the histoplasmosis incidence temporal fluctuations could be
19 explained by climatic factors and thus postulated that DH cases could be in a large proportion
20 due to new infections [4]. The description of the seasonal pattern of histoplasmosis could
21 potentially give both valuable information for clinicians and help to test the new
22 contamination hypothesis. Furthermore, the seasonality of histoplasmosis was never studied
23 in French Guiana or South America. In this perspective, we aimed in the present study to
24 determine the relation between first DH events and seasonality in a retrospective cohort of
25 HIV patients in French Guiana.

26
27
28 French Guiana is a large French overseas territory with a surface of 83,534 km². Its
29 latitude ranges from 2° to 5° and its longitude from 51° to 54°. The hygrometry rate is about
30 90%. The mean temperature is 27°C and the precipitation is heavy in particular during the
31 rainy season. There are four main seasons: The Long dry season (LDS) from July to
32 November, the Short wet season (SWS) from December to February, the Short dry season
33 (SDS) in March and the long wet season (LWS) from April to June.

34
35
36 Monthly first cases of disseminated histoplasmosis in HIV patients followed at the
37 Cayenne General Hospital between 1 January 1992 and 31 December 2007 were obtained
38 from the French Hospital Database for HIV (FHDH). Time independent variables such as sex,
39 nationality, mode of acquisition of HIV, and time-dependent variables such as age, CD4 and
40 CD8 cell counts, HIV-1 viral loads, treatments, and clinical events reported by the clinicians
41 following patients are routinely entered by trained clinical studies technicians. Patients
42 included in the FHDH give informed consent to the use of their data. Their identity is

1 encrypted before the data are sent to the Ministry of Health and the Institut National de la
2 Recherche Médicale, which centralizes data from COREVIHs (Regional Coordination of the
3 fight against HIV) throughout France. These data collection is approved by the Commission
4 Nationale Informatique et Libertés.
5

6
7 A confirmed case of DH was defined by an identification of *H. capsulatum* in tissues
8 or fluids (either direct microscopic examination of May–Grunwald–Giemsa-stained smears or
9 fungal culture or histopathological examination).
10

11
12 Single failure Cox proportional hazards models were used to evaluate the crude and
13 CD4 and antiretroviral treatment adjusted relationship between first DH failure and the two
14 main periods of cases occurrence (SWS – LDS versus SDS – LWS). The 9.0 version of the
15 STATA software was used to conduct all statistical analyses.
16
17
18
19
20

21
22 A total of 2275 subjects were followed for a total of 9201.9 years. The total number of
23 observations was 37668, the median follow up time was 2.8 years. There were 127 single
24 failure events. The incidence rate was 1.07 per 100 person years during the SDS – LWS and
25 1.77 per 100 person-years during the SWS – LDS. First DH event survival curves in HIV
26 patients for SWS – LDS and SDS – LWS are shown in figure 1. After adjusting for CD4
27 counts and antiretroviral treatment, the incidence of DH was still significantly higher during
28 the SWS – LDS than during the SDS – LWS (Adjusted Hazard ratio 1.7 (1.1-2.5), $P = 0.01$).
29
30
31
32
33
34
35

36
37 This is the first study to describe the seasonal pattern of histoplasmosis. We here show
38 evidence that histoplasmosis incidence in HIV patients was strongly related to seasonal cycles
39 with a significantly higher incidence during the long dry and short wet seasons than during the
40 long wet and short dry seasons. In HIV patients, during the SWS and LDS, there is a greater
41 risk of disseminated histoplasmosis than during the SDS and LWS.
42
43
44
45
46

47
48 In a previous study describing the correlation between climatic factors and
49 histoplasmosis incidence [4], one of the hypotheses raised to explain this correlation was that
50 the climate influences the level of immunodepression and so produces reactivation in
51 particular conditions. Thus, many recent studies have emphasized the role of vitamin D,
52 mainly obtained from sun exposure, in the susceptibility to respiratory infections and immune
53 system regulation [5]. However, here, this hypothesis seems unlikely because of the
54 contamination peak occurred during the dry season (the sunniest one). The role of ultraviolet-
55
56
57
58
59
60
61
62
63
64
65

mediated immunosuppression is another theoretical possibility but the seasonal pattern was not affected by adjustment on CD4 count.

The second, and, in our view, most plausible hypothesis was that cases were largely due to new infections instead of reactivation. In this situation, climatic factors could favour ideal conditions for the development of mold and its subsequent dispersion, thereby increasing the risk of exogenous exposure. In French Guiana, most human activities are practiced during the dry season and therefore are conducive to greater exposition to histoplasmosis molds. Furthermore, the role of the dry season as an activator of the development/ aerial dispersion of mold is plausible.

To conclude, this marked seasonality is in favour of the hypothesis that in endemic areas, HIV-associated DH is mainly due to recent infections rather than reactivations. This is therefore a supplementary argument for using primary prophylaxis in French Guiana.

Authors' contributions

MH performed data analysis and interpretation, and prepared the manuscript. AA participated in the interpretation, the preparation and the revision of the manuscript. BC and PC participated in the interpretation of data and manuscript revision as well. MN, initiated the study, provided guidance on the data analysis and was involved in the interpretation of data and manuscript redaction and revision. All authors read and approved the final manuscript.

Funding: Cayenne General Hospital and INSERM (Institut National de la Santé Et de la Recherche Médicale)

Conflicts of interest: None declared

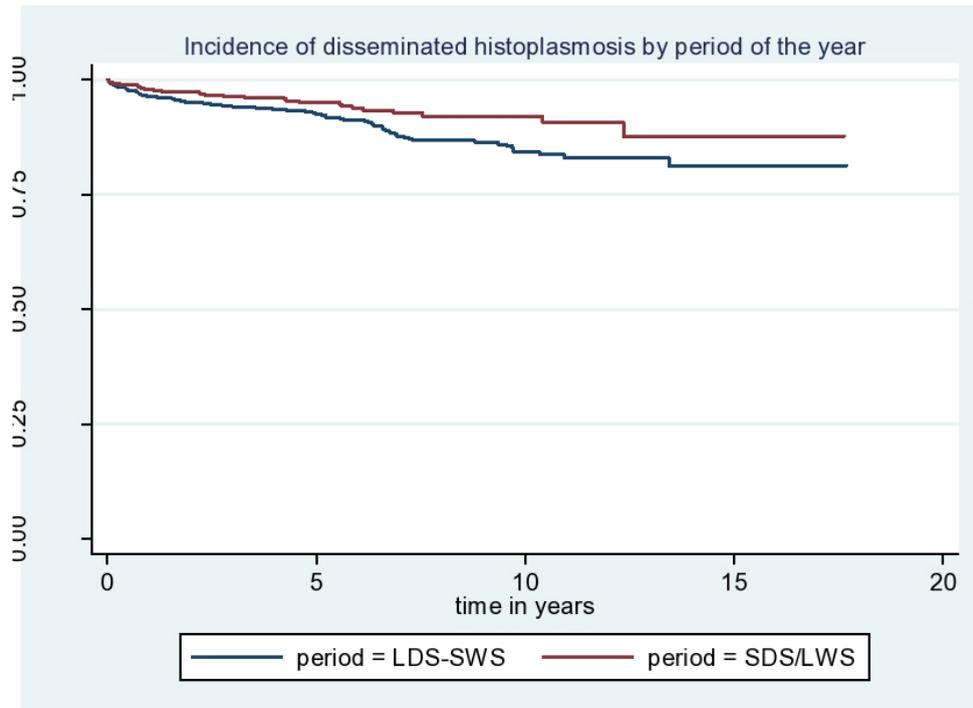
Ethical approval: Yes

REFERENCES

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1. Nacher M, Adenis A, Adriouch L, Dufour J, Papot E, Hanf M, Vantilcke V, Calvez M, Aznar C, Carme B et al: What is AIDS in the Amazon and the Guianas? Establishing the burden of disseminated histoplasmosis. *Am J Trop Med Hyg* 2010, (In press).
2. Couppie P, Sobesky M, Aznar C, Bichat S, Clyti E, Bissuel F, El Guedj, M, Alvarez F, Demar M, Louvel D, Pradinaud R, Carme B. Histoplasmosis and acquired immunodeficiency syndrome: a study of prognostic factors. *Clin Infect Dis*. 2004 Jan 1;38(1):134-8.
3. Wheat LJ. Histoplasmosis: a review for clinicians from non-endemic areas. *Mycoses*. 2006 Jul;49(4):274-82.
4. Hanf M, Adenis A, Couppie P, Carme B, Nacher M. HIV-associated histoplasmosis in French Guiana: recent infection or reactivation? *AIDS*. Jul 17;24(11):1777-8.
5. Norman AW. From vitamin D to hormone D: fundamentals of the vitamin D endocrine system essential for good health. *Am J Clin Nutr*. 2008 Aug;88(2):491S-9S.

FIGURE 1: Kaplan Meier Survival curves comparing the incidence of disseminated histoplasmosis between 2 climatic sequences: Long Dry Season-Short Wet Season and Short Dry Season-Long Wet Season



LDS: Long dry season from July to November
SWS: Short wet season from December to February
SDS: Short dry season in March
LWS: Long wet season from April to June

6.13. Etude 7 : Climat et leishmaniose cutanée en Guyane : étude à partir d'une série de 1302 patients vus entre 1994 et 2010.

Introduction : Lors d'une précédente étude, une corrélation inverse entre fréquence de la leishmaniose cutanée et pluviométrie élevée en Guyane a été mise en évidence sur la période 1994-2003. Dans cette nouvelle étude, l'analyse a été complétée en introduisant en plus de la pluviométrie d'autres variables climatiques et en allongeant la période d'étude jusqu'à 2010.

Matériels/Méthodes : Cette étude proposait d'investiguer la relation des variables météorologiques et des cas de leishmaniose cutanée recensés au centre hospitalier de Cayenne entre 1994 et 2010. Les corrélations des contaminations avec les précipitations, la température et l'humidité ont été estimées par un test de Spearman et par une modélisation des séries temporelles de type ARIMA. Un indicateur du phénomène El Niño, le Multivariate ENSO Index (MEI), était également intégré dans l'analyse.

Résultats : L'analyse des séries temporelles retrouvait que les contaminations étaient corrélées négativement aux précipitations (avec un décalage de 2 mois) ainsi qu'aux nombres de jours supérieurs à 50 mm de pluie (décalages de 4 et 7 mois). Les variables positivement corrélées étaient l'indice MEI et l'amplitude thermique, avec des décalages respectifs de 4 et 8 mois.

Conclusions : En Guyane, le climat a une forte influence sur les fluctuations annuelles et mensuelles des cas de leishmaniose cutanée. L'hypothèse principale serait une régulation de la densité vectorielle (phlébotomes) par la pluie et la température. Ces hypothèses nécessiteraient d'être étayées par des enquêtes entomologiques complémentaires.

Poster présenté au congrès de parasitologie, Strasbourg 2011

Climat et leishmaniose en Guyane



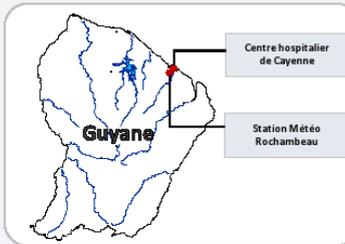
A. ROGER¹, M. HANF^{1,2}, C. BASURKO^{1,2}, M. NACHER^{1,2}, B. CARME^{1,2,3}, P. COUPIÉ^{2,4}

1 - Centre d'Investigation Clinique – Épidémiologie Clinique Antilles-Guyane (CIE 802 Inserm), CH de Cayenne
 2 - Équipe EPA-T (EA 3593), Labex CEBA (Centre Étude Biodiversité Amazonienne), Université Antilles-Guyane, Cayenne
 3 - Laboratoire Hospitalo-Universitaire de Parasitologie Mycologie, CH de Cayenne
 4 - Service de Dermato-Vénérologie, CH de Cayenne



Introduction

La leishmaniose cutanée est endémique en Guyane. Le parasite responsable, *Leishmania spp*, est transmis de l'animal à l'homme par un vecteur arthropode, le phlébotome. Le service de dermatologie de Cayenne est le principal centre civil du département pour la prise en charge de la maladie. Les variations saisonnières du nombre de cas diagnostiqués suggèrent un rôle du climat dans la dynamique de transmission de la maladie.



Matériels et méthodes

Cette étude proposait d'investiguer la relation des variables météorologiques de la station principale du département et des cas de leishmaniose recensés au centre hospitalier de Cayenne entre 1994 et 2010.

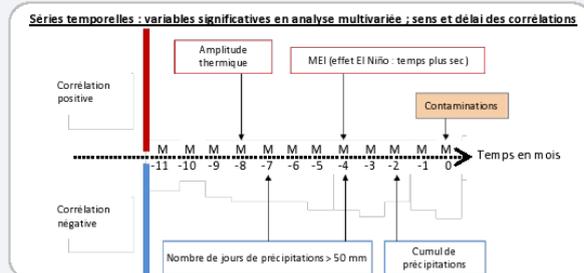
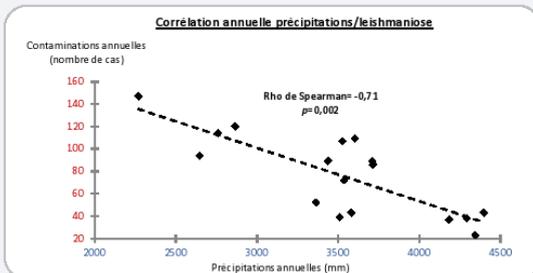
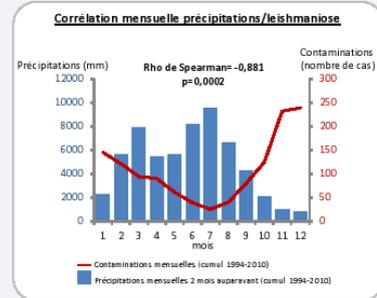
La corrélation des contaminations avec les précipitations, la température et l'humidité était estimée par un test de Spearman et par une modélisation des séries temporelles de type ARIMA (*Autoregressive Integrated Moving Average*). Un indicateur du phénomène El-Niño, le Multivariate ENSO Index (MEI), était également intégré dans l'analyse.

Résultats

Les contaminations annuelles étaient corrélées négativement aux précipitations ($\rho = -0,71$; $p = 0,002$).

Les contaminations mensuelles étaient corrélées négativement aux précipitations avec un décalage de 2 mois ($\rho = -0,88$; $p < 0,001$).

L'analyse des séries temporelles retrouvait aussi que les contaminations étaient corrélées négativement aux précipitations (avec un décalage de 2 mois) ainsi qu'aux nombres de jours supérieurs à 50 mm de pluie (décalages de 4 et 7 mois). Les variables positivement corrélées étaient l'indice MEI et l'amplitude thermique, avec des décalages respectifs de 4 et 8 mois.



modèle ARIMA final avec variables climatiques	Paramètres	Ecart type	p
Terme autorégressif	0,57720	0,11430	<0,00001
Terme de moyenne mobile	-0,11260	0,13670	0,20498
Terme de moyenne mobile saisonnière	-0,99990	0,12030	0,00000
Cumul de précipitations (décalage 2 mois)	-0,00100	0,00030	0,00012
Amplitude thermique (décalage 8 mois)	0,22270	0,07820	0,00221
Nombre de jours >50mm de précipitations (décalage 4 mois)	-0,06970	0,02800	0,00649
Nombre de jours >50 mm de précipitations (décalage 7 mois)	-0,10090	0,02840	0,00019
Index El Nino « MEI » (décalage 4 mois)	0,23910	0,08170	0,00172

Discussion

En Guyane, le climat semble donc avoir une influence sur les fluctuations annuelles et mensuelles des cas de leishmaniose. Une diminution des précipitations est associée à une augmentation des contaminations 2 mois plus tard. L'hypothèse principale serait une régulation de la densité vectorielle par la pluie. Toutefois, cette hypothèse nécessiterait d'être étayée par des enquêtes entomologiques complémentaires.

6.14. Conclusions

Le couplage de méthodes écologiques de type série temporelle avec les données météorologiques librement accessibles à la communauté scientifique a contribué à une meilleure compréhension de la transmission du paludisme dans deux villages situés en Guyane française : Camopi et Cacao. Ces différentes études ont montré que l'environnement proche mais aussi les paramètres météorologiques jouaient un rôle prépondérant dans la transmission chez les enfants de Camopi et chez les habitants du village de Cacao. Dans ces deux villages, les associations significatives retrouvées entre le niveau des fleuves voisins (fortement dépendants des paramètres climatiques) et le nombre cas de paludisme, ont permis de mieux comprendre la dynamique de la transmission du paludisme ainsi que les caractéristiques écologiques des vecteurs locaux. Cette méthodologie a également permis de quantifier globalement l'impact du phénomène El Niño sur les épidémies de paludisme à *P.falciparum* en Guyane française. Bien qu'un lien statistique significatif ait été retrouvé, la valeur prédictive d'El Niño pour le paludisme à *P.falciparum* semble marginale. Ce résultat est une information à prendre en compte pour le développement de futurs systèmes de surveillance capables de détecter précocement les épidémies de paludisme.

De manière identique, cette méthodologie a permis de montrer qu'en Guyane le climat avait une forte influence sur les fluctuations annuelles et mensuelles des cas de leishmaniose cutanée. L'hypothèse principale pour expliquer ce phénomène serait une régulation de la densité vectorielle (des phlébotomes) par la pluie et la température. Ces hypothèses nécessiteraient d'être étayées par des enquêtes entomologiques complémentaires. Grâce à la qualité du modèle statistique développé, le développement d'un futur système de surveillance capable de détecter précocement les épidémies de leishmanioses se basant sur les fluctuations du climat pourrait être envisagé.

La forte influence des variables climatiques sur l'incidence de l'histoplasmosse, suggère qu'une proportion importante des cas pourrait être due à une exposition récente. Ces résultats donnent de précieuses informations épidémiologiques pour les cliniciens et posent la question de la mise en place d'une prophylaxie primaire en Guyane.

Que ce soit pour le paludisme, la leishmaniose cutanée, ou l'histoplasmosse disséminée, l'utilisation de méthodes écologiques couplées à des données librement accessibles a permis, en l'absence de données plus objectives, de préciser, tout au moins pour une part, le rôle du climat sur la dynamique de maladies à infections parasitaires guyanaises. Une fois cet impact quantifié, certaines implications ont pu être mises en avant (pistes de réflexions sur les caractéristiques écologiques des vecteurs, possibilité de création de systèmes de prévisions des épidémies, et hypothèses cliniques sur la dynamique des processus infectieux...).

Bibliographie

1. Rees R. **Under the weather: climate and disease, 1700-1900.** *History Today* 1996,46:35-42.
2. Colwell R, Patz J. **Climate, infectious disease and health : an interdisciplinary perspective.** In. Washington, D.C.: American Academy of Microbiology; 1998.
3. Rose JB, Huq A, Lipp EK. **Health, climate and infectious disease: a global perspective.** In. Washington D.C.: American Academy of Microbiology; 2001.
4. World Health Organization. **Climate and health (working group meeting)** In. Geneva, Switzerland: World Health Organization; 1999.
5. Pascual M, Rodo X, Ellner SP, Colwell R, Bouma MJ. **Cholera dynamics and El Nino-Southern Oscillation.** *Science* 2000,289:1766-1769.
6. Altizer S, Dobson A, Hosseini P, Hudson P, Pascual M, Rohani P. **Seasonality and the dynamics of infectious diseases.** *Ecol Lett* 2006,9:467-484.
7. Grassly NC, Fraser C. **Seasonal infectious disease epidemiology.** *Proc Biol Sci* 2006,273:2541-2550.
8. Pascual M, Dobson A. **Seasonal patterns of infectious diseases.** *PLoS Med* 2005,2:e5.
9. Pascual M, Bouma MJ, Dobson A. **Cholera and climate:revisiting the quantitative evidence.** *Microbes Infect.* 2002,4:237-245.
10. Glantz MH. **Currents of change: El Niño's impact on climate and society.** Cambridge: Cambridge University Press; 1996.
11. Walker GT, Bliss EW. **World weather.** V. *Mem R Met Soc* 1932,4:53-84.
12. Ropelewski C, Halpert MS. **Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation.** *Monthly Weather Review* 1987,115:1606-1626.
13. Bouma MJ, Kovats RS, Goubet SA, Cox JS, Haines A. **Global assessment of El Nino's disaster burden.** *Lancet* 1997,350:1435-1438.
14. Dilley M, Heyman BN. **ENSO and disaster: droughts, floods and El Nino/Southern Oscillation warm events.** *Disasters* 1995,19:181-193.
15. Paz S. **Impact of temperature variability on cholera incidence in southeastern Africa, 1971-2006.** *Ecohealth* 2009,6:340-345.
16. Speelman EC, Checkley W, Gilman RH, Patz J, Calderon M, Manga S. **Cholera incidence and El Nino-related higher ambient temperature.** *JAMA* 2000,283:3072-3074.
17. Klempa B. **Hantaviruses and climate change.** *Clin Microbiol Infect* 2009,15:518-523.
18. Zhang Y, Bi P, Wang G, Hiller JE. **El Nino Southern Oscillation (ENSO) and dysentery in Shandong province, China.** *Environ Res* 2007,103:117-120.
19. Tipayamongkholgul M, Fang CT, Klinchan S, Liu CM, King CC. **Effects of the El Nino-southern oscillation on dengue epidemics in Thailand, 1996-2005.** *BMC Public Health* 2009,9:422.

Partie 2 : Travaux scientifiques

20. Johansson MA, Cummings DA, Glass GE. **Multiyear climate variability and dengue--El Nino southern oscillation, weather, and dengue incidence in Puerto Rico, Mexico, and Thailand: a longitudinal data analysis.** *PLoS Med* 2009,6:e1000168.
21. Rohani P. **The link between dengue incidence and El Nino southern oscillation.** *PLoS Med* 2009,6:e1000185.
22. Jones AE, Wort UU, Morse AP, Hastings IM, Gagnon AS. **Climate prediction of El Nino malaria epidemics in north-west Tanzania.** *Malar J* 2007,6:162.
23. Mabaso ML, Kleinschmidt I, Sharp B, Smith T. **El Nino Southern Oscillation (ENSO) and annual malaria incidence in Southern Africa.** *Trans R Soc Trop Med Hyg* 2007,101:326-330.
24. Gagnon AS, Smoyer-Tomic KE, Bush AB. **The El Nino southern oscillation and malaria epidemics in South America.** *Int J Biometeorol* 2002,46:81-89.
25. Schwarz TF, Hassler D. **[El nino and its consequences: outbreak of Rift Valley fever in Kenya].** *Dtsch Med Wochenschr* 2007,132:363-364.
26. Dushoff J, Plotkin JB, Levin SA, Earn DJ. **Dynamical resonance can account for seasonality of influenza epidemics.** *Proc Natl Acad Sci U S A* 2004,101:16915-16916.
27. Fine PE, Clarkson JA. **Measles in England and Wales--I: An analysis of factors underlying seasonal patterns.** *Int J Epidemiol* 1982,11:5-14.
28. London WP, Yorke JA. **Recurrent outbreaks of measles, chickenpox and mumps. I. Seasonal variation in contact rates.** *Am J Epidemiol* 1973,98:453-468.
29. Soper MA. **The interpretation of periodicity in disease prevalence.** *J. R. Stat. Soc. Ser. A* 1929,92:34-61.
30. Dowell SF. **Seasonal variation in host susceptibility and cycles of certain infectious diseases.** *Emerg. Infect. Dis.* 2001,7:369-373.
31. Nelson RJ, Demas GE, Klein SL, Kriegsfeld LJ. **Seasonal Patterns of Stress, Immune Function, and Disease.** New York: Cambridge University Press; 2002.
32. Gillett J. **Direct and indirect influences of temperature on the transmission of parasites from insects to man.** In: *The Effects of Meteorological Factors upon Parasites* Edited by Taylor AER, Muller R. London: Blackwell Scientific Publications; 1974:79-95.
33. Gordon RM, Davey TH, Peaston H. **The transmission of human bilharziasis in Sierra Leone, with an account of the life cycle of the schistosomes concerned, S. mansoni and S. haematobium.** *Ann. Trop. Med. Parasitol.* 1934,28:323-418.
34. Stromberg BE. **Environmental factors influencing transmission.** *Vet Parasitol* 1997,72:247-256; discussion 257-264.
35. Colwell RR. **Global climate and infectious disease: the cholera paradigm.** *Science* 1996,274:2025-2031.

Chapitre 6 - Apports du couplage études écologiques/données climatiques dans la compréhension de la dynamique des maladies infectieuses : exemple de la Guyane Française.

36. Lipp EK, Huq A, Colwell RR. **Effects of global climate on infectious disease: the cholera model.** *Clin Microbiol Rev* 2002,15:757-770.
37. Dobson A, Carper R. **Global warming and potential changes in host-parasite and disease-vector relationships.** In: *Global Warming and Biodiversity* Edited by Peters RL, Lovejoy TE. New Haven, CT: Yale University Press; 1992:201-207.
38. Hay SI, Myers MF, Burke DS, Vaughn DW, Endy T, Ananda N, *et al.* **Etiology of interepidemic periods of mosquito-borne disease.** *Proc Natl Acad Sci U S A* 2000,97:9335-9339.
39. Rogers DJ, Randolph SE. **Tsetse flies in Africa: bane or boon?** *Conserv. Biol.* 1988,2.
40. Ogden NH, Maarouf A, Barker IK, Bigras-Poulin M, Lindsay LR, Morshed MG, *et al.* **Climate change and the potential for range expansion of the Lyme disease vector *Ixodes scapularis* in Canada.** *Int J Parasitol* 2006,36:63-70.
41. Gage KL, Burkot TR, Eisen RJ, Hayes EB. **Climate and vectorborne diseases.** *Am J Prev Med* 2008,35:436-450.
42. Linthicum KJ, Anyamba A, Tucker CJ, Kelley PW, Myers MF, Peters CJ. **Climate and satellite indicators to forecast Rift Valley fever epidemics in Kenya.** *Science* 1999,285:397-400.
43. Muir D. **Anopheline mosquitoes: vector-reproduction, life cycle and biotope.** In: *Malaria: Principles and Practice of Malariology.* Edited by Wernsdorfer W, McGregor I. New York: Churchill Livingstone; 1988:431-452.
44. Focks DA, Daniels E, Haile DG, Keesling JE. **A simulation model of the epidemiology of urban dengue fever: literature analysis, model development, preliminary validation, and samples of simulation results.** *Am J Trop Med Hyg* 1995,53:489-506.
45. Hoshen MB, Morse AP. **A weather-driven model of malaria transmission.** *Malar J* 2004,3:32.
46. Hu W, Tong S, Mengersen K, Connell D. **Weather variability and the incidence of cryptosporidiosis: comparison of time series poisson regression and SARIMA models.** *Ann Epidemiol* 2007,17:679-688.
47. Box G, Jenkins G. **Time series analysis: Forecasting and control.** San Francisco: Holden-Day; 1970.
48. Franke CR, Ziller M, Staubach C, Latif M. **Impact of the El Nino/Southern Oscillation on visceral leishmaniasis, Brazil.** *Emerg Infect Dis* 2002,8:914-917.
49. Hu W, McMichael AJ, Tong S. **El Nino Southern Oscillation and the transmission of hepatitis A virus in Australia.** *Med J Aust* 2004,180:487-488.
50. Zhang Y, Bi P, Hiller JE. **Meteorological variables and malaria in a Chinese temperate city: A twenty-year time-series data analysis.** *Environ Int* 2010,36:439-445.
51. Météo-France. **Site internet de Météo-France.** In. Paris: Météo-France; 2011.
52. Aceituno P. **On the functioning of the Southern Oscillation in the South American sector. Part I: Surface climate.** *Monthly Weather Review* 1988,116:505-524.

Partie 2 : Travaux scientifiques

53. Kiladis GN, Diaz HF. **Global Climatic Anomalies Associated with Extremes in the Southern Oscillation.** *Journal of Climate* 1989,2:1069-1090.
54. Wilson EO. **The Diversity of Life:** Harvard University Press; 1992.
55. Guernier V, Hochberg ME, Guegan JF. **Ecology drives the worldwide distribution of human diseases.** *PLoS Biol* 2004,2:e141.
56. Carme B, Ardillon V, Girod R, Grenier C, Joubert M, Djossou F, Ravachol F. **[Update on the epidemiology of malaria in French Guiana].** *Med Trop (Mars)* 2009,69:19-25.
57. Carme B. **Substantial increase of malaria in inland areas of eastern French Guiana.** *Trop Med Int Health* 2005; 10: 154-159
58. CIRE Antilles Guyane. **Situation du paludisme dans les zones à fort risque de transmission. Guyane. Point épidémiologique.** In. Saint Maurice: Institut National de Veille Sanitaire; 2008.
59. Dusfour I, Carinci R, Gaborit P, Issaly J, Girod R. **Evaluation of four methods for collecting malaria vectors in French Guiana.** *J Econ Entomol* 2010,103:973-976.
60. Girod R, Gaborit P, Issaly J, Carinci R, Basurko C. **Anophèles et transmission du paludisme dans le village et les environs de Cacao (Guyane).** *Bull Soc Pathol Exot* 2009,102:199-205.
61. Kovats RS, Bouma MJ, Hajat S, Worrall E, Haines A. **El Nino and health.** *Lancet* 2003,362:1481-1489.
62. Bouma MJ, Dye C. **Cycles of malaria associated with El Nino in Venezuela.** *JAMA* 1997,278:1772-1774.
63. Bouma MJ, Poveda G, Rojas W, Chavasse D, Quinones M, Cox J, Patz J. **Predicting high-risk years for malaria in Colombia using parameters of El Nino Southern Oscillation.** *Trop Med Int Health* 1997,2:1122-1127.
64. Dedet JP. **Cutaneous leishmaniasis in French Guiana: a review.** *Am J Trop Med Hyg* 1990,43:25-28.
65. Roger A. **La leishmaniose cutanée en Guyane : études épidémiologique, clinique et thérapeutique des cas recensés au centre hospitalier de Cayenne de 1994 à 2010.** Cayenne: Université des Antilles et de la Guyane; 2011.
66. Rotureau B, Couppie P, Nacher M, Dedet JP, Carme B. **[Cutaneous leishmaniasis in French Guiana].** *Bull Soc Pathol Exot* 2007,100:251-260.
67. Rotureau B, Catzefflis F, Carme B. **Absence of leishmania in Guianan bats.** *Am J Trop Med Hyg* 2006,74:318-321.
68. Léger N, Depaquit J. **Les phlebotomes et leur role dans la transmission des leishmanioses.** *Revue Française des Laboratoires* 2001:41-48.
69. Gentile B, Le Pont F, Pajot FX, Besnard R. **Dermal leishmaniasis in French Guiana: the sloth (*Choloepus didactylus*) as a reservoir host.** *Trans R Soc Trop Med Hyg* 1981,75:612-613.
70. Feliciangeli MD, Rabinovich J. **Abundance of *Lutzomyia ovallesi* but not *Lu. gomezi* (Diptera: Psychodidae) correlated with cutaneous leishmaniasis incidence in north-central Venezuela.** *Med Vet Entomol* 1998,12:121-131.

Chapitre 6 - Apports du couplage études écologiques/données climatiques dans la compréhension de la dynamique des maladies infectieuses : exemple de la Guyane Française.

71. Salomon OD, Wilson ML, Munstermann LE, Travi BL. **Spatial and temporal patterns of phlebotomine sand flies (Diptera: Psychodidae) in a cutaneous leishmaniasis focus in northern Argentina.** *J Med Entomol* 2004,41:33-39.
72. Rabinovich JE, Feliciangeli MD. **Parameters of Leishmania braziliensis transmission by indoor Lutzomyia ovallesi in Venezuela.** *Am J Trop Med Hyg* 2004,70:373-382.
73. Nacher M, Carme B, Sainte Marie D, Couppie P, Clyti E, Guibert P, Pradinaud R. **Seasonal fluctuations of incubation, healing delays, and clinical presentation of cutaneous leishmaniasis in French Guiana.** *J Parasitol* 2001,87:1495-1498.
74. Nacher M, Couppie P, Carme B, Clyti E, Sainte Marie D, Guibert P, Pradinaud R. **Influence of meteorological parameters on the clinical presentation of cutaneous leishmaniasis in French Guiana and on the efficacy of pentamidine treatment of the disease.** *Ann Trop Med Parasitol* 2002,96:773-780.
75. Nacher M, El Guedj M, Vaz T, Nasser V, Randrianjohany A, Alvarez F, et al. **Risk factors for late HIV diagnosis in French Guiana.** *AIDS* 2005,19:727-729.
76. Nacher M, El Guedj M, Vaz T, Nasser V, Randrianjohany A, Alvarez F, et al. **Risk factors for follow-up interruption of HIV patients in French Guiana.** *Am J Trop Med Hyg* 2006,74:915-917.
77. Lewden C, Sobesky M, Cabie A, Couppie P, Boulard F, Bissuel F, et al. **[Causes of death among HIV infected adults in French Guyana and the French West Indies in the era of highly active antiretroviral therapy (HAART)].** *Med Mal Infect* 2004,34:286-292.
78. Huber F, Nacher M, Aznar C, Pierre-Demar M, El Guedj M, Vaz T, et al. **AIDS-related Histoplasma capsulatum var. capsulatum infection: 25 years experience of French Guiana.** *AIDS* 2008,22:1047-1053.
79. Nacher M, Adenis A, Adriouch L, Dufour J, Papot E, Hanf M, et al. **What is AIDS in the Amazon and the Guianas? Establishing the burden of disseminated histoplasmosis.** *Am J Trop Med Hyg* 2010,84:239-240.
80. Adenis A. **Tuberculose et histoplasmose chez les patients infectés par le VIH : étude comparative.** Cayenne: Université des Antilles et de la Guyane; 2009.
81. Kauffman CA. **Histoplasmosis: a clinical and laboratory update.** *Clin Microbiol Rev* 2007,20:115-132.
82. McKinsey DS, Spiegel RA, Hutwagner L, Stanford J, Driks MR, Brewer J, et al. **Prospective study of histoplasmosis in patients infected with human immunodeficiency virus: incidence, risk factors, and pathophysiology.** *Clin Infect Dis* 1997,24:1195-1203.
83. Carme B, Lecat J, Lefebvre P. **[Malaria in an outbreak zone in Oyapock (French Guiana): incidence of malaria attacks in the American Indian population of Camopi].** *Med Trop (Mars)* 2005,65:149-154.
84. Veron V, Legrand E, Yrinesi J, Volney B, Simon S, Carme B. **Genetic diversity of msp3alpha and msp1_b5 markers of Plasmodium vivax in French Guiana.** *Malar J* 2009,8:40.

Chapitre 7 - Utilisation des données issues de la littérature scientifique : Application à l'étude de la séroprévalence à *Toxoplasma gondii* dans les populations humaines

« Qui a jamais demandé à la thèse et à l'antithèse si elles étaient d'accord pour devenir synthèse ? »
Stanislaw Jerzy Lec

7.1. La révolution des bases de données bibliographiques

Nous avons vu dans l'introduction que les technologies de l'information et de la communication ont radicalement transformé ces 20 dernières années le monde de la recherche. Ainsi, consécutivement à ce phénomène, le besoin d'adapter les pratiques scientifiques pour tirer pleinement avantage des connaissances et résultats engrangés par les chercheurs s'est fait de plus en plus sentir. Une des avancées majeures dans ce domaine a été l'émergence de l'«e-science», que l'on peut définir brièvement comme un accès facilité à travers un poste informatique et via Internet à des ressources distribuées permettant la consultation d'une très grande partie de la création et de la production intellectuelle de la communauté scientifique ainsi que des résultats associés. La gestion et l'accessibilité de cette information sont devenues stratégiques. Dans le domaine des publications scientifiques, un des moyens mis en œuvre pour palier à ce problème a été la création de bases de données bibliographiques accessibles en ligne qui recensent un grand nombre de journaux publiant des recherches faites dans un domaine particulier.

Ainsi, la principale base de données regroupant la littérature relative aux sciences biologiques et biomédicales est MEDLINE (en anglais : Medical Literature Analysis and Retrieval System Online). Une interface, PubMed [1], permet de la consulter gratuitement à partir d'un simple navigateur Internet. Elle a été développée par le National Center for Biotechnology Information (NCBI), et est hébergée par la bibliothèque nationale de médecine américaine du « National Institutes of Health ».

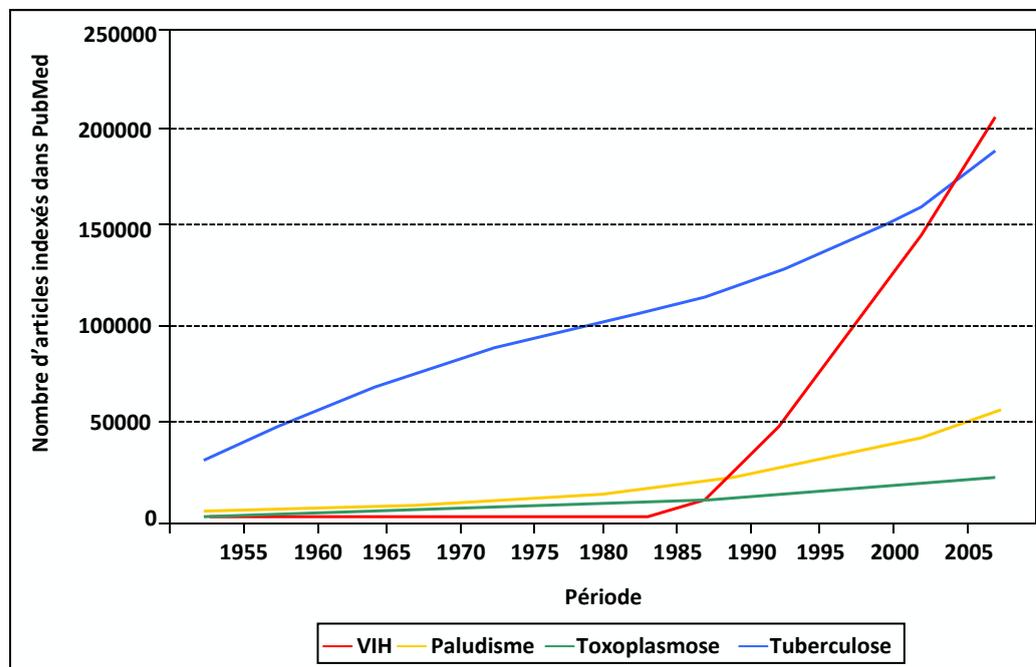
La base de données MEDLINE possédait en juillet 2011 plus de 18 millions de citations publiées depuis 1950 dans environ 5000 revues biomédicales (revues en biologie et en médecine) distinctes. Elle est la base de données de référence pour les sciences biomédicales. En plus de MEDLINE, PubMed offre aussi accès à OLDMEDLINE pour les articles d'avant 1966. Elle donne également accès aux citations de tous les articles, même « hors sujet » (c'est-à-dire couvrant des sujets aussi variés que la tectonique des plaques ou l'astrophysique) de certains journaux de MEDLINE, principalement ceux publiés dans les journaux généralistes d'importance de

Chapitre 7 - Utilisation des données issues de la littérature scientifique : Application à l'étude de la séroprévalence à *Toxoplasma gondii* dans les populations humaines

science ou de biochimie (comme Science et Nature, par exemple). La plupart des citations contiennent un lien vers l'article entier quand celui-ci est disponible gratuitement.

Cette masse d'information considérable ainsi que celles fournies par les autres bases bibliographiques disponibles en sciences médicales (ScienceDirect, Scielo, EMBASE,...) ou celles spécialisées dans un autre domaine sont maintenant à la portée de chaque chercheur disposant d'une simple connexion internet. Chaque année un nombre considérable d'articles sont nouvellement indexés. A titre d'exemple, 921706 articles publiés en 2010 ont été incorporés dans MEDLINE. Pour des chercheurs voulant s'investir sur une pathologie ou dans une discipline particulière, la lecture de la totalité des articles la concernant est devenue impossible. De plus, le nombre de nouveaux articles publiés chaque année augmente sans cesse et résulte en une masse considérable d'information à traiter, analyser et synthétiser (Figure 35).

Figure 35 : Evolution temporelle du nombre d'articles publiés et indexés dans PubMed [1] contenant les termes MESH : « VIH », « malaria », « toxoplasmosis », et « tuberculose ».



7.2. Utilisation et synthèse de la littérature scientifique en santé publique

Avec ces nouveaux modes d'accès à la littérature scientifique, un enjeu important se dessine maintenant pour la communauté scientifique : comment synthétiser et interpréter les résultats de cette masse d'information ?

Avec l'avènement de l'« evidence based medicine », les professionnels de santé ont de plus en plus la nécessité de fonder leurs pratiques sur les meilleures preuves scientifiques disponibles. Après une recherche documentaire sur une question spécifique, de très nombreux articles peuvent être récupérés. La qualité des études récupérées peut être variable et ces dernières peuvent avoir produit des résultats contradictoires. Ainsi, une recherche bibliographique effectuée rapidement, peut ne pas être représentative de l'ensemble de la littérature portant sur un sujet particulier. Il est pourtant primordial que les décisions prises dans le domaine de

la santé ne soient pas uniquement basées sur une ou deux études, et qu'elles tiennent compte de l'ensemble des données de recherche disponibles sur un sujet. Ainsi, les professionnels de santé utilisent très souvent des articles de synthèse comme source d'information récapitulative et synthétique sur un sujet particulier. Ces revues dans la littérature médicale étaient traditionnellement sous la forme d'un examen narratif où des experts dans un domaine particulier fournissaient ce qui était censé être un résumé des preuves et connaissances dans ce domaine. Ces revues narratives, encore très communes dans le domaine médical, ont été critiquées en raison du risque élevé de biais qu'elles présentaient [2]. D'autres méthodes ont été développées pour limiter au maximum ces biais et exploiter la masse d'information contenue dans les publications scientifiques.

7.2.1. Les revues systématiques de la littérature

La validité d'un article de revue dépend de sa qualité méthodologique. Bien que les revues traditionnelles ou revues narratives puissent être utiles quand elles sont bien menées, il a été démontré qu'elles sont généralement de mauvaise qualité. Les auteurs des revues narratives utilisent souvent des méthodes subjectives et informelles pour recueillir et interpréter les études et tendent à être sélectifs en ne citant que des rapports qui renforcent leurs idées préconçues ou promeuvent leurs propres opinions sur un sujet particulier [3]. Ils sont également rarement explicites sur la façon dont ils ont choisi, évalué, et analysé les études primaires, ce qui ne permet pas au lecteur d'évaluer les biais potentiels dans le processus de la revue. Les revues narratives sont donc souvent biaisées, et les recommandations formulées par ces dernières peuvent ainsi être inappropriées [4].

Contrairement à une revue narrative, une revue systématique est une forme de recherche qui fournit un résumé des articles et rapports médicaux sur une question médicale spécifique. Ces dernières utilisent des méthodes explicites pour la recherche, évaluent de façon critique, et synthétisent la littérature mondiale de manière systématique [5]. Cette méthode est particulièrement utile en rassemblant et en synthétisant un certain nombre d'études réalisées séparément ayant parfois des résultats contradictoires.

En fournissant de manière claire et explicite un résumé de toutes les études portant sur une question médicale particulière [5], les revues systématiques permettent de tenir compte de l'ensemble des conclusions pertinentes de recherche sur un sujet particulier, et de ne pas se restreindre seulement aux résultats d'une ou deux études. D'autre part, elles peuvent être aussi utilisées pour savoir si des résultats scientifiques sont cohérents et généralisables au sein des populations, de groupes socio-économiques différents, de variations de traitement, ou si au contraire les résultats varient considérablement selon les sous-groupes particuliers. En synthétisant les connaissances sur un problème de santé, les revues systématiques peuvent aussi mettre en lumière des champs de recherches où les connaissances manquent [6]. Cela peut être ensuite utilisé pour guider les futures recherches dans ce domaine [6].

Par ailleurs, les méthodes explicites utilisées dans les revues systématiques limitent les biais et permettent d'améliorer la fiabilité et l'exactitude des conclusions. Pour ces raisons, les revues systématiques d'essais contrôlés randomisés sont considérés comme une des preuves les plus fortes dans la hiérarchie des schémas d'étude évaluant l'efficacité des interventions [7].

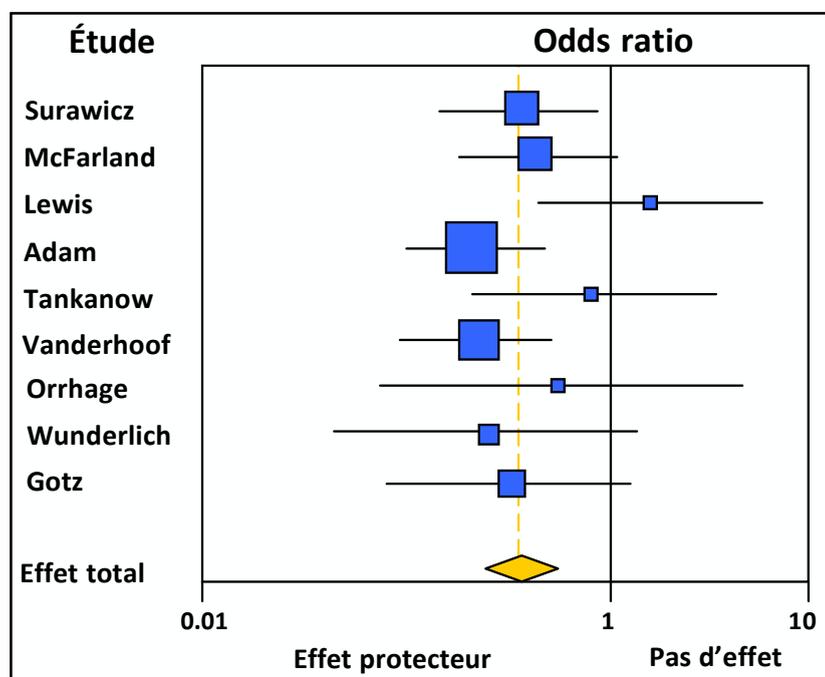
Chapitre 7 - Utilisation des données issues de la littérature scientifique : Application à l'étude de la séroprévalence à *Toxoplasma gondii* dans les populations humaines

Bien que les revues systématiques sont considérées comme la forme de preuve médicale la plus forte, un examen de 300 études a cependant constaté que les études systématiques n'avaient pas toutes la même fiabilité, et que ces revues pouvaient être encore améliorées par un ensemble de normes et de directives universellement appliquées [8]. Une étude plus approfondie a aussi constaté que sur un échantillon de 100 revues systématiques, 7% nécessitaient une mise à jour au moment de la publication, 4% une mise à jour dans les un an, et 11% dans les 2 ans. Ce chiffre était plus élevé dans les champs en évolution rapide comme la médecine [9]. Une étude de 2003 suggérait aussi que l'extension des recherches au-delà des principales bases de données bibliographiques, comme dans la littérature grise, augmenterait potentiellement l'efficacité de revues [10]. La réalisation seule d'une revue systématique est aussi sujette à une limitation majeure : les synthèses réalisées par ces dernières sont uniquement qualitatives. Des méthodes statistiques particulières ont été développées pour palier à ce problème.

7.2.2. Les méta-analyses

Suite à un examen systématique, les données provenant des études individuelles peuvent éventuellement être regroupées quantitativement et réanalysées en utilisant une méthode statistique établie. Cette technique est appelée méta-analyse. La justification d'une méta-analyse est que, en combinant les populations échantillonnées dans les études individuelles, la taille de l'échantillon globale est augmentée, améliorant ainsi la puissance statistique de l'analyse ainsi que la précision des estimations de l'effet d'une variable par exemple [5].

Figure 36 : Exemple de graphique typique d'une méta-analyse. Méta-analyse de l'effet des probiotiques sur les diarrhées associées à la prise d'antibiotiques. Adaptée de [11].



La méta-analyse est un processus qui s'effectue en deux étapes [5]. La première étape implique la récupération ou le calcul d'une mesure de l'effet de la variable étudiée avec son intervalle de confiance à 95% dans chacune des études individuelles. Les statistiques sommaires qui sont habituellement utilisées pour

Partie 2 : Travaux scientifiques

effectuer une mesure de l'effet d'une variable sont principalement les odds ratios, les risques relatifs, et les différences de risque. Dans la deuxième étape de la méta-analyse, un effet global de la variable est ensuite calculé comme la moyenne pondérée de statistiques individuelles. Il est à noter que, dans la méta-analyse, les données des études individuelles ne sont pas simplement combinées. Une plus grande importance est donnée aux résultats des études fournissant plus d'informations, car elles sont susceptibles d'être plus proches du "vrai" effet que l'on cherche à estimer. Cette importance est souvent quantifiée par l'inverse de la variance (le carré de l'erreur type) de l'effet de la variable, qui est étroitement lié à la taille de l'échantillon [5]. Un exemple de graphique typique pour afficher les résultats d'une méta-analyse est présenté dans la Figure 36.

La principale critique des méta-analyses réside principalement autour de la notion de biais de publication. À côté de la répercussion des biais des études, toute synthèse d'information est sujette à un biais qui lui est propre et qui est le biais de publication. Il est connu qu'un certain nombre de travaux ne font jamais l'objet d'une publication. Cela est particulièrement fréquent en cas d'études dont les résultats sont négatifs. Les raisons de cette censure sont multiples et peuvent provenir soit des comités de relecture, soit des firmes finançant l'étude, mais aussi d'une autocensure que s'inflige spontanément les auteurs [12]. Dans une méta-analyse, si aucune recherche poussée des études non publiées n'est entreprise, le risque couru est de ne travailler qu'avec les études positives, ce qui conduit à une surestimation de l'efficacité de la variable étudiée. De plus, avec la diffusion de la technique de la méta-analyse, sont apparues des répétitions de méta-analyses sur le même sujet, conduisant parfois à des résultats divergents. Cet état de fait pose problème car, l'un des buts de la méta-analyse est de fournir des résultats reproductibles. Devant une telle situation, la lecture critique de ces différentes méta-analyses s'attache à chercher les causes de ces divergences et à les analyser. Parmi ces causes, l'une des plus fréquentes tient à ce que les différentes méta-analyses n'ont pas inclus les mêmes essais. Pour limiter au maximum ces problèmes, la communauté scientifique s'est appliquée à développer une méthodologie stricte autour de la pratique de la méta-analyse [13].

Beaucoup de revues spécialisées dans la santé publient maintenant des revues systématiques avec ou sans méta-analyses, mais la source la plus connue est la « Collaboration Cochrane », un groupe de plus de 28000 spécialistes des questions de santé qui examinent systématiquement les études sur les effets de la prévention, des traitements ainsi que les interventions des systèmes de santé. Elles sont publiées dans la base de données Cochrane des revues systématiques de la librairie Cochrane [14]. Le facteur d'impact pour 2010 de cette base de données est 6.186 et fait partie maintenant des bases de données médicales les plus visitées.

7.2.3. Intérêt des études écologiques dans la synthèse de la littérature scientifique

Une autre limite des méta-analyses réside principalement dans le fait qu'elles ne peuvent être réalisées que sur des problématiques déjà étudiées par les études individuelles. Un troisième type de méthodes permet cependant d'utiliser les données issues de la littérature scientifique pour la réalisation d'études originales : Les méthodes écologiques. Ces dernières permettent de récupérer différents types d'informations contenues dans des publications scientifiques et qui, après combinaisons avec d'autres données extérieures non mesurées explicitement dans les études, ont la possibilité de réaliser de nouvelles analyses. Des applications majeures de ce type d'études utilisant des données de la littérature sont par exemple le « disease mapping » ou l'étude de

Chapitre 7 - Utilisation des données issues de la littérature scientifique : Application à l'étude de la séroprévalence à *Toxoplasma gondii* dans les populations humaines

l'influence du climat sur la répartition géographique des maladies. Avec la multiplication des données, ces dernières connaissent un succès grandissant de fait de leur relative simplicité de mise en œuvre et de leurs résultats originaux. Ces analyses écologiques représentent souvent l'unique moyen d'étudier certaines problématiques. Quoique fortement décriées à cause de forts biais écologiques possibles, leurs utilisations peuvent s'avérer très utiles dans certaines branches de l'épidémiologie. Ici aussi, un encadrement méthodologique strict est nécessaire lors de leur réalisation pour éviter l'apparition de certains biais inhérents à l'utilisation des données de la littérature scientifique.

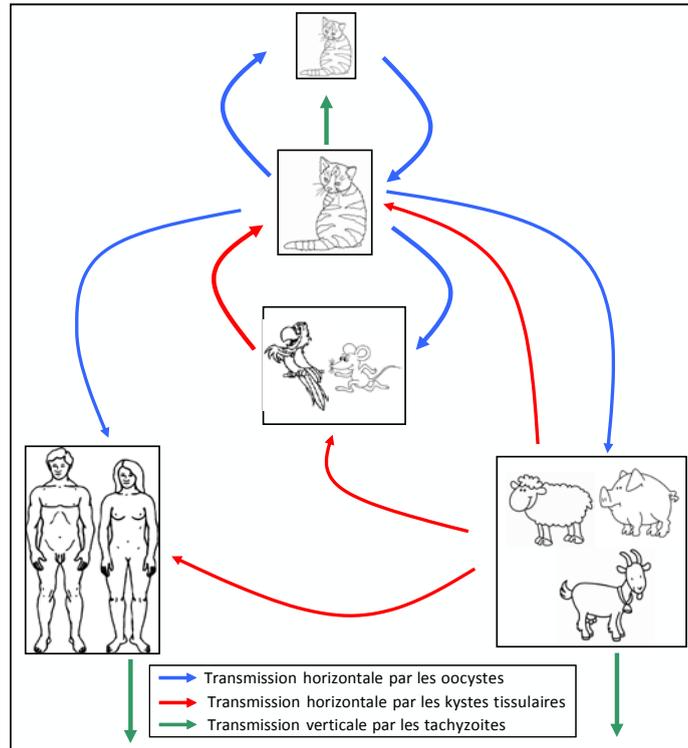
La partie qui suit illustre l'utilisation et l'application de ces méthodes écologiques à des données de la littérature scientifique concernant le niveau de séroprévalence de la toxoplasmose dans les populations humaines.

7.3. *Toxoplasma gondii* : un enjeu de santé publique chez la femme enceinte

Toxoplasma gondii est un protozoaire intracellulaire obligatoire appartenant à l'ordre des coccidies, phylum des apicomplexa. Il est considéré comme l'un des parasites les plus communs, capable de coloniser un grand nombre d'espèces d'hôtes homéothermes sur une vaste gamme d'habitats [15]. L'hôte définitif de ce parasite est un félin (en général le chat) et les hôtes intermédiaires sont des animaux, herbivores et carnivores (comprenant l'homme). *Toxoplasma gondii* se reproduit par multiplication sexuée dans l'intestin grêle du chat. Les oocystes produits sont alors disséminés dans l'environnement par les fèces du chat. Les oocystes sont infestants au bout de quelques jours de maturation dans le milieu extérieur. Ils sont très résistants et peuvent rester contaminants pendant plusieurs mois au niveau du sol. De nombreux animaux ainsi que l'homme peuvent se contaminer par ingestion d'oocystes. C'est sous forme de tachyzoïte, après multiplication et dissémination dans le sang, que le parasite envahit les cellules de l'hôte infesté. Quand les cellules meurent, elles dégagent des tachyzoïtes lesquels vont s'attaquer à d'autres cellules. Le microbe va ainsi se propager à l'ensemble de l'organisme. Quelques temps après son premier contact avec l'organisme (deux ou trois semaines), le parasite se divise plus lentement. Il reste latent sous forme de bradyzoïtes à l'intérieur de kystes disséminés dans l'organisme (cerveau, œil, muscles) (Figure 37).

Chez l'Homme, la toxoplasmose est une infection le plus souvent bénigne ou asymptomatique. Les formes graves sont avant tout observées en cas d'infection congénitale, chez les patients immunodéprimés et en fonction de la virulence des souches infectantes. En cas de contamination survenant chez une femme enceinte préalablement séronégative, il existe un risque de transmission materno-foetale et de toxoplasmose congénitale (TC) [16]. Une infection de la femme enceinte durant la grossesse peut se révéler grave et provoquer l'avortement, la mort du fœtus ou une naissance prématurée. Si la grossesse arrive à terme, il existe après la naissance la possibilité qu'un kyste cause la cécité, des troubles cardiaques ou cérébraux voire la mort de l'enfant. La fréquence de transmission materno-foetale est estimée globalement à 30 % des cas [17]. Elle s'effectue en moyenne 4 à 8 semaines après la colonisation du placenta. Elle est d'autant plus élevée que l'infection survient tardivement au cours de la grossesse. La période la plus dangereuse se situe entre la dixième et la vingt-quatrième semaine d'aménorrhée [17].

Figure 37: Les routes naturelles de transmission de *Toxoplasma gondii*. Adaptée de [15].

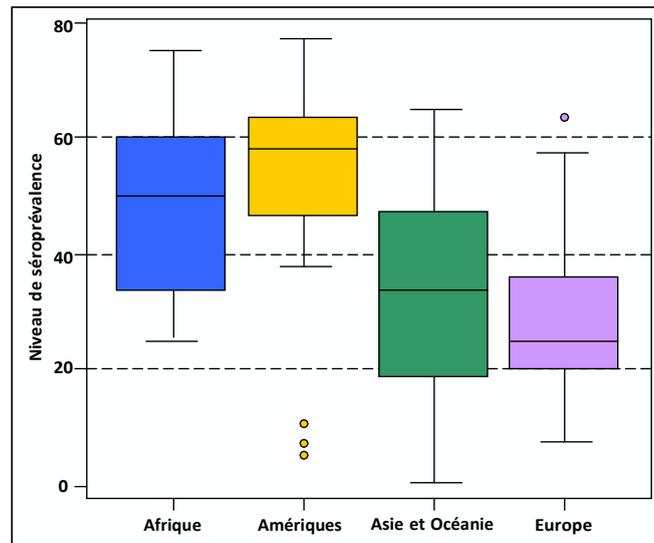


Le diagnostic de la toxoplasmose est avant tout sérologique car l'infection par le toxoplasme détermine une immunité humorale spécifique aisément dépistable apparaissant rapidement après la première infection et persistant toute la vie par la suite sous la forme d'anticorps de nature IgG. La mesure de la séroprévalence dans une population donnée permet d'évaluer son degré d'exposition à la toxoplasmose. La toxoplasmose est une zoonose cosmopolite. Il est estimé que plus de 50% des personnes au terme de leur existence sont contaminées par *Toxoplasma gondii*. Cependant, bien que la toxoplasmose humaine soit cosmopolite, sa prévalence est très variable selon les zones géographiques (Figure 38): elle diffère d'un pays à l'autre et parfois à l'intérieur même d'un pays [16]. En France, sa séroprévalence diminue régulièrement depuis une trentaine d'années mais reste actuellement assez élevée [18].

En Europe, seuls quatre pays (la France, le Danemark, l'Allemagne et l'Italie) ont organisé une surveillance active de la toxoplasmose congénitale [19]. Une telle surveillance n'existe pas ailleurs. Les pays développés en dehors de l'Europe ont généralement de faibles niveaux de séroprévalence toxoplasmose, et la toxoplasmose congénitale n'est donc pas une priorité. Les pays pauvres ont d'autres priorités sanitaires, même si le risque de toxoplasmose congénitale est élevé. La France apparaît comme le pays où la surveillance est la plus développée. Un programme national de prévention de la toxoplasmose congénitale existe depuis 1978. En plus de la vérification sérologique obligatoire en début de grossesse et de la récolte d'informations sur le risque de primo-infection, les femmes non immunes sont testées tous les mois jusqu'à l'accouchement depuis 1992. L'objectif de cette surveillance est de détecter les séroconversions le plus tôt possible pour mettre en œuvre des mesures diagnostiques et thérapeutiques sans délai en cas de séroconversions.

Chapitre 7 - Utilisation des données issues de la littérature scientifique : Application à l'étude de la séroprévalence à *Toxoplasma gondii* dans les populations humaines

Figure 38 : Variabilité de la prévalence à *Toxoplasma gondii* en fonction de la zone géographique. Adaptée de [16].



En agrégeant des données issues de la littérature scientifique sur la prévalence de la toxoplasmose, les deux études suivantes vont essayer de dessiner des réponses à une triple question :

- 1) Comment estimer le risque de toxoplasmose congénitale en fonction du niveau de prévalence globale dans une population ?
- 2) Quels sont les facteurs humains et environnementaux qui influencent ce niveau de prévalence global ?
- 3) Peut-on prédire, en croisant les données de la littérature recueillies à d'autres données sociétales et environnementales, le niveau de prévalences et le risque de toxoplasmose congénitale dans des zones où aucune donnée ne sont actuellement disponibles ?

7.4. Origine et contexte de ces recherches

Bien qu'il s'agisse d'une des rares parasitoses cosmopolites, la toxoplasmose constitue un problème de santé particulièrement important en Guyane française. Pour cela, trois raisons :

- 1) Une exposition théorique élevée au risque de toxoplasmose congénitale du fait de l'absence fréquente d'immunisation chez les femmes en âge de procréer, associée aux difficultés de prévention et de surveillance spécifique au cours des grossesses [20].
- 2) Sa gravité au cours du sida, la Guyane étant le département français le plus touché par l'infection à VIH [21].
- 3) L'identification ces dernières années de primo-infections toxoplasmiques sévères chez le sujet immunocompétent en rapport avec un cycle sauvage forestier de *Toxoplasma gondii* [22, 23].

L'exposition et le risque de toxoplasmose constituent depuis longtemps un sujet de réflexion et d'intérêt pour le Pr. Carme, chef de service du laboratoire de parasitologie de l'hôpital de Cayenne et médecin

Partie 2 : Travaux scientifiques

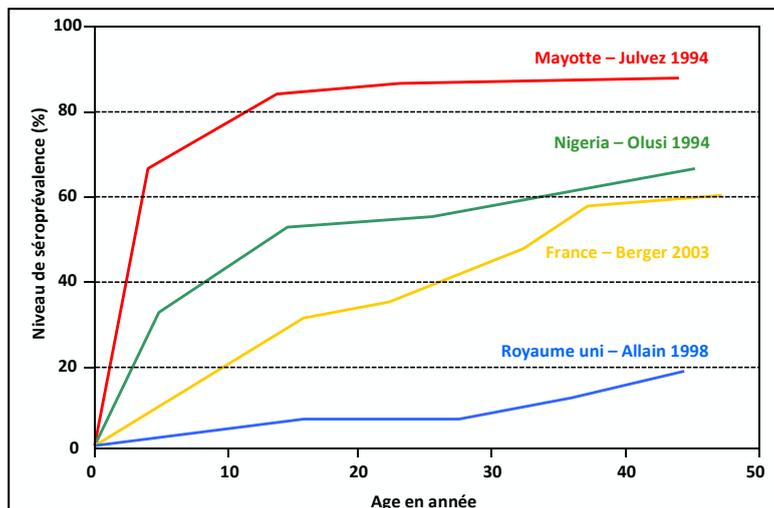
coordinateur du Centre d'investigation Clinique –Epidémiologie clinique Antilles Guyane. Son intérêt pour la toxoplasmose remontait à bien avant les différentes études réalisées sur ce sujet en Guyane.

Dans des travaux précédents, ce dernier avait proposé qu'une estimation rapide du risque de toxoplasmose congénitale pouvait être obtenue par l'observation du différentiel de séroprévalence entre 15 et 40 ans [24, 25].

A partir de cette hypothèse et en tenant compte des modes de contamination habituels et du contexte environnemental régional, 4 profils différents d'acquisition de l'immunité antitoxoplasmique reflétant la survenue des primo-infections au cours des âges dans une population avaient alors été théorisés ainsi que leur risque associé de toxoplasmose congénitale (Figure 39):

- Profil 1 : progression régulière, continue, importante de la séroprévalence au cours des âges pour atteindre des taux élevés chez les plus âgés (risque élevé de toxoplasmose congénitale).
- Profil 2 : progression régulière, continue mais limitée de la séroprévalence au cours des âges (risque faible de toxoplasmose congénitale).
- Profil 3: progression rapide et importante pour les enfants et adolescents puis évolution plus lente par la suite avec aplatissement de la courbe traduisant la rareté des contaminations chez les adultes (risque faible de toxoplasmose congénitale).
- Profil 4: même tendance mais moins marquée avec un plateau atteint pour des taux de positivité plus faible (risque faible ou élevé de toxoplasmose congénitale).

Figure 39 : Exemple de 4 profils différents d'acquisition de l'immunité antitoxoplasmique reflétant la survenue des primo-infections au cours des âges dans une population. Adaptée de [18, 26-28].



La question de valider ces différents modèles d'évolution de l'infection à *Toxoplasma gondii* au sein des populations se posait. L'idée d'une étude utilisant les données issues de la littérature scientifique fut évoquée et initia ce travail. La deuxième étude est le prolongement de cette réflexion et des résultats obtenus lors de la première étude.

7.5. Etude 8 : Classification des populations exposées à la toxoplasmose congénitale par l'analyse de leur séroprévalence globale

Introduction : *Toxoplasma gondii* aura infecté près la moitié des habitants de la planète au terme de leur existence. L'infection toxoplasmique varie toutefois beaucoup d'un pays à l'autre (10 à 80%). Chez l'homme l'infection est le plus souvent bénigne. Les formes graves sont observées en cas de contamination congénitale (risque d'atteintes fœtales) pouvant survenir lorsqu'une femme enceinte est contaminée pour la première fois ou chez les patients immunodéprimés. Cette étude visait à cerner les populations exposées à la toxoplasmose congénitale par l'analyse de l'évolution de leur séroprévalence au cours des âges.

Matériels/Méthodes : Un large revue de la littérature portant sur des études décrivant l'évolution de la séroprévalence selon les âges dans des populations a été faite. Toutes les études correspondant aux critères de sélections déterminés préalablement ont été incorporées dans l'analyse. Pour déterminer des profils types dans les évolutions de la séroprévalence selon l'âge, un algorithme de classification hiérarchique ascendante utilisant une distance euclidienne et un critère d'agrégation de Ward a été utilisé.

Résultats : 65 études décrivant l'évolution du niveau de séroprévalence de populations au cours des âges ont finalement pu être incluses dans cette analyse. Quatre groupes de profils d'acquisition de l'immunité anti – Toxoplasme exposant à différents risques de toxoplasmose congénitale ont pu être mis en évidence. Globalement ce risque concernait principalement les pays ou région et/ou les communautés qui présentent un fort différentiel de prévalence entre 15 et 40 ans (>15%). Dans l'étude, les groupes où une telle situation se retrouvait, correspondaient à ceux où la séroprévalence moyenne de la population était comprise entre 30% et 60%.

Conclusions : Malgré le fait qu'une telle classification peut varier dans le temps et l'espace pour un pays avec de forts contrastes géographiques et culturels, de simples indicateurs comme la prévalence globale et le différentiel de prévalence entre 15 et 40 ans peuvent permettre d'identifier les populations les plus à risque de toxoplasmose congénitale. Cette situation correspond à celle où le ratio avantages/inconvénients d'un programme de lutte contre la toxoplasmose congénitale est le meilleur.

Article soumis le 01 septembre 2011 à International Journal for Parasitology

1	Title page
2	
3	Running head
4	Exposition to congenital toxoplasmosis
5	
6	Title
7	Classification of populations exposed to congenital toxoplasmosis using a systematic review of
8	seroprevalences
9	
10	Authors and their affiliations
11	Bernard Carne * [£] , Matthieu Hanf * [£] , Magalie Demar [£] , Mathieu Nacher * [£]
12	
13	* : Centre d'Investigation Clinique Epidémiologie Clinique Antilles Guyane CIC-EC INSERM
14	CIE 802, Cayenne General Hospital, Cayenne, French Guiana.
15	[£] : Equipe EPaT EA3593, Cayenne General Hospital and UFR Médecine, Université des Antilles
16	et de la Guyane, Cayenne, French Guiana
17	

Chapitre 7 - Utilisation des données issues de la littérature scientifique : Application à l'étude de la séroprévalence à *Toxoplasma gondii* dans les populations humaines

18 **Footnote page**

19

20 **Statement: Conflict of interest**

21 We declare that the authors do not have a commercial or other association that might pose a
22 conflict of interest.

23

24 **Source of financial support**

25 No financial source other than those of the institutions (Inserm – Centre Hospitalier de Cayenne)

26

27 **Meeting where a part of the information has been presented**

28 Société Française de Parasitologie, Strasbourg Mai 2011, abstract number : CA38. CARME B,
29 M. HANF, M. DEMAR, M. NACHER. Profils de séroprévalence et exposition au risque de
30 toxoplasmose congénitale.

31

32 **Corresponding author :**

33 Bernard Carme : carme.bernard@wanadoo.fr, Tel -- 594 594 39 53 01, Fax –594 594 39 50 51,
34 Centre Hospitalier de Cayenne, BP 6006, Rue des Flamboyants, F-97306 Cayenne (French
35 Guiana)

36

37 **ABSTRACT** (200 words)

38

39 **Background:** *Toxoplasma* infection varies widely between countries. The infection is mostly
40 benign. Severe forms are observed during congenital infections that can occur when a pregnant
41 woman is infected for the first time.

42 **Methods:** The present study aimed to identify the populations exposed to congenital
43 toxoplasmosis through the analysis of seroprevalences at the global scale. A large selection of
44 published studies which described seroprevalence evolution by age in a healthy population was
45 incorporated in the analysis. To determine groups with specific particularities into seroprevalence
46 age profiles, a hierarchical clustering algorithm based on the average agglomeration method was
47 used. A Euclidian distance was defined as the similarity measure

48 **Results:** The profiles of the acquisition of anti-*Toxoplasma* immunity can integrate 4 groups
49 exposed to different risks of congenital toxoplasmosis. Overall; this risk mainly concerns
50 countries, regions and/or communities that present a high age prevalence rate differential between
51 40 years and 15 years (> 15%); this indicator being itself linked to a general mean seroprevalence
52 situated between 30 and 60%.

53 **Conclusion:** Although classifications by country may vary through time and space for countries
54 with geographic and cultural contrasts, simple indicators such as mean seroprevalence can help
55 identify populations most at risk for congenital toxoplasmosis.

56

57 **Key words**

58 Toxoplasmosis, seroprevalence, congenital toxoplasmosis, risk factors, prediction

59 TEXT: 3093 words

60

61 INTRODUCTION

62

63 Toxoplasmosis is the most widespread parasitic infection in the world with over 50% of people
64 infected by the end of their existence and an animal reservoir of parasites including most
65 homeotherms (mammals, birds). Felids, principally cats, are the only definitive hosts (sexual
66 maturation of the parasite), all other animals being intermediary hosts. In humans, the prevalence
67 varies widely between countries (from 10 to 80%) and often within a given country, or within
68 different communities living in a same region (Pappas et al., 2009). Toxoplasmosis is usually a
69 benign infection in immunocompetent subjects. Severe forms are mostly observed in congenital
70 infections that can occur when a pregnant woman is infected for the first time (fetal transmission
71 occurs in 30% of cases) (Dunn et al., 1999), and in immunocompromised patients. Apart from
72 these circumstances, severe forms can be exceptionally observed with particular virulent strains
73 except until now in French Guiana (Carme et al., 2002) where a wild Neotropical forest cycle is
74 described (Carme et al., 2009).

75 The diagnosis is mostly serological because the infection elicits a specific humoral response that
76 is easily detected shortly after the first infection and persists throughout life in the form of IgG
77 antibodies. A positive serology reflects that a person has been infected and the seroprevalence
78 (SP) rate of a population reflects the general level of infection.

79 For an appreciation of the potential risk of congenital toxoplasmosis it is essential to know the
80 importance of the risk group (non immune persons) among the female population of reproductive
81 age and its exposure to *Toxoplasma* during pregnancy. A rapid estimation can be obtained by the
82 observation in the general population of the SP differential between 15 and 40 years of age under
83 the triple premise: the probability of being infected for the first time by *T.gondii* is identical or
84 very similar: 1- between women and men, 2- whether the woman is pregnant or not, 3- and all
85 along the past 25 years (40 - 15 = 25) for the persons of the same age. Taking into account the
86 usual infection modes and the regional environmental context, different acquisition profiles for
87 anti*Toxoplasma* immunity can be proposed reflecting the incidence of primo-infections. Four
88 types of profiles were hypothesized in previous publications (Carme and Tirard-Fleury, 1996;
89 Carme, 2001). More explicitly:

90 - Profile 1: regular, continuous, important progression of SP to reach elevated SP rates among
91 the older age groups: mostly food borne infection by ingestion of cysts while eating undercooked
92 meat.

93 - Profile 2: regular, continuous but limited progression of SP (similar mode of infection but rarer
94 because dietary habits are less favorable for *Toxoplasma*).

95 - Profile 3: rapid and important progression of SP for children and teenagers then slower
96 evolution reflecting the rarity of adult infections: mostly through telluric (or water borne)
97 infection by oocyst ingestion in a context that is favorable to their survival on the ground
98 associated with poor hygiene conditions.

99 - Profile 4: same tendency but less pronounced with a plateau at lower SP rates.

100

101 In order to test this hypothetical framework, a review of the scientific literature of the past
102 decades was performed to describe the evolution of SP by age in a range of different populations.
103 The objective was to determine whether the above intuitive models could be generalized and if
104 so how to characterize them using a large data set. The quantification of the risk of CT of each
105 statistically identified group could subsequently support the implementation of CT prevention
106 programs, as performed nationally in France and Austria.

107

108

109

110

Chapitre 7 - Utilisation des données issues de la littérature scientifique : Application à l'étude de la séroprévalence à *Toxoplasma gondii* dans les populations humaines

59 **TEXT:** 3093 words

60

61 **INTRODUCTION**

62

63 Toxoplasmosis is the most widespread parasitic infection in the world with over 50% of people
64 infected by the end of their existence and an animal reservoir of parasites including most
65 homeotherms (mammals, birds). Felids, principally cats, are the only definitive hosts (sexual
66 maturation of the parasite), all other animals being intermediary hosts. In humans, the prevalence
67 varies widely between countries (from 10 to 80%) and often within a given country, or within
68 different communities living in a same region (Pappas et al., 2009). Toxoplasmosis is usually a
69 benign infection in immunocompetent subjects. Severe forms are mostly observed in congenital
70 infections that can occur when a pregnant woman is infected for the first time (fetal transmission
71 occurs in 30% of cases) (Dunn et al., 1999), and in immunocompromised patients. Apart from
72 these circumstances, severe forms can be exceptionally observed with particular virulent strains
73 except until now in French Guiana (Carne et al., 2002) where a wild Neotropical forest cycle is
74 described (Carne et al., 2009).

75 The diagnosis is mostly serological because the infection elicits a specific humoral response that
76 is easily detected shortly after the first infection and persists throughout life in the form of IgG
77 antibodies. A positive serology reflects that a person has been infected and the seroprevalence
78 (SP) rate of a population reflects the general level of infection.

79 For an appreciation of the potential risk of congenital toxoplasmosis it is essential to know the
80 importance of the risk group (non immune persons) among the female population of reproductive
81 age and its exposure to *Toxoplasma* during pregnancy. A rapid estimation can be obtained by the
82 observation in the general population of the SP differential between 15 and 40 years of age under
83 the triple premise: the probability of being infected for the first time by *T.gondii* is identical or
84 very similar: 1- between women and men, 2- whether the woman is pregnant or not, 3- and all
85 along the past 25 years ($40 - 15 = 25$) for the persons of the same age. Taking into account the
86 usual infection modes and the regional environmental context, different acquisition profiles for
87 anti*Toxoplasma* immunity can be proposed reflecting the incidence of primo-infections. Four
88 types of profiles were hypothesized in previous publications (Carne and Tirard-Fleury, 1996;
89 Carne, 2001). More explicitly:

90 - Profile 1: regular, continuous, important progression of SP to reach elevated SP rates among
91 the older age groups: mostly food borne infection by ingestion of cysts while eating undercooked
92 meat.

93 - Profile 2: regular, continuous but limited progression of SP (similar mode of infection but rarer
94 because dietary habits are less favorable for *Toxoplasma*).

95 - Profile 3: rapid and important progression of SP for children and teenagers then slower
96 evolution reflecting the rarity of adult infections: mostly through telluric (or water borne)
97 infection by oocyst ingestion in a context that is favorable to their survival on the ground
98 associated with poor hygiene conditions.

99 - Profile 4: same tendency but less pronounced with a plateau at lower SP rates.

100

101 In order to test this hypothetical framework, a review of the scientific literature of the past
102 decades was performed to describe the evolution of SP by age in a range of different populations.
103 The objective was to determine whether the above intuitive models could be generalized and if
104 so how to characterize them using a large data set. The quantification of the risk of CT of each
105 statistically identified group could subsequently support the implementation of CT prevention
106 programs, as performed nationally in France and Austria.

107

108

109

110

111 **METHODS**

112

113 *Reviewing methodology of Toxoplasmosis SP studies*

114

115 In view to confirm this hypothetical framework of four distinct acquisition profiles for
116 anti*Toxoplasma* immunity in populations, a large selection of published studies which described
117 age evolution of toxoplasmosis SP in human populations was reviewed in the analysis.

118 The following search strategies (flow diagram in figure 1) were used for this systematic review
119 of the toxoplasmosis SP age evolution literature: 1) computer search of databases, 2) review of
120 specialized reports or already made toxoplasmosis SP reviews, and 3) review of studies coming
121 from the grey literature identified and collected previously by the senior author. This search
122 strategy tries to be the most possible in accordance with expert recommendations on the subject.

123 **Computer search:** Toxoplasmosis SP studies were searched in MEDLINE and SCIENCE
124 DIRECT databases from January 1960 to December 2009 for articles in English, German,
125 French, Spanish, and Portuguese. Search words/terms were as follows: seroprevalence,
126 prevalence, rate, congenital toxoplasmosis, *toxoplasma gondii*, toxoplasmosis, pregnant women,
127 general population, epidemiology, blood donors, child-bearing age, antenatal, and TORCH.

128 • **Step 1:** Abstracts were reviewed by two of the authors (BC & M.H) and selected for
129 further review if they met the following 3 criteria: 1) Original investigations of prevalence
130 based on populations with no particular health problem (General population, pregnant
131 women, women in child bearing age, and blood donors); 2) a sample size > 200 persons; 3)
132 Studies made with one of the following tests : the Sabin-Feldman dye test, indirect
133 fluorescent antibody (IFA), agglutination tests, or Enzyme Linked ImmunoSorbent Assay
134 (ELISA). If a criterion was not met because not enough information was provided, the
135 abstract was set aside for further evaluation.

136 • **Step 2:** If full-text articles of abstracts selected in Step 1 were retrieved, these last were
137 reviewed by the two first authors. Studies were included if they met all the following criteria
138 : 1) Studies with a clearly reported study period 2) Studies describing SP with a number of
139 age categories superior or equal to 4. 3) Studies with a SP description at least up to 40 year
140 age 4) A sample size > 200 persons 5) Studies made with one of the following tests : the
141 Sabin-Feldman dye test, indirect fluorescent antibody (IFA), agglutination tests, or Enzyme
142 Linked ImmunoSorbent Assay (ELISA). For non retrieved articles, if all criteria described in
143 step two were specified in the abstract and respected, the study was finally include in the
144 analysis. Inclusion was based on the approbation of the senior investigator (B.C). For studies
145 describing several well identified geographical areas, a distinction was made if each sub-
146 studies met the inclusion criteria described in step 2.

147 **Review of specialized reviews or reports:** Four major reviews or reports were reviewed to
148 identify references to original surveys that might be incorporated in the review processing
149 (Tenter et al., 2000; Agence Francaise de Sécurité Sanitaire des Aliments (AFSSA), 2005;
150 Benard et al., 2008; Pappas et al., 2009).

151 **Review of literature collected previously by the senior author:** all articles were reviewed, and
152 the full texts of potentially interesting studies were examined.

153

154 *Statistical methods*

155

156 To reconstruct SP age profile of selected studies, the means of each age category were assigned
157 to the middle age of the category. SP at ages between these values were then linearly interpolated
158 between 1 year and 40 years. To determine groups with specific particularities into SP age
159 profiles, an agglomerative hierarchical clustering algorithm was performed.

160 Cluster analysis was used to divide the analyzed studies into a limited number of clusters with
161 maximally differing age SP patterns. The hierarchical agglomerative clustering method was
162 used, with Ward's minimum variance method and Euclidean distance. Cluster analysis was

163 performed on the 40 variables describing the level of SP for ages from 1 to 40. The hierarchical
164 agglomerative clustering method consists of merging the two closest individuals (here SP
165 studies) or clusters to form a new cluster that replaces them and iterating the process with the
166 new group until complete regrouping (Kaufman and Rousseeuw, 2008); it provides a
167 dendrogram (classification tree). The Ward's minimum variance method minimized the within-
168 cluster sum of squares at each step of the algorithm. The number of clusters was determined by
169 using the dendrogram. The squared multiple correlations R^2 , which is the proportion of variance
170 accounted for by the clusters, were then calculated. The number of clusters was then decided
171 based on two criteria: a sufficient proportion of variance accounted for by the clusters and large
172 enough groups to be further studied. A bootstrap analysis was finally conducted to determine the
173 confidence cut off for each branch. For the main clusters, approximately unbiased p-values were
174 calculated via multiscale bootstrap resampling. The P-value of a cluster is a value between 0 and
175 1, which indicates how strongly the cluster is supported by the data.

176
177 For each group mean SP evolution by age and associated confident interval, were then
178 computed. To evaluate congenital toxoplasmosis risk, mean percentages of seroconversion
179 during child bearing age were computed and crude relative risks were calculated using, as the
180 reference, the group with the weaker percentage of seroconversions during child bearing age.
181 All statistical analyses were conducted using R 2.12.0 (R Development Core Team, 2010).

182

183 **RESULTS**

184

185 3032 abstracts published since 1960 were screened. A total of 356 articles were reviewed,
186 and 75 studies were included in this systematic review (figure 1). Studies from all continents
187 were retrieved. 81 distinct SPs' populations described in these 75 studies reaching the selection
188 criteria were incorporated in the analysis (Table 1). Geographic area repartition of the
189 incorporated populations was: Africa (17%), Asia (16%), Europe (38%), North and Central
190 America (8%), South America (16%) and South West Pacific (5%). Forty studies (49%) were
191 made in the general population, 27 (33%) in the population of pregnant women, 8 (10%) in the
192 population of women of child bearing age and 6 (8%) in other populations (blood donors and
193 women).

194 Four main distinct profiles of SP evolution by age were identified by the cluster analysis (figure
195 2). The evolution of toxoplasmosis SP in these averaged age groups and confidence intervals
196 were drawn in figure 3. SP by age and associated infection rate of each identified group could be
197 summarized as follows:

198 -Group 1: The infection rate is constant and weak along ages (linear low or LL).

199 -Group 2: The overall SP is high with an infection period occurring mainly during childhood
200 (high curve or HC).

201 -Group 3: The infection rate is constant and high across age groups (linear high or LH).

202 -Group 4: The overall SP is intermediate with an infection period occurring mainly during
203 childhood (low curve or LC).

204 The characteristics of each group are summarized in table 2. The risk of congenital
205 toxoplasmosis reflected by the proportion of women who seroconvert during their child-bearing
206 age was weak for the groups 1 and 2 (13% and 13% respectively) and high in the groups 3 and 4
207 (20% and 20% respectively). In these last groups the risk to seroconvert during pregnancy was
208 1.5 times more important than in the first groups. The less exposed groups corresponding to low
209 mSP (< 30%) and high mSP (> 60%), the most exposed corresponding to situations in between.
210 Therefore, the sole knowledge of mSP could allow predicting the risk of CT globally.

211

212 **DISCUSSION**

213

Partie 2 : Travaux scientifiques

214 The analysis of 81 populations illustrates the cosmopolitan importance of this parasite with
215 marked differences between countries, and, at times, within a single country or between different
216 periods. Prevalences below 30% are principally observed in North America, Great Britain,
217 Scandinavia and South East Asia. Prevalences > 60 % are mostly encountered in Africa and
218 Latin America. Europe presently seems in a median situation. Climate, but also other risk factors
219 linked to lifestyles, hygiene levels and dietary habits are discussed in the literature to account for
220 these global differences (Tenter et al., 2000; Pappas et al., 2009).

221 The initial hypothesis of 4 SP dynamic profiles seemed to hold the test of the statistical analysis
222 of the literature. But some countries were in different groups and some groups contained
223 countries that seemed quite different (i.e. some studies in France were close to Iran and Benin).
224 These counter intuitive observations require further explanations. Regional variations within a
225 single country are more likely if the country is vast, associating seashore, continental, and
226 mountainous regions, and if populations have diverse origins and cultures. Hence, a study of
227 Swiss pregnant women will presumably be more representative of the whole country than a
228 study in the Amazon will be of the Brazilian situation. In France, for example there are notable
229 differences between regions (Berger et al., 2009). It is of note that for those with ≥ 4 studies
230 included in the analyses (Sweden, Venezuela, Brazil and France) their group varies but none of
231 them fitted in 3 different groups. Sweden and Brazil nevertheless appeared in distant groups:
232 group 1 (studies Sweden 1, 4, 5 and 6) and 3 (studies Sweden 2 and 3) for the first and group 2
233 (studies Brazil 1 and 3) and 4 (studies Brazil 2, 4 and 5) for the second. These differences can
234 also be explained by different regions, or populations, and different study periods.

235 The exactitude of results and their comparison also depend on the serological techniques used.
236 Globally, and mostly for old studies, the tendency is for an underestimation of prevalence rates
237 because of the lack of sensitivity of certain methods that may not diagnose very remote
238 infections in which antibody levels are low. The studies allowing a rigorous comparison between
239 different countries for a same period and with a similar serological technique are rare. This was
240 the case in 1985 with a prospective study of 800 women of childbearing age from 4 European
241 cities using a similar serological reference technique (dye test), performed in the same laboratory
242 (Remington JS et al., 1996). SPs were very different between London (21%), Stuttgart (36%),
243 Padua (56%) and Paris (72%), which reflected different epidemiological environments.

244 The profile of the study population is important to know: general population, women of child
245 bearing age, blood donors, or pregnant women. Three premises allowed selecting an important
246 number of studies for the analysis. The 2 first concerns the absence of difference between the
247 exposure of males and females 15 to 40 years and between pregnant and non pregnant women.
248 The first condition is debatable and salient differences tend to go in opposite directions and
249 neutralize their respective effects: house work (cooking, caring for the cat) may expose women
250 more, but a meat diet could also expose men more. The influence of pregnancy could be
251 discussed if there is a prevention program of CT in the country. However, the absence of notable
252 variation in time is contradicted, at least in Europe, by the trend towards an important reduction
253 of toxoplasmic SP during the past 4 decades.

254 Thus, comparative studies must take into account the period of the survey. In France, the SP in
255 pregnant women was about 80% in the early sixties, around 66% in the eighties, 54% in 1995,
256 and 44% in 2003 while at the same time the average age of pregnant women increased (Villena I
257 et al., 2010). But the 15-40 SP differentials has not significantly declined during this period. This
258 is also the case in all areas where this was studied in Europe. For example, in The Netherlands
259 the SP decreased from 35.2% in 1995/1996 to 18.5% in 2006/2007 in women of reproductive
260 age leaving the majority of pregnant women susceptible to primary infection with *T. gondii* and
261 their babies to congenital toxoplasmosis (Hofhuis et al., 2010). The necessity of periodical
262 studies for this population at a national level was successfully emphasized in 1994 (Carme et al.,
263 1994), and in 1995 French national perinatal studies included toxoplasmosis (Ancelle et al.,
264 1995) and still do (2003, 2010) (Villena et al., 2010).

Chapitre 7 - Utilisation des données issues de la littérature scientifique : Application à l'étude de la séroprévalence à *Toxoplasma gondii* dans les populations humaines

265 Concerning the probability of a first infection by *Toxoplasma gondii* in a determined population
266 and for a defined age, 2 elements must be taken in consideration: 1- the mean SP (mSP) , 2- the
267 SP differential (Δ SP) applied to the population at the end and the beginning of the considered
268 age groups. The level of mSP indicates the theoretical exposure by the fact that only
269 seronegative persons are susceptible to develop a toxoplasmic primo-infection. The value of the Δ
270 SP accounts for the theoretical exposure during a defined period, (i.e. for the 15-40 years of age
271 $\rightarrow \Delta$ SP 15-40).

272 Within the 4 groups described in figure 3, groups 3 and 4 had a high Δ SP 15-40 (showing a high
273 exposure to the risk of CT) and groups 1 & 2 had a lower Δ SP 15-40; the less exposed groups
274 corresponding to low mSP (< 30%) and high mSP (> 60%), the most exposed corresponding to
275 situations in between. Therefore, the sole knowledge of mSP could allow predicting the risk of
276 CT globally.

277 But the situation is more complex because the SP differential reflects the risk of a new infection
278 only if the general trend is stable. This supposes that the incidence of toxoplasmosis is constant
279 for all age groups, which is not accurate for numerous countries. The observed SPs correspond to
280 infections before the year of the study and the results reflect for a large part the incidence of
281 previous decades and not the incidence for the year of the study. However, the conditions of
282 *Toxoplasma* exposure may have varied, even within a homogenous population. A reduction or an
283 increase of the exposure underlying the recent primo-infections, the only ones of interest for CT
284 cases, only concern women that are not yet infected. These variations will have little effect on
285 the curve's profile at short term and even medium term, notably when the SP rate is globally
286 high; positive persons no longer being concerned by the present situation. The reactivity to an
287 increase in exposition will be stronger in low prevalence areas. On the contrary, a reduction of
288 exposure will be noticed more easily in high prevalence regions.

289 The opportunity of CT prevention programs should be determined weighing the pros (reduction
290 of CT and/or reduction of pathological consequences) and the cons of such programs (serological
291 surveillance costly and source of anxiety (Khoshnood et al., 2007), diagnostic and/or therapeutic
292 measures sometimes inappropriate). Systematic reviews found no good comparative data
293 measuring the potential harms and benefits of antiparasitic drugs used for presumed antenatal
294 *Toxoplasma* infection. Indeed the value of antenatal toxoplasmosis screening programs depends
295 on safe treatments that reduce the risk of congenital disease (Wallon et al., 1999). Some consider
296 that it is preferable not to generalize surveillance (Bader et al., 1997).

297

298 CONCLUSION

299

300 The present study confirmed the hypothesis that there were 4 distinct acquisition profiles for anti
301 *Toxoplasma* immunity. Although classifications by country may vary through time and space for
302 countries with geographic and cultural contrasts, the risk of CT mainly concerns populations
303 with a general mean seroprevalence between 30% and 60%. This situation corresponds to
304 situations where the pros of prevention programs outweigh the cons.

305

306

307 FUNDING

308

309 No financial source other than those of the institutions: Inserm (Institut National de la Santé et de
310 Recherche Médicale) – Cayenne General Hospital

311

312

313

314

315

316

317 **AKNOWLEDGMENTS**

318

319 **REFERENCES**

320

321 Abu-Madi, M.A., Al-Molawi, N., Behnke, J.M., 2008. Seroprevalence and epidemiological
322 correlates of *Toxoplasma gondii* infections among patients referred for hospital-based
323 serological testing in Doha, Qatar. *Parasit Vectors* 1, 39.

324 Agence Francaise de Sécurité Sanitaire des Aliments (AFSSA), 2005. *Toxoplasma gondii*:
325 present knowledge and risk assessment of foodborne toxoplasmosis. In, AFSSA, Maison
326 Alfort.

327 Ahlfors, K., Borjeson, M., Hultdt, G., Forsberg, E., 1989. Incidence of toxoplasmosis in pregnant
328 women in the city of Malmo, Sweden. *Scand J Infect Dis* 21, 315-321.

329 Al-Qurashi, A.R., Ghandour, A.M., Obeid, O.E., Al-Mulhim, A., Makki, S.M., 2001.
330 Seroepidemiological study of *Toxoplasma gondii* infection in the human population in
331 the Eastern Region. *Saudi Med J* 22, 13-18.

332 Allain, J.P., Palmer, C.R., Pearson, G., 1998. Epidemiological study of latent and recent
333 infection by *Toxoplasma gondii* in pregnant women from a regional population in the
334 U.K. *J Infect* 36, 189-196.

335 Ancelle, T., Goulet, V., Tirard-Fleury, V., Baril, L., Du Mazaubrun, C., Thulliez, P., et al., 1995.
336 La toxoplasmose chez la femme enceinte en France en 1995. *BEH* 51.

337 Asthana, S.P., Macpherson, C.N., Weiss, S.H., Stephens, R., Denny, T.N., Sharma, R.N., Dubey,
338 J.P., 2006. Seroprevalence of *Toxoplasma gondii* in pregnant women and cats in
339 Grenada, West Indies. *J Parasitol* 92, 644-645.

340 Avelino, M.M., Campos, D., Jr., Parada, J.B., Castro, A.M., 2004. Risk factors for *Toxoplasma*
341 *gondii* infection in women of childbearing age. *Braz J Infect Dis* 8, 164-174.

342 Bader, T.J., Macones, G.A., Asch, D.A., 1997. Prenatal screening for toxoplasmosis. *Obstet*
343 *Gynecol* 90, 457-464.

344 Bartolome Alvarez, J., Martinez Serrano, M., Moreno Parrado, L., Lorente Ortuno, S., Crespo
345 Sanchez, M.D., 2008. [Prevalence and incidence in Albacete, Spain, of *Toxoplasma*
346 *gondii* infection in women of childbearing age: differences between immigrant and non-
347 immigrant (2001-2007)]. *Rev Esp Salud Publica* 82, 333-342.

348 Benard, A., Petersen, E., Salamon, R., Chene, G., Gilbert, R., Salmi, L.R., 2008. Survey of
349 European programmes for the epidemiological surveillance of congenital toxoplasmosis.
350 *Euro Surveill* 13.

351 Berger, F., Goulet, V., Le Strat, Y., Desenclos, J.C., 2009. Toxoplasmosis among pregnant
352 women in France: risk factors and change of prevalence between 1995 and 2003. *Rev*
353 *Epidemiol Sante Publique* 57, 241-248.

354 Birgisdottir, A., Asbjornsdottir, H., Cook, E., Gislason, D., Jansson, C., Olafsson, I., Gislason,
355 T., Jogi, R., Thjodleifsson, B., 2006. Seroprevalence of *Toxoplasma gondii* in Sweden,
356 Estonia and Iceland. *Scand J Infect Dis* 38, 625-631.

357 Bobic, B., Nikolic, A., Djurkovic-Djakovic, O., 2003. [Identification of risk factors for infection
358 with *Toxoplasma gondii* in Serbia as a basis of a program for prevention of congenital
359 toxoplasmosis]. *Srp Arh Celok Lek* 131, 162-167.

360 Boia, M.N., Carvalho-Costa, F.A., Sodre, F.C., Pinto, G.M., Amendoeira, M.R., 2008.
361 Seroprevalence of *Toxoplasma gondii* infection among Indian people living in Iauarete,
362 Sao Gabriel da Cachoeira, Amazonas, Brazil. *Rev Inst Med Trop Sao Paulo* 50, 17-20.

363 Breurec, S., Berlioz-Arthaud, A., Baumann, E., Miegerville, M., Billaud, E., 2004. [Evaluation of
364 toxoplasmosis seroprevalence among 2416 women of childbearing age followed at the
365 Pasteur Institute of New Caledonia]. *Bull Soc Pathol Exot* 97, 271-273.

366 Carne, B., Lenne, E., Tirard, V., Hayette, M.P., Gondry, J., 1994. Etude épidémiologique de la
367 toxoplasmose chez les femmes enceintes à Amiens (Picardie). Nécessité d'une enquête
368 nationale. *Médecine et Maladies Infectieuses* 24, 1271-1273.

Chapitre 7 - Utilisation des données issues de la littérature scientifique : Application à l'étude de la séroprévalence à *Toxoplasma gondii* dans les populations humaines

- 369 Carme, B., Tirard-Fleury, V., 1996. [Toxoplasmosis among pregnant women in France :
370 Sero prevalence, seroconversion and knowledge levels. Trends 1965-1995.]. *Med Mal*
371 *Infect* 26 Suppl 3, 431-436.
- 372 Carme, B., 2001. [Exposure to *Toxoplasma gondii* and risk of congenital toxoplasmosis]. *Med*
373 *Trop (Mars)* 61, 550-551.
- 374 Carme, B., Bissuel, F., Ajzenberg, D., Bouyne, R., Aznar, C., Demar, M., Bichat, S., Louvel, D.,
375 Bourbigot, A.M., Peneau, C., Neron, P., Darde, M.L., 2002. Severe acquired
376 toxoplasmosis in immunocompetent adult patients in French Guiana. *J Clin Microbiol* 40,
377 4037-4044.
- 378 Carme, B., Demar, M., Ajzenberg, D., Darde, M.L., 2009. Severe acquired toxoplasmosis caused
379 by wild cycle of *Toxoplasma gondii*, French Guiana. *Emerg Infect Dis* 15, 656-658.
- 380 Chacin-Bonilla, L., Sanchez-Chavez, Y., Monsalve, F., Estevez, J., 2001. Seroepidemiology of
381 toxoplasmosis in amerindians from western Venezuela. *Am J Trop Med Hyg* 65, 131-
382 135.
- 383 Chacin-Bonilla, L., Sanchez-Chavez, Y., Estevez, J., Larreal, Y., Molero, E., 2003.
384 Seroepidemiology of Toxoplasmosis in amerindians from Western Venezuela.
385 *Interciencia* 28, 457-462.
- 386 Condorelli, F., Scalia, G., Stivala, A., Costanzo, M.C., Adragna, A.D., Franceschino, C.,
387 Santagati, M.G., Furneri, P.M., Marino, A., Castro, A., 1993. Sero prevalence to some
388 TORCH agents in a Sicilian female population of fertile age. *Eur J Epidemiol* 9, 341-343.
- 389 Contreras, M., Schenone, H., Salinas, P., Sandoval, L., Rojas, A., Villarroel, F., Solis, F., 1996.
390 Seroepidemiology of human toxoplasmosis in Chile. *Rev Inst Med Trop Sao Paulo* 38,
391 431-435.
- 392 Crucerescu, E., Lovin, D., Tutunaru, D., Chirila, A., Ilcenco, D., 2006. Sero prevalence of
393 *Toxoplasma gondii* in general population in moldavian provinces. *Les annales de*
394 *l'université "Dunarea de jos"* 5, 53-55.
- 395 Dar, F.K., Alkarmi, T., Uduman, S., Abdulrazzaq, Y., Grundsell, H., Hughes, P., 1997.
396 Gestational and neonatal toxoplasmosis: regional sero prevalence in the United Arab
397 Emirates. *Eur J Epidemiol* 13, 567-571.
- 398 Diza, E., Frantzidou, F., Souliou, E., Arvanitidou, M., Gioula, G., Antoniadis, A., 2005.
399 Sero prevalence of *Toxoplasma gondii* in northern Greece during the last 20 years. *Clin*
400 *Microbiol Infect* 11, 719-723.
- 401 Dromigny, J.A., Pecarrere, J.L., Leroy, F., Ollivier, G., Boisier, P., 1996. [Prevalence of
402 toxoplasmosis in Tananarive. Study conducted at the Pasteur Institute of Madagascar
403 (PIM) on a sample of 2354 subjects]. *Bull Soc Pathol Exot* 89, 212-216.
- 404 Dumas, N., Cazaux, M., Rivaillier, P., Seguela, J.P., 1985. [Toxoplasmosis in the African
405 tropical zone. Preliminary studies]. *Bull Soc Pathol Exot Filiales* 78, 795-800.
- 406 Dumas, N., Cazaux, M., Carme, B., Seguela, J.P., Charlet, J.P., 1990. [Toxoplasmosis in the
407 Congo Republic. Seroepidemiological study]. *Bull Soc Pathol Exot* 83, 349-359.
- 408 Dunn, D., Wallon, M., Peyron, F., Petersen, E., Peckham, C., Gilbert, R., 1999. Mother-to-child
409 transmission of toxoplasmosis: risk estimates for clinical counselling. *Lancet* 353, 1829-
410 1833.
- 411 Duong, T.H., Dufillot, D., Martz, M., Richard-Lenoble, D., Kombila, M., 1992.
412 [Seroepidemiological study of toxoplasmosis in Libreville, Gabon]. *Ann Soc Belg Med*
413 *Trop* 72, 289-293.
- 414 El Mansouri, B., Rhajaoui, M., Sebti, F., Amarir, F., Laboudi, M., Bchitou, R., Hamad, M.,
415 Lyagoubi, M., 2007. [Sero prevalence of toxoplasmosis in pregnant women in Rabat,
416 Morocco]. *Bull Soc Pathol Exot* 100, 289-290.
- 417 Espeillac, D., Malavaud, S., Bessieres, M., Grandjean, H., 1989. Etude séroépidémiologique vis-
418 à-vis de la Toxoplasmose chez la femme enceinte dans la région toulousaine. *Médecine*
419 *et Maladies Infectieuses* 19, 80-82.

Partie 2 : Travaux scientifiques

- 420 Fallah, M., Rabice, S., Matini, M., Taherkhani, H., 2008. Seroepidemiology of toxoplasmosis in
421 primigravida women in Hamadan, Islamic Republic of Iran, 2004. *East Mediterr Health J*
422 14, 163-171.
- 423 Fallahi, S., Badparva, E., Mohammadi, M., Ebrahimzadeh, F., Pournia, Y., 2009.
424 Seroepidemiological study of *Toxoplasma gondii* in women referred to Khorramabad
425 laboratory of health center for medical examination before marriage, Lorestan province,
426 Iran. *Asian J. Biol. Sci.*
- 427 Faye, O., Leye, A., Dieng, Y., Richard-Lenoble, D., Diallo, S., 1998. [Toxoplasmosis in Dakar.
428 Seroepidemiologic sampling of 353 women of reproductive age]. *Bull Soc Pathol Exot*
429 91, 249-250.
- 430 Fiedler, K., Hulsse, C., Straube, W., Briese, V., 1999. [Toxoplasmosis-antibody seroprevalence
431 in Mecklenburg-Western Pomerania]. *Zentralbl Gynakol* 121, 239-243.
- 432 Foulon, W., Naessens, A., Volckaert, M., Lauwers, S., Amy, J.J., 1984. Congenital
433 toxoplasmosis: a prospective survey in Brussels. *Br J Obstet Gynaecol* 91, 419-423.
- 434 Grimard, D., 1998. Prévalence et incidence chez la femme enceinte de la région du Saguenay.
435 Université Laval, Laval.
- 436 Guessous-Idrissi, N., Lahlou, D., Sefiani, R., Benmira, A., 1984. [Toxoplasmosis and rubella in
437 Moroccan women. Results of a serological survey]. *Pathol Biol (Paris)* 32, 761-765.
- 438 Guimaraes, A.C., Kawarabayashi, M., Borges, M.M., Tolezano, J.E., Andrade Junior, H.F.,
439 1993. Regional variation in toxoplasmosis seronegativity in the Sao Paulo metropolitan
440 region. *Rev Inst Med Trop Sao Paulo* 35, 479-483.
- 441 Hofhuis, A., W, V.A.N.P., YT, V.A.N.D., Nijhuis, C.D., Mollema, L., FR, V.D.K., Havelaar,
442 A.H., Kortbeek, L.M., 2010. Decreased prevalence and age-specific risk factors for
443 *Toxoplasma gondii* IgG antibodies in The Netherlands between 1995/1996 and
444 2006/2007. *Epidemiol Infect*, 1-9.
- 445 Jenum, P.A., Kapperud, G., Stray-Pedersen, B., Melby, K.K., Eskild, A., Eng, J., 1998.
446 Prevalence of *Toxoplasma gondii* specific immunoglobulin G antibodies among pregnant
447 women in Norway. *Epidemiol Infect* 120, 87-92.
- 448 Jones, J.L., Kruszon-Moran, D., Wilson, M., 2003. *Toxoplasma gondii* infection in the United
449 States, 1999-2000. *Emerg Infect Dis* 9, 1371-1374.
- 450 Julvez, J., Magnaval, J.F., Baixench, M.T., Maron, I., 1994. [Seroepidemiology of toxoplasmosis
451 on Mayotte (Comores Archipelago)]. *Med Trop (Mars)* 54, 343-344.
- 452 Kako, M., Kondo, Y., Seto, M., Shoda, M., Sato, K., 1993. Prevalence of anti-*Toxoplasma* and
453 hygienic problems among unmarried young women. 14, 165-169.
- 454 Kaufman, L., Rousseeuw, P.J., 2008. Finding groups in data: An introduction to cluster analysis.
455 John Wiley & Sons, Hoboken.
- 456 Khoshnood, B., De Vigan, C., Goffinet, F., Leroy, V., 2007. Prenatal screening and diagnosis of
457 congenital toxoplasmosis: a review of safety issues and psychological consequences for
458 women who undergo screening. *Prenat Diagn* 27, 395-403.
- 459 Kolbekova, P., Kourbatova, E., Novotna, M., Kodym, P., Flegr, J., 2007. New and old risk-
460 factors for *Toxoplasma gondii* infection: prospective cross-sectional study among
461 military personnel in the Czech Republic. *Clin Microbiol Infect* 13, 1012-1017.
- 462 Kortbeek, L.M., De Melker, H.E., Veldhuijzen, I.K., Conyn-Van Spaendonck, M.A., 2004.
463 Population-based *Toxoplasma* seroprevalence study in The Netherlands. *Epidemiol Infect*
464 132, 839-845.
- 465 Lago, E.G., Carvalho, R.L.D., Jungblut, R., Da Silva, V.B., Fiori, R.M., 2009. Screening for
466 *Toxoplasma gondii* antibodies in 2,513 consecutive parturient women and evaluation of
467 newborn infants at risk for congenital toxoplasmosis. *Scientia Medica* 19, 27-34.
- 468 Leao, P.R.D., Meirelles Fihlo, J., Medeiros, S.F.d., 2004. Toxoplasmosis: Seroprevalence in
469 Postpartum Women Attended by SUS (Brazilian Public Health System). *Rev. Bras.*
470 *Ginecol. Obstet.* 26, 627-632.

Chapitre 7 - Utilisation des données issues de la littérature scientifique : Application à l'étude de la séroprévalence à *Toxoplasma gondii* dans les populations humaines

- 471 Lebech, M., Larsen, S.O., Petersen, E., 1993. Prevalence, incidence and geographical
472 distribution of *Toxoplasma gondii* antibodies in pregnant women in Denmark. *Scand J*
473 *Infect Dis* 25, 751-756.
- 474 Lee, Y.H., Noh, H.J., Hwang, O.S., Lee, S.K., Shin, D.W., 2000. Seroepidemiological study of
475 *Toxoplasma gondii* infection in the rural area Okcheon-gun, Korea. *Korean J Parasitol*
476 38, 251-256.
- 477 Ljungstrom, I., Gille, E., Nokes, J., Linder, E., Forsgren, M., 1995. Seroepidemiology of
478 *Toxoplasma gondii* among pregnant women in different parts of Sweden. *Eur J*
479 *Epidemiol* 11, 149-156.
- 480 Luyasu, V., Robert, A., Lissenko, D., Bertrand, M., Bohy, E., Wacquez, M., De Bruyere, M.,
481 1997. A seroepidemiological study on toxoplasmosis. *Acta Clin Belg* 52, 3-8.
- 482 Martín-Hernández, I., García-Izquierdo, S.M., 2003. Prevalencia de anticuerpos IgG contra
483 *Toxoplasma gondii* en donantes de sangre cubanos. *Rev Biomed* 14, 247-251.
- 484 Mohamed, K., Rayah, I.E., Bilal, A., Gumaa, E.T., Magid, A.A., Maly, M., Kodym, P., 2009.
485 Immune-diagnosis of latent toxoplasmosis in childbearing age women in rural areas in
486 EL Geizera State, Sudan. *Int. J. Med. Med.Sci.* 1, 272-277.
- 487 Morris, A., Croxson, M., 2004. Serological evidence of *Toxoplasma gondii* infection among
488 pregnant women in Auckland. *N Z Med J* 117, U770.
- 489 Morvan, J.M., Mambely, R., Selekon, B., Coumanzi-Malo, M.F., 1999. [Toxoplasmosis at the
490 Pasteur Institute of Bangui, Central African Republic (1996-1998): serological data]. *Bull*
491 *Soc Pathol Exot* 92, 157-160.
- 492 Nabias, R., Ngouamizokou, A., Migot-Nabias, F., Mbou-Moutsimbi, R.A., Lansoud-Soukate, J.,
493 1998. [Serological investigation of toxoplasmosis in patients of the M.I.P. center of
494 Franceville (Gabon)]. *Bull Soc Pathol Exot* 91, 318-320.
- 495 Nabli, B., 1977. Séroépidémiologie de la toxoplasmose humaine et animale en Tunisie. *Arch.*
496 *Inst. Pasteur Tunis* 54, 39-50.
- 497 Olusi, T.A., Ajaya, J.A., Makinde, A.A., 1994. Antibodies to *Toxoplasma gondii* in a rat-eating
498 population on Benue State, Nigeria. *Ann Trop Med Parasitol* 88, 217-218.
- 499 Pappas, G., Roussos, N., Falagas, M.E., 2009. Toxoplasmosis snapshots: global status of
500 *Toxoplasma gondii* seroprevalence and implications for pregnancy and congenital
501 toxoplasmosis. *Int J Parasitol* 39, 1385-1394.
- 502 Paradisi, F., Bartoloni, A., Aquilini, D., Roselli, M., Nunez, L.E., Manzone, G., De Majo, E.,
503 Parri, F., 1989. Serological survey of toxoplasmosis in the Santa Cruz region of Bolivia.
504 *Trans R Soc Trop Med Hyg* 83, 213-214.
- 505 Pinlaor, S., Ieamviteevanich, K., Pinlaor, P., Maleewong, W., Pipitgool, V., 2000.
506 Seroprevalence of specific total immunoglobulin (Ig), IgG and IgM antibodies to
507 *Toxoplasma gondii* in blood donors from Loei Province, Northeast Thailand. *Southeast*
508 *Asian J Trop Med Public Health* 31, 123-127.
- 509 Punda-Polic, V., Tonkic, M., Capkun, V., 2000. Prevalence of antibodies to *Toxoplasma gondii*
510 in the female population of the County of Split Dalmatia, Croatia. *Eur J Epidemiol* 16,
511 875-877.
- 512 R Development Core Team, 2010. R: A Language and Environment for Statistical Computing.
513 In, R Foundation for Statistical Computing, Vienna, Austria.
- 514 Rai, S.K., Shibata, H., Sumi, K., Rai, G., Rai, N., Manandhar, R., Gurung, G., Ono, K., Uga, S.,
515 Matsuoka, A., Shrestha, H.G., Matsumura, T., 1998. *Toxoplasma* antibody prevalence in
516 Nepalese pregnant women and women with bad obstetric history. *Southeast Asian J Trop*
517 *Med Public Health* 29, 739-743.
- 518 Ramsewak, S., Gooding, R., Ganta, K., Seepersadsingh, N., Adesiyun, A.A., 2008.
519 Seroprevalence and risk factors of *Toxoplasma gondii* infection among pregnant women
520 in Trinidad and Tobago. *Rev Panam Salud Publica* 23, 164-170.

Partie 2 : Travaux scientifiques

- 521 Remington JS, McLeod R, G, D., 1996. Toxoplasmosis. In: J. S. Remington and J. O. Klein
522 (Ed.), *Infectious disease of the fetus and newborn infant*, WB Saunders, Philadelphia, pp.
523 140-268.
- 524 Rodier, M.H., Berthonneau, J., Bourgoin, A., Giraudeau, G., Agius, G., Burucoa, C., Hekpazo,
525 A., Jacquemin, J.L., 1995. Seroprevalences of *Toxoplasma*, malaria, rubella,
526 cytomegalovirus, HIV and treponemal infections among pregnant women in Cotonou,
527 Republic of Benin. *Acta Trop* 59, 271-277.
- 528 Salahi-Moghaddam, A., Hafizi, A., 2009. A serological study on *Toxoplasma gondii* infection
529 among people in south of Tehran, Iran. *Korean J Parasitol* 47, 61-63.
- 530 Sharifi-Mood, B., Hashemi-Shahri, M., Salehi, M., Naderi, M., Naser-poor, T., 2006.
531 Seroprevalence of *Toxoplasma* Infection in the Pregnant Women in Zahedan,
532 Southeast of Iran. *J. Res. Health. Sci.* 6, 1-3.
- 533 Signorell, L.M., Seitz, D., Merkel, S., Berger, R., Rudin, C., 2006. Cord blood screening for
534 congenital toxoplasmosis in northwestern Switzerland, 1982-1999. *Pediatr Infect Dis J*
535 25, 123-128.
- 536 Sobral, C.A., Amendoeira, M.R., Teva, A., Patel, B.N., Klein, C.H., 2005. Seroprevalence of
537 infection with *Toxoplasma gondii* in indigenous Brazilian populations. *Am J Trop Med*
538 *Hyg* 72, 37-41.
- 539 Studenicova, C., Bencaiova, G., Holkova, R., 2006. Seroprevalence of *Toxoplasma gondii*
540 antibodies in a healthy population from Slovakia. *Eur J Intern Med* 17, 470-473.
- 541 Sundar, P., Mahadevan, A., Jayshree, R.S., Subbakrishna, D.K., Shankar, S.K., 2007.
542 *Toxoplasma* seroprevalence in healthy voluntary blood donors from urban Karnataka.
543 *Indian J Med Res* 126, 50-55.
- 544 Svobodova, V., Literak, I., 1998. Prevalence of IgM and IgG antibodies to *Toxoplasma gondii* in
545 blood donors in the Czech Republic. *Eur J Epidemiol* 14, 803-805.
- 546 Tabbara, K.S., Saleh, F., 2005. Serodiagnosis of toxoplasmosis in Bahrain. *Saudi Med J* 26,
547 1383-1387.
- 548 Tenter, A.M., Heckeroth, A.R., Weiss, L.M., 2000. *Toxoplasma gondii*: from animals to humans.
549 *Int J Parasitol* 30, 1217-1258.
- 550 Terazawa, A., Muljono, R., Susanto, L., Margono, S.S., Konishi, E., 2003. High *Toxoplasma*
551 antibody prevalence among inhabitants in Jakarta, Indonesia. *Jpn J Infect Dis* 56, 107-
552 109.
- 553 Tonkic, M., Punda-Polic, V., Sardelic, S., Capkun, V., 2002. [Occurrence of *Toxoplasma gondii*
554 antibodies in the population of Split-Dalmatia County]. *Lijec Vjesn* 124, 19-22.
- 555 Triolo-Mieses, M., Traviezo-Valles, L., 2006. Seroprevalencia de anticuerpos contra
556 *Toxoplasma gondii* en gestantes del municipio Palavecino, estado Lara, Venezuela.
557 *Kasmera* 34, 07-13.
- 558 Valcavi, P.P., Natali, A., Soliani, L., Montali, S., Dettori, G., Cheezi, C., 1995. Prevalence of
559 anti-*Toxoplasma gondii* antibodies in the population of the area of Parma (Italy). *Eur J*
560 *Epidemiol* 11, 333-337.
- 561 Varela, I.C.D., Loyola, C.N.P., Borges, A.G., 2009. Comportamiento de la infección
562 toxoplásmica. *Rev. Elect. Cien. Med. Cien.* 7.
- 563 Velasco-Castrejón, O., Salvatierra-Izaba, B., Valdespino, J.L., Sedano-Lara, A.M., Galindo-
564 Virgent, S., Magos, C.L., et al, 1992. Seroprevalencia de la toxoplasmosis en México.
565 *Salud Publica Mex.* 34, 222-229.
- 566 Villena I, Ancelle T, Delmas C, Garcia P, Brezin AP, Thulliez P, Wallon M, King L, V, G.,
567 2010. Toxosurv network and National Reference Centre for Toxoplasmosis. Congenital
568 toxoplasmosis in France in 2007: first results from a national surveillance system. *Euro*
569 *Surveill* 15.
- 570 Wallon, M., Liou, C., Garner, P., Peyron, F., 1999. Congenital toxoplasmosis: systematic review
571 of evidence of efficacy of treatment in pregnancy. *BMJ* 318, 1511-1514.

Chapitre 7 - Utilisation des données issues de la littérature scientifique : Application à l'étude de la séroprévalence à *Toxoplasma gondii* dans les populations humaines

- 572 Wittersheim, P., Candolfi, E., Kien, T., 1994. Prévalence de la toxoplasmose humaine à
573 Strasbourg. *J. Méd. Strasbourg* 25, 171-174.
- 574 Yacoub, A.A., Bakr, S., Hameed, A.M., Al-Thamery, A.A., Fartoci, M.J., 2006.
575 Seroepidemiology of selected zoonotic infections in Basra region of Iraq. *East Mediterr*
576 *Health J* 12, 112-118.
577
578

Partie 2 : Travaux scientifiques

579 **Table legends**

580

581 Table 1: Characteristics of the 81 studies included in the analysis

582

583 Table 2: Main characteristics of the 4 identified groups

584

585

586 **Figure legends**

587

588 Figure 1: Flow diagram of study selection concerning toxoplasmosis seroprevalence

589

590 Figure 2: Clustering tree of the 81 studied populations with a 4 groups visualization

591

592 Figure 3: Mean evolution of toxoplasmosis seroprevalence and confident intervals of each group

Chapitre 7 - Utilisation des données issues de la littérature scientifique : Application à l'étude de la séroprévalence à *Toxoplasma gondii* dans les populations humaines

593
594

Table 1: Characteristics of the 81 studies included in the analysis

Author	Country	Period of sampling	Studied population	World region	Number of samples tested	Name	References
Ohisi	Nigeria	1994	General population	Africa	978	Nigeria	(Ohisi et al., 1994)
Dromigny	Madagascar	1990-1995	General population	Africa	2354	Madagascar	(Dromigny et al., 1996)
Dumas	Dem Rep of Congo	1991	General population	Africa	2539	Dem Rep Cong.	(Dumas et al., 1990)
Dumas	Togo	1985	General population	Africa	277	Togo	(Dumas et al., 1985)
Duong	Gabon	1992	General population	Africa	2324	Gabon 2	(Duong et al., 1992)
Faye	Senegal	1993	Child-bearing age	Africa	353	Senegal	(Faye et al., 1998)
Guessous-Idrissi	Morocco	1982	Pregnant women	Africa	200	Morocco 2	(Guessous-Idrissi et al., 1984)
Jalvez	Comoros (Mayotte)	1984	General population	Africa	871	Mayotte	(Jalvez et al., 1994)
Mohamed	Sudan	2009	Child-bearing age	Africa	255	Sudan	(Mohamed et al., 2009)
El Mansouri	Morocco	2007	Pregnant women	Africa	2456	Morocco 1	(El Mansouri et al., 2007)
Morvan	Central African Republic	1996-1998	General population	Africa	1953	Centr Afr Rep	(Morvan et al., 1999)
Nabias	Gabon	1995-1997	Pregnant women	Africa	767	Gabon 1	(Nabias et al., 1998)
Nabli	Tunisia	1977	General population	Africa	1060	Tunisia	(Nabli, 1977)
Rodier	Benin	1993	Pregnant women	Africa	211	Benin	(Rodier et al., 1995)
Salahi-Moghaddam	Iran	2009	General population	Asia	1187	Iran 3	(Salahi-Moghaddam and Hafizi, 2009)
Abu-Madi	Qatar	2005-2007	General population	Asia	1625	Qatar	(Abu-Madi et al., 2008)
Al-Qurashi	Saudi Arabia	2000	General population	Asia	1464	Saudi Arabia	(Al-Qurashi et al., 2001)
Dar	United Arab Emirates	1997	Pregnant women	Asia	1503	Unit Arab Emir.	(Dar et al., 1997)
Fallah	Iran	2004	Pregnant women	Asia	576	Iran 1	(Fallah et al., 2008)
Fallah	Iran	2008	Child-bearing age	Asia	465	Iran 4	(Fallah et al., 2009)
Lee	Korea	2000	General population	Asia	1119	Korea	(Lee et al., 2000)
Pinlaor	Thailand	1997	Blood donors	Asia	345	Thailand	(Pinlaor et al., 2000)
Rai	Nepal	1994-1995	General population	Asia	656	Nepal	(Rai et al., 1998)
Sharifi-Mood	Iran	2004-2005	Female population	Asia	980	Iran 2	(Sharifi-Mood et al., 2006)
Sundar	India	2004-2005	Blood donors	Asia	1000	India	(Sundar et al., 2007)
Tabbara	Bahrain	2000-2003	General population	Asia	4739	Bahrain	(Tabbara and Saleh, 2005)
Yacoub	Irak	2006	General population	Asia	419	Irak	(Yacoub et al., 2006)
Ahlfors	Sweden	1982-1983	Pregnant women	Europe	4351	Sweden 2	(Ahlfors et al., 1989)
Allain	United kingdom	1992	Pregnant women	Europe	13 328	Unit kingdom	(Allain et al., 1998)
Ancelle	France	1995	Pregnant women	Europe	12928	France 4	(Ancelle et al., 1995)
Bartolome Alvarez	Spain	2001-2007	Child-bearing age	Europe	2623	Spain	(Bartolome Alvarez et al., 2008)
birgisdottir	Estonia	1990-1994	General population	Europe	282	Estonia	(Birgisdottir et al., 2006)
birgisdottir	Iceland	1990-1994	General population	Europe	503	Iceland	(Birgisdottir et al., 2006)
birgisdottir	Sweden	1990-1994	General population	Europe	492	Sweden 1	(52)
bobic	Serbia	1988-1991	Child-bearing age	Europe	1157	Serbia	(Bobic et al., 2003)
Condorelli	Italy	1987-1990	General population	Europe	2993	Italy 1	(Condorelli et al., 1993)
Crucerescu	Romania	2003-2004	General population	Europe	424	Romania	(Crucerescu et al., 2006)
Diza	Greece	2004	General population	Europe	958	Greece	(Diza et al., 2005)
Espaillet	France	1980-1983-1986	Pregnant women	Europe	8829	France 3	(Espaillet et al., 1989)
Fiedler	Germany	1994-1996	General population	Europe	8397	Germany	(Fiedler et al., 1999)
Foulon	Belgium	1979-1982	Pregnant women	Europe	2986	Belgium 1	(Foulon et al., 1984)
invz	France	2003	Pregnant women	Europe	15108	France 1	(Berger et al., 2009)
Jenm	Norway	1992-1993	Pregnant women	Europe	35940	Norway	(Jenm et al., 1998)
Kolbekova	Czech Republic	2000-2004	General population	Europe	3290	Czech Rep. 1	(Kolbekova et al., 2007)
Kortbeek	Netherlands	1995	General population	Europe	9717	Netherlands	(Kortbeek et al., 2004)
Lebech	Denmark	1990	Pregnant women	Europe	5402	Denmark	(Lebech et al., 1993)
Ljungstrom	Sweden	1987	Pregnant women	Europe	467	Sweden 3	(Ljungstrom et al., 1995)
Ljungstrom	Sweden	1987	Pregnant women	Europe	937	Sweden 4	(Ljungstrom et al., 1995)
Ljungstrom	Sweden	1987	Pregnant women	Europe	1413	Sweden 5	(Ljungstrom et al., 1995)
Ljungstrom	Sweden	1987	Pregnant women	Europe	837	Sweden 6	(Ljungstrom et al., 1995)
Luyasu	Belgium	1991	Pregnant women	Europe	830	Belgium 2	(Luyasu et al., 1997)
Punda-Polic	Croatia	1994-1995	Female population	Europe	1109	Croatia 1	(Punda-Polic et al., 2000)
Signorell	Switzerland	1999	Pregnant women	Europe	24950	Switzerland	(Signorell et al., 2006)
Studenicova	Slovakia	2003	General population	Europe	508	Slovakia	(Studenicova et al., 2006)
Svobodova	Czech Republic	1992-1993	Blood donors	Europe	663	Czech Rep. 2	(Svobodova and Literak, 1998)
Tonkic	Croatia	1994-1996	General population	Europe	4022	Croatia 2	(Tonkic et al., 2002)
Valcavi	Italy	1987-1991	General population	Europe	28247	Italy 2	(Valcavi et al., 1995)
Wittersheim	France	1994	General population	Europe	1600	France 2	(Wittersheim et al., 1994)
Asthana	Grenada	1995	Pregnant women	N. & C. America	534	Grenada	(Asthana et al., 2006)
Grimard	Canada	1994-1995	Pregnant women	N. & C. America	2224	Canada	(Grimard, 1998)
Jones	United States	1999-2000	General population	N. & C. America	4875	Unit. States	(Jones et al., 2003)
Martin-Hernández	Cuba	2000	Blood donors	N. & C. America	922	Cuba	(Martin-Hernández and Garcia-Izquierdo, 2003)
Ramsewak	Trinidad and Tobago	2002-2003	Pregnant women	N. & C. America	450	Trinidad Tobago	(Ramsewak et al., 2008)
Velasco-Castrejon	Mexico	1985-1991	General population	N. & C. America	29 935	Mexico	(Velasco-Castrejon et al., 1992)
Boia	Brazil	2008	General population	S. America	230	Brazil 1	(Boia et al., 2008)
Chacin-Bonilla	Venezuela	2001	General population	S. America	447	Venezuela 1	(Chacin-Bonilla et al., 2001)
Chacin-bonilla	Venezuela	2001	General population	S. America	335	Venezuela 2	(Chacin-Bonilla et al., 2003)
Contreras	Chile	1982-1996	General population	S. America	76317	Chile	(Contreras et al., 1996)
Leao	Brazil	2000	Pregnant women	S. America	205	Brazil 3	(Leao et al., 2004)
Lago	Brazil	2002-2009	Pregnant women	S. America	2513	Brazil 5	(Lago et al., 2009)
Guimaraes	Brazil	1989-1990	Child-bearing age	S. America	1286	Brazil 4	(Guimaraes et al., 1993)
Avelino	Brazil	1997-1999	Child-bearing age	S. America	2242	Brazil 2	(Avelino et al., 2004)
Paradis	Bolivia	1989	General population	S. America	446	Bolivia	(Paradis et al., 1989)
Sobral	Venezuela	2000-2001	General population	S. America	302	Venezuela 5	(Sobral et al., 2005)
Sobral	Venezuela	2000-2001	General population	S. America	568	Venezuela 6	(Sobral et al., 2005)
Triolo-Mises	Venezuela	2001-2004	Pregnant women	S. America	446	Venezuela 3	(Triolo-Mises and Travezio-Valles, 2006)

Partie 2 : Travaux scientifiques

Varela	Venezuela	2006	General population	S. America	343	Venezuela 4	(Varela et al., 2009)
Breurec	New Caledonia	2000	Child-bearing age	S. W. Pacific	2416	New Caledonia	(Breurec et al., 2004)
Kako	Japan	1993	General population	S. W. Pacific	1402	Japan	(Kako et al., 1993)
Morris	New Zealand	2000	Pregnant women	S. W. Pacific	500	New Zealand	(Morris and Croxson, 2004)
Terazawa	Indonesia	2003	General population	S. W. Pacific	1693	Indonesia	(Terazawa et al., 2003)

595

596

597

598

Table 2: Main characteristics of the 4 identified groups

Group Number	Number of studies	Mean seroprevalence (95% CI)	Mean percentage Δ SP ¹⁵⁻⁴⁰ (95% CI) *	Relative risks **
1 (Linear Low)	28	22 (18-25)	13 (10-16)	1.0
2 (High Curved)	8	76 (66-85)	13 (4-23)	1 (0.3-1.8)
3 (Linear High)	15	41 (38-44)	20 (13-28)	1.5 (1.0-2.2)
4 (Low Curved)	30	58 (56-61)	20 (15-25)	1.5 (1.2-1.9)

599

600

601

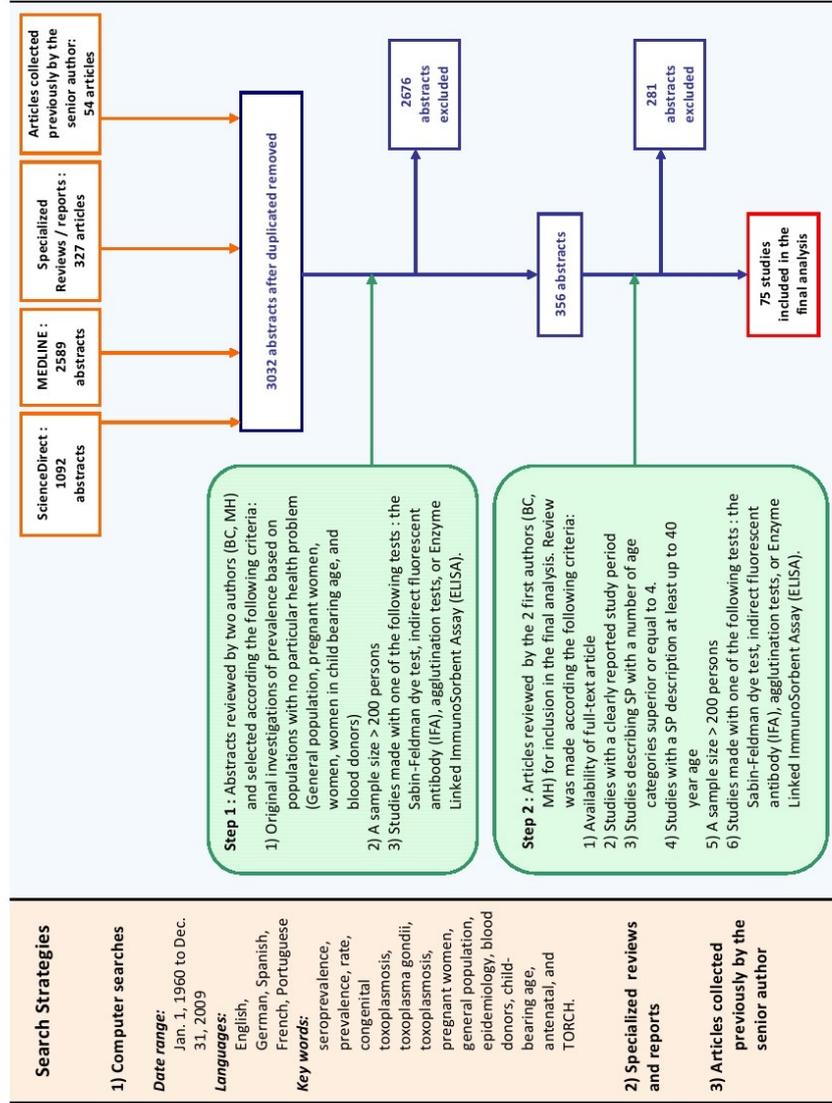
602

603

* Percentage of women who seroconvert during the childbearing age (15-40 years)

** Relative risks were calculated using as the reference, the group 1 which had the weaker percentage of seroconversions during child bearing age.

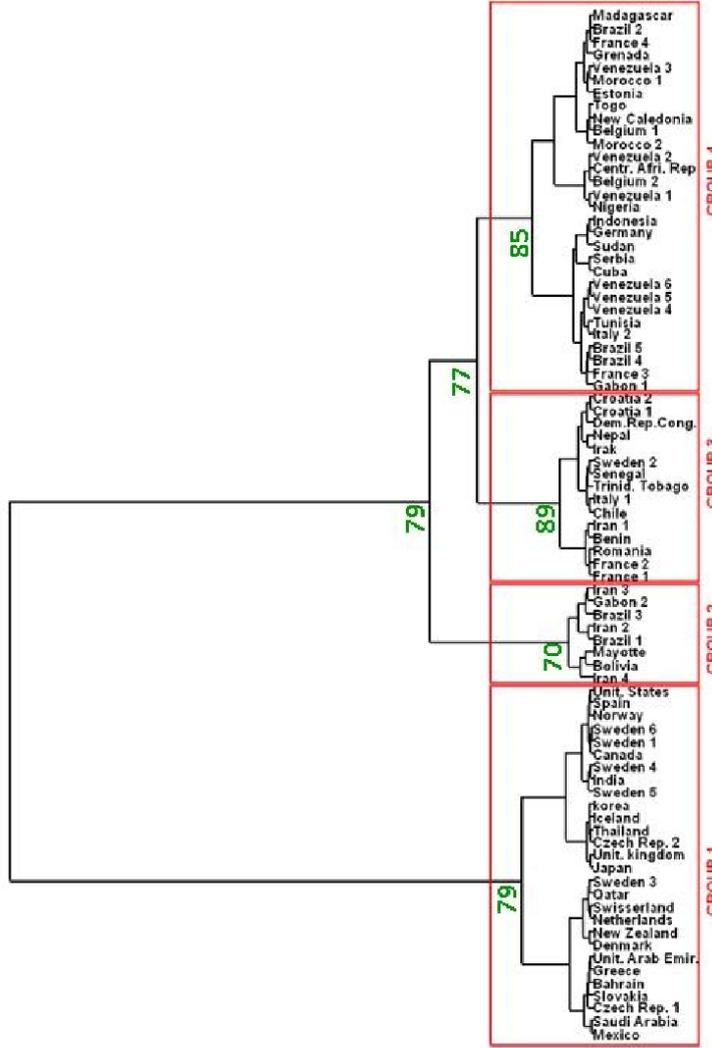
Figure 1: Flow diagram of study selection concerning toxoplasmosis seroprevalence



604
605

606
607

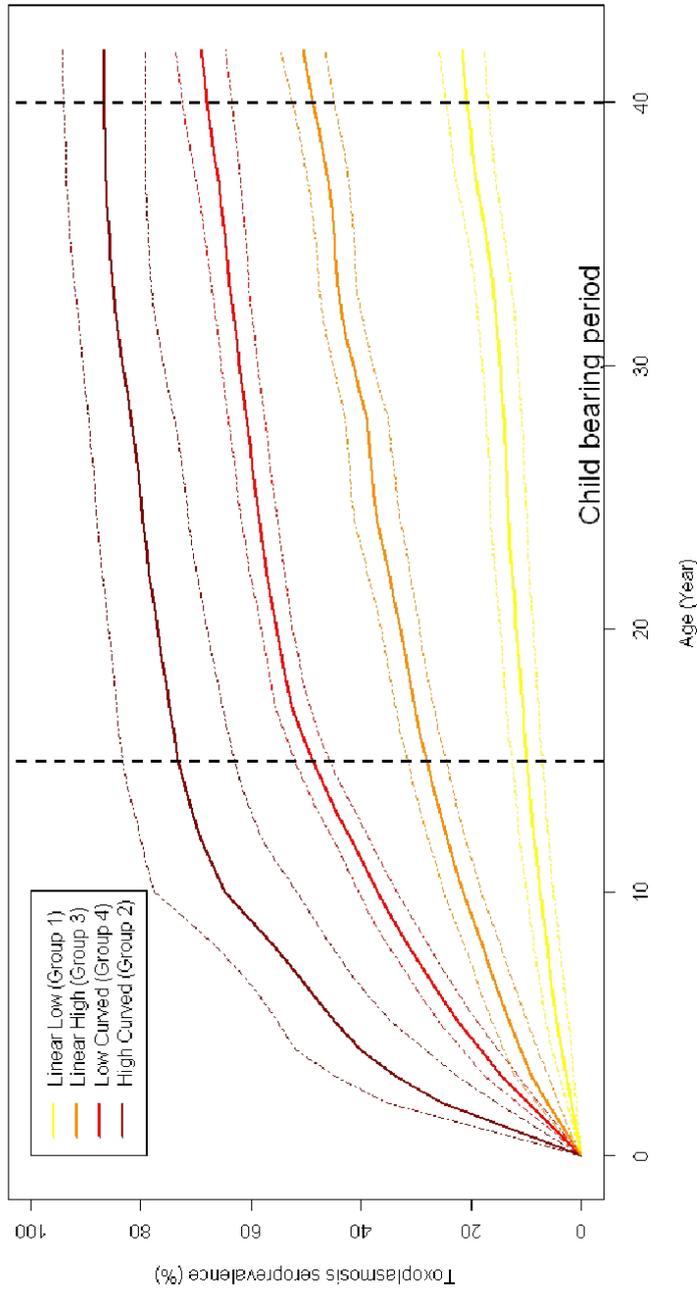
Figure 2: Clustering tree of the 81 studied populations with a 4 groups visualization *.



609 * A hierarchical agglomerative clustering method was used, with Ward's minimum variance method and Euclidean distance. Cluster analysis was performed on the 40 variables
 610 describing the level of SP for ages from 1 to 40. The number of clusters was determined by using the dendrogram: the squared multiple correlations R2, were then calculated. The
 611 number of clusters was then decided based on two criteria: a sufficient proportion of variance accounted for by the clusters and large enough groups to be further studied. A bootstrap
 612 analysis was conducted to determine the confidence cut off for each branch. For the main clusters, approximately unbiased p-values (in green) were calculated via multiscale
 613 bootstrap resampling.
 614

608

Figure 3: Mean evolution of toxoplasmosis seroprevalence and confidence intervals of each group



615
616

617

7.6. Etude 9 : Utilisation de l'environnement géographique et sociétal pour prédire le niveau de séroprévalence à *Toxoplasma gondii* et le risque de toxoplasmose congénitale associé

Introduction : Le premier article a montré que le risque de toxoplasmose congénitale pouvait être évalué dans une population en utilisant le niveau de séroprévalence global. En fonction de ce risque, des programmes de prévention nationaux peuvent être mis en place. Malheureusement, un certain nombre de pays ne disposent pas actuellement de données à jour permettant d'évaluer le niveau de séroprévalence de leurs populations. Cette étude vise à identifier les grands facteurs socio environnementaux expliquant l'hétérogénéité de la répartition mondiale de la séroprévalence afin d'évaluer et de prédire le risque de toxoplasmose congénitale dans des zones géographiques où aucune donnée ne sont disponibles.

Matériels/Méthodes : Une revue de la littérature recensant les études décrivant la séroprévalence globale de populations a d'abord été faite selon une méthodologie stricte. Après couplage avec des variables sociodémographiques nationales, une approche en deux étapes a été utilisée pour modéliser la relation entre la séroprévalence de la toxoplasmose et les variables sélectionnées. Une analyse par forêts aléatoires a été utilisée pour estimer et classer par ordre d'importance les variables explicatives. Une analyse par arbre de régression a été appliquée ensuite pour examiner la nature de la relation entre le taux de séroprévalence et les variables précédemment sélectionnées. En utilisant le modèle final, une carte mondiale du niveau de séroprévalence a finalement été tracée.

Résultats : 204 études décrivant le niveau de séroprévalence de populations ont finalement pu être incluses dans cette analyse. En utilisant les arbres de régression, les interactions entre le climat, la religion majoritaire du pays, l'origine anglo-saxonne du pays et la zone géographique sont apparues comme étant explicatives de la répartition mondiale du niveau de séroprévalence ($p < 0.01$).

Conclusions : Malgré certaines limitations, cette étude est la première à proposer un modèle permettant l'obtention d'une carte mondiale de la séroprévalence à *Toxoplasma gondii*. Quatre zones géographiques apparaissent comme étant particulièrement à risque pour la toxoplasmose congénitale : L'Europe de l'Ouest, l'Afrique du Nord, le Moyen Orient et la partie Ouest de l'Asie Centrale. Des données complémentaires sont cependant nécessaires dans les pays appartenant à ces zones géographiques pour l'obtention d'une vision claire de leur risque national face à la toxoplasmose congénitale.

En cours de soumission à International Journal for Parasitology

1 **Approaching the seroprevalence level and the risk of congenital**
2 **toxoplasmosis from geographical and societal environments: an**
3 **analytical review**

4
5 Matthieu Hanf *^ε, Mathieu Nacher*^ε, Antoine Adenis*, Magalie Demar^ε,
6 Bernard Carme*^ε

7
8 *: Centre d'Investigation Clinique Epidémiologie Clinique Antilles Guyane CIC-EC
9 INSERM CIE 802, Cayenne General Hospital, Cayenne, French Guiana.

10 ^ε: Equipe EA3593, Cayenne General Hospital and UFR Médecine, Université des
11 Antilles et de la Guyane, Cayenne, French Guiana

12
13 **Corresponding author :**

14 Bernard Carme, carme.bernard@wanadoo.fr, Tel -- 594 594 39 53 01, Fax : --
15 594 594 39 50 51, LHUPM Centre Hospitalier de Cayenne, BP 6006, Rue des
16 Flamboyants, F-97306 Cayenne (French Guiana)

17
18 **Key words:** Toxoplasmosis, seroprevalence, congenital toxoplasmosis, risk
19 factors, prediction

20
21 **Acknowledgments:** Funding: Cayenne General Hospital (French Guiana) and
22 INSERM: Institut National de la Santé et de la Recherche Médicale.

23
24 **No competing financial interest**

25
26

27 **SUMMARY:**

28

29 **Background:**

30 The risk of congenital toxoplasmosis in a given population can be approached by
31 the combination of its overall level of seroprevalence and the aspect of the
32 curve throughout age groups. Prevention programs may be proposed depending on
33 the risk level. But a number of countries do not have available data.

34 **Objectives:**

35 Through an analytical review, this ecological study aimed to identify variables
36 that could explain the global pattern of toxoplasmosis seroprevalence in view to
37 be able to roughly estimate seroprevalence levels for geographic areas where no
38 data was currently available.

39 **Methods:**

40 A review of the literature was performed to compile seroprevalence data in the
41 world and available geographical and societal variables measured at a country
42 scale which could potentially explain it. Random forest analyses were first
43 implemented to rank the importance of predictors and unbiased conditional
44 inference trees (RTA) were then used to examine the nature of relationships
45 between the response and important predictor variables. Using this model,
46 estimation of the current world distribution of toxoplasmosis seroprevalence
47 was finally made.

48 **Results:**

49 204 populations where toxoplasmosis seroprevalence was available and
50 corresponding to selection criteria were finally included in the analysis.
51 According to RTA, Climate, religions, belonging to Anglosphere and study
52 geographical area appeared to significantly ($P < 0.01$) design the world
53 seroprevalence pattern.

54 **Conclusions:**

55 Although there are some limitations, this is the first model that leads to a world
56 map of the seroprevalence of toxoplasmosis. 4 geographical areas appeared to
57 be particularly at risk of CT: West Europe, North Africa, Middle East and West
58 part of Central Asia. Complementary data and country specificities regarding
59 toxoplasmosis in these high risk subgroups are needed to have a clear vision of
60 their national risk of congenital toxoplasmosis.

61

62

63 INTRODUCTION

64

65 In Europe, only 4 countries (France, Denmark, Germany and Italy) have
66 organized surveillance of congenital toxoplasmosis (CT) [1]. Such surveillance
67 does not exist elsewhere. Developed countries outside of Europe usually have
68 low toxoplasmosis seroprevalence (SP) levels throughout life, CT is not a
69 priority. Poor countries have other health priorities even when the risk of CT is
70 high. France appears like the country where surveillance is the most developed.
71 A national official programme of CT prevention exists since 1978. In addition to
72 the mandatory serologic screening in early pregnancy and to the informations on
73 the risk of primoinfection, non immune women are tested every month until
74 delivery since 1992. The objective is to detect seroconversions early in order to
75 implement diagnostic and therapeutic measures without delay. But the efficacy
76 of such a programme has only been recently evaluated at a national level with
77 the estimation of the number and severity of CT cases [2]. It is with objective
78 data on the efficacy of the implemented measures and on the SP level of women
79 of reproductive age that economical analyses destined to evaluate the best
80 screening measures become credible [3]. The example of France can only be
81 useful for countries that are geographically, economically and culturally
82 comparable. An evaluation of the interest of CT prevention programme at the
83 global scale is however necessary. A first approach applicable to all countries
84 should in priority analyze already existing data on concerned populations.
85 Globally, as shown in the previous review [4], this risk concerns mostly countries
86 or regions and/or communities presenting a high differential (>15%) between SP
87 at age 40 and 15 years of age ; this indicator being itself linked to a general SP
88 between 30 and 60%. This situation corresponds to the best ratio between the
89 pros and the cons of a national program against CT. But this is not possible for a
90 number of countries or regions with unknown SP levels.

91

92 Despite regular efforts to collect and describe available data on world
93 toxoplasmosis SP patterns and the associated risk factors [5,6,7], we are not
94 aware of any work attempting to clearly identify and quantify the driving forces
95 of these SP world patterns, their interactions, and explain the observed
96 geographical disparities. To do so, we therefore reviewed studies on SP of
97 toxoplasmosis throughout the world, and collected data on the environmental,
98 geographical, and societal aspects of their respective countries. An ecological
99 model was then constructed to merge together all these variables. Modeling the
100 relation between SP and the other covariates offers the possibility to roughly
101 estimate SP levels in countries where SP data is unavailable.

102

103 The objectives of this study were to 1) identify the main predictors of
104 world toxoplasmosis SP patterns, 2) to describe their interactions and
105 associated toxoplasmosis SP levels, 3) to construct, using data, a model able to
106 explain geographical disparities of toxoplasmosis SP, and finally draw a world
107 map of current toxoplasmosis SP and thus determining the areas where the risk
108 of CT is high.

109

110 **METHODS**

111

112 **Reviewing methodology of Toxoplasmosis SP studies**

113 In view to model the relation between SP and the environmental,
114 geographical, and societal variables, a large selection of published studies which
115 described toxoplasmosis SP in human populations was reviewed in the analysis.

116

117 The following search strategies were used for this systematic review of
118 the toxoplasmosis SP literature: 1) computer search of databases, 2) review of
119 specialized reports or already made toxoplasmosis SP reviews, and 3) Review of
120 literature collected previously by the senior author. As much as possible, this
121 search strategy tried to be in accordance with expert recommendations on the
122 subject.

123

124 *Computer search:* Toxoplasmosis SP studies were searched in MEDLINE and
125 SCIENCE DIRECT databases from January 1960 to December 2009 for articles
126 in English, German, French, Spanish, and Portuguese. Search words/terms were
127 as follows: seroprevalence, prevalence, rate, congenital toxoplasmosis,
128 toxoplasma gondii, toxoplasmosis, pregnant women, general population,
129 epidemiology, blood donors, child-bearing age, antenatal, and TORCH.

130

131 • Step 1: Abstracts were reviewed by one of the authors (M.H) and
132 selected for further review if they met the following 3 criteria: 1)
133 Original investigations of prevalence based on populations with no
134 particular health problem (General population, pregnant women, women in
135 child bearing age, and blood donors); 2) a sample size > 200 persons; 3)
136 Studies made with one of the following serological tests for toxoplasmosis
137 : the Sabin-Feldman dye test, indirect fluorescent antibody (IFA),
138 agglutination tests, or Enzyme Linked ImmunoSorbent Assay (ELISA). If
139 a criterion was not met because not enough information was provided, the
140 abstract was set aside for further evaluation.

141

142 •Step 2: If full-text articles of abstracts selected in Step 1 were
143 retrieved, they were then reviewed by one author. Studies were included

144 if they met all the following criteria: 1) Studies with a clearly reported
145 study period 2) A sample size > 200 persons 3) Studies made with one of
146 the following serological tests for toxoplasmosis: the Sabin-Feldman dye
147 test, indirect fluorescent antibody (IFA), agglutination tests, or Enzyme
148 Linked ImmunoSorbent Assay (ELISA). For non retrieved articles, if all
149 criteria described in step two were specified in the abstract and
150 respected, the study was finally include in the analysis. Inclusion was
151 based on the approbation of the senior investigator (B.C). For studies
152 describing several well identified geographical areas, a distinction was
153 made if each sub-studies met the inclusion criteria described in step 2.

154

155 ***Review of specialized reviews or reports:*** Four major reviews or reports were
156 reviewed to identify references to original surveys that might be incorporated
157 in the review processing [1, 7-9].

158

159 ***Review of literature collected previously by the senior author:*** all articles
160 were reviewed, and the full texts of potentially interesting studies were
161 examined.

162

163 **Toxoplasmosis SP study data**

164 Included studies were classified according the study period (<1980's,
165 1980's, 1990's, and 2000's), their world region localization (Europe, Asia, Africa,
166 South West Pacific, North and Central America, and South America) as defined
167 by the World meteorological classification and were defined as female
168 population or gender mixed population depending on the surveyed population. The
169 percentage of females was also registered when available. Furthermore, the
170 minimum, maximum and mean age of the population were recorded when available.

171 Study sites were also categorized as rural, mixed or urban based on the
172 description in the original study. The percentage of surveyed people living in
173 urban area was also used when available.

174 Studies of which localization were in countries of the world in which the
175 English language and cultural values predominate were grouped in a category we
176 called Anglosphere. Countries fitting this criterion were USA, Canada, United
177 Kingdom, Ireland, Australia, and New Zealand.

178

179 **Food consumption, Environmental and socio-economic data**

180 Climate, temperature, and precipitation were categorized according to the
181 revised Köppen-Geiger climate classification [8]. For studies defined on large
182 areas, the main climate was assigned. When the pattern of precipitations or
183 temperature was not clear, a non identified criterion (NI) was assigned.

184 Data on gross domestic product (GDP) per capita were included. This was
185 available at the national level from the UN statistics database [9] and the
186 average annual GDP per capita for the study period was calculated.

187 Water and sanitation coverage (as proportion of the population with
188 access to improved sources or services) were also used in this analysis. Data is
189 compiled by the WHO/UNICEF Joint Monitoring Programme [10], and is
190 available at the national level, divided into rural and urban settings. Despite the
191 uneven quality of this data before the 1980's, it was considered that the
192 influence of water and sanitation was of particular interest to describe
193 toxoplasmosis SP and was therefore included in the analysis.

194 Data on the main religions for each studied country were obtained
195 through the CIA World Factbook [11]. For each study, the main religion of the
196 country was categorized as follows: Christianity, Islam, others.

197 Country data on per capita meat and milk consumption (bovine meat, pork,
198 poultry meat, other meats, total meat, and milk) and data on crop foods
199 consumption (roots and tubercles, fruits, vegetables, milled rice, potatoes, and
200 cereals) were also included. This was available at the national level from the
201 World Food and agriculture Organization database [12]. The average annual
202 consumption in kg per capita for the study period was calculated.

203

204 *Statistical Analysis*

205

206 To summarize world distribution of toxoplasmosis, crude median SPs by
207 country were plotted on a world map using Geographical Information System
208 function of the R statistical language. Median values and distributions of SP
209 within gender, setting, climate and religion were boxplotted. For qualitative
210 predictors, the number of sampled studies and the associated percentages were
211 calculated by category. For quantitative predictors the mean, median, standard
212 deviations, minimum and maximum were calculated.

213 A two-stage approach was taken to model relationships between
214 toxoplasmosis SP and the selected predictors.

215 In order to limit the inherent reporting bias, we first used random forest
216 analysis (RFA) [13] to estimate and rank the importance of predictors. A
217 regression tree analysis (RTA) [14] was then used to examine the nature of
218 relationships between the response and important predictor variables. These
219 non-parametric methods are ideally suited for the analysis of high dimensional
220 ecological or epidemiological data with hierarchical and complex relationships
221 among the predictor variables (colinearity) and between predictor and response
222 variables.

223 RFA is an ensemble learning algorithm that averages predictions over
224 multiple bootstrapped regression trees; it consistently out-performs other
225 statistical methods for ranking predictor importance [15,16].

226 The party package within the R statistical language was used to implement
227 RFA [17]. This package provides a random forest implementation based on
228 unbiased trees that enables learning unbiased forests [18]. A random selection
229 of 1500 bootstrap samples each containing two-thirds of the data was made. To
230 guarantee unbiased variable selection, conditional permutation importance
231 measure was used to rank predictors in terms of the strength of their
232 relationship to the response [19].

233 After identifying important predictor variables with RFA, RTA was used
234 to better understand the nature of relationships between the 10 top-ranked
235 predictors and the response variable.

236 RTA is a non-parametric technique that recursively partitions a dataset
237 into subsets that are increasingly homogeneous with regard to the response [14].
238 RTA produces a set of decision rules on predictor variables that are easily
239 interpreted as a dendrogram. Most implementations of RTA have a selection bias
240 towards predictors with many possible splits and tend to overfit to a given
241 dataset by creating splits that do not significantly reduce the variance [20].
242 Trees are typically pruned back to include only those partitions assumed to be
243 valuable beyond the sample data. An implementation of RTA called unbiased
244 conditional inference trees was used. It requires a statistically significant
245 difference ($p < 0.01$), adjusted with Bonferroni correction, in order to partition
246 the data, which minimizes bias and prevents over-fitting and the need for
247 pruning [20].

248

249 RESULTS

250

251 *Overall patterns of world SP*

252 3032 abstracts published since 1960 were screened. A total of 356
253 articles were reviewed, and 139 studies were included in this systematic review.
254 These studies described 212 geographically well distinct populations. Only the
255 204 studied populations that did not have missing data for economic and food
256 variables were finally included in the study sample (Figure 1) (see annexe 1 for
257 references)

258 Study characteristics are summarized in table 1 for quantitative variables
259 and in table 2 for qualitative ones.

260 The overall SP in the sample was 39.3 % (minimum: 2%; maximum: 97.2%).
261 Figure 2 shows the crude world map of toxoplasmosis SP (median percentage)
262 reflecting the sampled studies and the country where they were performed.

263 Strong visual disparities in SP appeared when SP was studied by world
264 regions, religion, and climate. For gender and setting there were no obvious
265 differences (Figure 3).

266

267 ***SP model***

268 Climate was, by far, the most important predictor variable (Figure 4);
269 Religion was similarly important and was ranked second. Milled rice consumption,
270 precipitation, world region, GDP, cereal consumption, temperature, belonging to
271 the Anglosphere, and root/tubercle consumption were ranked third, fourth,
272 fifth, sixth, seventh, eighth, ninth and tenth and were also found to be related
273 to toxoplasmosis SP levels.

274 The first partition was based on climate: if the study climate belonged to
275 Arid/snow or polar climate, the level of SP was relatively low compared with
276 warm temperate and equatorial climates, where a strong disparity in
277 toxoplasmosis SP level was observed. In each of the 2 previous nodes defined by
278 climate the second partition was based on religion. For studies having an
279 arid/snow or polar climate, studies in countries where the main religion is Islam
280 had an intermediate SP level (mean: 30.4%). On the other branch, the level was
281 low for studies where Christian or others religions predominate (mean: 18.3%).
282 For studies located in warm temperate or tropical areas, populations where
283 others religions predominate were also characterized by a low level (mean:
284 17.5%).

285 The third node split the sample according to belonging to the Anglosphere
286 and assigned a low level of SP (mean: 17.7%) for studies made in these countries.
287 The fourth node used world regions to discriminate the sample studies. For
288 studies made outside Europe, a high SP was assigned (mean: 57.8%), although for
289 studies in Europe an intermediate SP level was found (mean: 41.8%). The model
290 mean error rate was 13%.

291

292 ***World map of toxoplasmosis SP predictions***

293 Using the previous model, religion and belonging to Anglosphere variables
294 were compiled and coupled with climate world repartition in order to predict the
295 current world distribution of toxoplasmosis SP. The resulting map is shown in
296 figure 5.

297

298 **DISCUSSION**

299

300 How to estimate the risk of toxoplasmosis? The main difficulty resides in
301 the uncertainty on the respective importance of foodborne contaminations, and
302 the direct telluric contamination by "dirty hands". Most precise studies on the
303 risk of toxoplasmosis agree on the predominance of the foodborne risk, without

304 being able to quantify it. In Europe and in the USA, the consumption of
305 undercooked meat is the main suspect. Cook et al estimates that 30 to 63% of
306 infections can be attributed to meat consumption, but this study did not take
307 into account the consumption of vegetables [21]. Far from Europe but still only
308 considering meat consumption, the New Zealand Food Safety Agency in 2002
309 attempted a Quantitative Risk Appreciation (QRA) in order to determine risk
310 profiles [22]. The authors concluded that this appreciation is not possible
311 because of the lack of data on food contamination.

312 In the present study, food consumption habits were less important
313 predictors of toxoplasmosis SP variations than were the climate. This result
314 underlines the fact that, at world scale, the variability in toxoplasmosis SP is
315 driven more by pathogen survival in the environment than by food habits or
316 sanitary conditions. In areas with polar/snow or arid climates a lower
317 toxoplasmosis SP was found compared to areas with warm temperate and
318 equatorial climates. As shown elsewhere, extreme environmental conditions
319 decrease the environmental resistance of *T. gondii* oocysts, and consecutively
320 the number of human infections [6,23].

321 A surprising result was that religion appeared to be a major predictor of
322 toxoplasmosis world SP levels. Explanations could be found in the fact that food
323 is an important part of religious practice and spiritual rituals for many faiths
324 including Christianity, Judaism, Islam, Hinduism and Buddhism. The role of food
325 in cultural practices and religious beliefs is complex and varies among individuals
326 and communities. In this study, others religions than Islam or Christianity,
327 whatever the climate, were associated with low SP levels. In the study sample,
328 others religions were essentially represented by Hinduism and Buddhism.

329 In these two religions, a vegetarian dietary is widespread. Knowing that
330 meat consumption is a known important transmission route for *Toxoplasma gondii*
331 [6], it was hypothesized that religious vegetarian diets could partly explain the
332 low SP found in populations practicing these religions.

333 Others differences between religions may be explained by different
334 dietary habits (method of cooking meat, hand washing, kinds of meat or
335 vegetables consumed, vegetable cleaning, level of meat contamination, ...).
336 Similarly, these reasons could also explain the significant relation between the
337 Anglosphere and toxoplasmosis SP retrieved in this analysis.

338 According to this analysis time variations of SP due to economic, social or
339 cultural habits evolution in a given population, which is reflected here by the
340 period of the survey, was not found as a major driver of world SP pattern. Its
341 effect seems to be diluted by others significant drivers evolving with time and
342 retrieved as significant in this study.

343 Similarly, the mean age of surveyed populations, quality of water, type of
344 studied population, study scale and sanitation coverage were not retrieved as

345 major drivers of world seroprevalence repartition. Their effects did not seemed
346 to be important enough to model in a large proportion the world seroprevalence
347 pattern.

348 None of national food consumption quantities were retrieved in the final
349 model selected by the RTA. Several potential explanations could be put forward.
350 The first one is that the national level of these data does not fit well with the
351 reality at the local scale where the study was made. Non negligible disparities
352 can be easily found between the national level and different populations living
353 inside the same countries. However, the results seemed to indicate an
354 alternative hypothesis. The fact that religion and belonging to an Anglo-Saxon
355 country were strongly significant in the final model, suggested that the risk of
356 toxoplasmosis is more linked with the sanitary quality of products and cooking
357 methods than the actual absorbed quantities. This trend was demonstrated by
358 others who explained that processed meats for example, which are produced
359 using commercial procedures killing oocysts, are an unlikely source of infections
360 for humans. Furthermore, uncooked meat was known as a classic risk factor for
361 toxoplasmosis contamination [6].

362 In the USA, where SP is low, the main source of contamination is the
363 consumption of insufficiently cooked meat. A recent study also finds this notion
364 [24], and other less usual factors such as the consumption of mussels, oysters,
365 and clams, or more classical but rarely demonstrated the possession of several
366 young cats. [25]. Indeed, the passing of non-infective oocysts, and the short
367 duration of oocyst shedding, explain why direct contact with cats is not thought
368 to be a primary risk for human infection [26].

369 Isolated studies in a given country, even when sample sizes are large, may
370 not permit to identify particular contamination risk factors. A study in Togo did
371 not find any difference between ethnic groups, religion, or food habits [27].
372 More generally, in developing tropical regions, the most probable source of
373 contamination is the reservoir from the soil [28,29]. Near maximal SPs may be
374 reached early in life but with very variable rates depending on the climate: low in
375 dry Sahelian areas (Niger), higher where the climate is more humid [30]. The
376 preponderance of the geographic component is mirrored by the climatic data.

377 The current efforts to collect and describe available data on world SP
378 patterns and associated risk factors have their limits [6,31]. This study
379 suggests that this simple alternative method, using modern statistical tools
380 coming from ecology coupled with GIS, could be helpful to roughly describe and
381 quantify the world distribution of toxoplasmosis SP.

382 The previous study [4] demonstrated that CT Risk in a given population is
383 important when general mean SP is situated between 30 and 60%. Here, by
384 applying this result, 4 geographical world areas appeared to be particularly at

385 risk of CT (in orange on the final map): West Europe, North Africa, Middle East
386 and West part of Central Asia.

387 Such results could help shaping health policies towards toxoplasmosis in
388 low income countries or countries where no national SP studies are available.

389 However, this method has its own limitations.

390 Firstly, the model error was relatively intermediate (13.5%) and does not
391 allow to design public health interventions solely on the base of these results.
392 Complementary data and country specificities regarding toxoplasmosis are
393 needed to have a clear vision of the national risk of congenital toxoplasmosis. At
394 a country scale, these results could be used as an argument to evaluate national
395 toxoplasmosis SP and the risk of congenital toxoplasmosis.

396 Secondly, no homogeneous data describing the world repartition of felid
397 populations and their toxoplasmosis prevalence were available, and therefore
398 could not be included in the model. Thus, this study was not able to evaluate the
399 importance of felids in the epidemiology. Further researches could fix this
400 problem to improve the model's precision.

401 Thirdly, the present study used seroprevalence as the endpoint of
402 analysis, which is subject to many factors in particular health intervention.
403 Interventions may vary from place to place, and this variability was not taken
404 into account in this analysis.

405 Fourthly, worldwide, the publications of infectious diseases are heavily
406 dependent on health resources (finance, science and technology) in each country,
407 which may cause a so-called "reporting bias" in the characterization of the global
408 distribution. However, the large diversity of populations (in terms of geographic
409 area and socio-economic level) included here coupled with the first step of
410 statistical analysis (RFA) minimized it as possible.

411 Fifthly, the exactitude of seroprevalences and their comparison also
412 depend on the serological techniques used. Globally, and mostly for old studies,
413 the tendency is for an underestimation of prevalence rates because of the lack
414 of sensitivity of certain methods that may not diagnose very remote infections
415 in which antibody levels are low.

416 Sixthly, ecological studies have their limitations (Bias, confounding). The
417 data used here may have been heterogeneous in quality. The available national
418 data may not have been sufficient to reflect the local variations of all the study
419 incorporated in this analysis.

420 However, this is the only data that is available to study the problem at
421 this scale. The fact that the present analytical review came to similar
422 observations on suspected and described phenomenon acting on toxoplasmosis SP
423 level than others studies is unlikely to have been due to a statistical artefact.

424

425

426 **CONCLUSIONS**

427

428 Although there are some limitations, this is the first model that leads to a
429 world map of the seroprevalence of toxoplasmosis. 4 geographical areas
430 appeared to be particularly at risk of CT: West Europe, North Africa, Middle
431 East and West part of Central Asia. Complementary data and country
432 specificities regarding toxoplasmosis in these high risk subgroups are needed to
433 have a clear vision of their national risk of congenital toxoplasmosis.

434

435 **BIBLIOGRAPHY**

436

- 437 1. Benard A, Petersen E, Salamon R, Chene G, Gilbert R, et al. (2008) Survey
438 of European programmes for the epidemiological surveillance of congenital
439 toxoplasmosis. *Euro Surveill* 13.
- 440 2. Villena I, Ancelle T, Delmas C, Garcia P, Brezin AP, et al. (2010) Congenital
441 toxoplasmosis in France in 2007: first results from a national surveillance
442 system. *Euro Surveill* 15.
- 443 3. Ancelle T, Yera H, Talabani H, Lebuissou A, Thulliez P, et al. (2009) [How
444 can the cost of screening for toxoplasmosis during pregnancy be
445 reduced?]. *Rev Epidemiol Sante Publique* 57: 411-417.
- 446 4. Carme B, Hanf M, Demar M, Nacher M (2011) Classification of populations
447 exposed to congenital toxoplasmosis: An analytical review of world
448 seroprevalences. *JID* (In submission).
- 449 5. Montoya JG, Liesenfeld O (2004) Toxoplasmosis. *Lancet* 363: 1965-1976.
- 450 6. Tenter AM, Heckeroth AR, Weiss LM (2000) *Toxoplasma gondii*: from
451 animals to humans. *Int J Parasitol* 30: 1217-1258.
- 452 7. Pappas G, Roussos N, Falagas ME (2009) Toxoplasmosis snapshots: global
453 status of *Toxoplasma gondii* seroprevalence and implications for
454 pregnancy and congenital toxoplasmosis. *Int J Parasitol* 39: 1385-1394.
- 455 8. Kottek M, Grieser J, Beck C, Rudolf B, Rubel F (2006) World map of the
456 Köppen-Geiger climate classification updated. *Meteorol Z* 15: 259-263.
- 457 9. ONU (2010) UN Common Database available.
- 458 10. WHO UNICEF (2003) Joint Monitoring Programme for Water Supply and
459 Sanitation Accessed: August 18, 2006.
- 460 11. CIA (2010) The World Fact Book available at:
461 <https://www.cia.gov/library/publications/the-world-factbook> Accessed:
462 July 08,2010. 2008.
- 463 12. FAO (2010) Common statistics available at: <http://faostat.fao.org>
464 Accessed: July 08,2010. 2010.
- 465 13. Breiman L (2001) Random forests. *Machine Learning* 45: 5-32.

Chapitre 7 - Utilisation des données issues de la littérature scientifique : Application à l'étude de la séroprévalence à *Toxoplasma gondii* dans les populations humaines

- 466 14. De'ath G, Fabricius KE (2000) Classification and regression trees: a
467 powerful yet simple technique for ecological data analysis. *Ecology* 81:
468 3178-3192.
- 469 15. Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, et al. (2007)
470 Random forests for classification in ecology. *Ecology* 88: 2783-2792.
- 471 16. Prasad AM, Iverson LR, Liaw A (2006) Newer classification and regression
472 tree techniques: bagging and random forests for ecologic prediction.
473 *Ecosystems* 9: 181-199.
- 474 17. Hothorn T, Buehlmann P, Dudoit S, Molinaro A, Van Der Laan M (2006)
475 Survival Ensembles. *Biostatistics* 7: 355-373.
- 476 18. Strobl C, Boulesteix A-L, Zeileis A, Hothorn T (2007) Bias in random
477 forest variable importance measures: Illustrations, sources and a solution.
478 *BMC Bioinformatics* 8: 25.
- 479 19. Strobl C, Boulesteix A-L, Kneib T, Augustin T, Zeileis A (2008) Conditional
480 variable importance for random forests. *BMC Bioinformatics* 9: 307.
- 481 20. Hothorn T, Hornik K, Zeileis A (2006) Unbiased Recursive Partitioning: A
482 Conditional Inference Framework. *Journal of Computational and Graphical
483 Statistics* 15: 651-674.
- 484 21. Cook AJ, Gilbert RE, Buffolano W, Zufferey J, Petersen E, et al. (2000)
485 Sources of toxoplasma infection in pregnant women: European multicentre
486 case-control study. *European Research Network on Congenital
487 Toxoplasmosis. BMJ* 321: 142-147.
- 488 22. Lake R, Hudson A, Cressey P (2002) Risk profile: *Toxoplasma gondii* in red
489 meat and meat products. Christchurch: New Zealand Food Safety
490 Authority. 36 p.
- 491 23. Dubey JP (1998) *Toxoplasma gondii* oocyst survival under defined
492 temperatures. *J Parasitol* 84: 862-865.
- 493 24. Jones JL, Dargelas V, Roberts J, Press C, Remington JS, et al. (2009)
494 Risk factors for *Toxoplasma gondii* infection in the United States. *Clin
495 Infect Dis* 49: 878-884.
- 496 25. Shuhaiber S, Koren G, Boskovic R, Einarson TR, Soldin OP, et al. (2003)
497 Seroprevalence of *Toxoplasma gondii* infection among veterinary staff in
498 Ontario, Canada (2002): implications for teratogenic risk. *BMC Infect Dis*
499 3: 8.
- 500 26. Elmore SA, Jones JL, Conrad PA, Patton S, Lindsay DS, et al. (2010)
501 *Toxoplasma gondii*: epidemiology, feline clinical aspects, and prevention.
502 *Trends Parasitol* 26: 190-196.
- 503 27. Deniau M, Tourte-Schaefer C, Agbo K, Dupouy-Camet J, Heyer C, et al.
504 (1991) [Evaluation of the risk of congenital toxoplasmosis in Togo]. *Bull
505 Soc Pathol Exot* 84: 664-672.

Partie 2 : Travaux scientifiques

- 506 28. Adou-Bryn KD, Ouhon J, Nemer J, Yapo CG, Assoumou A (2004)
507 [Serological survey of acquired toxoplasmosis in women of child-bearing
508 age in Yopougon (Abidjan, Cote d'Ivoire)]. Bull Soc Pathol Exot 97: 345-
509 348.
- 510 29. Duong TH, Dufillot D, Martz M, Richard-Lenoble D, Kombila M (1992)
511 Seroepidemiological study of toxoplasmosis in Libreville, Gabon. Ann Soc
512 Belg Med Trop 72: 289-293.
- 513 30. Dumas N, Cazaux M, Seguela JP, Charlet JP (1991) [Epidemiology of
514 toxoplasmosis in mothers and children in tropical Africa]. Bull Soc Pathol
515 Exot 84: 645-658.
- 516 31. Porter SB, Sande MA (1992) Toxoplasmosis of the central nervous system
517 in the acquired immunodeficiency syndrome. N Engl J Med 327: 1643-
518 1648.

Figure 1 : Flow Diagram of Study Selection

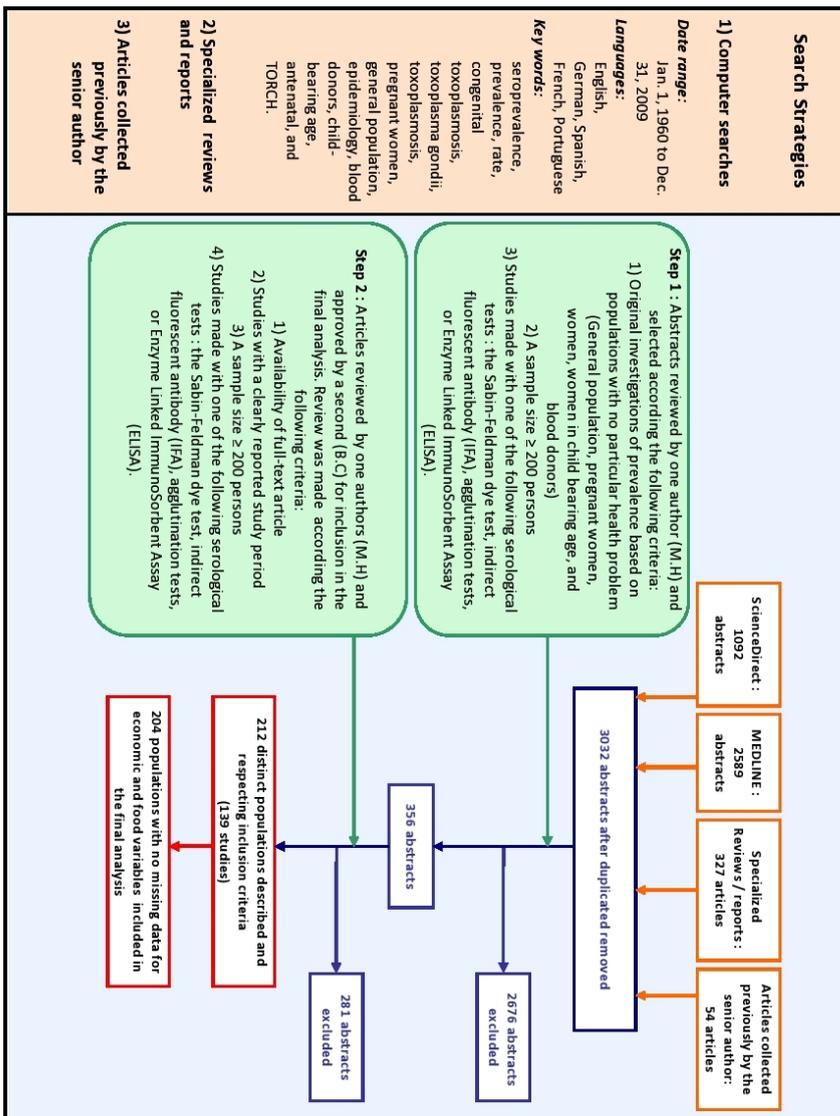


Figure 2 : Geographic distribution of study sites and crude toxoplasmosis seroprevalence (median percentage) by sampled country.

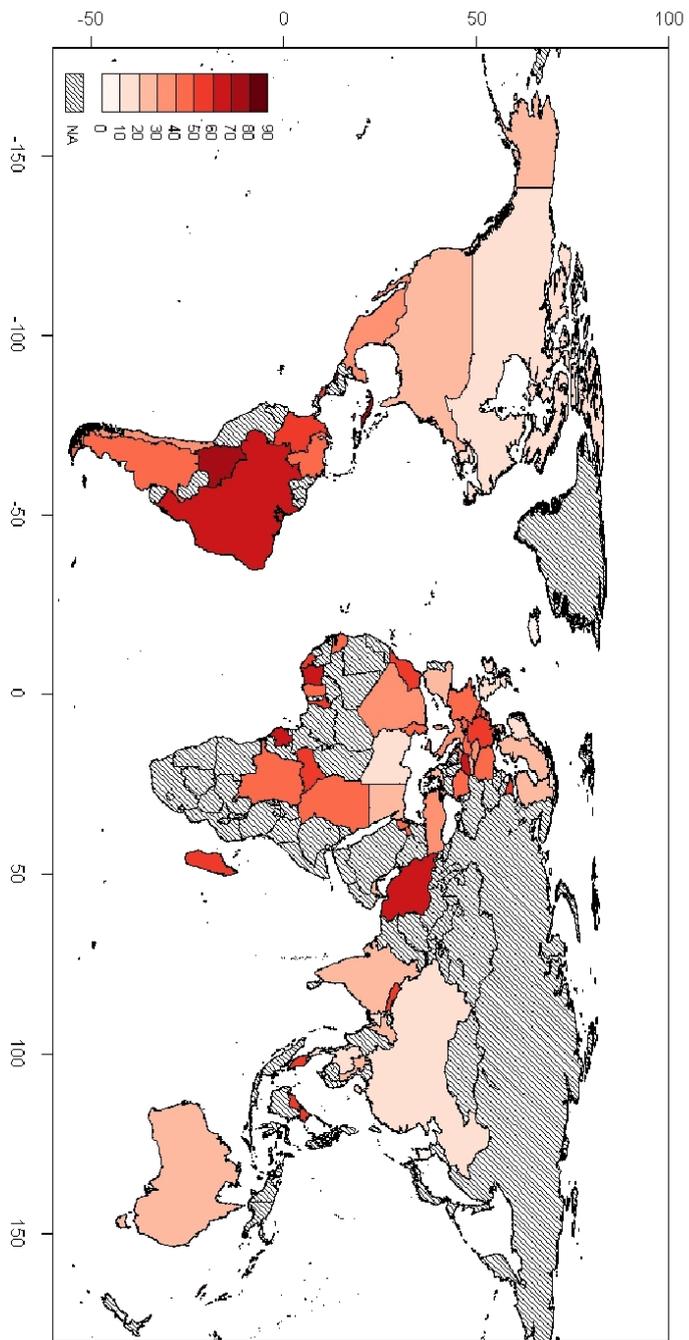


Figure 3 : Distributions of Seroprevalence by gender, setting, religions, climate, and world regions

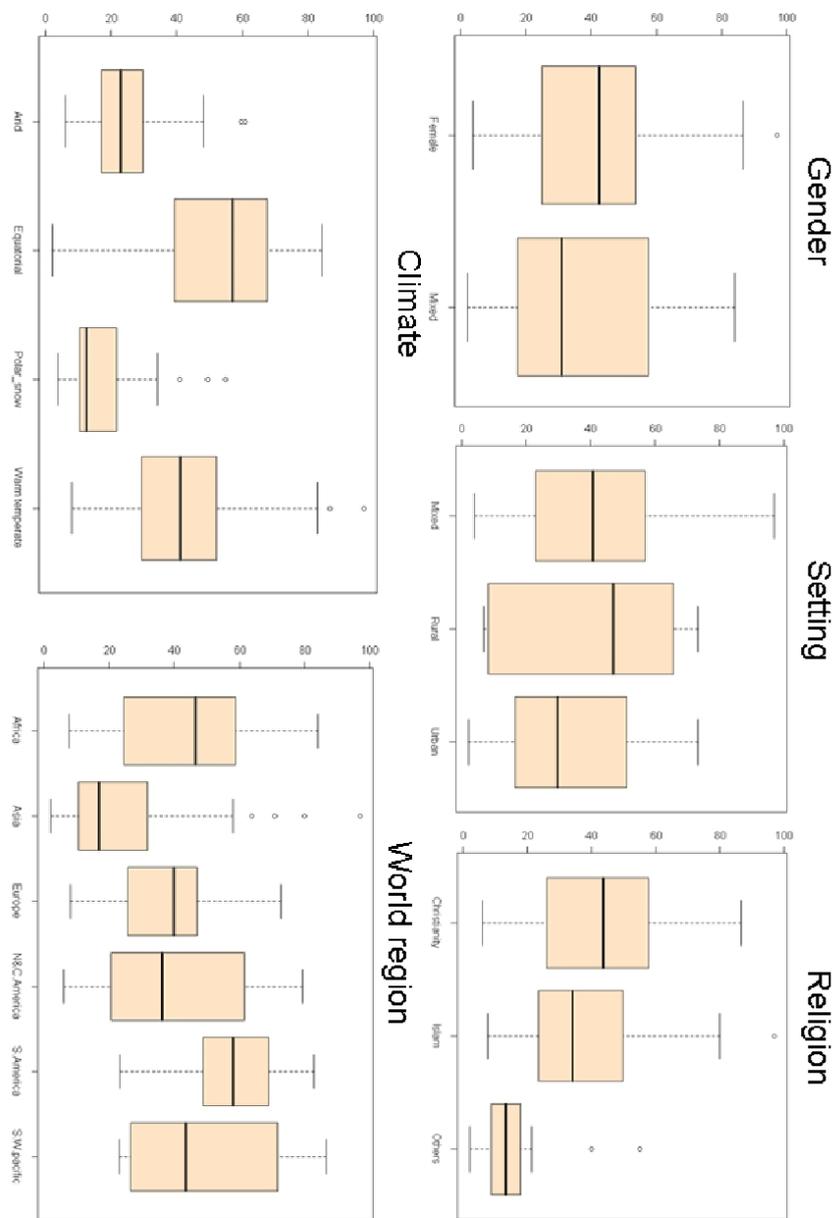
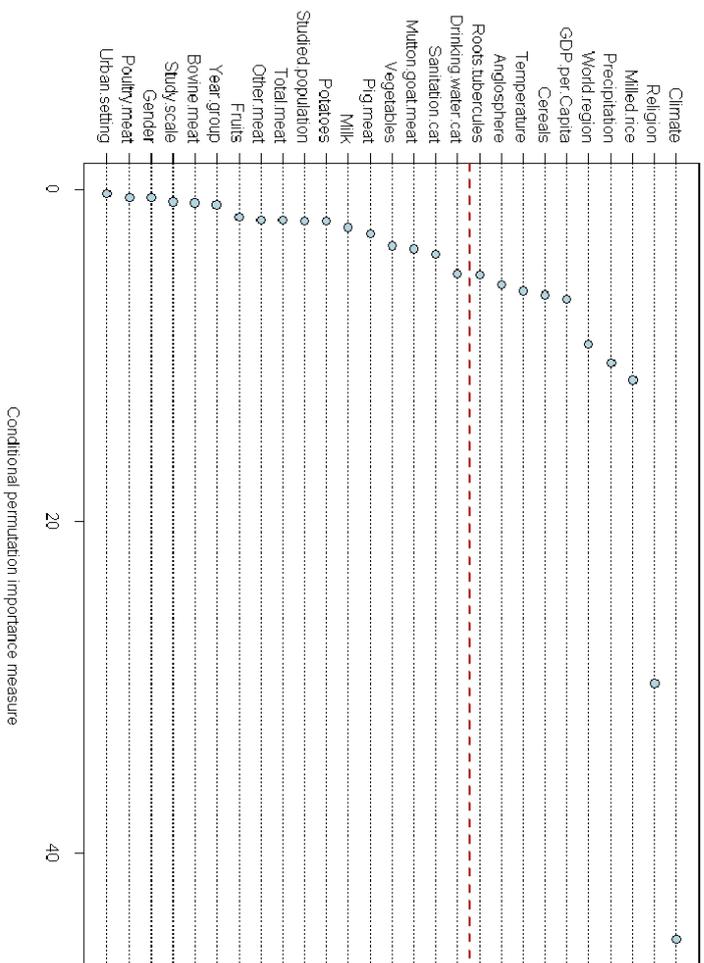


Figure 4: Variable importance plots for predictor variables from random forests models for toxoplasmosis seroprevalence.



Predictor variables are along the y-axis. More the conditional importance measure is high more the associated variable is important to predict the toxoplasmosis SP level

Figure 5: Regression tree for toxoplasmosis seroprevalence level

Regression tree for toxoplasmosis SP level was obtained using the 10 most important predictor variables as determined by the random forest analysis (see Fig. 3). P-values at each node are adjusted with Bonferroni correction. In order for a split to occur the p-value must be <0.01 and the node size must not be less than 20. Box plots at terminal nodes show the distribution of the data within that branch of the tree. Boxes represent inter-quartile range; horizontal lines within the boxes represent median values; whiskers extend to the most extreme data point that is no more than 1.5 times the inter-quartile range. Dots above and below whiskers represent more extreme data points

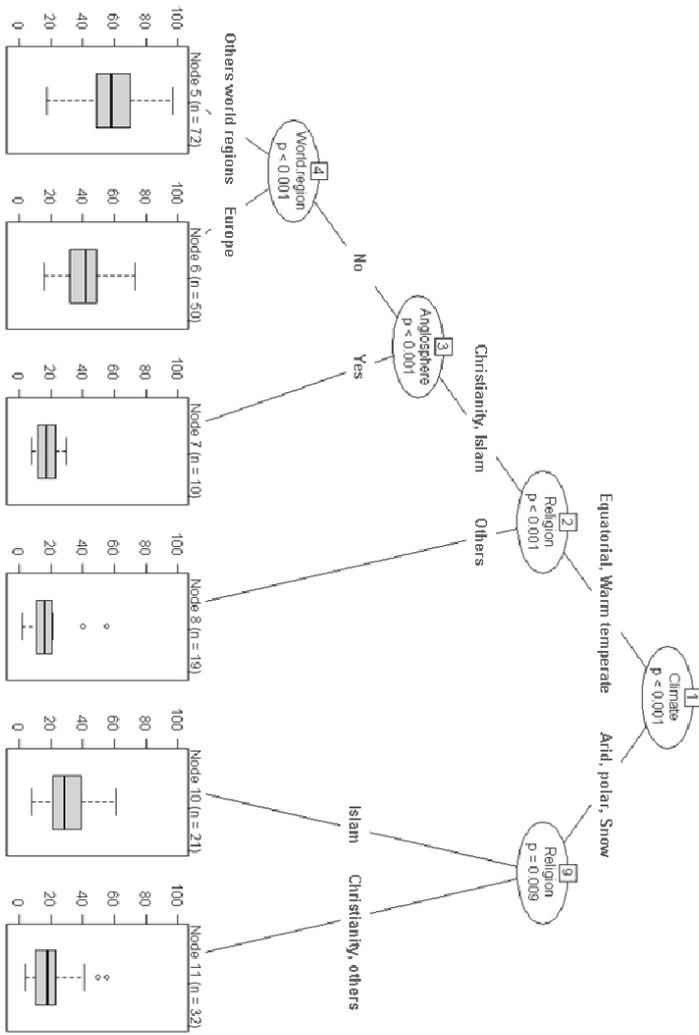
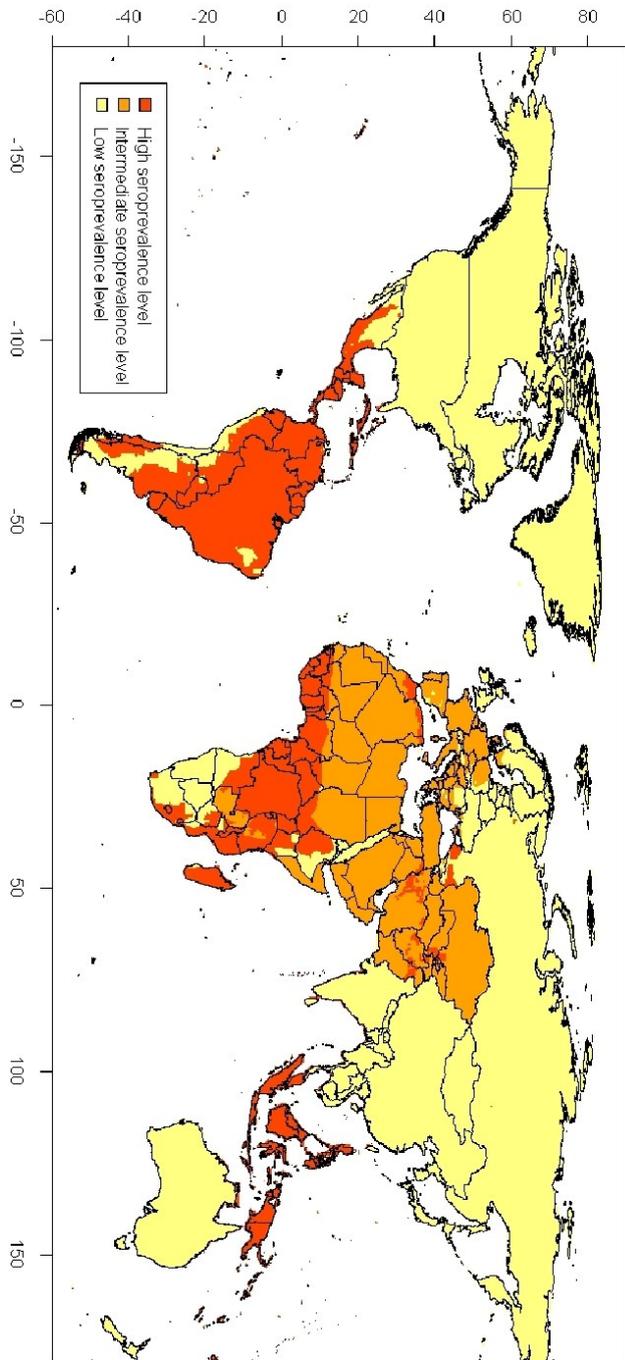


Figure 6 : Estimation of the world distribution of toxoplasmosis seroprevalence level

SP level was estimated using the final model obtained with the regression tree described in the figure 5. According to estimated SP levels, 3 groups were defined: low SP level (mean SP < 30%), intermediate SP level (mean SP \geq 30% and <50%) and high SP level (mean SP \geq 50%)



Chapitre 7 - Utilisation des données issues de la littérature scientifique : Application à l'étude de la séroprévalence à *Toxoplasma gondii* dans les populations humaines

Tables 1: Summary statistics for qualitative predictor variables used in the analysis of toxoplasmosis seroprevalence.

Qualitative predictors		N	%
Study population	Blood donors	18	8,8
	General population	76	37,3
	Pregnant women	77	37,7
	Women	8	3,9
	Women in child-bearing Age	25	12,3
Gender	Female	110	53,9
	Mixed	94	46,1
Setting	Mixed	146	71,6
	Rural	5	2,5
	Urban	53	26,0
Study scale	City	103	50,5
	Country	22	10,8
	Region	79	38,7
Study period	< 1980's	27	13,2
	1980's	54	26,5
	1990's	60	29,4
	2000's	63	30,9
World region	Africa	27	13,2
	Asia	35	17,2
	Europe	69	33,8
	Amérique du N.C	46	22,5
	Amérique du Sud	23	11,3
	Pacifique S.W	4	2,0
Climate	Arid	34	16,7
	Equatorial	57	27,9
	Polar / snow	19	9,3
	Warm temperate	94	46,1
Precipitation	Desert	13	6,4
	Fully humid	69	33,8
	Monsoonal	3	1,5
	Steppe	15	7,4
	Summer dry	23	11,3
	Winter dry	60	29,4
	Not identified	21	10,3
Temperature	Cold arid	7	3,4
	Cool summer	4	2,0
	Hot arid	18	8,8
	Hot summer	37	18,1
	Polar tundra	1	0,5
	Warm Summer	62	30,4
	Not identified	75	36,8
Religion	Christianity	148	72,5
	Islam	31	15,2
	Others	25	12,3
Drinking water coverage	≤ 80%	13	6,4
	81-85%	11	5,4
	86-90%	19	9,3
	91-95%	58	28,4
	> 95%	89	43,6
	Not identified	14	6,9
Sanitation coverage	≤ 30%	11	5,4
	31-70%	23	11,3
	71-80%	18	8,8
	81-90%	50	24,5
	> 90%	67	32,8
	Not identified	35	17,2
Anglosphere	Yes	13	6,4
	No	191	93,6

Partie 2 : Travaux scientifiques

Tables 2: Summary statistics for quantitative predictor variables used in the analysis of toxoplasmosis seroprevalence.

Quantitative predictors	Median	Mean	Min	Max
GDP / year / capita (\$) *	3558	9163	90	68988
Bovine meat consumption **	16,2	16,6	0,3	81,6
Mutton and goat meat consumption **	0,8	2,5	0,0	27,7
Pig meat consumption **	11,6	18,0	0,0	73,5
Poultry meat consumption **	13,1	15,0	0,5	47,1
Other meat consumption **	0,9	1,8	0,0	21,3
Total meat consumption **	49,7	53,9	2,7	129,9
Milk consumption (without butter) **	108,0	136,6	1,2	369,5
Potatoes consumption **	21,7	37,7	0,1	136,7
Milled rice consumption **	6,0	24,1	0,4	163,4
Roots and tubercles consumption **	11,7	13,1	1,2	131,2
Cereals (without beer) consumption **	123,6	137,2	34,3	499,7
Fruits (without grapes) consumption **	96,1	91,1	6,9	188,7
Vegetables consumption **	68,4	88,8	11,2	302,2

* Average GDP in \$ /year /capita during study period

** Average consumption in Kg / year / capita during study period

7.7. Conclusions

L'approche « étude écologique et données libres » des populations exposées à la toxoplasmose congénitale mise en œuvre ici ainsi que les déductions qui peuvent être tirées suite à la première partie de cette analyse ont permis de définir des groupes pour lesquels des indicateurs simples (prévalence sérologique moyenne et différentiel de prévalence 15 – 40 ans) permettent de juger grossièrement du risque de toxoplasmose congénitale dans une population. Globalement le risque de toxoplasmose congénitale concerne surtout les pays présentant un différentiel de sérologie toxoplasmique femme 15 – 40 ans élevé (> 15%), indice relié lui-même à une prévalence générale moyenne située entre 30 et 60%. Cette situation correspond à celle où le ratio avantage / inconvénient d'un programme national de lutte contre la toxoplasmose congénitale est le meilleur. Cependant, des indicateurs plus précis et plus fiables car ciblés doivent être mis en place à l'échelle nationale surtout si un programme d'action spécifique est envisagé.

La deuxième partie de cette étude en agrégeant des données de séroprévalence provenant de la littérature scientifique très facilement accessibles avec des données libres décrivant les caractéristiques géographiques, sociétales et alimentaires des populations étudiées avait pour objectif final d'élaborer un modèle permettant de prédire grossièrement la répartition géographique mondiale du niveau de séroprévalence à *Toxoplasma gondii*. Malgré les limitations de cette méthode, l'utilisation de méthodes écologiques couplées à des données issues de la littérature a permis d'identifier des grands facteurs influençant le niveau de séroprévalence globale et donc du risque de toxoplasmose congénitale associé. En extrapolant ces résultats, une première carte mondiale du risque de toxoplasmose congénitale a ainsi pu être dessinée.

Bibliographie

1. National Center for Biotechnology Information (NCBI). **PubMed**. In. Rockville Pike, Bethesda MD; 2011.
2. Cook DJ, Mulrow CD, Haynes RB. **Systematic reviews: synthesis of best evidence for clinical decisions**. *Ann Intern Med* 1997,126:376-380.
3. Pai M, McCulloch M, Gorman JD, Pai N, Enanoria W, Kennedy G, *et al*. **Systematic reviews and meta-analyses: an illustrated, step-by-step guide**. *Natl Med J India* 2004,17:86-95.
4. McAlister FA, Clark HD, van Walraven C, Straus SE, Lawson FM, Moher D, Mulrow CD. **The medical review article revisited: has the science improved?** *Ann Intern Med* 1999,131:947-951.
5. Akobeng AK. **Understanding systematic reviews and meta-analysis**. *Arch Dis Child* 2005,90:845-848.
6. Mulrow CD. **Rationale for systematic reviews**. *BMJ* 1994,309:597-599.
7. Greenhalgh T. **Papers that summarise other papers (systematic reviews and meta-analyses)**. *BMJ* 1997,315:672-675.
8. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. **Epidemiology and reporting characteristics of systematic reviews**. *PLoS Med* 2007,4:e78.
9. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. **How quickly do systematic reviews go out of date? A survival analysis**. *Ann Intern Med* 2007,147:224-233.
10. Savoie I, Helmer D, Green CJ, Kazanjian A. **Beyond Medline: reducing bias through extended systematic review search**. *Int J Technol Assess Health Care* 2003,19:168-178.
11. D'Souza AL, Rajkumar C, Cooke J, Bulpitt CJ. **Probiotics in prevention of antibiotic associated diarrhoea: meta-analysis**. *BMJ* 2002,324:1361.
12. Dickersin K. **The existence of publication bias and risk factors for its occurrence**. *JAMA* 1990,263:1385-1389.
13. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, *et al*. **The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration**. *J Clin Epidemiol* 2009,62:e1-34.
14. The Cochrane Collaboration. **The Cochrane Library**. In. London; 2011.
15. Tenter AM, Heckerroth AR, Weiss LM. **Toxoplasma gondii: from animals to humans**. *Int J Parasitol* 2000,30:1217-1258.
16. Pappas G, Roussos N, Falagas ME. **Toxoplasmosis snapshots: global status of Toxoplasma gondii seroprevalence and implications for pregnancy and congenital toxoplasmosis**. *Int J Parasitol* 2009,39:1385-1394.
17. Dunn D, Wallon M, Peyron F, Petersen E, Peckham C, Gilbert R. **Mother-to-child transmission of toxoplasmosis: risk estimates for clinical counselling**. *Lancet* 1999,353:1829-1833.
18. Berger F, Goulet V, Le Strat Y, de Valk H, Désenclos JC. **La toxoplasmose en France chez la femme enceinte en 2003 : : séroprévalence et facteurs associés**. In. Saint Maurice: Institut de veille sanitaire; 2007.

Chapitre 7 - Utilisation des données issues de la littérature scientifique : Application à l'étude de la séroprévalence à *Toxoplasma gondii* dans les populations humaines

19. Benard A, Petersen E, Salamon R, Chene G, Gilbert R, Salmi LR. **Survey of European programmes for the epidemiological surveillance of congenital toxoplasmosis.** *Euro Surveill* 2008,13.
20. Carme B, Demar-Pierre M. **Toxoplasmosis in French Guiana. Atypical (neo-) tropical features of a cosmopolitan parasitosis.** *Med Trop (Mars)* 2006,66.
21. Nacher M, El Guedj M, Vaz T, et al. **Risk factors for late HIV diagnosis in French Guiana.** *AIDS* 2005,19:727-729.
22. Carme B, Bissuel F, Ajzenberg D, Bouyne R, Aznar C, Demar M, et al. **Severe acquired toxoplasmosis in immunocompetent adult patients in French Guiana.** *J Clin Microbiol* 2002,40:4037-4044.
23. Carme B, Demar M, Ajzenberg D, Darde ML. **Severe acquired toxoplasmosis caused by wild cycle of *Toxoplasma gondii*, French Guiana.** *Emerg Infect Dis* 2009,15:656-658.
24. Carme B. **[Exposure to *Toxoplasma gondii* and risk of congenital toxoplasmosis].** *Med Trop (Mars)* 2001,61:550-551.
25. Carme B, Tirard-Fleury V. **[Toxoplasmosis among pregnant women in France : Sero-prevalence, seroconversion and knowledge levels. Trends 1965-1995.].** *Med Mal Infect* 1996, 26:431-436.
26. Olusi TA, Ajaya JA, Makinde AA. **Antibodies to *Toxoplasma gondii* in a rat-eating population on Benue State, Nigeria.** *Ann Trop Med Parasitol* 1994,88:217-218.
27. Julvez J, Magnaval JF, Baixench MT, Maron I. **[Seroepidemiology of toxoplasmosis on Mayotte (Comores Archipelago)].** *Med Trop (Mars)* 1994,54:343-344.
28. Allain JP, Palmer CR, Pearson G. **Epidemiological study of latent and recent infection by *Toxoplasma gondii* in pregnant women from a regional population in the U.K.** *J Infect* 1998,36:189-196.

Chapitre 8 - Utilisation des statistiques onusiennes : Applications à l'étude des déterminants de la dynamique des résistances aux antituberculeux, du paludisme et de la mortalité juvéno-infantile.

« Pour que les activités de développement donnent des résultats, il faut d'abord que des données statistiques sur la pauvreté, l'accès à l'enseignement et l'incidence des maladies soient réunies et analysées. Les statistiques sous-tendent presque tous les aspects des budgets et des programmes qui permettent de nourrir les enfants affamés et d'offrir un abri et des soins d'urgence aux victimes de catastrophes naturelles ».

Ban Ki-Moon

8.1. Le 20 avril 2010 une date historique

20 avril 2010. Peu de gens connaissent la signification particulière de cette date et son coté historique, mais dans le petit monde des données libres (ou « Open Data » en anglais), cette dernière représente une petite révolution. Mais pourquoi?

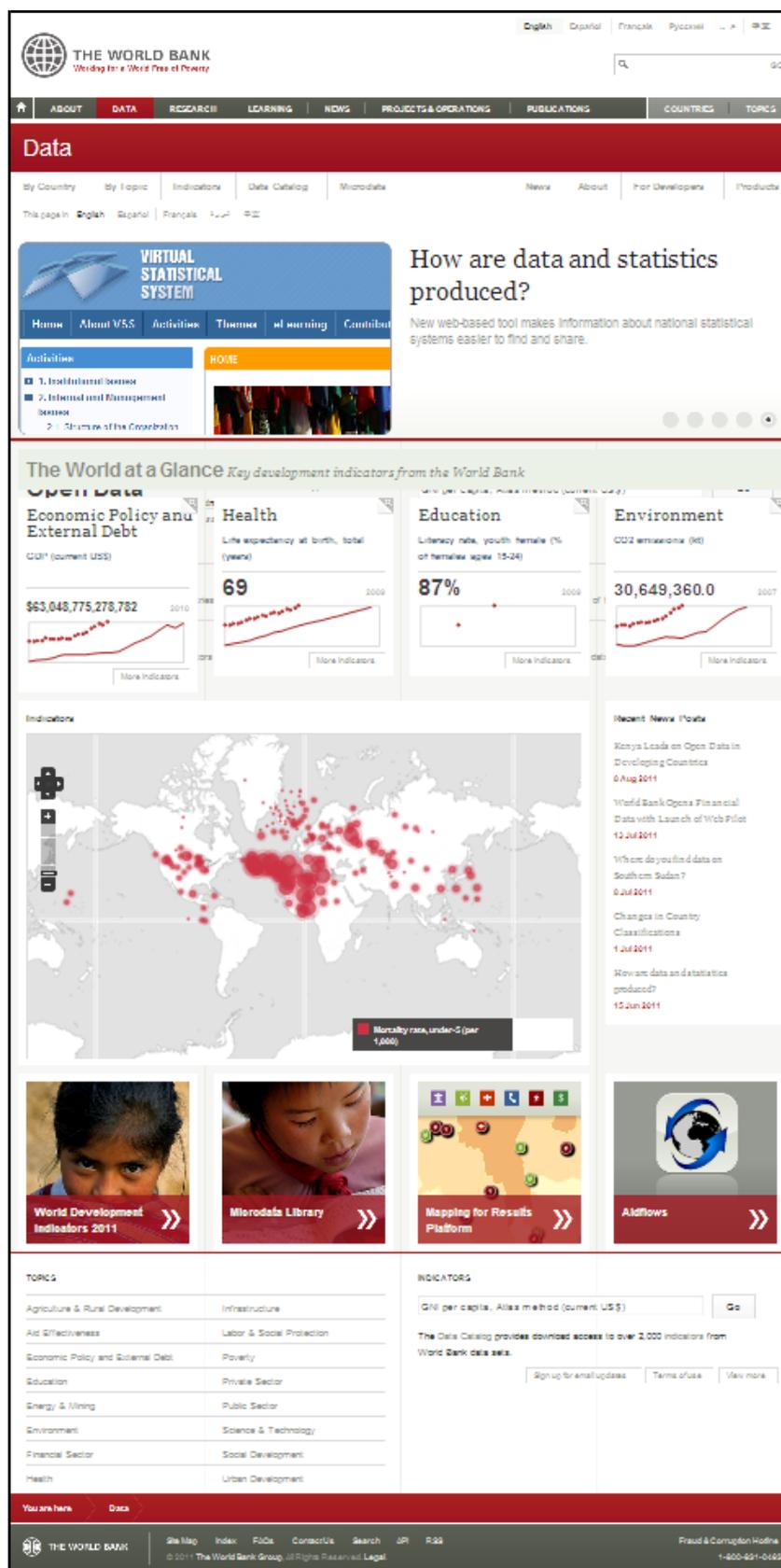
Le 20 avril 2010 la banque mondiale a révélé son intention de rendre librement accessible l'ensemble des statistiques de développement qu'elle collecte quotidiennement. Cette étape attendue depuis très longtemps donna enfin accès aux étudiants, chercheurs, journalistes, économistes et organisations du monde entier à une colossale masse d'information sur l'ensemble des pays du monde. Dans la foulée, la banque mondiale lança aussi un nouveau site Internet worldbank.org permettant un accès facilité à l'ensemble de ces données [1] (Figure 40). Ce nouveau site Internet donne désormais accès à toute personne qui le souhaite à des statistiques diverses et variées sur 209 pays du monde et sur une période de temps remontant jusqu'à 50 ans. Grâce à ce site Internet, l'utilisateur peut maintenant simplement télécharger des jeux de données brutes entiers portant sur un pays ou un indicateur particulier. Les données proposées proviennent de plusieurs sources qui incluent plus de 30 agences internationales et privées ou organisations non gouvernementales. En ouvrant ses données, la banque mondiale a ainsi rejoint le mouvement grandissant de la « donnée libre ».

8.2. Les statistiques onusiennes : des informations stratégiques

La Société des Nations (SDN) n'ayant pas pu empêcher la seconde guerre mondiale, l'Organisation des Nations Unies (ONU) est fondée le 26 juin 1945 à San Francisco, pour la remplacer et fournir une nouvelle plate-forme de dialogue international.

Chapitre 8- Utilisation des statistiques onusiennes : Applications à l'étude des déterminants de la dynamique des résistances aux antituberculeux, du paludisme et de la mortalité juvéno-infantile.

Figure 40 : Page d'accueil du site Internet de la banque mondiale donnant accès à l'ensemble de ses statistiques sur 209 pays du monde et sur une période de temps remontant jusqu'à 50 ans (<http://data.worldbank.org/>).



Partie 2 : Travaux scientifiques

L'ONU est une organisation qui regroupe, à quelques exceptions près, tous les États de la planète. Distincte des États qui la composent, l'organisation a pour finalité la paix internationale. Ses objectifs sont de faciliter la coopération dans les domaines du droit international, de la sécurité internationale, du développement économique, du progrès social et des droits de l'homme.

Le système des Nations unies est constitué de ces six organes principaux de l'ONU, auxquels ont été ajoutés au fil du temps divers organismes, institutions et programmes ayant une vocation plus spécifique. Dans ces organisations périphériques on peut citer l'OMS (l'institution spécialisée de l'ONU pour la santé), et la FAO (l'institution spécialisée pour l'alimentation et l'agriculture).

La figure publique la plus visible de l'ONU est le Secrétaire général. Ce poste est occupé depuis 2007 par le Sud-coréen Ban Ki-moon. A l'occasion de la journée mondiale de la statistique, le 20 octobre 2010, ce dernier a déclaré :

« Pour que les activités de développement donnent des résultats, il faut d'abord que des données statistiques sur la pauvreté, l'accès à l'enseignement et l'incidence des maladies soient réunies et analysées. Les statistiques sous-tendent presque tous les aspects des budgets et des programmes qui permettent de nourrir les enfants affamés et d'offrir un abri et des soins d'urgence aux victimes de catastrophes naturelles ».

Cette phrase à elle seule résume parfaitement bien l'importance stratégique pour l'ONU d'édition de statistiques fiables sur l'ensemble des pays appartenant à son système.

Pour relever ce défi, une Commission de statistique à l'ONU a été créée dès 1947 pour contribuer à imposer des normes et des directives méthodologiques internationales dans l'ensemble des domaines de la statistique. Cette commission avait aussi pour but d'aider les États à renforcer leurs capacités d'établissement de rapports statistiques et leur permettre de comparer leurs données aux niveaux régional et international.

En dehors des Nations Unies, les statistiques onusiennes sont aussi de première nécessité pour les chercheurs, chefs de projet, politiques ou législateurs du monde entier qui à l'aide de ces dernières peuvent avoir des informations chiffrées sur une situation dans une région ou un pays donné et y implémenter des programmes spécifiques.

Pour pérenniser la diffusion et l'utilisation de ces données, chaque organe des Nations Unies met à disposition au travers de son site Internet les chiffres et statistiques qu'il compile dans sa spécialité [2-4].

8.3. Autour de la qualité des statistiques éditées par l'ONU

L'ensemble des données de l'ONU est compilé à partir de publications et de bases de données produites et maintenues par les programmes techniques et les bureaux régionaux. Les indicateurs qui y figurent ont été sélectionnés sur la base de leur pertinence, de la disponibilité et de la qualité des données et de la fiabilité et de la comparabilité des estimations.

Pris dans leur ensemble, les indicateurs fournissent un résumé complet de l'état actuel d'un pays, dans de nombreux domaines.

Les estimations proviennent de sources multiples, selon l'indicateur et la disponibilité et la qualité des données. Dans nombre de pays, les systèmes d'information statistique présentent des faiblesses et les données

Chapitre 8- Utilisation des statistiques onusiennes : Applications à l'étude des déterminants de la dynamique des résistances aux antituberculeux, du paludisme et de la mortalité juvénile-infantile.

empiriques sous-jacentes peuvent être indisponibles ou de qualité médiocre. Cependant, tous les efforts ont été consentis pour garantir un usage optimal des données rapportées par les pays ajustées si nécessaire pour tenir compte des valeurs manquantes, pour corriger les biais connus et pour assurer une comparabilité maximale des statistiques entre les pays et au cours du temps. En outre, des méthodes statistiques et de modélisation ont été appliquées pour combler les lacunes en termes de données.

Dans le domaine de l'épidémiologie ces données sont précieuses pour l'établissement d'un bilan mondial de la situation sanitaire vis-à-vis de certaines pathologies et leurs suivis au cours du temps. Cependant ces dernières ne sont utilisées le plus souvent que de manière descriptive et sans mise en lumière avec les autres chiffres édités par les Nations Unies. Pour combler cette lacune, l'utilisation de méthodes écologiques peut s'avérer très utile pour caractériser et surtout expliquer certaines de ces statistiques.

8.4. Méthodes écologiques et données onusiennes : un éclairage des épidémies à l'échelle populationnelle

Comme nous l'avons vu dans la première partie, comprendre et maîtriser les grandes pandémies infectieuses actuelles requiert de raisonner à différents niveaux d'analyse [5]:

- Au niveau moléculaire/cellulaire: L'utilisation de la biologie moléculaire à cette échelle est nécessaire pour déterminer les mécanismes et le timing de la transmission de l'infection et ainsi identifier un moyen de la stopper.
- Au niveau individuel: A ce niveau, la médecine se charge d'étudier les conséquences de l'infection sur les individus dans une optique de soins.
- Au niveau populationnel: L'épidémiologie permet d'identifier à ce niveau certains groupes de la population les plus à risque vis-à-vis d'une infection et aider à la mise en place d'actions de santé publique spécifiques.
- Au niveau écologique : A ce niveau, l'épidémiologie « écologique » est nécessaire pour étudier la dynamique des épidémies dans de larges populations en se focalisant sur l'étude de la prévalence/incidence et des caractéristiques populationnelles. Elle va permettre de définir et d'évaluer des plans d'actions à grande échelles.

Les données onusiennes en nous renseignant de manière assez homogène et fiable sur le niveau national des grandes maladies infectieuses, sont des données uniques pour étudier à l'échelle écologique la dynamique et les déterminants importants des grandes maladies infectieuses. La partie qui suit illustre ce concept en appliquant des méthodes écologiques à des données libres concernant la mortalité juvénile-infantile, le paludisme et les multi-résistances aux antituberculeux pour identifier et/ou hiérarchiser leurs déterminants.

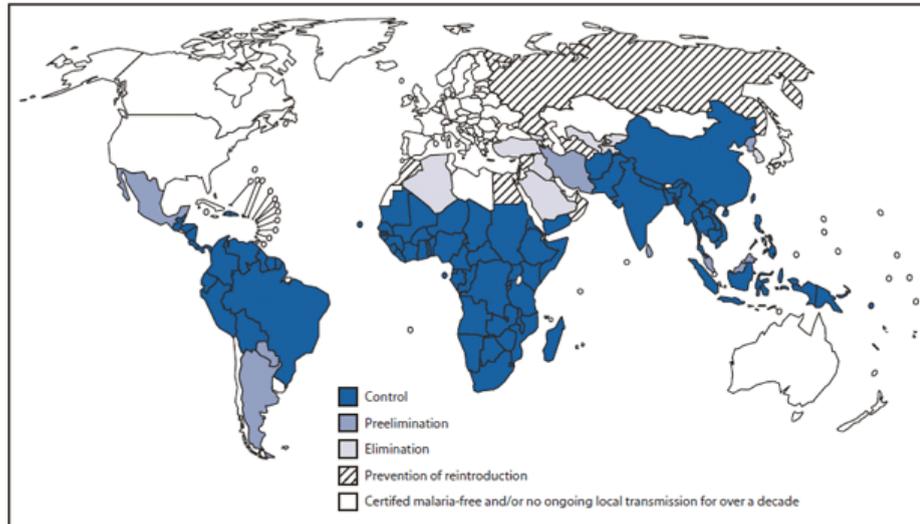
8.5. Etude des interactions paludisme / géohelminthiases à l'échelle populationnelle

Le paludisme reste depuis longtemps un des problèmes majeurs de santé publique à l'échelle mondiale et touche particulièrement les pays du Sud (Figure 41). En 2008, cette maladie a causé près de 247

Partie 2 : Travaux scientifiques

millions de cas et près d'un million de morts [6]. *Plasmodium falciparum* et *Plasmodium vivax* sont les plasmodiums les plus communs, *P.falciparum* étant le plus mortel.

Figure 41: Carte en 2008 des pays exempts de paludisme et des pays endémiques en phases de contrôle, de pré-élimination, d'élimination et de prévention d'une réintroduction. Adaptée de [7].



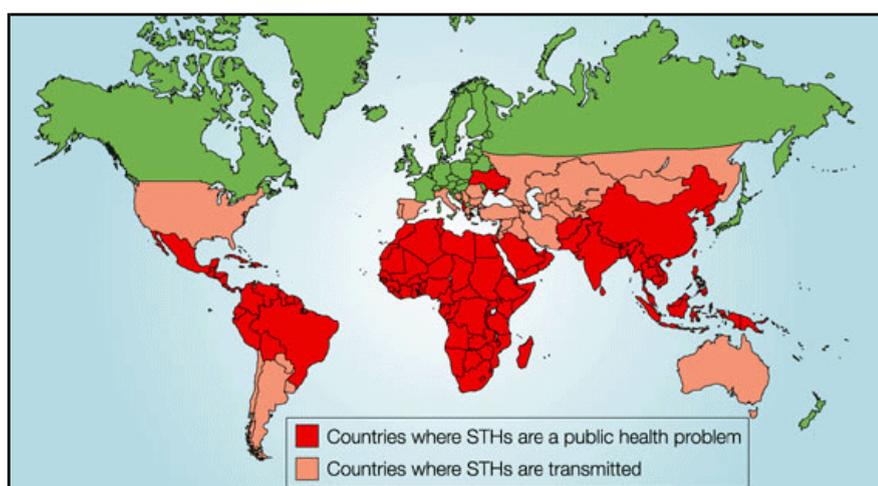
Il a été montré que le paludisme est à l'origine de pertes économiques importantes, et diminue fortement le produit intérieur brut (PIB) des pays ayant des niveaux élevés de transmission (jusqu'à 1,3%). Sur le long terme, ces pertes annuelles cumulées sont à l'origine de différences substantielles de PIB entre les pays avec et sans paludisme [8]. Les coûts de santé du paludisme sont considérables. Dans certains pays fortement touchés, le paludisme représente :

- Jusqu'à 40% des dépenses de santé publique
- 30% à 50% des hospitalisations
- Jusqu'à 60% des consultations externes.

Le paludisme affecte de manière disproportionnée les populations pauvres qui ne peuvent s'offrir un traitement ou ayant un accès limité aux soins de santé.

Parallèlement à cette maladie, Les géohelminthiases, plus communément appelées vers intestinaux, sont les infections les plus répandues dans le monde et touchent préférentiellement les communautés les plus démunies (Figure 42). Les agents causaux des géohelminthiases sont les vers suivants: *Ascaris lumbricoides*, *Trichuris trichiura* et les ankylostomes. Des estimations récentes suggèrent que *A. lumbricoides* infecte plus de 1 milliard de personnes, *T. trichiura* 795 millions et les ankylostomes (*Ancylostoma duodenale* et *Necator americanus*) 740 millions de personnes [9]. Le plus grand nombre des transmissions des helminthiases surviennent en Afrique sub-saharienne, dans les Amériques, en Chine et en Asie Orientale. L'infection est causée par l'ingestion d'œufs provenant de sols contaminés (*A. lumbricoides* et *T. trichiura*) ou par la pénétration active de la peau par des larves dans le sol (*ankylostomes*). Les géohelminthiases produisent un large éventail de symptômes, comprenant des manifestations intestinales (diarrhées, douleurs abdominales), des malaises et une faiblesse générale, qui peuvent affecter les capacités de travail et d'apprentissage et nuire à la croissance physique. Les ankylostomes provoquent des saignements intestinaux chroniques pouvant entraîner des anémies sévères.

Figure 42 : Répartition mondiale et statut de contrôle des géohelminthiases en 2004. Adaptée de [10].



Les nématodes gastro-intestinaux ainsi que le paludisme à *Plasmodium falciparum* ont partagé le même réservoir humain depuis la nuit des temps [11]. Chez l'homme, les nématodes gastro-intestinaux sont connus pour être des immuno-modulateurs déclenchant une réponse immunitaire Th2 ainsi qu'une réponse immunitaire de type T-reg [12]. Bien que les réponses immunitaires déclenchées par les nématodes gastro-intestinaux semblent stéréotypées, leurs amplitudes et modalités varient au sein du réservoir humain. Les infections par ces nématodes gastro-intestinaux sont rarement associées à une mortalité importante mais peuvent cependant induire une morbidité non négligeable dans le cadre de la malnutrition, de l'anémie due à des saignements chroniques, ou de complications opératoires.

Le paludisme, le parasite humain le plus mortel, a évolué dans ce contexte hématologique et immunologique particulier. Durant ces dix dernières années, un nombre important d'études se sont attachées à décrire les conséquences cliniques consécutives de l'interaction entre le paludisme et ces nématodes gastro-intestinaux [13, 14]. Il semblerait actuellement que tous les parasites intestinaux n'aient pas la même action sur le paludisme. Bien que certaines études n'observent aucune interactions [15, 16], un nombre important tendent pourtant à montrer une tendance à une certaine protection vis-à-vis du paludisme induite par les nématodes gastro-intestinaux, et plus particulièrement associée à *Ascaris lumbricoides* [17-22], ou au contraire à une possible augmentation de l'incidence du paludisme [23, 24]. Les mécanismes précis de cette interaction sont toujours insuffisamment compris. De nombreux mécanismes immunologiques hypothétiques ainsi que d'autres hypothèses concernant les conséquences hématologiques de l'anémie sur l'attractivité de l'homme vis-à-vis des moustiques anophèles, sont avancés [14].

Malgré un intérêt récent de la communauté scientifique pour ce sujet, des études de terrain supplémentaires sont encore nécessaires pour élucider l'orientation exacte de ces interactions. En attendant d'autres résultats, l'étude de ces interactions à l'échelle populationnelle pourrait permettre de donner un éclairage nouveau sur cette thématique.

Dans cette perspective, des données de prévalence des nématodes gastro-intestinaux publiées dans la littérature et les données officielles de l'OMS sur les incidences nationales de paludisme de 108 pays aux alentours de l'année 2003 ont été agrégées. A l'aide de ces dernières, l'hypothèse que le niveau des incidences

nationales du paludisme sont elles aussi influencées par les prévalences des différents helminthes gastro-intestinaux, après ajustement par le niveau socio-économique et le climat, a été testée.

8.6. Etude du rôle de la corruption sur la mortalité juvéno-infantile en 2008

La corruption est la perversion ou le détournement d'un processus ou d'une interaction avec une ou plusieurs personnes dans le dessein, pour le corrupteur, d'obtenir des avantages ou des prérogatives particulières ou, pour le corrompu, d'obtenir une rétribution en échange de sa bienveillance. Elle conduit en général à l'enrichissement personnel du corrompu ou à l'enrichissement de l'organisation corruptrice. C'est une pratique qui peut être tenue pour illicite selon le domaine considéré (commerce, affaires, politique...).

Selon Transparency International, une organisation internationale surveillant et agissant dans le domaine de la corruption, « la corruption consiste en l'abus d'un pouvoir reçu en délégation à des fins privées » [25].

Plus particulièrement dans le domaine de la santé, la corruption est un problème complexe qui menace l'impact des investissements publics, la performance, l'accès et l'équité des soins et des services de santé [26, 27]. De plus en plus, les dirigeants, et les citoyens du monde entier, reconnaissent les effets pernicieux de la corruption, et la nécessité de prendre des mesures [27-29]. Au niveau des individus et des ménages, il existe des preuves des effets négatifs de la corruption sur la santé et le bien-être des citoyens [30, 31]. L'impact sur la santé de la corruption est une résultante complexe des nombreuses opportunités de corruption qui peuvent survenir à chaque étape du processus de soin et de la mise en place des politiques de santé (Tableau 5).

Le rapport de 2006 de « Transparency International » affirme que la corruption est un des éléments principaux perturbant la bonne réalisation des Objectifs de Développement du Millénaire (ODM) et qu'elle serait une des causes principales du retard et de l'impossibilité de la communauté mondiale à atteindre ces même ODM [27].

Par-dessus cela, les impacts sur la santé de la corruption ne peuvent pas se résumer uniquement à ceux agissant directement sur l'accès et la qualité des services de santé. Par exemple, près de 1,2 milliard de personnes dans le monde n'ont pas un accès suffisant à l'eau potable et plus de 2,6 milliards ne disposent pas d'un système d'assainissement adéquat, avec des conséquences dévastatrices pour le développement et la réduction de la pauvreté. Dans les pays en développement, environ 80 % des problèmes de santé peuvent être rattachées à un approvisionnement en eau et/ou à un systèmes d'assainissement inadéquat [32]. Les conséquences humaines de cette crise de l'eau sont dévastatrices et affectent majoritairement les pauvres et les femmes. Les experts s'accordent à dire que cette crise de l'eau est principalement une crise de gouvernance, dont la corruption pourrait être l'une des principales causes et catalyseurs affectant le secteur de l'eau et de l'assainissement dans son ensemble, de la gestion des ressources jusqu'aux services [33]. Plus globalement la corruption affecte tous les domaines de la vie, allant de l'éducation jusqu'aux performances économiques [28].

Chapitre 8- Utilisation des statistiques onusiennes : Applications à l'étude des déterminants de la dynamique des résistances aux antituberculeux, du paludisme et de la mortalité juvéno-infantile.

Tableau 5: Types de corruption dans le secteur de la santé. Adapté de [26].

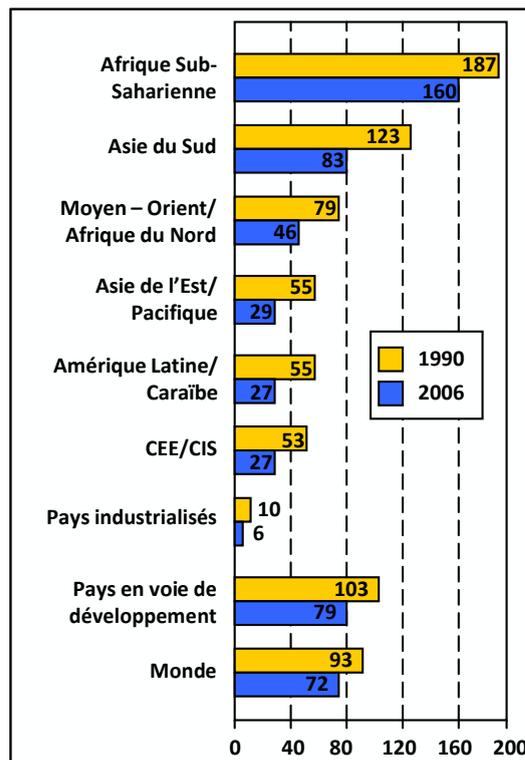
Area or process	Types of corruption and problems	Results
Construction and rehabilitation of health facilities	<ul style="list-style-type: none"> ▶ Bribes, kickbacks and political considerations influencing the contracting process ▶ Contractors fail to perform and are not held accountable 	<ul style="list-style-type: none"> ▶ High cost, low quality facilities and construction work ▶ Location of facilities that does not correspond to need, resulting in inequities in access ▶ Biased distribution of infrastructure favouring urban- and elite-focused services, high technology
Purchase of equipment and supplies, including drugs	<ul style="list-style-type: none"> ▶ Bribes, kickbacks and political considerations influence specifications and winners of bids ▶ Collusion or bid rigging during procurement ▶ Lack of incentives to choose low cost and high quality suppliers ▶ Unethical drug promotion ▶ Suppliers fail to deliver and are not held accountable 	<ul style="list-style-type: none"> ▶ High cost, inappropriate or duplicative drugs and equipment ▶ Inappropriate equipment located without consideration of true need ▶ Sub-standard equipment and drugs ▶ Inequities due to inadequate funds left to provide for all needs
Distribution and use of drugs and supplies in service delivery	<ul style="list-style-type: none"> ▶ Theft (for personal use) or diversion (for private sector resale) of drugs/supplies at storage and distribution points ▶ Sale of drugs or supplies that were supposed to be free 	<ul style="list-style-type: none"> ▶ Lower utilization ▶ Patients do not get proper treatment ▶ Patients must make informal payments to obtain drugs ▶ Interruption of treatment or incomplete treatment, leading to development of anti-microbial resistance
Regulation of quality in products, services, facilities and professionals	<ul style="list-style-type: none"> ▶ Bribes to speed process or gain approval for drug registration, drug quality inspection, or certification of good manufacturing practices ▶ Bribes or political considerations influence results of inspections or suppress findings ▶ Biased application of sanitary regulations for restaurants, food production and cosmetics ▶ Biased application of accreditation, certification or licensing procedures and standards 	<ul style="list-style-type: none"> ▶ Sub-therapeutic or fake drugs allowed on market ▶ Marginal suppliers are allowed to continue participating in bids, getting government work ▶ Increased incidence of food poisoning ▶ Spread of infectious and communicable diseases ▶ Poor quality facilities continue to function ▶ Incompetent or fake professionals continue to practice
Education of health professionals	<ul style="list-style-type: none"> ▶ Bribes to gain place in medical school or other pre-service training ▶ Bribes to obtain passing grades ▶ Political influence, nepotism in selection of candidates for training opportunities 	<ul style="list-style-type: none"> ▶ Incompetent professionals practicing medicine or working in health professions ▶ Loss of faith and freedom due to unfair system
Medical research	<ul style="list-style-type: none"> ▶ Pseudo-trials funded by drug companies that are really for marketing ▶ Misunderstanding of informed consent and other issues of adequate standards in developing countries 	<ul style="list-style-type: none"> ▶ Violation of individual rights ▶ Biases and inequities in research
Provision of services by medical personnel and other health workers	<ul style="list-style-type: none"> ▶ Use of public facilities and equipment to see private patients ▶ Unnecessary referrals to private practice or privately owned ancillary services ▶ Absenteeism ▶ Informal payments required from patients for services ▶ Theft of user fee revenue, other diversion of budget allocations 	<ul style="list-style-type: none"> ▶ Government loses value of investments without adequate compensation ▶ Employees are not available to serve patients, leading to lower volume of services and unmet needs, and higher unit costs for health services actually delivered ▶ Reduced utilization of services by patients who cannot pay ▶ Impoverishment as citizens use income and sell assets to pay for health care ▶ Reduced quality of care from loss of revenue ▶ Loss of citizen faith in government

Parmi les problèmes de santé majeurs dans le monde, la mortalité juvéno-infantile est l'un des plus préoccupants. En 2006, pour la première fois depuis que l'on calcule ces données, le nombre d'enfants qui meurent avant leur cinquième anniversaire est passé en dessous de la barre des 10 millions, à 9,7 millions. Ce

Partie 2 : Travaux scientifiques

Le tournant décisif est l'aboutissement d'un recul régulier, au niveau mondial, du taux de mortalité des moins de 5 ans depuis 1960. En 2008 on estimait ainsi à 8,795 millions le nombre de décès d'enfant de moins de 5 ans dans le monde entier. De nombreux pays affichent toujours cependant des taux élevés de mortalité infantile, en particulier en Afrique subsaharienne et en Asie du Sud. Ces dernières années, ils ont peu progressé, ou pas du tout, dans la réduction du nombre de décès d'enfants (Figure 43).

Figure 43: Tendances du taux de mortalité des moins de 5 ans (pour 1 000 naissances vivantes), par région (1990-2006). Adaptée de [1].



La réduction de la mortalité est cruciale pour les ODM qui visent à diminuer le taux de mortalité infantile de deux tiers entre 1990 et 2015 [34, 35]. Par son influence sur les services de santé, sur l'accès et la distribution d'eau potable / assainissement, sur l'éducation et sur le développement économique, la corruption pourrait potentiellement avoir un grand impact sur la dynamique de la mortalité juvéno-infantile. Toutefois, les données quantitatives sur ce sujet à une échelle populationnelle sont pauvres. L'étude la plus récente quantifiant le lien spécifique entre ces deux variables, est celle de Abed et de ses collaborateurs [29]. Cette étude, réalisée en 2002, a observé un lien significatif entre le niveau de corruption et la mortalité juvéno-infantile, mais n'incorporait qu'un nombre limité de variables d'ajustement et avait été faite sur des données vieilles de 10 ans. Une vision plus récente de l'impact mondial de la corruption sur la mortalité juvéno-infantile est nécessaire pour la mise en place de programmes efficaces d'amélioration de la survie des enfants.

Dans cette perspective, les taux nationaux de la mortalité des enfants de moins de 5 ans, les facteurs de risque classiques de la mortalité juvéno-infantile et le niveau de corruption ont été réunis aux alentours de l'année 2008. En utilisant des méthodes de régression linéaire, cette étude visait à 1) évaluer à l'échelle pays, les impacts actuels de la corruption sur le taux de mortalité globale des enfants, après ajustement avec un

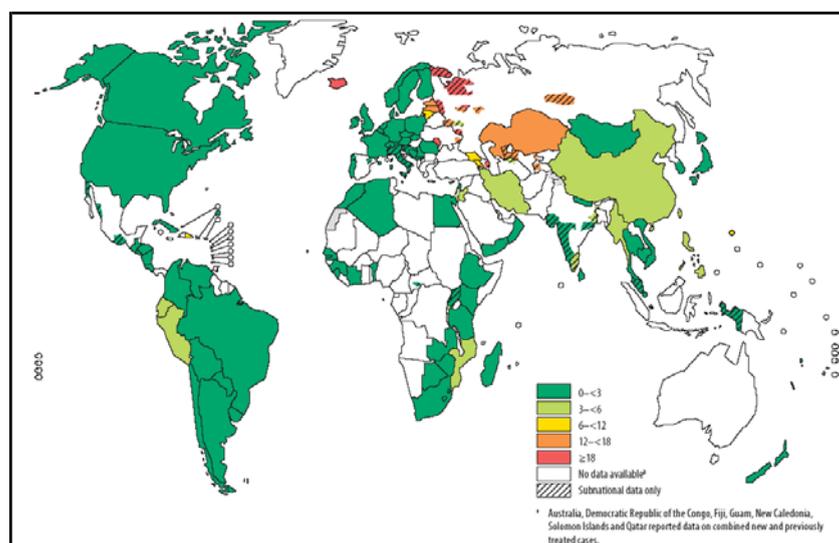
Chapitre 8- Utilisation des statistiques onusiennes : Applications à l'étude des déterminants de la dynamique des résistances aux antituberculeux, du paludisme et de la mortalité juvénile-infantile.

grand ensemble de variables, et 2) calculer approximativement le nombre de décès d'enfants attribuables à la corruption.

8.7. Etude du rôle de la corruption sur les multirésistances aux anti-tuberculeux en 2008

La tuberculose multirésistante (MDR-TB) est un autre problème de santé majeur pour lequel l'étude d'un potentiel lien avec la corruption serait intéressante en termes de santé publique. Les MDR-TB figurent parmi les éléments les plus préoccupants de la pandémie de résistances antimicrobiennes en raison du risque élevé de décès chez les patients infectés [36, 37]. Les MDR-TB sont causées par des mycobactéries de type *M.tuberculosis* qui sont résistantes à au moins l'isoniazide et la rifampicine, les médicaments antituberculeux les plus efficaces. Les MDR-TB résultent d'une primo-infection par des bactéries résistantes ou peuvent se développer dans le cadre du traitement d'un patient. Avec l'apparition récente des tuberculoses extrêmement résistantes (XDR-TB), le contrôle des MDR-TB est devenu l'un des défis les plus importants de santé globale [38-40]. En 2008, 440 000 cas de MDR-TB a émergé dans le monde. Les MDR-TB représentent maintenant 3,6% de l'ensemble des cas de tuberculose et sont à l'origine de 150 000 décès. La plupart des cas surviennent en Chine ou en Inde [41]. Les MDR-TB se retrouvent partout mondialement, mais des variations significatives de la prévalence ont été observées dans différentes régions géographiques (Figure 44).

Figure 44 : Distribution de la proportion des cas de MDR-TB parmi les nouveaux cas de tuberculose, 1994-2009. Adaptée de [42].



De manière globale, l'émergence des résistances aux antimicrobiens est un problème complexe provoqué par de nombreux facteurs interdépendants, dont beaucoup sont reliés à l'utilisation des antimicrobiens. Il est maintenant admis que l'utilisation des antimicrobiens est le facteur le plus important dans l'explication de l'augmentation des résistances. L'utilisation d'antimicrobiens est influencée par une interaction complexe entre les connaissances, les attentes des patients et prescripteurs, les incitations économiques, les caractéristiques des systèmes de santé, l'environnement réglementaire et la disponibilité des ressources [43]. Ainsi, selon Caminero et collaborateurs [40] la prévalence des MDR-TB peut être fortement

liée à la gestion sanitaire actuelle de la tuberculose et à la dynamique de la transmission des tuberculoses multirésistantes. Au niveau communautaire, ces derniers distinguent quatre grands groupes de facteurs potentiels associés à la sélection de résistance et de génération de MDR-TB dans des conditions épidémiques qui sont: 1) une non-application des stratégies DOTS 2) une offre inadéquate ou mauvaise concernant les médicaments, 3) une prise inadéquate des médicaments par les patients et 4) un ensemble d'autres facteurs comprenant par exemple l'ampleur de l'épidémie VIH ou le faible contrôle de l'infection dans les centres de santé ou les hôpitaux.

En raison d'une forte imbrication entre les conséquences de la corruption et les facteurs potentiels favorisant la diffusion des tuberculoses multirésistantes, Certains auteurs ont avancé dans la littérature que la corruption pourrait avoir un impact profond sur les prévalences des résistances aux antituberculeux dans certaines régions du monde [44, 45]. Malheureusement, les données quantitatives au niveau populationnel sur ce sujet sont très pauvres. Dans cette perspective et de manière identique à l'étude des déterminants de la mortalité juvéno-infantile, les données de prévalence des MDR-TB, les facteurs potentiels de risque d'une forte prévalence des MDR-TB, et le niveau de corruption ont été recueillies aux alentours de l'année 2008.

8.8. Origine et contexte de ces études

Avant son arrivée en Guyane, le Pr. Mathieu Nacher, a réalisé une thèse de science ainsi qu'un post doctorat en Thaïlande sur le paludisme. Durant cette période, il s'est focalisé plus particulièrement sur l'étude du lien possible entre infection par les helminthes gastro-intestinaux et sévérité des crises à *Plasmodium falciparum*. Il fut l'un des premiers à mettre en évidence une tendance à une certaine protection vis-à-vis du paludisme induite par les nématodes gastro-intestinaux, et plus particulièrement par *Ascaris lumbricoides*.

Etant toujours très impliqué et passionné par cette thématique, l'idée d'étudier le rôle des nématodes gastro-intestinaux sur les incidences du paludisme à une échelle populationnelle pour apporter de nouveaux éléments aux débats actuels sur les interactions helminthes/paludisme germa petit à petit et se concrétisa sous la forme de cette étude.

Les deux articles sur la corruption découlent directement de différentes réflexions faites avec Le Pr Bernard Carme, le Pr Mathieu Nacher et d'autres collègues travaillant au sein du CIC-EC Antilles Guyane. Dans le cadre de l'étude des déterminants de la mortalité juvéno-infantile, consécutivement à la découverte du « CPI », un indice national fiable du niveau de corruption perçue, et suite à de nombreuses discussions avec Astrid Van-Melle, démographe de formation, la zone d'ombre autour de la quantification de l'impact de la corruption sur la mortalité juvéno-infantile nous est rapidement apparue. Dans un deuxième temps, des discussions informelles de cette étude avec le Dr. Antoine Adenis, très investis dans l'étude épidémiologique de la tuberculose, nous ont persuadés de la justesse d'une « réplique » de la méthodologie utilisée précédemment pour quantifier dans un deuxième temps l'impact de la corruption sur les prévalences nationales des tuberculoses multirésistantes.

8.9. Etude 10: Une analyse par arbre de régression au niveau national retrouve une association significative entre de fortes prévalences d'*Ascaris lumbricoides* et des incidences plus faibles de paludisme

Introduction: Un débat persiste actuellement sur l'impact des nématodes gastro-intestinaux sur le paludisme, et en particulier sur la possibilité d'une augmentation des formes graves du paludisme après un traitement anti-helminthique pour *Ascaris lumbricoides*.

Matériel/méthodes: Dans le but de tester cette hypothèse, des données nationales sur la prévalence du parasitisme de différents types d'helminthes ainsi que des données nationales sur l'incidence du paludisme de 108 pays au alentour de l'année 2003 ont été analysées en utilisant une méthode d'arbre de régression. Le niveau économique ainsi que le type de climat des pays ont aussi été inclus dans le modèle comme variables d'ajustement.

Résultats: Les résultats de cette analyse montre que dans les pays ayant un climat équatorial et un produit intérieur brut par personne et par an inférieur à 660 dollars US, une prévalence d'ascaris $\geq 25\%$ est associée avec une réduction par 10 de l'incidence du paludisme ($p=0.01$)

Discussion: Bien qu'il y ait des biais inhérents aux études écologiques, cette analyse suggère que l'amplitude de l'effet d'ascaris sur le paludisme pourrait être considérable et renforce le besoin d'une réflexions autour de la mise en place de grandes campagnes de vermifugeage dans les zones endémiques pour le paludisme à faibles ressources médicales.

Article soumis le 7 juillet 2011 à Plos One

***Manuscript**

[Click here to download Manuscript: main document.doc](#)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

**Regression tree analysis of national gastrointestinal
nematode prevalences and malaria incidences: is there a
link?**

Running title: Worms and malaria

Matthieu HANF* °, Bernard CARME* °, Mathieu NACHER*‡ °

* Centre d'Investigation Clinique Epidémiologie Clinique Antilles Guyane CIC-EC INSERM

CIE 802, Cayenne General Hospital, Cayenne, French Guiana.

° Department of Parasitology and Mycology – Team EA 3593, Université des Antilles et de la

Guyane, Cayenne, French Guiana

‡ COREVIH, Cayenne General Hospital, Cayenne, French Guiana.

Corresponding Author : Matthieu HANF, Centre d'Investigation Clinique Epidémiologie
Clinique Antilles Guyane CIC-EC INSERM CIE 802, Cayenne General Hospital, EHPAD,
Avenue des Flamboyants, BP 6006 97300 Cayenne, French Guiana, France. E-mail address:

matthieu@hanf.fr

Abstract

Background.

There has been a debate about the impact of gastrointestinal nematodes infections on malaria, notably on the possibility of an increase of severe malaria after anthelmintic treatment of *Ascaris lumbricoides*.

Materials and methods

In order to test this hypothesis, data on the prevalence of different geohelminths and data on malaria incidence from 108 countries *circa* 2003 were analyzed using regression trees analysis. Data on GDP per capita and the type of climate in the country were also included in the model.

Results

The results showed that in countries with an equatorial climate and a GDP per capita <660 USD per year, an *Ascaris* prevalence $\geq 25\%$ was associated with a tenfold reduction of malaria incidence, ($p=0.01$).

Discussions

Although there are limitations in ecological studies, this suggests that the magnitude of the effect of *Ascaris* may be considerable, and reinforces the need for caution when deworming in malaria endemic areas with poor medical facilities.

Keywords

Epidemiology, Statistics, Malaria, Helminthiasis-Intestinal, Modeling, Plasmodium, Public Health

Introduction

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Gastrointestinal nematodes and *falciparum* malaria have shared the same human host for thousands of years[1]. GI nematodes are potent immunomodulators, leading to a bias towards a Th2 immune response and a significant T-reg response[2]. Although the immune responses induced by different GI nematodes seem stereotypical, their size and their particular behaviour within the human host vary. They rarely lead to death but are known to induce significant morbidity such as malnutrition, anaemia due to chronic bleeding, and surgical complications. Malaria, the deadliest human parasite, has evolved in this immunological and haematological context. For the past decade, a growing number of studies have studied the outcome of this interaction between GI nematodes and malaria[3, 4]. Although the immune response profile seems stereotypical, all worms do not have the same outcome when it comes to malaria. While some studies have not observed any interactions [5, 6], a growing number of results have shown some trends towards protection from malaria, notably by *Ascaris* [7-12], and possibly increased malaria incidence [13, 14]. The precise mechanisms of these interactions are still poorly understood. There are a variety of immunological hypothetical mechanisms and other hypotheses regarding the haematological consequences of anaemia on host attractiveness for anopheline mosquitoes [4].

Despite the recent interest in the subject, there is still a need for more field data to determine the broad trends of this interaction. In the same perspective, we gathered data from published studies on gastrointestinal nematode prevalence and on malaria incidence in 108 countries, *circa* 2003. From these studies, we aimed to test the hypothesis that, after taking into account climate and socioeconomic level, ecological data on malaria incidence trends could also be predicted by the prevalence of different gastrointestinal nematodes. The analysis showed that *Ascaris lumbricoides* indeed seemed to have a negative association with malaria.

Materials and methods

This study was performed at a population level (country scale) by merging freely available data from the published literature with data compiled by international organisations.

National Helminth data

To assess the national helminth prevalence, data computed in 2003 by de Silva et al. were used [15]. In this study, the national prevalences of Ascariasis, Trichuriasis and hookworm were categorized by Silva et al. as follow: 0 - 4.9%, 5-24.9%, 25-44.9%, 45-59.9%, $\geq 60\%$. Unfortunately, we could not obtain global national data for China and for India. For unavailable national prevalences (NA) a supplementary category was created for the prevalence of each helminth species.

National Malaria data

National numbers of reported malaria cases were extracted from the WHO Statistical Information System (WHOSIS) for 2003 [16]. The data did not specify plasmodial species. For some countries, when 2003 data was not available, the data closest to 2003 was used.

To compute the national incidence, the national population at the time of the malaria counts computation was assessed using data extracted from the International Data Base (IDB) of the U.S Census Bureau [17].

Due to the high variability of the incidence between countries, statistical calculations were made using the base 10 logarithm of malaria incidence.

National Climate and socio-economic status

It is important to consider the risk factors associated with both types of infections in studies describing helminth-malaria interactions [3]. Identifying common risk factors for

1 malaria and helminth infections is important because apparent associations between the two
2 infections may be due to common social or environmental factors rather than a true biological
3 interaction. Factors that are thought to be involved in the geographical distributions of malaria
4 and helminth distributions are mainly climate and the socio-economic status (SES).
5
6
7
8
9

10 For each country, the presence of each type of climate on the national territory was
11 included in the analysis. Climates were categorized based on the revised Köppen-Geiger
12 climate classification [18]. Three derived variables were created: presence of an equatorial
13 climate on the national territory, presence of an arid climate on the national territory, and
14 presence of other climates (warm temperate, snow and polar climates)
15
16
17
18
19
20

21 Data on the 2003 gross domestic product (GDP) per capita were included to adjust for
22 national SES status. This was available at the national level from the UN statistics database
23 [19] and the annual GDP per capita for each country was calculated.
24
25
26
27
28
29
30

31 **Statistical methods**

32 To summarize the world distributions of malaria and helminth infections, prevalence
33 and incidence by country were plotted on a world map. The distributions of malaria incidence
34 within each helminth, climate, and SES groups, were plotted.
35
36
37
38
39
40

41 To examine the nature of the relationships between malaria incidence and predictor
42 variables, a regression tree analysis (RTA) was used [20]. These non-parametric methods are
43 ideally suited for the analysis of ecological or epidemiological data with hierarchical and
44 complex relationships among the predictor variables (colinearity) and between predictor and
45 response variables [20].
46
47
48
49
50
51

52 RTA is a non-parametric technique that recursively partitions a dataset into subsets
53 that are increasingly homogeneous with regard to the response [21]. RTA produces a set of
54 decision rules on predictor variables that are easily interpreted as a dendrogram. Most
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

implementations of RTA have a selection bias towards predictors with many possible splits and tend to overfit to a given dataset by creating splits that do not significantly reduce the variance [21]. Trees are typically pruned back to include only those partitions assumed to be valuable beyond the sample data [20]. An implementation of RTA called unbiased conditional inference trees was used. It requires a statistically significant difference ($p < 0.05$) adjusted with Bonferroni correction, in order to partition the data, which minimizes bias and prevents over-fitting and the need for pruning [21].

The party package within the R statistical language was used to implement RTA [21].

Results

National data on malaria levels were found for 108 countries on the WHO Statistical Information System. Figure 1 shows the data used in this analysis per country for malaria incidence, and the prevalence of Ascariasis, Trichuriasis and hookworm.

Strong visual disparities in malaria incidence appeared when it was stratified by each nematodes prevalence, by GDP per capita and by climate (Figure 2).

After RTA calculation, the regression tree in figure 3 was obtained. In this tree, the first partition was based on the presence of an equatorial climate in the country ($p < 0.001$).

In the node defined by the absence of equatorial climate, the second split was again based on climate ($p < 0.038$). In countries where other climates could be found, low malaria incidence was reported (median: 6 cases / year * 100 000 persons). Intermediate incidence was reported for countries where other climates were absent (median: 599 cases / year * 100 000 persons). For countries with equatorial climates the second partition was based on national SES status ($p = 0.002$). If the GDP per capita was superior to 660 dollars / person / year, the national malaria incidence level was intermediate (median: 177 cases / year * 100 000 persons). For countries with a lower SES status, the last partition was based on the

1 national *Ascaris* prevalence ($p = 0.012$). A high malaria incidence level was associated with
2 an *Ascaris* prevalence $<25\%$ (median: 11019 cases / year * 100 000 persons) and an
3
4 intermediate malaria incidence was associated with an *Ascaris* prevalence $\geq 25\%$ (median: 803
5
6 cases / year * 100 000 persons).
7
8
9

10 Discussion

11 Ecological studies have their limitations (Bias, confounding). Because the malaria
12 incidences used in this analysis are reported and not estimated, the data may have been
13 heterogeneous in quality. The available data in a country may not have been sufficient to
14 reflect the local variations of all the studied variables. Furthermore, the present study used
15 incidence as the endpoint of analysis, which is subject to many factors in particular health
16 intervention. Interventions may vary from place to place, and this variability was not taken
17 into account in this analysis. However, this is the only data that is available to study the
18 problem at this scale.
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33

34 Although there are limitations, the present study has the merit to approach a
35 controversial problem from a different and innovative angle.
36
37

38 The regression trees analysis presented here, after taking into account climate and
39 GDP per capita, seems to point towards a significant protective link between *Ascaris*
40 *lumbricoides* and malaria. Populations with less than 25% prevalence of *Ascaris* had a malaria
41 incidence that was 10 fold that of populations with *Ascaris* prevalence $\geq 25\%$. Simple plots
42 showed that, on the contrary, hookworm prevalences $\geq 25\%$ seemed to be associated with
43 increased malaria incidence. Other ecological studies have suggested a geographical overlap
44 of hookworm and malaria[22]. Individual studies have also raised this possibility [4, 12, 23,
45 24]. This is compatible with the idea that there are different mechanisms at work, the strong
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Partie 2 : travaux scientifiques

1 immunomodulation by *Ascaris* and haematological cues increasing attractiveness of the
2 hookworm-infected host for malaria.
3

4
5 However, the model including all the variables only identified *Ascaris* prevalence as
6
7 predictive of malaria incidence. Due to the small sample size (n=108), the model may not
8
9 have detected weaker relations. In addition, cerebral malaria was 6.6 times less frequent in
10
11 *Ascaris* infected patients relative to those without *Ascaris* [4]. This was stronger than
12
13 hookworm's link with malaria (twofold increase)[4]. This difference of magnitude, and the
14
15 fact that the present data is not data on coinfections *per se* but ecological data on single
16
17 parasite prevalences, could have made it more difficult to demonstrate an impact of
18
19 hookworm on malaria incidence. In addition, the possible complexity of hookworm effects on
20
21 the host (immunomodulation and sometimes anemia), depending on the relative magnitude of
22
23 each effect, could lead to contrasting results. The intense immunomodulation of *Ascaris*, and
24
25 the absence of blood spoliation by this parasite may explain this finding. There have been
26
27 generalizations on the negative impact of worms on malaria[25] and debates on the risk of
28
29 overlooking a protective effect of *Ascaris* [26, 27]. Studies showing protection from malaria
30
31 by *Ascaris* had different designs and alternatively focussed on severe malaria [8, 28, 29],
32
33 parasitemia [9], incidence or prevalence [7], [11] of malaria.
34
35

36
37 Here, the detection of a negative association of *Ascaris* with malaria, of this
38
39 magnitude, and at this observation scale, suggests a very significant protective role of *Ascaris*
40
41 in poor countries with an equatorial climate, which often bear the brunt of malaria. This,
42
43 again, raises the question of unforeseen effects of mass deworming.
44
45

46
47 However, other helminth prevalences, such as filariases and schistosomiases should also be
48
49 included in further models. A study in West Africa, showed a negative association between
50
51 *Wuchereria bancrofti* and malaria[30]. One possible explanation for this observation was the
52
53 immunological interaction between filarial and malarial parasites.
54
55
56
57
58
59
60
61
62
63
64
65

1 The range of epidemiological, medical, socioeconomic, and climatic conditions,
2 further complicated by different study designs, leads to discrepancies between publications.
3
4 Given the potential importance of parasitic interactions, studies such as the present one are an
5 important addition to the debate on the public health consequences of multiparasitism. The
6
7 internet and the increasing availability of data on parasitic diseases should facilitate this.
8
9

10 11 12 13 14 **Competing interests**

15 The authors declare that they have no competing interests.
16
17

18 19 20 21 **Ethical approval**

22 No need.
23
24

25 26 27 28 29 **Author's contributions**

30 MH was responsible for the research design and data collection, performed data analysis and
31 interpretation, and prepared the manuscript. BC participated in the interpretation of data and
32 manuscript revision as well. MN, initiated the study, provided guidance on the data analysis
33 and was involved in the interpretation of data and manuscript revision. All authors read and
34 approved the final manuscript.
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

References

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1. Jackson JA, Friberg IM, Little S, Bradley JE: **Review series on helminths, immune modulation and the hygiene hypothesis: immunity against helminths and immunological phenomena in modern human populations: coevolutionary legacies?** *Immunology* 2009, **126**(1):18-27.
2. Bradley JE, Jackson JA: **Immunity, immunoregulation and the ecology of trichuriasis and ascariasis.** *Parasite Immunol* 2004, **26**(11-12):429-441.
3. Mwangi TW, Bethony JM, Brooker S: **Malaria and helminth interactions in humans: an epidemiological viewpoint.** *Ann Trop Med Parasitol* 2006, **100**(7):551-570.
4. Nacher M: **Worms and malaria: blind men feeling the elephant?** *Parasitology* 2008, **135**(7):861-868.
5. Shapiro AE, Tukahebwa EM, Kasten J, Clarke SE, Magnussen P, Olsen A, Kabatereine NB, Ndyomugenyi R, Brooker S: **Epidemiology of helminth infections and their relationship to clinical malaria in southwest Uganda.** *Trans R Soc Trop Med Hyg* 2005, **99**(1):18-24.
6. Bejon P, Mwangi TW, Lowe B, Peshu N, Hill AV, Marsh K: **Helminth infection and eosinophilia and the risk of Plasmodium falciparum malaria in 1- to 6-year-old children in a malaria endemic area.** *PLoS Negl Trop Dis* 2008, **2**(1):e164.
7. Murray MJ, Murray AB, Murray MB, Murray CJ: **Parotid enlargement, forehead edema, and suppression of malaria as nutritional consequences of ascariasis.** *Am J Clin Nutr* 1977, **30**(12):2117-2121.

Chapitre 8- Utilisation des statistiques onusiennes : Applications à l'étude des déterminants de la dynamique des résistances aux antituberculeux, du paludisme et de la mortalité juvéno-infantile.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
8. Nacher M, Gay F, Singhasivanon P, Krudsood S, Treeprasertsuk S, Mazier D, Vouldoukis I, Looareesuwan S: **Ascaris lumbricoides infection is associated with protection from cerebral malaria.** *Parasite Immunol* 2000, **22**(3):107-113.
 9. Brutus L, Watier L, Briand V, Hanitrasoamampionona V, Razanatsoarilala H, Cot M: **Parasitic co-infections: does Ascaris lumbricoides protect against Plasmodium falciparum infection?** *Am J Trop Med Hyg* 2006, **75**(2):194-198.
 10. Brutus L, Watier L, Hanitrasoamampionona V, Razanatsoarilala H, Cot M: **Confirmation of the protective effect of Ascaris lumbricoides on Plasmodium falciparum infection: results of a randomized trial in Madagascar.** *Am J Trop Med Hyg* 2007, **77**(6):1091-1095.
 11. van Eijk AM, Lindblade KA, Odhiambo F, Peterson E, Rosen DH, Karanja D, Ayisi JG, Shi YP, Adazu K, Slutsker L: **Geohelminth Infections among pregnant women in rural western Kenya; a cross-sectional study.** *PLoS Negl Trop Dis* 2009, **3**(1):e370.
 12. Degarege A, Animut A, Legesse M, Erko B: **Malaria severity status in patients with soil-transmitted helminth infections.** *Acta Trop* 2009, **112**(1):8-11.
 13. Nacher M, Singhasivanon P, Yimsamran S, Manibunyong W, Thanyavanich N, Wuthisen R, Looareesuwan S: **Intestinal helminth infections are associated with increased incidence of Plasmodium falciparum malaria in Thailand.** *J Parasitol* 2002, **88**(1):55-58.
 14. Spiegel A, Tall A, Raphenon G, Trape JF, Druilhe P: **Increased frequency of malaria attacks in subjects co-infected by intestinal worms and Plasmodium falciparum malaria.** *Trans R Soc Trop Med Hyg* 2003, **97**(2):198-199.

Partie 2 : travaux scientifiques

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
15. de Silva NR, Brooker S, Hotez PJ, Montresor A, Engels D, Savioli L: **Soil-transmitted helminth infections: updating the global picture.** *Trends Parasitol* 2003, **19**(12):547-551.
16. WHO: **WHO Statistical Information System (WHOSIS)** In.; 2010.
17. US Census Bureau: **International Data Base (IDB)** In.; 2010.
18. Kottek M, Grieser J, Beck C, Rudolf B, Rubel F: **World map of the Köppen-Geiger climate classification updated.** *Meteorol Z* 2006, **15**:259–263.
19. ONU: **UN Common Database.** In.; 2010.
20. De'ath G, Fabricius KE: **Classification and regression trees: a powerful yet simple technique for ecological data analysis.** *Ecology* 2000, **81**:3178-3192.
21. Hothorn T, Hornik K, Zeileis A: **Unbiased Recursive Partitioning: A Conditional Inference Framework.** *Journal of Computational and Graphical Statistics* 2006, **15**(3):651-674.
22. Brooker S, Clements AC, Hotez PJ, Hay SI, Tatem AJ, Bundy DA, Snow RW: **The co-distribution of Plasmodium falciparum and hookworm among African schoolchildren.** *Malar J* 2006, **5**:99.
23. Hillier SD, Booth M, Muhangi L, Nkurunziza P, Kihembo M, Kakande M, Sewankambo M, Kizindo R, Kizza M, Muwanga M *et al*: **Plasmodium falciparum and helminth coinfection in a semi urban population of pregnant women in Uganda.** *J Infect Dis* 2008, **198**(6):920-927.
24. Yatich NJ, Yi J, Agbenyega T, Turpin A, Rayner JC, Stiles JK, Ellis WO, Funkhouser E, Ehiri JE, Williams JH *et al*: **Malaria and intestinal helminth co-infection among pregnant women in Ghana: prevalence and risk factors.** *Am J Trop Med Hyg* 2009, **80**(6):896-901.

Chapitre 8- Utilisation des statistiques onusiennes : Applications à l'étude des déterminants de la dynamique des résistances aux antituberculeux, du paludisme et de la mortalité juvéno-infantile.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

25. Druilhe P, Tall A, Sokhna C: **Worms can worsen malaria: towards a new means to roll back malaria?** *Trends Parasitol* 2005, **21**(8):359-362.

26. Nacher M: **Worms and malaria: noisy nuisances and silent benefits.** *Parasite Immunol* 2002, **24**(7):391-393.

27. Nacher M: **Worms and malaria: resisting the temptation to generalize.** *Trends Parasitol* 2006, **22**(8):350-351; author reply 351-352.

28. Nacher M: **Comment on: Severe malaria attack is associated with high prevalence of Ascaris lumbricoides infection among children in rural Senegal.** *Trans R Soc Trop Med Hyg* 2005, **99**(2):161-163; discussion 164-165.

29. Le Hesran JY, Akiana J, Ndiaye el HM, Dia M, Senghor P, Konate L: **Severe malaria attack is associated with high prevalence of Ascaris lumbricoides infection among children in rural Senegal.** *Trans R Soc Trop Med Hyg* 2004, **98**(7):397-399.

30. Kelly-Hope LA, Diggle PJ, Rowlingson BS, Gyapong JO, Kyelem D, Coleman M, Thomson MC, Obsomer V, Lindsay SW, Hemingway J *et al*: **Short communication: Negative spatial association between lymphatic filariasis and malaria in West Africa.** *Trop Med Int Health* 2006, **11**(2):129-135.

Figure
[Click here to download Figure: Figure 1.doc](#)

Figure 1 : The global distribution of A) malaria reported cases, B) *Ascaris lumbricoides*, C) *Trichuris trichiura* and D) hookworm according to the data used in the analysis.

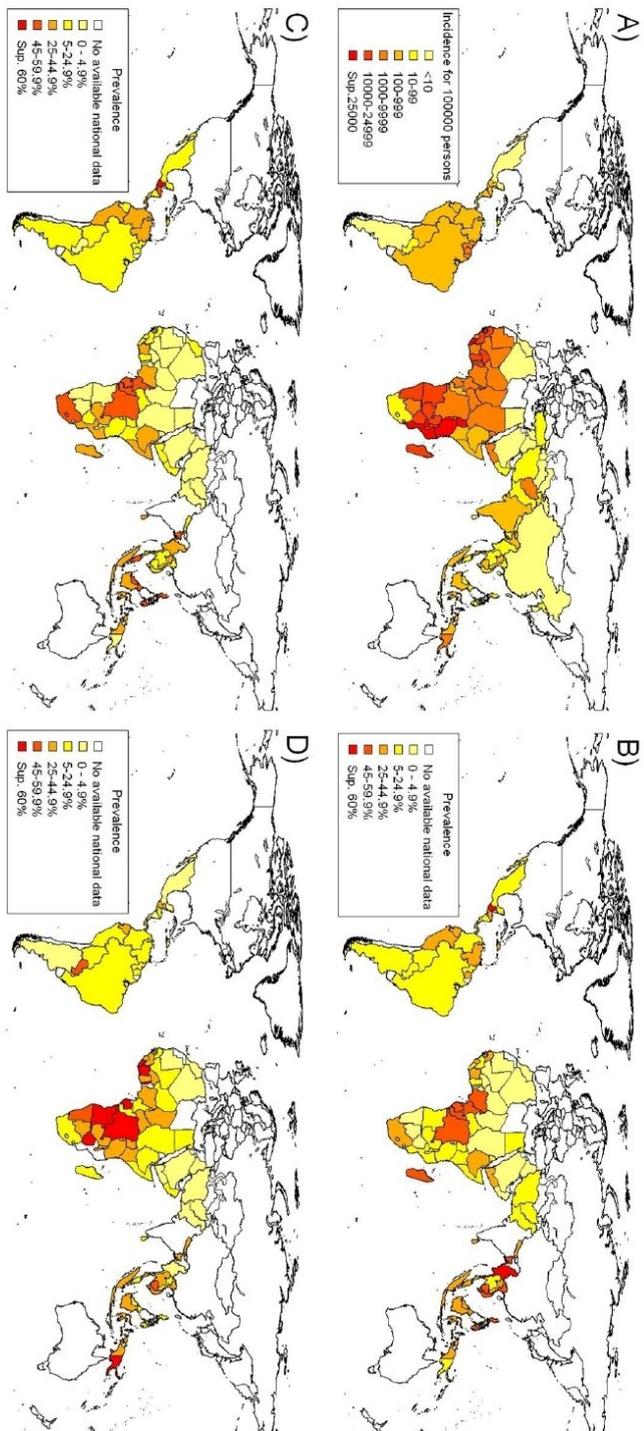


Figure
[Click here to download Figure: Figure 2.doc](#)

Figure 2: Distributions of national malaria incidence (per 100 000 persons / year) in a base 10 logarithm scale by helminths species, climate, and GDP.

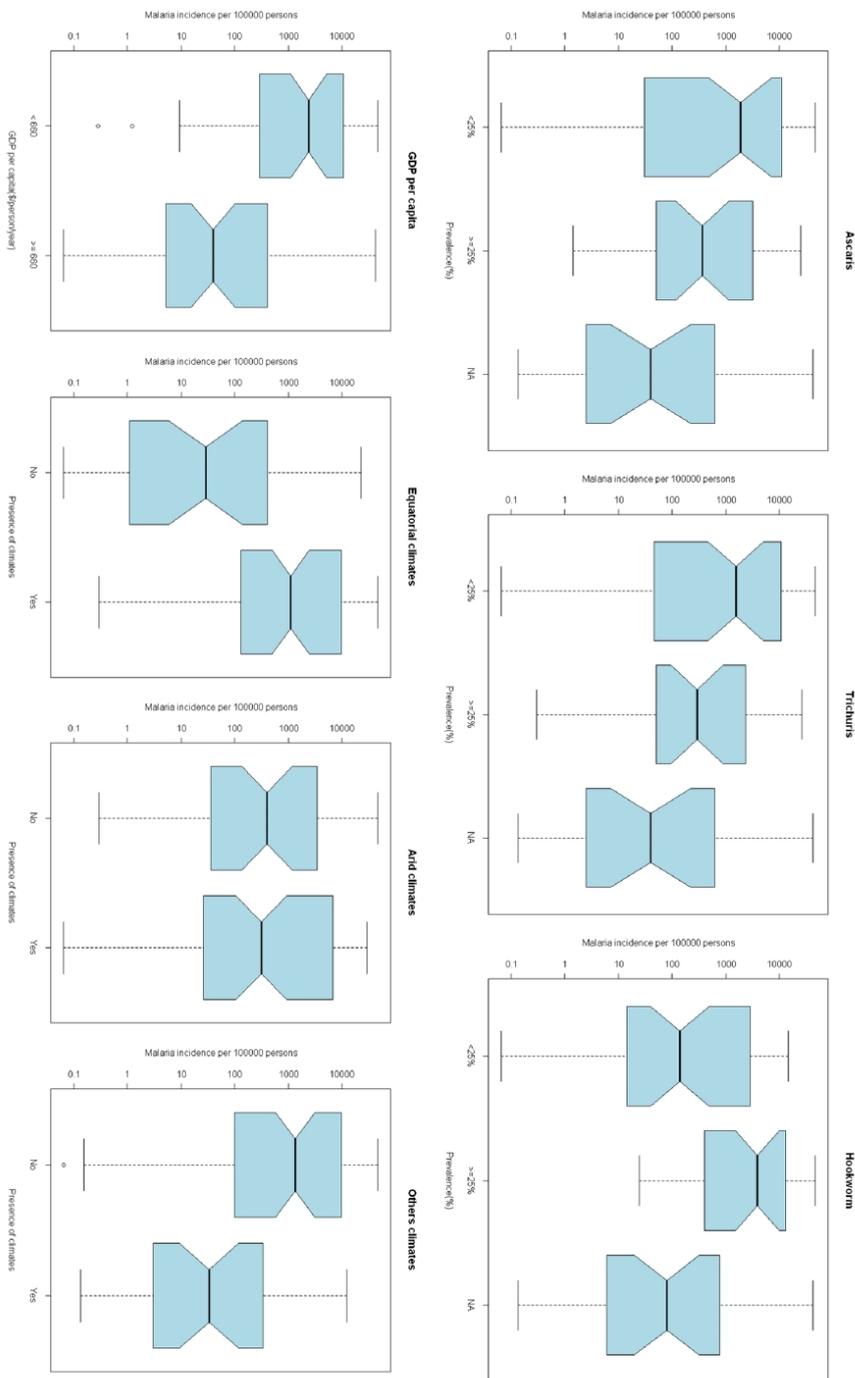
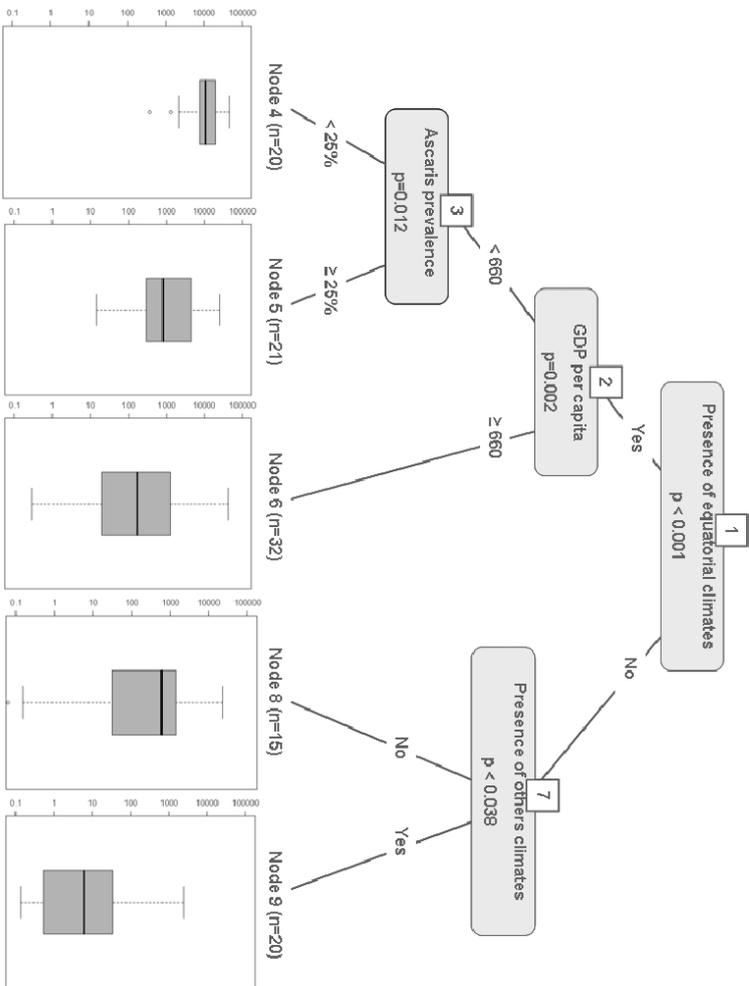


Figure
[Click here to download Figure: Figure 3.doc](#)

Figure 3: Regression tree for base 10 logarithm of malaria incidence using the predictor variables as determined by the methodology. P-values at each node are adjusted with Bonferroni correction. In order for a split to occur the p-value must be <0.05 and the node size must not be less than 20. Box plots at terminal nodes show the distribution of the data within that branch of the tree. Boxes represent inter-quartile range; horizontal lines within the boxes represent median values; whiskers extend to the most extreme data point that is no more than 1.5 times the inter-quartile range. Dots above and below whiskers represent more extreme data points.



8.10. Etude 11: la corruption tue. Estimation de l'impact global de la corruption sur la mortalité juvéno-infantile

Introduction: Des connaissances précises sur les facteurs de risque globaux de la mortalité juvéno-infantile sont essentielles pour guider les efforts mondiaux visant à améliorer la survie des enfants. L'impact significatif de la corruption sur la mortalité juvéno-infantile a déjà été démontré. Cependant, aucune quantification récente n'est disponible.

Matériel/méthodes: L'impact de la corruption a été évalué en utilisant des corrélations brutes de Pearson, et des modèles linéaires univariés et multivariés couplant les taux nationaux de mortalité juvéno-infantile en 2008 avec le "niveau perçu de corruption" (CPI) et un grand nombre de variables d'ajustement mesurées durant la même période.

Résultats: Le modèle multivarié final (R^2 ajusté = 0,89) comprenait les variables suivantes: le pourcentage de personnes ayant un système d'assainissement amélioré (p.value <0,001), le logarithme des dépenses de santé totales (p.value = 0,006), l'indice de perception de la corruption (p.value <0,001), la présence d'un climat aride sur le territoire national (p.value = 0,006), et le taux de dépendance (p.value <0,001). Une diminution du CPI d'un point (c'est-à-dire une corruption perçue plus importante) était associée à une augmentation nationale du logarithme du taux de mortalité juvéno-infantile de 0.0644. Selon ce résultat, il peut être estimé grossièrement que plus de 140 000 décès annuels d'enfants pourraient être indirectement attribués à la corruption.

Discussion: La réponse globale à la mortalité des enfants doit nécessairement impliquer une augmentation des fonds disponibles pour développer l'accès à l'eau et l'assainissement et pour mettre en place de nouvelles méthodes de prévention, de gestion et de traitements face aux principales maladies dessinant l'actuelle répartition globale de la mortalité juvéno-infantile. Cependant sans réfléchir aux mécanismes de contrôle de la corruption nécessaires pour assurer leur bonne utilisation, cela fournira aussi des opportunités supplémentaires à la corruption. Les interventions de santé publique soutenues par les gouvernements et les donateurs doivent intégrer des initiatives qui reconnaissent comment ces deux phénomènes interagissent.

Article publié le 2 novembre 2011 dans Plos One

Corruption Kills: Estimating the Global Impact of Corruption on Children Deaths

Matthieu Hanf^{1,2*}, Astrid Van-Melle^{1,2}, Florence Fraisse¹, Amaury Roger¹, Bernard Carne^{1,2}, Mathieu Nacher^{1,2}

1 Centre d'Investigation Clinique Epidémiologie Clinique Antilles Guyane CIC-EC INSERM CIE 802, Cayenne General Hospital, Cayenne, French Guiana, **2** Team EA 3593, Department of Parasitology and Mycology, Université des Antilles et de la Guyane, Cayenne, French Guiana

Abstract

Background: Information on the global risk factors of children mortality is crucial to guide global efforts to improve survival. Corruption has been previously shown to significantly impact on child mortality. However no recent quantification of its current impact is available.

Methods: The impact of corruption was assessed through crude Pearson's correlation, univariate and multivariate linear models coupling national under-five mortality rates in 2008 to the national "perceived level of corruption" (CPI) and a large set of adjustment variables measured during the same period.

Findings: The final multivariable model (adjusted $R^2 = 0.89$) included the following significant variables: percentage of people with improved sanitation (p.value<0.001), logarithm of total health expenditure (p.value = 0.006), Corruption Perception Index (p.value<0.001), presence of an arid climate on the national territory (p = 0.006), and the dependency ratio (p.value<0.001). A decrease in CPI of one point (i.e. a more important perceived corruption) was associated with an increase in the log of national under-five mortality rate of 0.0644. According to this result, it could be roughly hypothesized that more than 140000 annual children deaths could be indirectly attributed to corruption.

Interpretations: Global response to children mortality must involve a necessary increase in funds available to develop water and sanitation access and purchase new methods for prevention, management, and treatment of major diseases drawing the global pattern of children deaths. However without paying regard to the anti-corruption mechanisms needed to ensure their proper use, it will also provide further opportunity for corruption. Policies and interventions supported by governments and donors must integrate initiatives that recognise how they are inter-related.

Citation: Hanf M, Van-Melle A, Fraisse F, Roger A, Carne B, et al. (2011) Corruption Kills: Estimating the Global Impact of Corruption on Children Deaths. PLoS ONE 6(11): e26990. doi:10.1371/journal.pone.0026990

Editor: James G. Beeson, Burnet Institute, Australia

Received: June 17, 2011; **Accepted:** October 7, 2011; **Published:** November 2, 2011

Copyright: © 2011 Hanf et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by Cayenne General Hospital (French Guiana) and INSERM (Institut National de la Santé et de la Recherche Médicale). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: matthieu@hanf.fr

Introduction

Corruption is a complex problem which threatens the impact of public investments, health care access and services, equity and outcomes [1,2]. Increasingly, health sector leaders, and citizens of all countries, are recognizing the pernicious effects of corruption both at micro and macro levels, and the need to take action [1,3,4]. At the level of individuals and households, there is mounting evidence of the negative effects of corruption on the health and welfare of citizens [5,6]. The 2006 report of transparency International asserts that the corruption level undermines the achievement of the Millennium Development Goal (MDGs) and that the corruption is one of the primary causes of the fact that the global community is already off target to meet the MDGs [1]. Based on literature and previous works, Vian et al. [2] distinguished 7 areas in the health sector where corruption could have pejorative health consequences: construction and rehabilitations of health facilities, purchase of equipment and supplies, distribution and use of drugs, regulation of qualities in

products, education of health professionals, medical research, and provision of services by medical personnel.

Health impacts of corruption can not be only summarized as those acting directly on health access and services. For example, nearly 1.2 billion people in the world do not have guaranteed access to water and more than 2.6 billion are without adequate sanitation, with devastating consequences for development and poverty reduction. In developing countries, about 80 per cent of health problems can be linked back to inadequate water and sanitation [7]. The human consequences of this water crisis are devastating and affect the poor and women most of all. Experts concur that the water crisis is a crisis of water governance where, as suggested elsewhere, corruption could be one of the main cause and catalyst for this crisis and that it could affect all aspects of the water and sanitation sector, from resources management to services [8]. More globally corruption was shown to affect all domains of life from education to economic performances [4] and thought these ways could also impact health outcomes.

Among major health problems, child mortality is one of the most preoccupying. In 2006, for the first time since records have been kept, the number of children dying before their fifth birthday fell below 10 million, to 9.7 million. In 2008, it was estimated that 8.795 million deaths under 5 years of age occurred worldwide (figure 1). This milestone follows a long-term decline in the global under-five mortality rate since 1960. However, many countries still have high levels of child mortality, particularly in sub-Saharan Africa and South Asia, and in recent years have made little or no progress in reducing the number of child deaths. This global progress is insufficient to achieve MDGs in which reduction of mortality is crucial. The United Nations aim to decrease the child mortality rate by two-thirds between 1990 and 2015 [9,10]. By its influences on health, water/sanitation services, education and economic development, corruption could potentially have a great impact on the dynamics of child mortality. However quantitative data on this subject at a population scale are poor. To our knowledge, the most recent study quantifying the specific link between these two variables was the one realised by Abed and colleagues [3]. This study, in 2002, observed a significant link between corruption level and child mortality but only incorporated a limited set of adjustment variables and was made using 10 years old data. A more recent vision of current global impacts of corruption on child mortality is needed to implement efficient programs and improve children survival.

In this perspective, data on national under five mortality rates, the classical risk factors of child mortality, and corruption level were gathered circa 2008. By using linear regressions, this ecological study aimed to 1) assess at a country scale the current impacts of corruption on global children mortality rate after adjusting with a large set of variables, and 2) roughly calculate the number of children deaths attributable to corruption.

Materials and Methods

Data source

In this population based study, national data on under five mortality presented as the number of deaths per 1000 live births in 2008 were compiled from estimations calculated by Black and colleagues [11] and freely available on the Gapminder project website [12]. National corruption level was assessed through the 2008 CPI compiled by Transparency International which orders the countries of the world according to "the degree to which corruption is perceived to exist among public officials and politicians" [13]. This index is ranging from 0 to 10 with a higher score meaning less (perceived) corruption.

According to the literature, the main risk factors of children deaths are access to drinkable water and sanitation facilities [9,14,15], socio-economic conditions [9,16,17], vaccination coverage [18], health services level [9,19], undernutrition [20,21], climate [22,23], war and natural disasters [24], and political context and corruption [1,3]. Adjustment variables were categorized into four groups:

Socio economic factors. To reflect the economic and development level of studied countries, the Gross Domestic Product (GDP) per capita in 2008 and the dependency ratio (percentage of working-age population) in 2008 were used [25].

National level of water and sanitation facilities were assessed through the two following World Bank indicators [25]: the 2008 percentage of the population with at least adequate access to excreta disposal facilities and 2008 percentage of the population with reasonable access to an adequate amount of water from an improved source.

Education level was assessed by the literacy rate in 2008 which was compiled from 2 different sources of data [25,26].

Two major population characteristics obtained from the World Bank [25] were also used: the percentage of rural population in 2008 and the population density per square km in 2008.

Health and medical factors. The national health services level was assessed through the total health expenditure per capita in 2008 and through the percentage of GDP invested in health expenditure [25].

Vaccination coverage was assessed through the percentage of children ages 12–23 months immunised against DTP and measles [25].

The nutritional level was measured by the 2007 food supply (kilocalories/person/day) compiled by the FAO [27].

Environmental factors. National climates were categorized according to the revised Köppen-Geiger climate classification [28]. Four derived variables were created: presence on the national territory of an equatorial climate, an arid climate, a warm temperate climate and a snow or polar climate.

Impacts of floods were measured through the percentage of national population affected by flood compiled by the Center for Research on the Epidemiology of Disasters (CRED) [29].

Political and societal factors. Impact of war on children deaths was measured through the national battle-related deaths/person derived from World Bank indicators [25].

Democracy was measured through two perceived scores on civil liberties and political rights calculated by Freedom House, an international non-governmental organization that conducts research and advocacy on democracy, political freedom and human rights [30]. These indexes are ranging from 1 to 7 with a higher score meaning less (perceived) civil liberties and political rights.

Statistical analysis

Logarithm transformations of national data on deaths of children younger than 5 years per 1000 live births in 2008 were used in this study as the outcome variable. Data on 178 countries were provided.

To understand the crude correlation structure between all the quantitative variables incorporated in the model (dependent and independent variables), a Focused Principal Components Analysis was performed [31]. Compared to a traditional crude correlation matrix, the FPCA gives a multidimensional graphical display allowing, in one look, to simply understand the correlation structure between the outcome and exploratory variables and between exploratory variables themselves. The representation obtained with this method is close to a Principal Components Analysis (PCA). Contrary to PCA, correlations (Pearson's correlation) between the dependent variable and the other variables are represented faithfully. The relationships between non dependent variables are interpreted like in a PCA: correlated variables are close or diametrically opposite (for negative correlations), independent variables make a right angle with the origin. The package "psy" of the R statistical language was used to perform the Focused Principal Components Analysis.

Unweighted univariate linear regressions were then performed. In view to identify the dominant factors in those that were statistically significant in univariate regression ($p < 0.05$), significant variables were included in a multivariate linear regression model. To determine the final multivariate model, a backward stepwise linear regression was then used to simultaneously adjust for various variables of interest. Statistical significance was set at $p < 0.1$.

As a step towards stabilizing the residual variation in regression analysis, logarithmic transformations of the data on health expenditure per capita, GDP per capita, battle-related deaths/

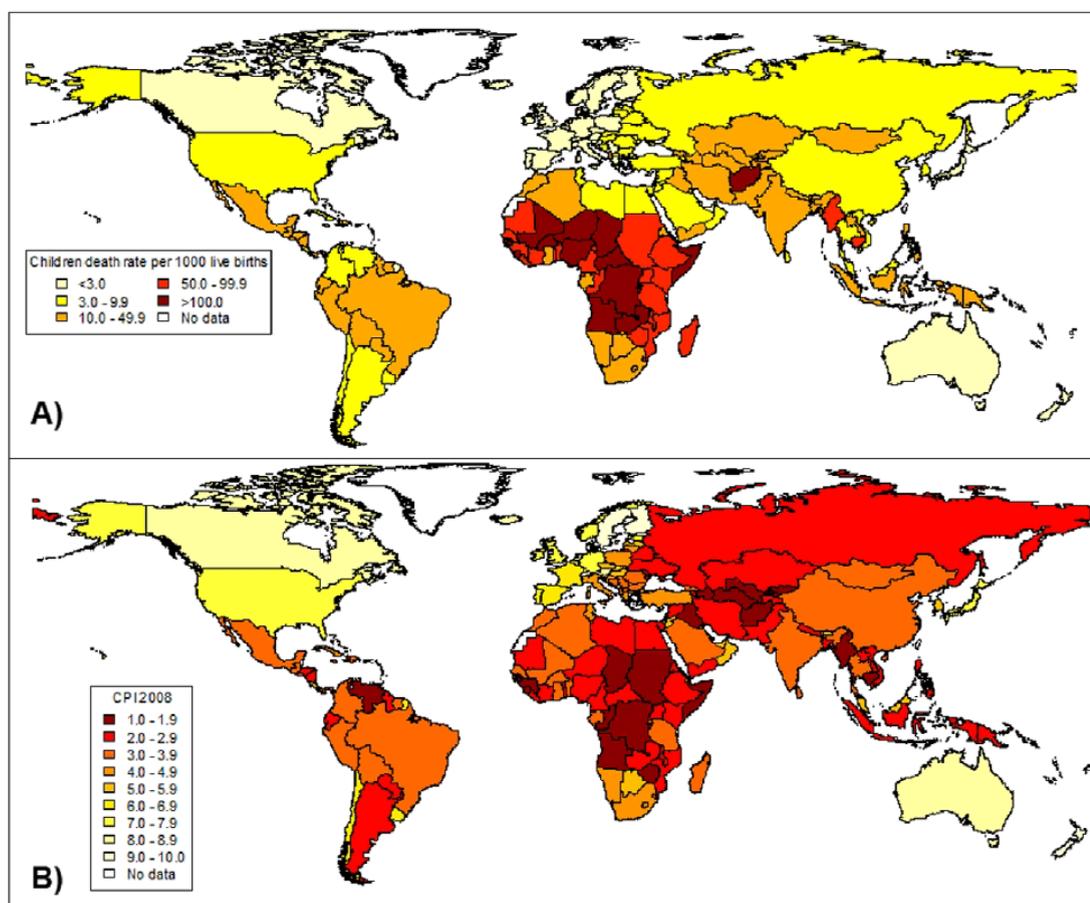


Figure 1. World repartition of children deaths and perceived corruption. A) Deaths of children younger than 5 years per 1000 live births in 2008. Derived from data compiled by Black and colleagues [11]. B) Corruption Perceptions Index (CPI) in 2008. Derived from data compiled by Transparency International [10].

doi:10.1371/journal.pone.0026990.g001

person, percentage of national population affected by floods were used. Homoscedasticity was verified graphically and the model including corruption was compared to the model without corruption using the adjusted R^2 . Multicollinearity and variance inflation factors were also checked in the final linear model to assess its robustness. All statistical analyses were conducted using R 2.12.0 [32].

Results

National data on under five years deaths per 1000 live births and CPI in 2008 are shown in figure 1. Variables that can potentially explain inter-country variation in the child mortality rate were summarized in table 1. In 2008, the child mortality rate varied from 1.027 (Slovenia) to 195.188 (Afghanistan) per 1000 live births with a world median at 11.789.

A synthesis of Pearson's correlations between all quantitative variables used in this analysis was presented in figure 2 (obtained with a Focused Principal Components Analysis). The six variables the most correlated with log of mortality rate in children were: the log of health expenditure per capita ($r = -0.87$), the log of GDP

per capita ($r = -0.86$), the percentage of people with access to improve sanitation ($r = -0.84$), the dependency ratio ($r = 0.81$), the food supply ($r = -0.78$) and the perceived corruption ($r = -0.76$). The six variables the most correlated with the perceived corruption index were: the health expenditure per capita ($r = 0.82$), the GDP per capita ($r = 0.78$), civil liberties index ($r = -0.69$), the food supply ($r = 0.66$), the political rights index ($r = -0.64$) and percentage of rural people ($r = -0.58$).

In the univariate analysis, all predictor variables (excepted battle-related deaths/person) were significantly associated ($p < 0.001$) with child mortality (Table 2). The coefficient for the corruption perception index [$\beta = -0.2482$; 95% C.I.: -0.2482 ; -0.1906] in the univariate model would mean that an increase in 1 point of the CPI would decrease the logarithm transformation of children deaths per 1000 live births by 0.2482.

For child mortality rates, the final multivariable model (adjusted $R^2 = 0.890$) included the following significant variables: percentage of people with improved sanitation ($p.value < 0.001$), logarithm of total health expenditure ($p.value = 0.006$), CPI score ($p.value < 0.001$), presence of an arid climate on the national territory ($p = 0.006$), and the dependency ratio ($p.value < 0.001$). In this

Table 1. Available variables that can potentially explain inter-country variation in the rate of deaths in children in 178 countries, 2008.

	No. of countries with data (%)	Value range	Reference
Socio economic factors			
GDP per capita (2008)	171 (96%)	144.8–117954.7	[25]
People with access to improved water source (2008) (%)	172 (97%)	45–100	[25]
People with access to improved sanitation (2008) (%)	161 (90%)	9–100	[25]
Percentage of rural population (2008) (%)	178 (100%)	0–89.6	[25]
Literacy rate (circa 2008) (%)	167 (94%)	27.6–100	derived from [25] and [26]
Dependency ratio (2008)	178 (100%)	20.89–106.95	[25]
Density per square km (2008)	178 (100%)	1.7–6757.0	[25]
Health and medical factors			
Total health expenditure per capita (2008)	171 (96%)	10.3–8071.1	[25]
Health expenditure (2008) (% GDP)	175 (98%)	1.949–15.180	[25]
DTP coverage (2008) (%)	178 (100%)	23–99	[25]
Measles coverage (2008) (%)	178 (100%)	23–99	[25]
Food supply (2007) (kilocalories/person/day)	166 (93%)	1605–3819	[27]
Environmental factors			
Presence of an equatorial climate on the national territory	178 (100%)	-	derived from [28]
Presence of an arid climate on the national territory	178 (100%)	-	derived from [28]
Presence of a warm temperate climate on the national territory	178 (100%)	-	derived from [28]
Presence of a polar or snow climate on the national territory	178 (100%)	-	derived from [28]
National population affected by flood (2008) (%)	178 (100%)	0–13.1	Derived from [29]
Political and societal factors			
Civil liberties (2008)	177 (99%)	1–7	[30]
Political rights (2008)	177 (99%)	1–7	[30]
Corruption perception index (2008)	168 (94%)	1.0–9.3	[13]
National battle-related deaths/person (2008)	178 (100%)	0.00–0.05	[25]

doi:10.1371/journal.pone.0026990.t001

model, presence of a warm temperate climate on the national territory and DTP vaccination coverage were also found as almost significant ($p < 0.1$). Incorporation of the CPI in the model increased the adjusted R^2 from 0.8736 to 0.8897 (+1.61%). The coefficient for the corruption perception index [$\beta = -0.0644$; 95% C.I.: -0.0954 ; -0.0334] in the final multivariate model would mean that a decrease in 1 point of the CPI score would increase the log of child deaths per 1000 live birth by 0.0644.

Plot of residuals versus predicted values and qqplot did not revealed particular pattern and indicated that homoscedasticity and normal hypothesis were respected in the final linear model. Variance inflation factors of variables included in the final model were also found to be lower than 5 showing that the magnitude of multicollinearity was low.

Discussions

Ecological studies have their limitations (i.e. bias, confounding) [33]. The available data in a country may not have been sufficient to reflect the local variations of all the studied variables. This study was limited to a one year period (2008) and, due to the lack of data, some important variables to explain child mortality (as breastfeeding for example [34]) could not be included in the present analysis. Furthermore, several variables were based on estimations made by the United Nations and their quality could vary greatly compared to other objective data used in this analysis.

However, all the data used in this analysis is the only available to study the problem at this scale. Although there are limitations, the fact that the present analysis came to similar observations on suspected and described phenomenon acting on child mortality rates is unlikely to have been due to a statistical artefact.

The linear regression analysis presented here, after taking into account known significant predictors of children deaths (mainly health expenditure, sanitation level, dependency ratio and arid climate) seems to point towards a significant pejorative link between perceived corruption and death rate at national level. In this model, a decrease in CPI of one point (i.e. a more important perceived corruption) was associated with an increase in log of national child mortality rate of 0.0644. The adjustment of the final model was very good (Pseudo $R^2 = 0.89$) meaning that 89% of the data variability was captured by modelisation. Incorporation of CPI in the final model made a rise in the Pseudo R^2 of 1.61%. According to this result, it could be hypothesized that roughly 1.6% of world deaths in children could be explained by corruption meaning that, of the annual 8.795 million children deaths, more than 140000 annual children deaths could be indirectly attributed to corruption.

By its simplistic nature, the model constructed here, has probably a non negligible tendency to underestimate the effects of corruption on children deaths. As shown elsewhere, across different country contexts, corruption has been a cause and consequence of poverty and so could possibly directly impacts the

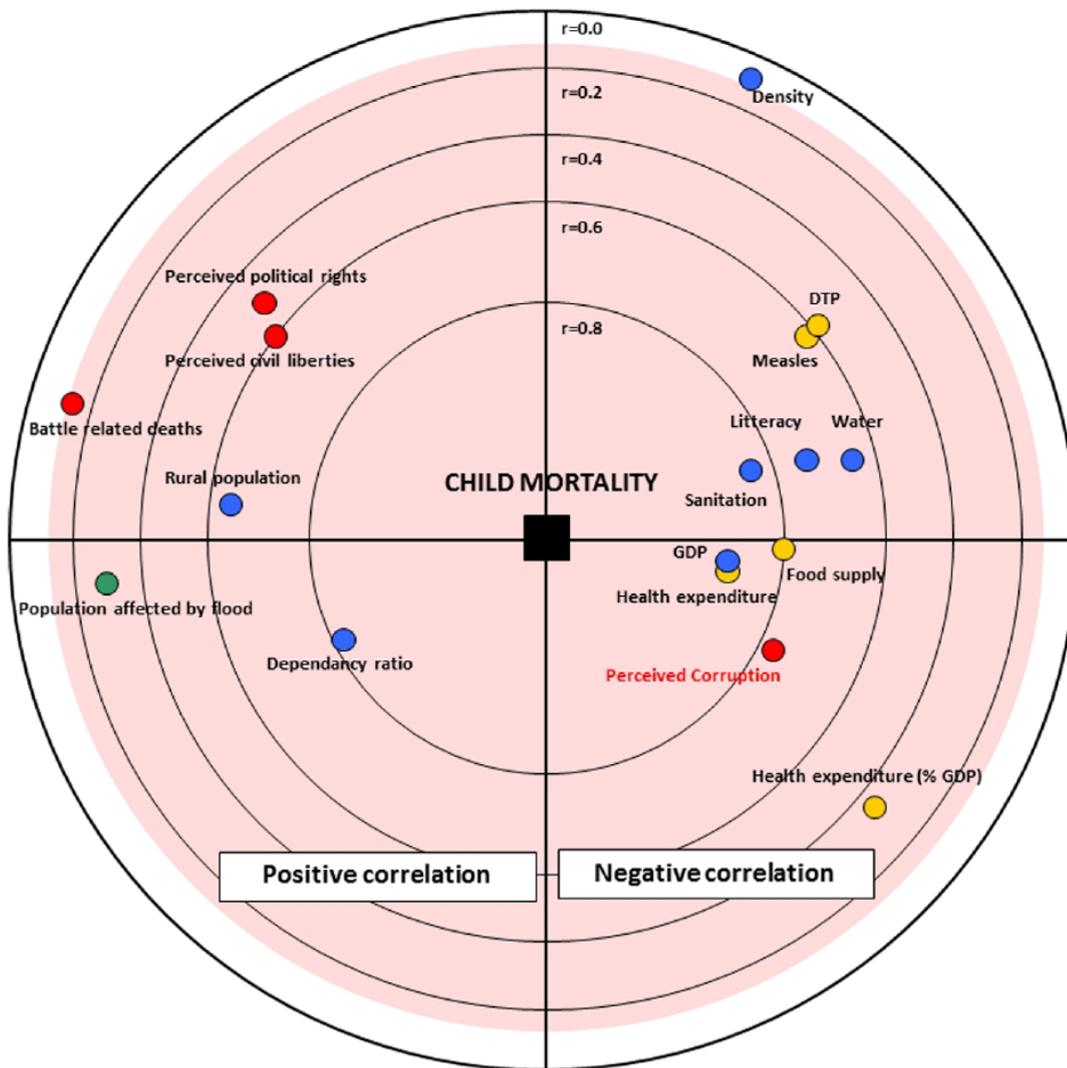


Figure 2. Pearson's correlations between the mortality rate in children and the other variables used in the analysis. Representation obtained with a Focused Principal Components Analysis. Pearson's correlations between the mortality rate in children and the other variables are represented faithfully. Positive correlations with child mortality are plotted on the left and negative ones on the right. Correlations which are significantly different to zero are inside the pink circle. The relationships between the other variables are interpreted like in a PCA: correlated variables are close or diametrically opposite (for negative correlations), independent variables make a right angle with the origin. Colours indicate the nature of explicative variables (red: political and societal factors, blue: socio economic factors, orange: health and medical factors and green: environmental factors).
doi:10.1371/journal.pone.0026990.g002

others significant parameters retrieved by the analysis, especially the percentage of population with adequate access to sanitation [35] and the public health expenditure [36]. Growing evidence from around the world indicates that corruption, fraud, and abuse are also resulting in significant losses of public money and denial of good quality health and sanitation services to millions of people [1,2,4,7]. The diversity of health and water/sanitation systems worldwide, the multiplicity and complexity of parties involved make it difficult to determine the overall impacts of corruption in children deaths around the globe. Local studies have to be made to elucidate the underground complexity of the relation between child health, socio-economic conditions, and corruption.

This study is a first step in the quantification of the current global impact of corruption on child deaths in human populations and shows that annually at least 140000 children deaths could be related to corruption, a total that largely exceeds the conspicuous pooled total of cholera, rabies, Ebola and combat-related deaths. But still, because the equation corruption = deaths is seldom explicit, corruption only seems like a nuisance.

The global response to child deaths must involve a necessary increase in funds available to 1) develop water and sanitation access and 2) purchase new methods for prevention, management, and treatment of major diseases killing children around the globe (principally pneumonia, diarrhoea and malaria). However,

Table 2. Univariate and multivariate linear regressions predicting Log of children deaths per 1000 live births in 178 countries, 2008 (£).

	Univariate analysis				Multivariate analysis			
	Estimate	Std Error	t value	Pr(> t)	Estimate	Std Error	t value	Pr(> t)
Intercept	-	-	-	-	1.7925	0.2242	7.997	<0.001
Socio economic factors								
Log10(GDP per capita)	-0.7486	0.0348	-21.490	<0.001	-	-	-	-
People with access to improved water source	-0.0404	0.0036	-11.290	<0.001	-	-	-	-
People with access to improved sanitation	-0.0169	0.0009	-19.590	<0.001	-0.0050	0.0010	-4.786	<0.001
Percentage of rural population	0.0166	0.0016	10.703	<0.001	-	-	-	-
Literacy rate	-0.0243	0.0017	-14.180	<0.001	-	-	-	-
Dependency ratio	0.0637	0.0054	11.800	<0.001	0.0096	0.0016	6.162	<0.001
Density	-0.0002	0.0001	-2.173	0.0311	-	-	-	-
Health and medical factors								
Log10(Total health expenditure per capita)	-0.7075	0.0309	-22.920	<0.001	-0.1714	0.0619	-2.769	0.006
Health expenditure (% GDP)	-0.0723	0.0167	-4.323	<0.001	-	-	-	-
DTP coverage (%)	-0.0233	0.0025	-9.207	<0.001	-0.0027	0.0016	-1.720	0.0877
Measles coverage (%)	-0.0242	0.0024	-10.000	<0.001	-	-	-	-
Food supply	-0.0009	0.0001	-16.070	<0.001	-	-	-	-
Environmental factors								
Presence of an equatorial climate on the national territory	0.5572	0.0802	6.949	<0.001	-	-	-	-
Presence of an arid climate on the national territory	0.4718	0.0838	5.634	<0.001	0.1012	0.0366	2.763	0.006
Presence of a warm temperate climate on the national territory	-0.5316	0.0815	-6.524	<0.001	-0.0749	0.0388	-1.928	0.0558
Presence of a polar or snow climate on the national territory	-0.4359	0.0966	-4.514	<0.001	-	-	-	-
Log10(Percentage of national population affected by flood)	0.0892	0.0213	4.178	<0.001	-	-	-	-
Political and societal factors								
Civil liberties	0.1880	0.0210	8.959	<0.001	-	-	-	-
Political rights	0.1404	0.0185	7.596	<0.001	-	-	-	-
Corruption perception index	-0.2194	0.0147	-14.980	<0.001	-0.0644	0.0158	-4.088	<0.001
Log10(National battle-related deaths/person)	0.0802	0.0433	1.853	0.0656	-	-	-	-

£: Analysis of deaths in children from a multivariate linear regression model. Only variables with a p.value<0.1 in univariate regression were incorporated in the final multivariate analysis. Backward stepwise procedure with a stop criterion fixed at 0.1 was then applied to determinate the final multivariate model. doi:10.1371/journal.pone.0026990.t002

without paying attention to the anti-corruption mechanisms needed to ensure their proper use, it will also provide further opportunity for corruption. In practice, donors and governments still treat health, water/sanitation access and corruption as separate rather than integral components of the same strategy. To address these obstacles, designing dedicated indicators at micro and macro levels which monitor efficiently corruption impacts on health and health related services, is urgently needed. Policies and interventions supported by governments and donors must integrate initiatives that recognise how health and corruption are inter-related.

References

1. Transparency International (2006) Global corruption report 2006: Corruption and Health. Berlin: Transparency International (TI).
2. Vian T (2008) Review of corruption in the health sector: theory, methods and interventions. Health Policy Plan 23: 83–94.
3. Abed GT, Gupta S (2002) Governance, Corruption, & Economic Performance Washington, D.C: International Monetary Fund.
4. World Bank (2000) Helping countries combat corruption: progress at the World Bank since 1997. Washington, DC: World Bank.
5. Azfar O, Gurgur T (2008) Does Corruption Affect Health and Education Outcomes in the Philippines? College Park: Institutional Reform and Informal Sector.
6. Maestad O, Mwisongo A (2011) Informal payments and the quality of health care: Mechanisms revealed by Tanzanian health workers. Health Policy 99: 107–115.
7. United Nations (2000) The Millennium Development Goals Report 2000. New York: United Nations.
8. Transparency International (2008) Global corruption report 2008: Corruption in the water sector. Berlin: Transparency International (TI).
9. Black RE, Morris SS, Bryce J (2003) Where and why are 10 million children dying every year? Lancet 361: 2226–2234.
10. Santosham M, Chandran A, Fitzwater S, Fischer-Walker C, Baqui AH, et al. (2010) Progress and barriers for the control of diarrhoeal disease. Lancet 376: 63–67.

11. Black RE, Cousens S, Johnson HL, Lawn JE, Rudan I, et al. (2010) Global, regional, and national causes of child mortality in 2008: a systematic analysis. *Lancet* 375: 1969–1987.
12. Gapminder (2011) Gapminder database. Stockholm.
13. Transparency International (TI) (2011) Transparency International website. Berlin.
14. Fewtrell L, Kaufmann RB, Kay D, Enanoria W, Haller L, et al. (2005) Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis. *Lancet Infect Dis* 5: 42–52.
15. Rutstein SO (2000) Factors associated with trends in infant and child mortality in developing countries during the 1990s. *Bull World Health Organ* 78: 1256–1270.
16. Aber JL, Bennett NG, Conley DC, Li J (1997) The effects of poverty on child health and development. *Annu Rev Public Health* 18: 463–483.
17. Walker SP, Wachs TD, Gardner JM, Lozoff B, Wasserman GA, et al. (2007) Child development: risk factors for adverse outcomes in developing countries. *Lancet* 369: 145–157.
18. World Health Organization (WHO), United Nations Children's Fund (UNICEF), World Bank (2009) State of the world's vaccines and immunization. Geneva: World Health Organization.
19. Bryce J, el Arifeen S, Pariyo G, Lanata C, Gwatkin D, et al. (2003) Reducing child mortality: can public health deliver? *Lancet* 362: 159–164.
20. Calder PC, Jackson AA (2000) Undernutrition, infection and immune function. *Nutr Res Rev* 13: 3–29.
21. Caulfield LE, de Onis M, Blossner M, Black RE (2004) Undernutrition as an underlying cause of child deaths associated with diarrhea, pneumonia, malaria, and measles. *Am J Clin Nutr* 80: 193–198.
22. Cazelles B, Hales S (2006) Infectious diseases, climate influences, and nonstationarity. *PLoS Med* 3: e328.
23. Lloyd SJ, Kovats RR, Armstrong BG (2007) Global diarrhoea morbidity, weather and climate. *Clim Res* 34: 119–127.
24. Watson JT, Gayer M, Connolly MA (2007) Epidemics after natural disasters. *Emerg Infect Dis* 13: 1–5.
25. World Bank (2011) The World Bank Database.
26. United Nations Statistics Division (2011) United Nations statistical databases. New York, NY.
27. Food and Agricultural Organisation (2011) FAO statistic database (FAOSTAT). Roma.
28. Kottek M, Grieser J, Beck C, Rudolf B, Rubel F (2006) World map of the Köppen-Geiger climate classification updated. *Meteorol Z* 15: 259–263.
29. Centre for Research on the Epidemiology of Disasters (CRED) (2011) Emergency Events Database (EM-DAT).
30. Freedom House (2011) Freedom in the World 2008 survey. Washington, DC.
31. Falissard B (1996) A spherical representation of a correlation matrix. *Journal of Classification* 13: 267–280.
32. R Development Core Team (2010) R: A Language and Environment for Statistical Computing. 2.12 ed. Vienna, Austria: R Foundation for Statistical Computing.
33. Morgenstern H (1982) Uses of ecologic analysis in epidemiologic research. *Am J Public Health* 72: 1336–1344.
34. Ip S, Chung M, Raman G, Chew P, Magula N, et al. (2007) Breastfeeding and maternal and infant health outcomes in developed countries. *Evid Rep Technol Assess (Full Rep)*. pp 1–186.
35. Anbarci N, Escaleras M, Register CA (2007) Collective (In) Action and Corruption: Access to Improved Water and Sanitation. Miami: Florida International University.
36. Mauro P (1998) Corruption and the composition of government expenditure. *Journal of Public Economics* 69: 263–279.

8.11. Etude 12: L'impact de la corruption sur la tuberculose multirésistante: une évaluation quantitative

Introduction: Une identification des facteurs de risque globaux de la tuberculose multirésistante (MDR-TB) est cruciale pour guider les efforts mondiaux cherchant à contenir efficacement son expansion. La corruption est très fortement soupçonnée d'impacter significativement la dynamique des tuberculoses multirésistantes. Cependant, aucune quantification globale n'est actuellement disponible.

Matériel et méthodes: L'impact de la corruption a été évalué par des modèles linéaires univariés et multivariés couplant le pourcentage national de MDR-TB parmi les nouveaux cas de tuberculose en 2008, le "niveau perçu de corruption" et un ensemble de variables d'ajustement mesurées pendant la même période .

Résultats: Le modèle final multivarié incluait les variables significatives suivantes: le taux national de réussite du traitement antituberculeux (p.value = 0,034) et le niveau perçu de corruption (p.value = 0,001). L'incorporation de l'indice de corruption dans le modèle augmentait le R² ajusté de 5,5%.

Discussion: Cette étude montre que 5,5% de la variabilité inter-pays de la proportion des cas de MDR-TB parmi les nouveaux cas de tuberculose pourrait être imputable à la corruption, un total qui dépasse largement l'effet des facteurs de santé retrouvés comme significatifs dans l'analyse (stratégies de gestion actuelles de la tuberculose). Les politiques et les interventions soutenues par les gouvernements doivent intégrer des initiatives qui identifient comment la corruption et la MDR-TB interagissent pour contenir efficacement son épidémie.

Article soumis le 30 octobre 2011 à European Respiratory Journal

Partie 2 : travaux scientifiques

European Respiratory Journal



The impact of corruption on multidrug-resistant tuberculosis: a quantitative assessment

Journal:	<i>European Respiratory Journal</i>
Manuscript ID:	Draft
Manuscript Type:	Letter
Date Submitted by the Author:	n/a
Complete List of Authors:	Hanf, Matthieu; Cayenne General Hospital, Centre d'Investigation Clinique Epidémiologie Clinique Antilles Guyane CIC-EC INSERM CIE 802 Adenis, Antoine; Cayenne General Hospital, Centre d'Investigation Clinique Epidémiologie Clinique Antilles Guyane CIC-EC INSERM CIE 802 Couppié, Pierre; Cayenne General Hospital, Department of Dermatology Carne, Bernard; Cayenne General Hospital, Centre d'Investigation Clinique Epidémiologie Clinique Antilles Guyane CIC-EC INSERM CIE 802 Nacher, Mathieu; Cayenne General Hospital, Centre d'Investigation Clinique Epidémiologie Clinique Antilles Guyane CIC-EC INSERM CIE 802
Key Words:	Corruption, epidemiology, global health, multidrug-resistant tuberculosis, risk factor, Tuberculosis

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

COVER LETTER

Dear Sir or Madam,

We would be grateful if you could consider the following letter "The impact of corruption on multidrug-resistant tuberculosis: a quantitative assessment" for publication in the European Respiratory Journal.

Multidrug-resistant tuberculosis (MDR-TB) is among the most preoccupying aspects of the pandemic of antimicrobial resistance. In 2008, an estimated 440 000 cases of MDR-TB emerged globally and caused 150 000 deaths. MDR-TB is globally distributed, but significant variations in prevalence have been observed in different geographical regions.

Due to a strong overlap between corruption consequences and potential MDR-TB drivers, it was argued in the literature that corruption may have a deep impact on drug resistance and particularly on MDR-TB dynamic in some world areas. Accurate information on the global risk factors of MDR-TB is needed to guide global efforts to contain efficiently its burden.

In this article, by coupling estimations of the national percentage of MDR among new TB cases in 2008 in 190 countries to the national "perceived level of corruption" and a set of adjustment variables measured during the same period, we shown that corruption impacts significantly the national percentage of MDR among new TB cases and that 5.5% of its inter-country variability could be related to corruption, a total that largely exceeds the effect of other health related factors retrieved as significant in the analysis (TB management strategies). Practical implications were discussed. We feel that the European Respiratory Journal would be the journal for this.

Due to our deep feeling that both the table and the figure are important to understand this research and its implications, we took the liberty to include the two in the submitted paper, although only one is permitted in the letter format of the European Respiratory Journal. However, if this is a blocking point we could easily reconsider our view.

We hope you will find it interesting.

Sincerely,

The authors

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Title: The impact of corruption on multidrug-resistant tuberculosis: a quantitative assessment

Running title: Corruption and multidrug-resistant tuberculosis

Authors : Matthieu Hanf ¹⁻², Antoine Adenis ¹, Pierre Couppié ²⁻³, Bernard Carme ¹⁻², Mathieu Nacher¹⁻²

Affiliations : ¹ Centre d'Investigation Clinique Epidémiologie Clinique Antilles Guyane CIC-EC INSERM CIE 802, Cayenne General Hospital, Cayenne, French Guiana. ² Department of Parasitology and Mycology – Team EA 3593, Université des Antilles et de la Guyane, Cayenne, French Guiana ³ Department of Dermatology, Cayenne General Hospital, Cayenne, French Guiana.

Corresponding Author : Matthieu HANF, Centre d'Investigation Clinique Epidémiologie Clinique Antilles Guyane CIC-EC INSERM CIE 802, Cayenne General Hospital, EHPAD, Avenue des Flamboyants, BP 6006 97300 Cayenne, French Guiana, France. E-mail address: matthieu@hanf.fr

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ABSTRACT:

Background: Information on the global risk factors of multidrug-resistant tuberculosis (MDR-TB) is crucial to guide global efforts to contain efficiently its burden. Corruption has been suspected to significantly impact on MDR-TB dynamic. However no quantification is available.

Materials and methods: : The impact of corruption was assessed through univariate and multivariate linear models coupling national percentage of MDR-TB among new tuberculosis cases in 2008 to the national "perceived level of corruption" and a set of adjustment variables measured during the same period.

Results: The final multivariable model included the following significant variables: tuberculosis treatment success rate (p.value = 0.034) and perceived level of corruption (p.value = 0.001). Incorporation of the corruption index in the model increased the adjusted R2 of 5.5%.

Conclusion: This study shows that 5.5% of the inter-country variability of the percentage of MDR-TB among new tuberculosis cases could be related to corruption, a total that largely exceeds the effect of health related factors retrieved as significant in the analysis (current tuberculosis management strategies). Policies and interventions supported by governments must integrate initiatives that recognise how corruption and MDR-TB are inter-related to contain efficiently its burden.

Key words: Corruption; epidemiology; global health; multidrug-resistant tuberculosis; risk factor; tuberculosis

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

TEXT (1072 words)

Multidrug-resistant tuberculosis (MDR-TB) is among the most preoccupying aspects of the pandemic of antimicrobial resistance. In 2008, an estimated 440 000 cases of MDR-TB emerged globally and caused 150 000 deaths (1). MDR-TB is globally distributed, but significant variations in prevalence have been observed in different geographical regions.

The emergence of antimicrobial resistance is a complex problem driven by numerous interconnected factors, many of which are linked to the use of antimicrobials (2). At a community level, according to Caminero et al. (3), 4 main groups of potential factors associated with the selection of resistance and the generation of MDR-TB under epidemic conditions exist: 1) Non-implementation of DOTS and DOTS expansion strategies, 2) Inadequate supply or poor quality of drugs, 3) Patients' inadequate drug intake and 4) other factors such as HIV infection magnitude.

Corruption is a complex problem which threatens the impact of public investments, health care access and services, equity and outcomes (4). There is mounting evidence on the negative effects of corruption on the health and welfare of citizens and on the consequences of corruption in multiple health areas or processes especially in the regulation of product quality and the distribution and use of drugs. More globally, corruption was shown to affect all domains of life from education to economic performance (5) and it could also impact health outcomes.

Due to a strong overlap between corruption consequences and potential MDR-TB drivers, it was argued in the literature that corruption may have a deep impact on drug resistance and particularly on MDR-TB prevalence in some parts of the world (6). However quantitative data at a country level on this subject are very poor. In this perspective, data on MDR-TB

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

prevalence, the potential current risk factors of MDR-TB prevalence, and current corruption levels were gathered.

National MRD-TB data in 2008 were compiled from estimations calculated by the 2010 Global report on surveillance and response of MDR-TB and XDR-TB of the World Health Organisation (7). The studied indicator was the percentage of MDR among new TB cases. National corruption level was assessed through the 2008 Corruption Perception Index (CPI) compiled by Transparency International which orders the countries of the world according to "the degree to which corruption is perceived to exist among public officials and politicians" (8). This index ranges from 0 to 10 with a higher score meaning less (perceived) corruption.

To reflect the economic and development level of studied countries, the Gross Domestic Product (GDP) per capita, the dependency ratio, the literacy rate and the percentage of rural population in 2008 were used (9). The National health and medical level circa 2008 was assessed through the public health expenditure per capita, the HIV prevalence, the national TB case detection rate, the population DOTS coverage, and the TB treatment success rate, which were obtained from the World Bank and the WHO (9).

Logarithm transformations of % MDR among new TB cases in 2008 were used in this study as the outcome variables. Data on 190 countries were provided. For each outcome of interest, unweighted univariate linear regressions were then performed. In order to identify the dominant factors in those that were statistically significant in univariate regression ($p < 0.2$), significant variables were included in a multivariate linear regression model. To determine the final multivariate model, a backward stepwise linear regression was then used to simultaneously adjust for various variables of interest. Statistical significance was set at $p < 0.1$. As a step towards stabilizing the residual variation in regression analysis, logarithmic transformations of the data on health expenditure per capita, GDP per capita were used.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Homoscedasticity was verified graphically and the model including corruption was compared to the model without corruption using the adjusted R^2 . All statistical analyses were conducted using R 2.12.0.

In 2008, the percentage of MDR among new TB cases varied from 0 to 22.3 % with a world median at 1.9 %. For percentage of MDR among new TB cases, the final multivariable model (adjusted $R^2 = 0.068$) included the following significant variables: TB treatment success rate (p.value = 0.034) and CPI score (p.value = 0.001) which were both negatively correlated with the percentage of MDR among new TB cases (Table 1).

Incorporation of the CPI in the model increased the adjusted R^2 from 0.013 to 0.068 (+5.5 %). The coefficient for the CPI [$\beta = -0.0635$] in the final multivariate model would mean that a decrease in 1 point of the CPI (i.e. a more important perceived corruption) would increase the log of percentage of MDR among new TB cases by 0.0635.

After verification, the homoscedasticity and normal assumptions were shown to be respected in the final linear model. A simultaneous bubble plot for the CPI, the TB treatment success rate and the percentage of MDR among new TB cases are shown in the figure 1.

Ecological studies have their limitations (i.e. bias, confounding) (10). Furthermore, this study was limited to a one year period (2008) and, due to lack of objective data was based on rough estimations made by the World Health Organisation. However, all the data used in this analysis is the only available to study the problem at this scale.

The linear regression analysis presented here, after taking into account known significant predictors of MRD-TB level (TB treatment success rate) seems to point towards a significant pejorative link between perceived corruption and the percentage of MDR among new TB cases at national level. In this model, a decrease in CPI of one point (i.e. a more important perceived corruption) was associated with an increase in the logarithm transformation of the

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

percentage of MDR-TB of 0.0635. The adjustment of the final model was poor (Pseudo $R^2 = 0.068$) meaning that only 6.8% of the data variability was captured by the model. Incorporation of CPI in the final model led to a 5.5% rise of the Pseudo R^2 . According to this result, it could be hypothesized that roughly 5.5% of MDR-TB rate variability could be explained by corruption. This study is a first step in the quantification of the current global impact of corruption on MDR-TB in human populations and shows that 5.5% of the inter-country variability of the percentage of MDR among new TB cases could be related to corruption, a total that largely exceeds the effect of health related factors significantly associated with the outcome in the analysis (TB management strategies). But still, because the equation corruption=death is seldom explicit, corruption only seems like a nuisance. In practice, governments still treat health and corruption as separate rather than integral components of the same strategy. To address these obstacles, policies and interventions supported by governments must integrate initiatives that recognise how they are inter-related to fight and contain efficiently the burden of MDR-TB.

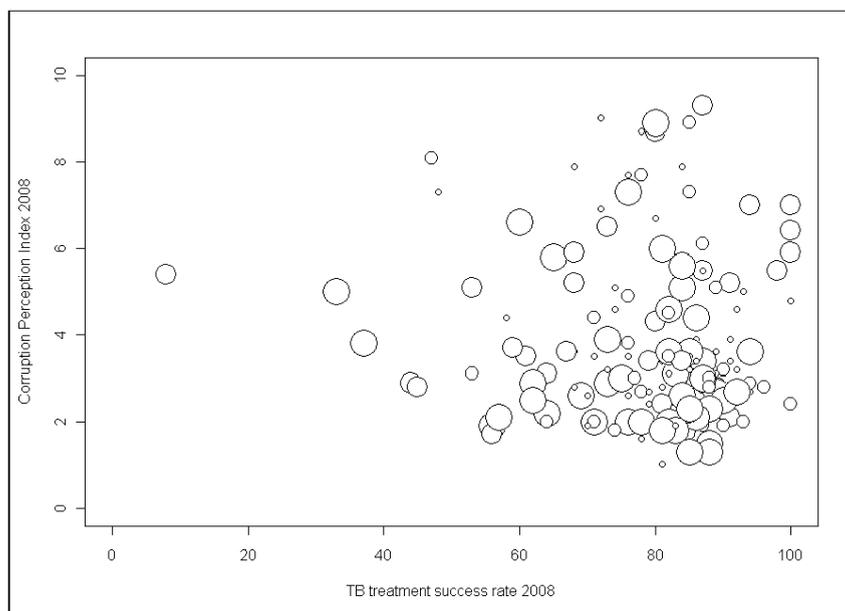
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

REFERENCES

1. Dye C, Williams BG, Espinal MA, Raviglione MC. Erasing the world's slow stain: strategies to beat multidrug-resistant tuberculosis. *Science*. 2002 Mar 15;295(5562):2042-6.
2. Byarugaba DK. A view on antimicrobial resistance in developing countries and responsible risk factors. *Int J Antimicrob Agents*. 2004 Aug;24(2):105-10.
3. Caminero JA. Multidrug-resistant tuberculosis: epidemiology, risk factors and case finding. *Int J Tuberc Lung Dis*. 2010 Apr;14(4):382-90.
4. Vian T. Review of corruption in the health sector: theory, methods and interventions. *Health Policy Plan*. 2008 Mar;23(2):83-94.
5. Abed GT, Gupta S. *Governance, Corruption, & Economic Performance* Washington, D.C: International Monetary Fund; 2002.
6. Cohen JC, Mrazek M, Hawkins L. Tackling corruption in the pharmaceutical systems worldwide with courage and conviction. *Clin Pharmacol Ther*. 2007 Mar;81(3):445-9.
7. World Health Organisation. *Multidrug and extensively drug-resistant TB (M/XDR-TB) : 2010 global report on surveillance and response*. Geneva: World Health Organisation; 2010.
8. Transparency International (TI). *Transparency International website*. 2011 [cited 01/02/2011]; Available from: <http://www.transparency.org/>
9. World Bank. *The World Bank Database*. 2011 [cited 01/04/2011]; Available from: <http://data.worldbank.org/>
10. Morgenstern H. Uses of ecologic analysis in epidemiologic research. *Am J Public Health*. 1982 Dec;72(12):1336-44.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37

Figure 1: Bubble plot of the percentage of MDR among new TB cases according to the CPI score and the TB treatment success rate in 2008. For a given country, the greater the bubble size the greater its percentage of MDR among new TB cases is important



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41

Table 1: Univariate and multivariate linear regression models predicting Log of the percentage of MDR among new TB cases in 190 countries, 2008 (£)

	Univariate analysis				Multivariate analysis			
	Estimate	Std Error	t value	Pr(> t)	Estimate	Std Error	t value	Pr(> t)
Intercept	-	-	-	-	0.8924	0.2469	3.6140	<0.001
Corruption perception index (2008)	-0.0635	0.0183	-3.4703	0.001	-0.0635	0.01930	-3.2920	0.001
Log10(GDP per capita) (2008)	-0.0617	0.0550	-1.1230	0.263	-	-	-	-
Literacy rate (circa 2008)	0.0020	0.0022	0.9100	0.364	-	-	-	-
Dependency ratio (2008)	-0.0022	0.0022	-0.9975	0.320	-	-	-	-
Percentage of rural population (2008)	0.0013	0.0016	0.8261	0.410	-	-	-	-
Log10(Total health expenditure per capita) (2008)	-0.0710	0.0549	-1.2920	0.198	-	-	-	-
Percentage of HIV+ adults (2007) (%)	0.0032	0.0099	0.3223	0.748	-	-	-	-
TB treatment success rate (2008)	-0.0049	0.0027	-1.8436	0.067	-0.0061	0.0028	-2.1410	0.034
TB case detection rate (2008)	-0.0003	0.0014	-0.2143	0.831	-	-	-	-
DOTS coverage (2007)	0.0012	0.0013	0.8667	0.387	-	-	-	-

£: Analysis of the percentage of MDR among new TB cases from a multivariate linear regression model. Only variables with a p.value < 0.1 in univariate regression were incorporated in the final multivariate analysis. Backwards stepwise procedure with a stop criterion fixed at 0.1 was then applied to determine the final multivariate model.

8.12. Conclusions

L'approche populationnelle mise en place ici et portant sur les données onusiennes libres concernant la mortalité juvéno-infantile, le paludisme et les multi-résistances aux antituberculeux a permis de répondre, ou tout du moins, d'ébaucher des éléments de réponse à des questions précises concernant ces pathologies. Ces statistiques, malgré leurs défauts, sont uniques en leur genre et permettent une approche globale des épidémies et de leurs déterminants.

Concernant les interactions paludisme-helminthiases, les résultats de l'analyse populationnelle montrent que dans les pays ayant un climat équatorial et à faible revenu, une prévalence importante d'ascaridiose est associée avec une réduction par 10 de l'incidence du paludisme. Bien qu'il y ait des biais inhérents à cette étude écologique, elle suggère que l'amplitude de l'effet d'ascaris sur le paludisme pourrait être considérable et renforce le besoin d'une réflexion autour de la mise en place de grandes campagnes de vermifugeage dans les zones endémiques pour le paludisme et à faibles ressources médicales.

Pour l'étude corruption et mortalité juvéno-infantile, il a été montré que, après ajustement avec de nombreux déterminants connus de la mortalité, l'impact de la corruption était considérable et que plus de 140 000 décès annuels d'enfants pourraient être indirectement attribués à la corruption. Un constat similaire peut être fait avec les tuberculoses multirésistantes. Ces deux analyses soulignent que la réponse globale à ces deux problèmes majeurs de santé publique doit nécessairement impliquer une augmentation des fonds disponibles pour mettre en place de nouvelles méthodes de prévention, de gestion et de traitements. Cependant sans réfléchir aux mécanismes de contrôle de la corruption nécessaires pour assurer leur bonne utilisation, cela fournira aussi des opportunités supplémentaires à celle-ci. Les interventions de santé publique soutenues par les gouvernements et les donateurs doivent donc maintenant intégrer des initiatives qui reconnaissent comment ces deux phénomènes interagissent.

Ces trois études ne sont que des exemples de l'utilisation des données onusiennes qui peut être faite. Du fait de leur schéma, les études utilisant les statistiques onusiennes sont sujettes aux biais écologiques classiques soulevés en introduction. Malgré tout, en nous renseignant à l'échelle populationnelle, ces dernières sont une source d'information unique pour la communauté scientifique et rendent indispensable leur valorisation pour un éclairage global des grandes pandémies infectieuses.

Bibliographie

1. World Bank. **The World Bank Database**. In; 2011.
2. ONU. **UN Common Database** available at: <http://unstats.un.org> Accessed: July 08,2010.
3. WHO. **WHO Statistical Information System (WHOSIS)** available at: <http://www.who.int/whosis/en/> Accessed: July 08,2010.
4. World Bank. **The World Bank Database** available at:<http://data.worldbank.org/> Accessed: July 08,2010.
5. Susser M, Susser E. **Choosing a future for epidemiology: II. From black box to Chinese boxes and eco-epidemiology**. *Am J Public Health* 1996,86:674-677.
6. World Health Organisation. **Malaria fact sheet**. In. Geneva: World Health Organisation; 2010.
7. World Health Organization. **World malaria report 2009**. In. Geneva: World Health Organization; 2009.
8. Sachs J, Malaney P. **The economic and social burden of malaria**. *Nature* 2002,415:680-685.
9. World Health Organisation. **Soil-transmitted helminths fact sheet**. In. Geneva: World Health Organisation; 2010.
10. Savioli L, Albonico M. **Soil-transmitted helminthiasis**. *Nat Rev Microbiol* 2004,2:618-619.
11. Jackson JA, Friberg IM, Little S, Bradley JE. **Review series on helminths, immune modulation and the hygiene hypothesis: immunity against helminths and immunological phenomena in modern human populations: coevolutionary legacies?** *Immunology* 2009,126:18-27.
12. Bradley JE, Jackson JA. **Immunity, immunoregulation and the ecology of trichuriasis and ascariasis**. *Parasite Immunol* 2004,26:429-441.
13. Mwangi TW, Bethony JM, Brooker S. **Malaria and helminth interactions in humans: an epidemiological viewpoint**. *Ann Trop Med Parasitol* 2006,100:551-570.
14. Nacher M. **Worms and malaria: blind men feeling the elephant?** *Parasitology* 2008,135:861-868.
15. Shapiro AE, Tukahebwa EM, Kasten J, Clarke SE, Magnussen P, Olsen A, *et al*. **Epidemiology of helminth infections and their relationship to clinical malaria in southwest Uganda**. *Trans R Soc Trop Med Hyg* 2005,99:18-24.
16. Bejon P, Mwangi TW, Lowe B, Peshu N, Hill AV, Marsh K. **Helminth infection and eosinophilia and the risk of Plasmodium falciparum malaria in 1- to 6-year-old children in a malaria endemic area**. *PLoS Negl Trop Dis* 2008,2:e164.
17. Brutus L, Watier L, Briand V, Hanitrasoamampionona V, Razanatsoarilala H, Cot M. **Parasitic co-infections: does Ascaris lumbricoides protect against Plasmodium falciparum infection?** *Am J Trop Med Hyg* 2006,75:194-198.

Partie 2 : travaux scientifiques

18. Brutus L, Watier L, Hanitrasoamampionona V, Razanatosarilala H, Cot M. **Confirmation of the protective effect of *Ascaris lumbricoides* on *Plasmodium falciparum* infection: results of a randomized trial in Madagascar.** *Am J Trop Med Hyg* 2007,77:1091-1095.
19. Murray MJ, Murray AB, Murray MB, Murray CJ. **Parotid enlargement, forehead edema, and suppression of malaria as nutritional consequences of ascariasis.** *Am J Clin Nutr* 1977,30:2117-2121.
20. Nacher M, Gay F, Singhasivanon P, Krudsood S, Treeprasertsuk S, Mazier D, *et al.* ***Ascaris lumbricoides* infection is associated with protection from cerebral malaria.** *Parasite Immunol* 2000,22:107-113.
21. van Eijk AM, Lindblade KA, Odhiambo F, Peterson E, Rosen DH, Karanja D, *et al.* **Geohelminth Infections among pregnant women in rural western Kenya; a cross-sectional study.** *PLoS Negl Trop Dis* 2009,3:e370.
22. Degarege A, Animut A, Legesse M, Erko B. **Malaria severity status in patients with soil-transmitted helminth infections.** *Acta Trop* 2009,112:8-11.
23. Nacher M, Singhasivanon P, Yimsamran S, Manibunyong W, Thanyavanich N, Wuthisen R, Looareesuwan S. **Intestinal helminth infections are associated with increased incidence of *Plasmodium falciparum* malaria in Thailand.** *J Parasitol* 2002,88:55-58.
24. Spiegel A, Tall A, Raphenon G, Trape JF, Druilhe P. **Increased frequency of malaria attacks in subjects co-infected by intestinal worms and *Plasmodium falciparum* malaria.** *Trans R Soc Trop Med Hyg* 2003,97:198-199.
25. Transparency International (TI). **Transparency International website.** In. Berlin; 2011.
26. Vian T. **Review of corruption in the health sector: theory, methods and interventions.** *Health Policy Plan* 2008,23:83-94.
27. Transparency International. **Global corruption report 2006 : Corruption and Health.** In. Berlin: Transparency International (TI); 2006.
28. World Bank. **Helping countries combat corruption: progress at the World Bank since 1997.** In. Washington, DC: World Bank; 2000.
29. Abed GT, Gupta S. **Governance, Corruption, & Economic Performance** Washington, D.C: International Monetary Fund; 2002.
30. Maestad O, Mwisongo A. **Informal payments and the quality of health care: Mechanisms revealed by Tanzanian health workers.** *Health Policy* 2011,99:107-115.
31. Azfar O, Gurgur T. **Does Corruption Affect Health and Education Outcomes in the Philippines?** In. College Park: Institutional Reform and Informal Sector; 2008.
32. United Nations. **The Millennium Development Goals Report 2000.** In. New York: United Nations; 2000.
33. Transparency International. **Global corruption report 2008 : Corruption in the water sector.** In. Berlin: Transparency International (TI); 2008.

Chapitre 8- Utilisation des statistiques onusiennes : Applications à l'étude des déterminants de la dynamique des résistances aux antituberculeux, du paludisme et de la mortalité juvéno-infantile.

34. Santosham M, Chandran A, Fitzwater S, Fischer-Walker C, Baqui AH, Black R. **Progress and barriers for the control of diarrhoeal disease.** *Lancet* 2010,376:63-67.
35. Black RE, Morris SS, Bryce J. **Where and why are 10 million children dying every year?** *Lancet* 2003,361:2226-2234.
36. Dye C, Williams BG, Espinal MA, Raviglione MC. **Erasing the world's slow stain: strategies to beat multidrug-resistant tuberculosis.** *Science* 2002,295:2042-2046.
37. Anderson RM. **The pandemic of antibiotic resistance.** *Nat Med* 1999,5:147-149.
38. Upshur R, Singh J, Ford N. **Apocalypse or redemption: responding to extensively drug-resistant tuberculosis.** *Bull World Health Organ* 2009,87:481-483.
39. Marahatta SB. **Multi-drug resistant tuberculosis burden and risk factors: an update.** *Kathmandu Univ Med J (KUMJ)* 2010,8:116-125.
40. Caminero JA. **Multidrug-resistant tuberculosis: epidemiology, risk factors and case finding.** *Int J Tuberc Lung Dis* 2010,14:382-390.
41. World Health Organisation. **Multidrug and extensively drug-resistant TB (M/XDR-TB): 2010 global report on surveillance and response.** In. Geneva: World Health Organisation; 2010.
42. World Health Organisation. **Multidrug and extensively drug-resistant TB (M/XDR-TB) : 2010 global report on surveillance and response.** In. Geneva: World Health Organisation; 2010.
43. Byarugaba DK. **A view on antimicrobial resistance in developing countries and responsible risk factors.** *Int J Antimicrob Agents* 2004,24:105-110.
44. Udawadia ZF. **India's multidrug-resistant tuberculosis crisis.** *Ann N Y Acad Sci* 2001,953:98-105.
45. Cohen JC, Mrazek M, Hawkins L. **Tackling corruption in the pharmaceutical systems worldwide with courage and conviction.** *Clin Pharmacol Ther* 2007,81:445-449.

Chapitre 9 - Discussion générale, perspectives et conclusions

« C'est peut-être en prenant les problèmes de santé actuels sous l'angle d'une vision écologique intégrative, où l'homme est un élément essentiel et indissociable d'un système très complexe et interactif, qu'il faut aujourd'hui reconsidérer notre analyse de la maladie et des épidémies »
Jean François Guégan

Les différentes études développées dans cette thèse ne sont que des illustrations de l'utilité de l'association des méthodes écologiques aux données publiques pour éclairer sous un angle nouveau des problématiques des maladies infectieuses. Que ce soit à un niveau local ou global, de par le grand nombre d'associations possibles entre les différents jeux de données existants, ce type d'étude fournit la flexibilité nécessaire à l'étude des interactions complexes des nombreux déterminants de la santé, et cela quel que soit le niveau d'étude. En permettant d'appréhender les différents niveaux d'analyses de l'épidémiologie, les méthodes écologiques permettent de prendre en compte simultanément dans un même modèle autant les facteurs populationnels et écologiques que les facteurs individuels, tous deux importants pour expliquer la dynamique des maladies chez les individus.

Nous avons vu que la notion de complexité est au cœur des questions épidémiologiques actuelles, et qu'elle est cruciale pour la compréhension des phénomènes liés à l'émergence d'agents pathogènes et à la dynamique des infections. La recherche contemporaine est maintenant confrontée à une nécessité de plus en plus contraignante : continuer de produire des connaissances nouvelles dans des domaines de plus en plus spécialisés. Cela est nécessaire mais pas suffisant. Elle doit aussi affronter la réalité des systèmes complexes dont la simple description des composants ne suffit plus.

D'autre part nous avons aussi vu que l'épidémiologie tend à être de plus en plus une discipline intégrative et collaborative, amplifiant la nécessité d'un accès ouvert aux données. Ce type d'études nécessite presque toujours des interactions fortes avec des disciplines voisines. Ce vaste champ de compétences et de connaissances génère des défis majeurs pour trouver des moyens efficaces pour découvrir, accéder, intégrer, et analyser la portée et le volume des informations intéressantes. L'association des méthodes écologiques aux données libres constitue un élément de réponse méthodologique prometteur à ces défis.

Quelles peuvent être les recommandations pratiques à faire pour que cette nouvelle démarche et appréhension de l'épidémiologie se développe et diffuse le plus facilement et rapidement possible dans le monde de la santé publique ?

1. Garder à l'esprit la complexité naturelle des systèmes vivants dans lesquels les pathogènes, leurs vecteurs et réservoirs évoluent, et en tenir compte dans les études scientifiques et les programmes de contrôle.

Chapitre 9 - Discussion générale, perspectives et conclusions

2. Prendre conscience des limites d'une vision réductionniste naïve. Les gens ne sont pas simplement des collections aléatoires de cellules ou de molécules, et les populations ne sont pas simplement des collections aléatoires d'individus. Les systèmes adaptatifs complexes ont une «vie» qui est plus que la somme de leurs composantes.
3. Mieux prendre en compte les dynamiques temporelles et spatiales des agents pathogènes et de leurs espèces hôtes dans les politiques de contrôle et de santé publique.
4. Mieux considérer l'environnement physique, économique et social et leurs conséquences sur la diffusion et la propagation d'agents pathogènes, de leurs vecteurs et/ou de leurs réservoirs.
5. Valoriser dans les formations universitaires en santé publique la découverte et l'apprentissage des schémas d'étude autres que les études cas-témoins ou de cohorte ainsi que les approches intégratives et multidisciplinaires.
6. Valoriser les parcours autres que médicaux en recherche épidémiologique et encourager les équipes transdisciplinaires intégrant une grande variété de parcours universitaires (biologiste, mathématicien, sociologue, médecin, démographe, géographe, agronome, statisticien, informaticien...)
7. Elaborer des programmes de recherches dans la dimension spatiale à laquelle se posent les problèmes et considérer attentivement les interrelations possibles entre les différents niveaux d'analyses.
8. Développer et promouvoir les méthodes statistiques adaptées à un schéma d'étude écologique permettant de prendre en compte dans un même modèle des variables définies sur plusieurs niveaux (comme les modèles multi-niveaux par exemple) et de les sortir des domaines spécialisés les utilisant.
9. Exploiter plus profondément les bases de données existantes et généraliser leurs recoupages pour l'obtention de résultats originaux.
10. Continuer à développer activement les sciences de l'information en épidémiologie (comme l'utilisation des systèmes d'informations géographiques par exemple)
11. Développer la culture de l'Open Data dans la communauté de chercheurs travaillant en santé publique. Une grande partie des infrastructures, des normes techniques, et des incitations qui sont nécessaires pour soutenir le partage des données manquent.
12. Appliquer prioritairement cette démarche dans les territoires, comme la Guyane Française, caractérisés par une forte complexité épidémiologique et où la diversité des hommes et des pathogènes, et l'intensité des facteurs qui les poussent à se croiser, sont particulièrement remarquables et en font des sites d'études multidisciplinaires particulièrement intéressants.

Partie 2 : travaux scientifiques

13. Généraliser cette démarche aux maladies chroniques non infectieuses dont la multiplicité des déterminants en font, au même titre que les maladies infectieuses, un terrain d'application idéal de ce type de méthodologie.

En appliquant ces recommandations, une ouverture de l'épidémiologie à des approches plus «intégratives» pourrait se développer et permettre de mieux répondre aux grands enjeux de santé publique actuels et futurs.

ANNEXES

Publications et communications personnelles

Publications

1. **Hanf M**, Adenis A, Couppie P, Carme B, Nacher M. HIV-associated histoplasmosis in French Guiana: recent infection or reactivation? *AIDS* 2010;24:1777-1778.
2. **Hanf M**, Stephani A, Basurko C, Nacher M, Carme B. Determination of the Plasmodium vivax relapse pattern in Camopi, French Guiana. *Malar J.* 2009;8:278.
3. Basurko C, **Hanf M**, Han-Sze R, Rogier S, Heritier P, et al. (2011) Influence of climate and river level on the incidence of malaria in Cacao, French Guiana. *Malar J* 10: 26.
4. **Hanf M**, Adenis A, Nacher M, Carme B (2011) The role of El Nino southern oscillation (ENSO) on variations of monthly Plasmodium falciparum malaria cases at the cayenne general hospital, 1996-2009, French Guiana. *Malar J* 10: 100.
5. Stephani A, **Hanf M**, Girot R, Nacher M, Carme B. (2011) Environmental, meteorological and behavioural risk factors for malaria attacks in children in Camopi, French Guiana *Malaria J.* 2011.
6. **Hanf M**, Van-Melle A, Fraisse F, Roger A, Carme B, et al. (2011) Corruption kills: Estimating the global impact of corruption on children deaths. *Plos One.*
7. **Hanf M**, Adenis A, Couppié P, Carme B, Nacher M. The impact of corruption on multidrug-resistant tuberculosis: a quantitative assessment. *ERJ* 2011 (in submission)
8. **Hanf M**, Carme B, Nacher M. Regression tree analysis of national gastrointestinal nematode and malaria data shows that greater Ascaris lumbricoides prevalences are associated with lower malaria incidences. *Plos One* 2011 (in submission).
9. **Hanf M**, Adenis A, Carme B, Couppie P, Nacher M. Disseminated Histoplasmosis seasonal incidence variations: an additional argument for recent infection? *TRSTMH* 2011 (In submission).
10. Carme B, **Hanf M**, Nacher M. Classification of populations exposed to congenital toxoplasmosis: An analytical review of world seroprevalences. *IJP.* 2011 (in submission).
11. **Hanf M**, Nacher M, Adenis A, Carme B. Prediction of the seroprevalence level and the risk of congenital toxoplasmosis from geographical and societal environments: an analytical review. *IJP.* 2011 (in submission)

Communications affichées

1. Roger A, **Hanf M**, Basurko C, Nacher M, Carme B, Couppié P. Climat et leishmaniose en Guyane Congrès de la société Française de Parasitologie, Strasbourg mai 2011
2. Basurko C, **Hanf M**, Han-Sze R, Rogier S, Heritier P, Grenier C, Joubert M, Nacher M, Carme B. Influence du niveau des cours d'eau sur l'incidence à Plasmodium vivax à Cacao (Guyane) Congrès de la Société Française de Parasitologie, Strasbourg mai 2011
3. Stephani A, **Hanf M**, Nacher M, Carme B, Facteurs de risque environnementaux et comportementaux d'accès palustres chez les enfants amérindiens de Camopi, Région de l'Oyapock, Guyane Congrès de la Société Française de Parasitologie, Strasbourg mai 2011
4. Carme B, **Hanf M**, Demar M, Nacher M. Profil de séroprévalence et exposition au risque de toxoplasmose congénitale Congrès de la Société Française de Parasitologie, Strasbourg mai 2011
5. **Hanf M**, Nacher M, Demar M, Carme B. Evaluation du niveau de séroprévalence et du risque de toxoplasmose congénitale à partir de l'environnement géographique et sociétal Congrès de la Société Française de Parasitologie, Strasbourg mai 2011
6. Roger A, **Hanf M**, Dufour J, Basurko C, Lazar M, Sainte-Marie D, Simon S, Nacher M, Carme B, Couppié P. Climat et leishmaniose cutanée en Guyane. Etude à partir d'une série de 1302 patients vus entre 1994 et 2010. Décembre 2011, Journées Dermatologiques de Paris

Résumé

Introduction: En épidémiologie autant les facteurs populationnels et écologiques que les facteurs individuels sont importants pour expliquer la dynamique des maladies. Les études écologiques sont maintenant considérées comme prometteuses en raison de leur capacité à intégrer dans un même modèle l'ensemble de ces facteurs. D'autre part, l'accès aux sources de données nécessaires pour mettre en place des approches multidisciplinaires en épidémiologie est devenu stratégique. Le mouvement récent de l'open data pourrait ainsi jouer un rôle important dans la pérennisation de ces démarches. Les différentes études développées dans cette thèse montrent que l'association des méthodes écologiques aux données publiques permet d'éclairer sous un angle nouveau des problématiques des maladies infectieuses.

Matériels et méthodes / résultats: Pour étudier l'impact du climat et de ses perturbations sur la santé, il est possible de mettre en œuvre des méthodes écologiques particulières : les analyses des séries temporelles. Du fait de la grande disponibilité des données climatiques, ces méthodes sont très facilement mises en œuvre. En Guyane Française, ces dernières ont contribué à une meilleure compréhension de la dynamique de la transmission du paludisme et de la leishmaniose cutanée ainsi que des caractéristiques écologiques de leurs vecteurs locaux. Elles ont également permis de quantifier globalement l'impact du phénomène El Niño sur leur niveau épidémique. De même, la forte influence des variables climatiques sur les incidences d'une autre pathologie, l'histoplasme disséminée, suggère qu'une proportion importante des cas pourrait être due à une exposition récente. Ces résultats donnent de précieuses informations épidémiologiques pour les cliniciens.

Les méthodes écologiques peuvent permettre d'utiliser les données issues de la littérature scientifique pour la réalisation d'études originales. L'utilisation et l'application de ces méthodes écologiques à des données concernant la séroprévalence toxoplasmose dans les populations humaines a permis de définir des indicateurs simples permettant de juger grossièrement du risque de toxoplasmose congénitale dans une population. Dans un deuxième temps elles ont aussi permis d'identifier les principaux facteurs influençant le niveau de séroprévalence globale et donc le risque de toxoplasmose congénital associé. En extrapolant ces résultats, une première carte mondiale du risque de toxoplasmose congénitale a ainsi pu être dessinée.

Les données onusiennes sont uniques pour étudier à l'échelle populationnelle la dynamique et les déterminants importants des grandes maladies infectieuses. 3 exemples illustrant ce concept et utilisant des méthodes écologiques ont été développés. La première analyse montre que dans les pays à faible revenu et ayant un climat équatorial, une prévalence importante d'ascaridiose est associée avec une réduction par 10 de l'incidence du paludisme. Ces résultats renforcent le besoin d'une réflexion autour de la mise en place des grandes campagnes de vermifugeage. Les deux autres études démontrent l'impact significatif de la corruption sur la mortalité juvénile-infantile et l'importance des résistances aux anti-tuberculeux et soulignent elles aussi la nécessité d'une réflexion sur les mécanismes de contrôle de la corruption nécessaires à une amélioration durable de la santé des populations.

Conclusions : Les différentes études développées dans cette thèse montrent que l'association des méthodes écologiques aux données publiques permet d'éclairer sous un angle nouveau des problématiques des maladies infectieuses. Que ce soit à un niveau local ou global, de par le grand nombre d'associations possibles entre les différents jeux de données existants, ce type d'étude fournit la flexibilité nécessaire à l'étude des interactions complexes des nombreux déterminants de la santé. Ces motifs sont à l'origine ces dernières années de son succès grandissant en épidémiologie environnementale et sociale. Cette démarche peut facilement se généraliser aux maladies chroniques non infectieuses dont la multiplicité des déterminants en font, au même titre que les maladies infectieuses, un terrain d'application idéal de ce type de méthodologie.

Mots clefs : épidémiologie ; complexité ; étude écologique ; données libres ; maladies infectieuses

Abstract

Introduction: In epidemiology, populational and ecological factors are as important as individual factors to explain disease dynamics. Ecological studies are now considered promising because of their ability to integrate in the same model all these factors. On the other hand, access to the different sources of data needed to develop multidisciplinary approaches in epidemiology has become strategic. The recent open data movement could thus play an important role in the sustainability of these approaches. The various studies developed in this thesis show that the combination of ecological methods with public data could give original results in the issues of infectious diseases.

Materials and methods / results: To study the health impacts of climate, it is possible to implement specific ecological methods called time series analysis. Because of the wide availability of climate data, these methods could be easily implemented. In French Guiana, they have contributed to a better understanding of the dynamics of transmission of malaria and cutaneous leishmaniasis and of the ecological characteristics of their local vectors. They also helped to quantify the overall impact of El Niño on their epidemic level. Similarly, the strong influence of climatic variables on the incidence of disseminated histoplasmosis, suggests that a significant proportion of cases could be due to recent exposure. These results provide valuable epidemiological information for clinicians.

Ecological methods have the flexibility to use data from the scientific literature for the realization of original studies. The application of ecological methods on data focused on toxoplasmosis seroprevalence in human populations has identified simple indicators that could be used to roughly evaluate the risk of congenital toxoplasmosis in a population. They have also identified the main factors influencing the overall level of seroprevalence and associated risk of congenital toxoplasmosis. With an extrapolation of these results, a first world map of the risk of congenital toxoplasmosis has been drawn.

The UN data are unique to study the dynamic and determinants of major infectious diseases at a population scale. Three examples of this concept, applying ecological methods, have been developed. The first analysis showed that in low-income countries with an equatorial climate, a high prevalence of Ascariasis was associated with a reduction by 10 of the incidence of malaria. These results reinforce the need for reflection on the establishment of large deworming campaigns. The other two studies demonstrate the significant impact of corruption on child mortality and on multi-resistant tuberculosis and emphasize the need for reflection on the mechanisms of corruption control necessary to significantly improve the health of populations.

Conclusions: The various studies developed in this thesis show that the combination of ecological methods to public data sheds a new light on the issues of infectious diseases. Both at local and global level, with the large number of possible associations between the various existing datasets, this type of methodology provides the flexibility to study the complex interactions of many health determinants. These reasons are behind his growing success in environmental and social epidemiology these last years. This can easily be generalized to the field of non-infectious chronic diseases, where multiple determinants represent, as in infectious diseases, an ideal field of application of such methodology.

Key words: epidemiology; complexity; ecological study; open data; infectious diseases