

Université de Montréal

Analyse exploratoire de la dynamique des maladies infectieuses  
communes de l'enfance au Canada

Par  
Helen Trottier  
Département de médecine sociale et préventive  
Faculté de médecine

Thèse présentée à la Faculté des études supérieures  
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D)  
en santé publique  
option épidémiologie

Août 2004

© Helen Trottier, 2004



WA

5

U58

2004

V. 010

**Direction des bibliothèques**

**AVIS**

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

**NOTICE**

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal  
Faculté des études supérieures

Cette thèse intitulée :

Analyse exploratoire de la dynamique des maladies infectieuses  
communes de l'enfance au Canada

Présentée par  
Helen Trottier

a été évaluée par un jury composé des personnes suivantes :

D<sup>re</sup> Maria Victoria Zunzunegui, présidente du jury

D<sup>r</sup> Pierre Philippe, directeur de recherche

D<sup>r</sup> Roch Roy, membre du jury

D<sup>re</sup> Hélène Carabin, examinatrice externe

Thèse acceptée le : .....

## RÉSUMÉ

---

On est loin d'avoir éliminé les vieilles maladies infectieuses communes de l'enfance, même dans les pays développés. Aussi, l'importante morbidité et mortalité associée à ces maladies infectieuses nécessite-t-elle d'en comprendre les modalités d'existence et de propagation. En dépit des efforts d'élimination déployés depuis plusieurs années au Canada, les maladies infectieuses transmissibles qu'on peut prévenir par la vaccination continuent de frapper périodiquement.

Ces maladies se prêtent bien à des études incorporant le temps comme variable d'exploration. Les données sur ces maladies sont souvent obtenues à la suite d'une surveillance épidémiologique effectuée par les instances officielles des gouvernements. Ces données se présentent souvent sous forme de séries chronologiques. On peut voir ces séries comme l'expression d'un système dynamique dont le mouvement dans le temps révèle, et son histoire propre, et celle d'autres systèmes en interaction. L'épidémiologiste qui étudie la structure de séries chronologiques est à la recherche de *patterns* d'émergence et de structures temporelles qui peuvent contribuer à expliquer les tâtonnements de la série dans son périple au cours du temps.

Ce projet de recherche exploratoire a consisté à analyser les séries chronologiques des maladies infectieuses communes (rougeole, rubéole, coqueluche et oreillons), au Canada, afin de caractériser leur dynamique à l'échelle populationnelle et d'explorer les tenants de leur « comportement » en termes de loi d'organisation. Outre le temps qui a servi de variable d'intérêt, nous avons privilégié l'étude de maladies différentes ainsi que les périodes de pré et postvaccination en guise de variables

contextuelles. De plus, une analyse de la qualité des données a permis d'interpréter adéquatement les résultats de cette recherche.

L'analyse de la qualité des données a montré que, en dépit de lacunes importantes comme la sous-estimation, les séries chronologiques utilisées ne posent pas de problème important pour l'étude de la dynamique des maladies sélectionnées. Ces données ont été soumises à deux méthodes de modélisation stochastique, celle de Box-Jenkins et celle de la loi de puissance inverse, l'une et l'autre appropriées pour caractériser les dynamiques populationnelles. La modélisation de Box-Jenkins des séries de rougeole, de coqueluche, d'oreillons et de rubéole a permis de dégager une structure stochastique unique à l'origine des séries, et ce malgré les différences biologiques importantes de ces infections et des facteurs contextuels présumés importants comme la vaccination de masse. Cette observation inattendue a permis de conclure que ce type de modélisation se limite à l'identification de la dynamique collective des infections et reste insensible aux caractéristiques biologiques spécifiques des maladies de même qu'aux effets de la vaccination. Par ailleurs, l'analyse des propriétés d'échelle des épidémies par une loi de puissance inverse a montré un ajustement satisfaisant de l'émergence des épidémies de rubéole, et de rougeole (dans une population hautement vaccinée). Quant aux oreillons et à la coqueluche (et à la rougeole dans un contexte de non ou faible vaccination), ils ont plutôt été ajustés par une loi exponentielle des tailles d'épidémies. L'ajustement par l'une ou l'autre loi a permis d'inférer la topologie du réseau de transmission des maladies à l'échelle individuelle. Nous avons alors conclu que l'existence d'une invariance d'échelle pour la rubéole et la rougeole (vaccination élevée) suggère une propagation de la maladie par l'entremise d'un réseau hétérogène de contacts interindividuels. Au contraire, les oreillons et la coqueluche (et la rougeole dans

un contexte de population peu ou non vacciné), qui adhèrent à une loi exponentielle, sont caractéristiques d'un réseau homogène de contacts interindividuels. La connaissance de la topologie des réseaux de contacts a des implications directes sur le choix d'une politique de vaccination.

Cette thèse est un exemple d'arrimage des échelles populationnelle et individuelle des infections transmissibles par une investigation de la « boîte noire » qui les relie. Plus spécifiquement, nous avons suggéré que des dynamiques individuelles distinctes de transmission peuvent produire des manifestations collectives ou populationnelles similaires. Comme les interventions de santé publique ciblent les individus, il n'est pas étonnant qu'une intervention universelle, la même dans tous les contextes, présente des effets discordants sur le plan collectif. Il apparaît donc nécessaire d'ajuster l'intervention à la structure spécifique des réseaux de contacts qui configurent l'ordre populationnel. Les modalités d'existence et de propagation des maladies transmissibles à l'échelle populationnelle, une fois arrimées à l'échelle individuelle, permettront d'opter pour des politiques préventives plus appropriées.

**MOTS CLÉS :** séries chronologiques, dynamique, infections communes de l'enfance, rougeole, coqueluche, oreillons, rubéole, vaccination de masse, structure stochastique, loi de puissance inverse, système complexe, Box-Jenkins, transmission interindividuelle.

## ABSTRACT

---

Childhood infectious diseases are far from being eliminated even in developed countries. Moreover, the significant mortality and morbidity patterns associated with these diseases require that one understands the specifics of their existence and propagation. Despite multiple efforts to eliminate these diseases in Canada since a few years, vaccine-preventable diseases continue to emerge.

Such diseases are amenable to studies incorporating time as an exploratory variable. Time-related disease data are often obtained following epidemiologic surveillance by governmental authorities. These data are called time series. One can see a time series as a dynamic system, the expression of which in time reveals the history of this particular system as well as that of the other systems with which it interacts. Therefore, the epidemiologist who studies time series is seeking for emergent patterns and temporal structures that could contribute to explain the behaviour of the series in the course of time.

This exploratory research has consisted of analyzing the time series of common infectious diseases (measles, rubella, pertussis and mumps) in Canada in order to characterize their dynamics at the population scale and explore their behaviour in terms of organisational law. Beside time as a central variable, we have considered different types of diseases, and pre- and post-vaccination periods as contextual variables. Moreover, an analysis of data quality has been undertaken to adequately interpret the study results.

The quality analysis has shown that, despite significant shortcomings such as disease underreporting, the time series used for studies of dynamics are not subject to bias. The data available were studied with two stochastic modelling methods: Box-Jenkins modelling and inverse power-law modelling. These methods are appropriate to characterize population dynamics. Box-Jenkins modelling of measles, rubella, pertussis and mumps has allowed to extract a single stochastic dynamics despite the biological differences of the infectious diseases and despite the different vaccination conditions involved. This unexpected observation has led us to conclude that the Box-Jenkins methodology is not sensitive enough to rescue the finely-tuned biological dynamics of the various types of infections; it has rather been concluded that the methodology is sensitive to the generic organizational principle of the diseases at the population scale. Moreover, the scaling analysis of the epidemics sizes by an inverse power-law has shown good fit for rubella, and for measles in a highly vaccinated population. Mumps and pertussis (and measles before mass vaccination) were rather adjusted by an exponential law. Adjustment by either law has allowed us to make inferences about the network topology of transmission of these diseases at the individual scale. We have therefore concluded that the scale-invariance of rubella and measles (with high vaccination coverage) epidemic sizes is in line with a heterogeneous network of inter-individual contacts. By contrast, mumps and pertussis (and measles before mass vaccination), which comply with an exponential law, were rather characterized by a homogenous network of inter-individual contacts. Network topology has direct implications for the appropriateness of a vaccination strategy.

This thesis allowed us to distinctly characterize the population and individual scales of the transmissible infections by investigating the in-between “black box”. More

specifically, we have suggested that different individual transmission dynamics may entail similar collective or population manifestations. As public health interventions target individuals, it is not surprising that a universal intervention scheme, the same in all contexts, presents dissimilar impact at the collective scale. It seems necessary to adjust intervention to the specifics of the network topology that makes up the population scale. The modalities of existence and propagation of common infectious diseases at the population scale are not independent of the individual scale. Studies of such interactions will allow for more appropriate vaccination policies.

**KEY WORDS :** Time series, dynamics, childhood infectious diseases, measles, pertussis, mumps, rubella, mass vaccination, stochastic structure, Box-Jenkins, inverse power-law, complex system, inter-individual transmission.

## TABLE DES MATIÈRES

---

RÉSUMÉ.....	i
ABSTRACT.....	iv
TABLE DES MATIÈRES.....	vii
LISTE DES TABLEAUX.....	xii
LISTE DES FIGURES.....	xiv
LISTE DES SIGLES ET ABRÉVIATIONS.....	xvii
REMERCIEMENTS.....	xviii
INTRODUCTION.....	1
CHAPITRE 1 : État des connaissances.....	6
1.1 Généralités.....	6
1.2 Les patterns d'émergence des maladies infectieuses communes de l'enfance.....	7
1.3 Développement de modèles théoriques relatifs aux patterns épidémiques.....	7
1.4 Facteurs à l'origine des patterns des maladies infectieuses.....	9
1.4.1 Cycles, nombre reproductif de base et variation saisonnière des taux de contacts.....	9
1.4.2 Natalité, immigration et vaccination de masse.....	9
1.4.3 Facteurs climatiques et sociaux et structure par âge.....	10
1.4.4 Facteurs contextuels et aléatoires.....	10
1.5 Types de dynamique.....	11
1.6 Méthodes.....	11
1.6.1 Simulations déterministes non linéaires.....	12

1.6.2	Simulations stochastiques et modélisation empirique.....	15
1.6.3	Modèles stochastiques à implication déterministe.....	16
1.7	Objectif de la thèse .....	20
1.8	Organisation de la thèse par articles.....	21
CHAPITRE 2 : Données et méthodes.....		23
2.1	Données.....	23
2.1.1	Provenance des données.....	23
2.1.2	Épidémiologie des maladies transmissibles analysées.....	24
2.1.2.1	Rougeole.....	24
2.1.2.2	Coqueluche.....	25
2.1.2.3	Oreillons.....	26
2.1.2.4	Rubéole.....	26
2.1.3	Structure des séries chronologiques analysées.....	27
2.1.4	Qualité des données.....	29
2.2	Méthodologie.....	35
2.2.1	Méthode de Box-Jenkins.....	35
2.2.2	Méthode de loi de puissance inverse.....	37
CHAPITRE 3 : Articles.....		38
ARTICLE 1 :		
Qualité des données de surveillance de la coqueluche, de la rougeole, de la rubéole et des oreillons au Canada.....		39
	Résumé.....	40
	Introduction.....	42

Bref historique du système de surveillance des MADO au Canada .....	43
Méthode .....	45
Résultats .....	47
Variation de la déclaration en fonction de la hiérarchie du système de surveillance .....	47
1) Consultation médicale .....	47
2) Diagnostic de la maladie .....	47
3) Déclaration des cas par les intervenants .....	49
Évaluation de la proportion des cas déclarés au système de surveillance .....	51
Évaluation des variations de l'incidence des maladies au cours du temps .....	54
Discussion .....	54
Conclusion .....	60
Bibliographie .....	62
Annexe .....	65

## ARTICLE 2 :

Time series analysis of some childhood infectious diseases before and after mass vaccination in Canada: Box-Jenkins modelling .....	66
Abstract .....	67
Introduction .....	68
Data .....	71

Methods.....	74
Results.....	79
Discussion.....	85
Conclusion.....	90
References.....	92
Appendix A: Time series data.....	96
Appendix B: Box-Jenkins (ARIMA) modelling.....	98
Appendix C: Formulas.....	100

### ARTICLE 3 :

#### Scaling properties of childhood infectious diseases before and after mass vaccination in Canada.....

Abstract.....	102
Introduction.....	103
Data.....	105
Method.....	106
Epidemic size.....	106
Modelling.....	108
Results.....	110
Discussion.....	115
Conclusion.....	123
References.....	126
Appendix.....	131

CHAPITRE 4 : Discussion et conclusion.....	132
BIBLIOGRAPHIE.....	143
ANNEXE I.....	149
1. Séries chronologiques d'infections annuelles, 1924-2002.....	150
2. Séries chronologiques analysées.....	152
ANNEXE II : Graphique « moyenne-étendue » pour chaque série chronologique .....	156
ANNEXE III : Analyse de Box-Jenkins détaillée (identification, estimation des paramètres et diagnostic) pour chaque série chronologique.....	163

## LISTE DES TABLEAUX

---

### CHAPITRE 1 : État des connaissances

Tableau 1.1	Nombre reproductif de base et périodes de latence et de contagiosité de certaines maladies infectieuses communes.....	8
-------------	---	---

### CHAPITRE 2 : Données et méthodes

Tableau 2.1	Proportion de cas déclarés au système de surveillance, Canada et pays industrialisés.....	31
-------------	---	----

Tableau 2.2	Nombre de provinces qui ne déclaraient pas les cas de rougeole, de rubéole et d'oreillons de 1953 à 1959 et pourcentage de la population canadienne que ces provinces représentent.....	32
-------------	---	----

Tableau 2.3	Nombre de provinces qui ne déclaraient pas les cas de rougeole et de rubéole de 1970 à 1976 et pourcentage de la population canadienne que ces provinces représentent.....	33
-------------	--	----

Tableau 2.4	Sommaire des caractéristiques des séries d'infections analysées.....	35
-------------	--	----

### CHAPITRE 3 : Articles

#### ARTICLE 1

Tableau 1	Évaluation de la déclaration des cas en fonction de la hiérarchie du système de surveillance selon différentes études.....	50
-----------	--	----

Tableau 2	Variations de la proportion des cas déclarés selon différentes études.....	53
-----------	--	----

## ARTICLE 2

Table 1a	Estimates of parameters for the pertussis series (1953-1962). ARIMA output SPSS Trend 11.5.....	82
Table 1b	Diagnosis checking for the pertussis series (1953-1962). Residual ACF and Ljung-Box statistics by time lag. ARIMA output SPSS Trend 11.5.....	82
Table 2	Estimates of parameters (and standard errors).....	84

## ARTICLE 3

Table 1	Measles. Goodness-of-fit ( $R^2$ ) to IPL and exponential distribution functions.....	111
Table 2	Rubella. Goodness-of-fit ( $R^2$ ) to IPL and exponential distribution functions.....	112
Table 3	Pertussis. Goodness-of-fit ( $R^2$ ) to IPL and exponential distribution functions.....	113
Table 4	Mumps. Goodness-of-fit ( $R^2$ ) to power law and exponential distribution functions.....	114
Table 5	The IPL slope, $b$ , for series represented by IPL with their corresponding estimated standard error (in parenthesis).....	121
Table A	Epidemic size distribution, measles (1990-2002).....	131

## LISTE DES FIGURES

---

### CHAPITRE 1 : État des connaissances

Figure 1.1	Modèle SEIR. Nombre de nouveaux cas de rougeole à travers le temps pour une population non vaccinée d'un million de personnes à la suite de l'introduction d'un cas infectieux introduit au temps $t_0$ .....	13
------------	---	----

### CHAPITRE 3 : Articles

#### ARTICLE 1

Figure 1	Organisation du système de surveillance des MADO au Canada.....	45
Figure 2	Étapes menant à la déclaration des cas de MADO .....	46

#### ARTICLE 2

Figure 1	Mean-Range Plot for Pertussis (1953-1962). (a) Mean and range for non-transformed data (X). (b) Mean and range applying the square root transformation (SRX). (c) Mean and Range applying the logarithmic transformation (LX).....	79
Figure 2	Pertussis (1953-1962): (a) autocorrelation function (ACF) of $\ln x_t$ ; (b) autocorrelation function (ACF) of $\nabla_{13} \ln x_t$ ; (c) partial autocorrelation function (PACF) of $\nabla_{13} \ln x_t$ ; (d) autocorrelation function of the residuals.....	81
Figure A1	4-weekly notifications of measles, mumps, rubella (1953-1959) and pertussis (1953-1962) in Canada.....	96

Figure A2	4-weekly notifications of measles, rubella and pertussis (1970-1976) in Canada. Adjusted for annual non-stated data	96
Figure A3	Monthly notifications of measles, mumps, rubella and pertussis (1990-2002) in Canada adjusted on a 28-day basis	96
Figure A4	Seasonal differenced and logarithm transformed time series of 4-weekly (1953-1959, 1970-1976) and monthly (1991-2002) notifications of pertussis, mumps, measles and rubella in Canada. )	97
ARTICLE 3		
Figure 1a	Measles epidemic size distribution for the Canada data (log-log plot)	111
Figure 1b	Measles epidemic size distribution (linear plot) fitted with power-law (solid line) and exponential (dashed line) functions	111
Figure 2a	Rubella epidemic size distribution for Canadian data (log-log plot)	112
Figure 2b	Rubella epidemic size distribution (linear plot) fitted with inverse power-law (solid line) and exponential (dashed line) functions	112
Figure 3a	Pertussis epidemic size distribution for Canadian data (log-log plot)	113
Figure 3b	Pertussis epidemic size distribution (linear plot) fitted with inverse power-law (solid line) and exponential (dashed line) functions	113

Figure 4a	Mumps epidemic size distribution for Canadian data (log-log plot).....	114
Figure 4b	Epidemic size distribution (linear plot) fitted with inverse power-law (solid line) and exponential (dashed line) functions.....	114
Figure 5	Exponential network vs scale-invariant network .....	117

## LISTE DES SIGLES ET ABRÉVIATIONS

---

Abréviation	Description
AIC	Aikake's information criterion
ARIMA	AutoRegressive, Integrated, Moving Average
ACF	Autocorrelation function
BFS	Bureau fédéral de la statistique
CCE	Comité consultatif de l'épidémiologie
CDC	Centers for Disease Control and Prevention
EXP	Exponential
IPL	Inverse power-law
LCDC	Laboratory Centre for Disease Control
LLCM	Laboratoire du lutte contre la maladie
MADO	Maladie à déclaration obligatoire
PACF	Partial autocorrelation function
$R_0$	Taux de reproduction de la maladie
SARIMA	Seasonal AutoRegressive Moving Average
SAR	Seasonal AutoRegressive
SBC	Schwarz's Bayesian Criterion
SEIR model	Susceptibles, Exposed, Infected and Recovered model
SMA	Seasonal Moving Average

## REMERCIEMENTS

---

En premier lieu, je tiens à remercier très sincèrement D<sup>r</sup> Pierre Philippe, directeur de cette thèse, pour son impressionnant savoir-faire scientifique, son précieux talent de pédagogue, sa très grande disponibilité et ses incomparables qualités professionnelles et personnelles. Je vous adresse un merci sincère pour votre rigueur scientifique, votre respect, votre écoute et votre confiance, et pour vos judicieux conseils qui ont grandement enrichi ce projet de recherche de même que mon apprentissage scientifique.

Je remercie aussi le Conseil de la recherche en sciences humaines du Canada (CRSH) pour l'obtention d'une bourse d'études qui m'a permis de me consacrer entièrement à ce projet de recherche sans souci financier.

Je suis également reconnaissante au Laboratoire de lutte contre la maladie (LLCM) de Santé Canada qui m'a aidée à mener à bien ce projet en me fournissant les données nécessaires, plus particulièrement à Madame Carole Scott pour sa gentillesse et sa disponibilité lors de nos nombreux échanges.

Je remercie aussi tous les professeurs que j'ai côtoyés et qui m'ont aidée de près ou de loin dans le cadre de ce projet, et plus particulièrement, D<sup>re</sup> Maria Victoria Zunzunegui, D<sup>re</sup> Nicole Leduc et D<sup>re</sup> Lise Goulet pour leurs conseils et la qualité de leur enseignement.

Je tiens également à dire un merci particulier à mon conjoint, Pierre McNeil. Ses encouragements, sa présence, sa compréhension et son support moral m'ont été bénéfiques tout au cours de ce projet de recherche. Merci pour ta présence et pour être la personne si exceptionnelle que tu es.

## INTRODUCTION

---

« Les grands écarts se produisent avec une probabilité très petite. Mais si l'on est patient, on finira par les observer; ce sont eux qui font souvent l'histoire et qui sont associés à des changements qualitatifs... Ils sont au centre de l'évolution temporelle des systèmes complexes. »

Dacunha-Castelle, 1996, p.54

On est loin d'avoir éliminé les vieilles pathologies infectieuses comme la tuberculose, les maladies sexuellement transmises (MST) ou les infections communes de l'enfance, qui frappent toujours dans les pays en développement. Par exemple, en 2001, la rougeole a provoqué environ quarante millions de cas et décimé près d'un million de personnes dans le monde (WHO, 2002). Dans les pays industrialisés, malgré la transition épidémiologique et le recul de certaines maladies infectieuses, le sida, tout comme les différentes MST, l'influenza, les infections aux streptocoques ou encore la méningococcémie, rappellent brutalement que le monde occidental n'est pas à l'abri des maladies contagieuses ou de nouveaux fléaux infectieux (Garrett, 1994). En effet, malgré la diminution marquée de la mortalité associée aux maladies infectieuses au XX<sup>e</sup> siècle, il n'est pas impossible de voir apparaître des foyers infectieux alarmants de maladies traditionnelles comme la tuberculose, ou rares comme le choléra, ou encore de nouvelles maladies jusqu'à récemment inconnues comme le sida ou le syndrome respiratoire aigu sévère (SRAS). De plus, et en dépit des efforts d'élimination mis en œuvre depuis plusieurs années au Canada, tout comme dans bon nombre d'autres pays,

les maladies infectieuses communes que l'on peut prévenir par vaccination, telles la coqueluche, les oreillons, la rougeole et la rubéole, continuent de frapper. Certaines, comme la coqueluche connaissent, depuis quelques années, une remontée et d'autres, comme les oreillons et la rubéole, font l'objet d'épidémies sporadiques (Santé Canada, 1997a).

Les facteurs contextuels responsables de l'émergence et de la réémergence des maladies infectieuses sont nombreux (Morse, 1995; Feldmann et al., 2002; Lashley, 2003). Il y a d'abord les facteurs sociodémographiques. La population mondiale continue d'augmenter avec des proportions importantes de jeunes enfants dans les pays en développement et de personnes âgées dans les pays développés, groupes particulièrement vulnérables aux maladies infectieuses. Les patterns socioculturels et comportementaux sont aussi en cause. Par exemple, la concentration des enfants en bas âge en garderie augmente les taux de contacts propices à la transmission des maladies infectieuses. Il y a également une demande croissante dans les pays occidentaux pour la nourriture biologique et exotique. L'importation de produits agricoles peut compromettre la sécurité. Par exemple, à la suite d'importation de lait de coco congelé, on a observé une épidémie de choléra dans le Maryland (Centers for Disease Control and Prevention, 1991). Par ailleurs, la liberté sexuelle a accéléré la propagation des MST et du VIH, et les drogues injectables ont augmenté le risque de contracter les maladies transmissibles par le sang comme l'hépatite B. L'avancement des techniques médicales a aussi eu un impact sur l'occurrence des maladies infectieuses par l'entremise du développement de thérapies immunosuppressives et de techniques sophistiquées comme les transplantations d'organes ou encore de l'utilisation répandue des antibiotiques (Lashley, 2003). De plus, l'accélération accrue des mouvements de populations à

l'échelle internationale a augmenté les risques d'introduction de nouvelles pathologies infectieuses dans des pays qui en étaient autrefois exempts (Morse, 1995).

L'importance de la morbidité et de la mortalité associée aux maladies infectieuses nécessite de comprendre leur modalité d'existence. Les modèles descriptifs/phénoménologiques et explicatifs constituent un instrument privilégié. En effet, la détermination objective des modalités d'existence d'une infection dans une population peut générer des hypothèses quant au mode de survivance sous forme épidémique ou autrement (endémie, croissance, décroissance, dérive, cycle, etc.).

Les maladies infectieuses sont souvent étudiées en fonction de variables substitués (proxy), comme le temps et/ou l'espace. Pour les séries chronologiques, les événements (diagnostics) sont accumulés au cours du temps, dans le cadre, souvent, d'une surveillance épidémiologique. Il s'agit alors du cumul du nombre de cas incidents sur une période de temps, à intervalle régulier (par exemple, par mois). Comme les facteurs de risque dûment identifiés sont généralement absents dans ce genre d'entreprise, le chercheur qui utilise des séries chronologiques est le plus souvent à la recherche de patterns temporels dans ces séries. Les patterns détectés représentent les modalités d'organisation des systèmes infectieux dans le temps, d'une part, et en relation avec d'autres systèmes infectieux, d'autre part. Ces patterns révèlent donc l'interaction des acteurs des systèmes : interaction de la maladie avec elle-même dans le temps et interaction avec d'autres systèmes infectieux sur le plan écologique et spatial. L'étude de ces interactions a été nommée pathocénose par Grmek (1990) ou encore l'étude « syndémique » des épidémies par le *Centers for Disease Control and Prevention* (2002). Chaque série chronologique pourrait donc être considérée comme un système complexe dominé par des interactions et dont les tenants ne sont pas immédiatement

observables sous forme de facteurs de risque conventionnels. L'approche la plus appropriée à une telle conceptualisation serait donc dynamique et systémique.

En épidémiologie, on cherche à comprendre la dynamique de la transmission des maladies infectieuses et à en prévoir l'importance quantitative. L'étude des systèmes infectieux dans le temps et dans l'espace pose ainsi des problèmes hors de l'ordinaire. L'absence de facteurs de risque autres que le temps complique l'approche qui porte sur les seules interactions. Il ne s'agit plus de penser en fonction du paradigme de la « boîte noire » où l'accent est mis sur la recherche de facteurs de risque à l'échelle individuelle. Les systèmes infectieux évoluent, changent et s'adaptent à chaque instant; rien n'est statique et toute réponse du système est historique et interdépendante. L'évolution d'une série chronologique d'infections dans le temps reflète la norme de réaction du système, telle qu'elle est déterminée par le passé du système et l'impact du milieu environnant actuel (Philippe, 1999a). Chaque pic, chaque creux d'une série correspond à des ajustements dynamiques et interactifs qui relèvent des événements passés et présents.

En général, les systèmes dynamiques comme les infections dans la population comportent une structure hiérarchisée qui va du niveau microscopique des gènes (des souches virales) au niveau macroscopique des institutions sociales en passant par celui des individus. La dynamique collective qui en résulte est le reflet des interactions entre ces différentes échelles d'observation. L'analyse multi-échelles nécessite la prise en compte des relations non linéaires des variables entre elles. En somme, l'analyse des séries chronologiques d'infections semble donc orienter l'investigateur, quasi naturellement, vers une approche systémique et dynamique, mettant de côté, par là, les approches plus traditionnelles, tout en favorisant d'autres, plus novatrices.

L'objectif de la présente recherche consiste à analyser les séries chronologiques de maladies infectieuses communes (rougeole, oreillons, rubéole et coqueluche) au Canada dans le but précis de caractériser leur dynamique à l'échelle populationnelle, et d'examiner les tenants de leur « comportement » en termes de loi d'organisation (nous reviendrons sur ces objectifs de manière détaillée plus loin).

Cette thèse par articles comprend quatre chapitres. Le premier chapitre montre comment le sujet s'inscrit dans le domaine de la recherche en dressant l'état des connaissances, et traite des objectifs de la thèse. Le deuxième chapitre aborde les données et la méthodologie utilisées. Le troisième chapitre, rédigé sous forme d'articles, résume les aspects importants de la qualité des données des maladies analysées et présente les principaux résultats de nos analyses. Le dernier chapitre fait état d'une discussion et d'une conclusion générale de nos travaux.

# CHAPITRE 1

---

## ÉTAT DES CONNAISSANCES

Ce chapitre constitue un résumé des divers aspects fondamentaux de la dynamique des maladies infectieuses communes de l'enfance. Il dresse l'état des connaissances et traite des objectifs de la thèse.

### 1.1 Généralités

Les maladies infectieuses, particulièrement les maladies transmissibles communes de l'enfance comme la rougeole, ont intéressé les mathématiciens et les épidémiologistes, dès le début du XX<sup>e</sup> siècle, non seulement en raison de la morbidité et de la mortalité qu'elles entraînent, mais également en raison de la remarquable périodicité de leur incidence. Ces travaux ont donné lieu à l'élaboration de modèles théoriques qui ont grandement influencé l'épidémiologie moderne comme le principe d'« action de masse » (*mass action*) (Hamer, 1906; Ross, 1908; Soper, 1929), le théorème de Kermack et McKendrick (1927), le seuil critique de population au-delà duquel la transmission peut persister (Bartlett, 1957; 1960), ou encore le seuil critique de vaccination requis pour stopper la transmission au sein de la population (Dietz, 1976). Ces travaux ont rapidement mené à une littérature grandissante concernant l'épidémiologie mathématique. Les aspects probabilistes du processus de transmission ont été considérés comme un important déterminant de l'étendue et de la propagation des maladies

infectieuses, et ils ont mené à l'élaboration de modèles stochastiques (Bailey, 1975). L'élaboration de ces concepts a également conduit au développement de modèles déterministes. Ces deux types de modèles sont décrits dans Bailey (1975). Anderson et al. (1991) fournissent également une référence complète sur la dynamique des maladies infectieuses et sur les modèles associés.

## **1.2 Les patterns d'émergence des maladies infectieuses communes de l'enfance**

Dans les grandes populations, l'incidence des maladies infectieuses communes de l'enfance témoigne d'une remarquable périodicité (Anderson et al., 1991). Les maladies infectieuses peuvent être caractérisées, d'une part, par un cycle saisonnier (cycle à l'intérieur d'une même année) et, d'autre part, par des cycles épidémiques (oscillations) pouvant revenir environ tous les deux à dix ans (Gottman, 1981; Shaffer et al., 1990; Anderson et al., 1991; Gomes et al., 1999). Par exemple, avant l'introduction de la vaccination de masse dans les grands centres urbains, la rougeole montre un pattern d'émergence régulier caractérisé par des cycles bisannuels et des cycles saisonniers. L'analyse de ces patterns a permis de développer des modèles théoriques concernant la dynamique des maladies infectieuses et de mieux comprendre les facteurs à l'origine de cette dynamique.

## **1.3 Développement de modèles théoriques relatifs aux patterns épidémiques**

Les travaux précurseurs de Hamer (1906), Ross (1908) et Soper (1929) permettent de représenter mathématiquement, de façon très simple, la propagation des maladies

infectieuses. Il s'agit du principe d'action de masse qui détermine que la propagation des infections se fait en fonction d'interactions homogènes (taux de contacts aléatoires) entre les susceptibles et les infectieux dans la population. Cette notion est devenue un important postulat de l'épidémiologie des maladies infectieuses.

La continuité de ces travaux permet à Kermack et al. (1927) d'introduire le *threshold theorem* selon lequel l'introduction d'infectieux au sein d'une communauté de susceptibles ne donnera pas lieu à un phénomène épidémique à moins que le nombre de susceptibles se situe au-delà d'une valeur critique. La notion du nombre reproductif de base ( $R_0$ ) de la maladie est cruciale. Cette notion définit le nombre moyen de cas secondaires d'infections générés par un cas nouvellement infectieux dans une population totalement susceptible. La valeur de  $R_0$  est déterminée par les paramètres biologiques de la maladie (temps de latence et période de contagiosité) et par les paramètres populationnels qui déterminent la transmission (probabilité de transmission et densité des susceptibles) (Anderson, 1994). Plusieurs auteurs se sont intéressés à la mesure de  $R_0$  pour différentes périodes dans les pays industrialisés (Anderson et al., 1991). Le tableau 1.1 ventile ce paramètre avant l'introduction de la vaccination de masse dans les pays industrialisés en regard de certaines maladies communes, ainsi que les périodes de latence et de contagiosité qui leur sont associées.

**Tableau 1.1**  
Nombre reproductif de base et périodes de latence et de contagiosité  
de certaines maladies infectieuses communes

Paramètres	rougeole	coqueluche	oreillons	rubéole
Nombre reproductif de base ( $R_0$ )	11-18	10-18	7-14	6-7
Période de latence	6-9 jours	21-23 jours	12-18 jours	7-14 jours
Période de contagiosité	6-7 jours	7-10 jours	4-8 jours	11-12 jours

Source : Anderson et al. (1991), p.31, 70.

## **1.4 Facteurs à l'origine des patterns des maladies infectieuses**

### 1.4.1 Cycles, nombre reproductif de base et variation saisonnière des taux de contacts

Les cycles des maladies transmissibles sont générés par la diminution du nombre de susceptibles en raison de l'infection et par l'apport de susceptibles (natalité) introduits dans la population. Ainsi, les cycles sont généralement plus prononcés pour des infections qui ont un nombre reproductif de base élevé et de courtes périodes de latence et de contagiosité comme la rougeole (Anderson et al., 1991). Ces cycles peuvent se perpétuer par une variation saisonnière des taux de contacts introduits par la rentrée des classes (Anderson, 1994). La variation saisonnière des taux de contacts crée les cycles saisonniers de l'émergence des maladies infectieuses et constitue également un facteur déterminant de l'émergence des oscillations (London et al., 1973; Anderson et al., 1991; Earn et al., 2000).

### 1.4.2 Natalité, immigration et vaccination de masse

Les périodes interépidémiques peuvent être également influencées par les tendances de la natalité. Une baisse de natalité, par exemple, viendra diminuer le nombre de susceptibles, se répercutant ainsi sur la chaîne de transmission qui sera plus longue (Earn et al., 2000). Évidemment, la vaccination aura un impact considérable dans la mesure où elle vient également diminuer le nombre de susceptibles. Plusieurs études montrent que la vaccination de masse augmente l'âge moyen à l'infection, augmente la durée des périodes interépidémiques (crée des oscillations plus espacées) et peut également changer l'ampleur des cycles saisonniers (Anderson et al., 1984; Anderson et al., 1991). L'immigration pourra également avoir une influence en augmentant le

nombre de susceptibles et en introduisant des individus infectés dans la population (Rangel et al., 1999).

#### 1.4.3 Facteurs climatiques et sociaux et structure par âge

Différents autres éléments comme les facteurs climatiques (London et al., 1973; Shaffer et al., 1990; Gomes et al., 1999) ou la structure par âge de la population (Schenzle, 1984; Anderson et al., 1991) pourraient être importants dans l'explication de l'émergence de maladies infectieuses. De plus, les cycles sont bien souvent dépendants de facteurs locaux comme le mouvement des infectés au sein de la population, et ils peuvent varier dans chaque population (Bartlett, 1960).

#### 1.4.4 Facteurs contextuels et aléatoires

Dans les petites communautés, des facteurs aléatoires ou contingents influencent de façon plus considérable la dynamique des maladies infectieuses. Pour mieux comprendre ces différences, certains auteurs ont analysé la dynamique des infections en mettant l'accent sur les intervalles interépidémiques (Shaffer et al., 1990). Cette technique a permis de voir que la moyenne des intervalles entre chaque épidémie augmente avec la diminution du nombre d'habitants d'une population. En effet, dans les petites communautés, les infectés disparaissent (*fade-out*) jusqu'à ce qu'un infectieux s'introduise à nouveau dans la population, élevant ainsi l'intervalle moyen entre chaque épidémie ainsi que l'âge moyen des infectés. Dans les petites populations, les cycles épidémiques sont moins régulés, car ils sont sous l'emprise plus grande des facteurs contextuels et aléatoires.

### **1.5 Type de dynamique**

Pour la rougeole, Bartlett (1957; 1960) estime que le seuil critique de population au-delà duquel la chaîne de transmission peut persister est de 250 000-300 000 personnes. La dynamique des populations sous ce seuil critique est caractérisée par des extinctions fréquentes de la maladie et par des épidémies irrégulières. Selon Rhodes et al. (1998), il existe trois types de dynamique. Les épidémies de type I surviennent dans les grandes populations qui évoluent au-delà du seuil critique. L'infection est alors endémique et caractérisée par une émergence périodique sans extinction. L'émergence des maladies infectieuses montre alors des patterns oscillatoires plus réguliers. Les épidémies de type II surviennent dans les populations proches du seuil critique. Les épidémies sont fréquentes, mais font occasionnellement état d'une extinction complète. Les épidémies de type III surviennent dans les petites populations nettement en-dessous du seuil critique; elles sont irrégulières et leur nature imprévisible les rend difficiles à analyser par les modèles usuels.

### **1.6 Méthodes**

Différentes méthodes ont été proposées pour l'analyse de la dynamique des maladies infectieuses (Anderson et al., 1991; Koopman, 2004). Il y a les méthodes déterministes et les méthodes stochastiques. Ces dernières sont basées soit sur des simulations, soit sur des méthodes d'analyse de données empiriques (séries réelles). Parmi les modèles déterministes de simulation, le modèle SEIR est le mieux connu. L'effet des différents paramètres (période de latence, contagiosité, etc.) à l'origine des patterns épidémiques a souvent été établi à l'aide du modèle SEIR. Ce modèle ne permet de modéliser que

qualitativement la dynamique, c'est-à-dire le « profil » de la dynamique sans espérer y voir une correspondance exacte avec la réalité. L'examen de la figure 1.1 (voir plus loin dans ce chapitre) montre néanmoins une ressemblance étonnante avec la réalité. De surcroît, il ne permet pas de modéliser adéquatement la dynamique des maladies infectieuses au sein des petites populations. Nous y reviendrons plus loin.

Les modèles de simulations stochastiques ou de Monte-Carlo permettent de simuler la dynamique des maladies infectieuses en tenant compte des aspects probabilistes. Ces modèles sont souvent utilisés pour modéliser la dynamique dans les petites populations qui est plus dépendante des facteurs aléatoires. Enfin, il y a les modèles stochastiques d'analyse de données empiriques comme le modèle de Box-Jenkins qui permet de modéliser la structure d'une série chronologique réelle en cherchant les paramètres qui pourront reproduire le plus fidèlement possible les variations de la série. Il y a également le modèle de puissance inverse de Rhodes et al. (1996a; 1996b; 1997; 1998) qui permet, malgré sa base probabiliste, d'aller un peu loin dans l'explication des patterns. Cette méthode est stochastique, mais l'ajustement du modèle permet de reconnaître l'existence d'une structure déterministe sous-jacente à la dynamique de la transmission dans la population.

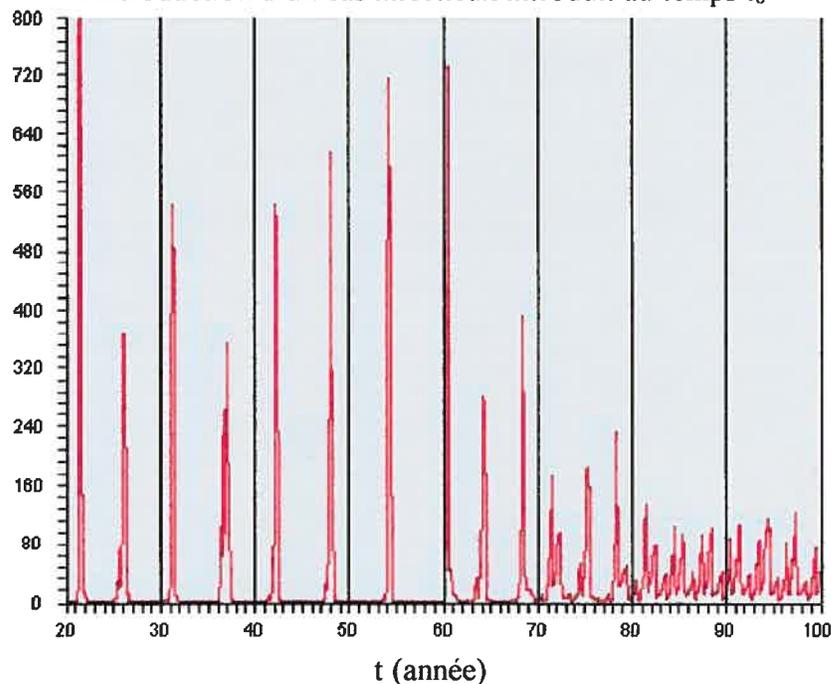
### 1.6.1 Simulations déterministes non linéaires

Les équations du modèle SEIR sont non linéaires. Ces équations sont capables de reproduire les séries d'infection en fonction de différents paramètres de la maladie (nombre reproductif de base ( $R_0$ )), temps de latence, période de contagiosité) (Anderson et al., 1991; Koopman, 2004). Les simulations déterministes fournissent des explications de la dynamique des maladies en dépit du fait qu'elles ne sont pas élaborées à partir de

données empiriques. À partir d'équations différentielles simples mais couplées, on est en mesure de reproduire des séries chronologiques d'infection très semblables à celles que l'on voit dans la réalité. Par exemple, la figure 1.1 montre le résultat d'une simulation à l'aide du modèle SEIR. Il s'agit du nombre de nouveaux cas de rougeole dans une population d'un million de personnes non vaccinées suite à l'introduction d'un cas infectieux au temps  $t_0$ .

**Figure 1.1.**

Modèle SEIR. Nombre de nouveaux cas de rougeole à travers le temps pour une population non vaccinée d'un million de personnes à la suite de l'introduction d'un cas infectieux introduit au temps  $t_0^*$



\* $R_0=15$ , période de latence= 8 jours, période de contagiosité=7, taux de natalité et de mortalité = 0,0000351 pers/jours, incluant une variation saisonnière dans les taux de contacts.

Source : Trottier et al. (2003).

Les cycles sont générés par la diminution du nombre de susceptibles en raison de l'infection et par l'apport de susceptibles (natalité) introduits dans la population. Les cycles sont généralement plus prononcés pour des infections qui ont de courtes périodes

de latence et de contagiosité et qui induisent une immunité à long terme comme la rougeole. Les cycles peuvent se perpétuer par la variation saisonnière des taux de contacts introduits par la rentrée des classes (Anderson, 1994). Le modèle SEIR qui incorpore une variation saisonnière dans les taux de contacts génère un riche éventail de dynamiques incluant une dynamique chaotique (Anderson, 1994).

Par contre, ce type d'approche n'est pas exempt de difficultés. Un des problèmes importants lors de la modélisation des séries chronologiques est la présence de « stochasticité démographique » et de perturbations externes (Shaffer et al., 1990). Ces dernières peuvent entretenir les oscillations et empêcher l'atteinte d'un équilibre. En effet, la dynamique de la rougeole et de toute maladie infectieuse commune est influencée par les grandes tendances de la natalité et autres phénomènes liés aux tendances démographiques séculaires (Earn et al., 2000; Grenfell, 2000). Ces phénomènes introduisent du bruit dans l'analyse des séries, ce qui rend difficile l'investigation de la dynamique. La diminution du nombre de susceptibles induits par la vaccination amène également la présence de bruit qui complique l'interprétation des dynamiques (Grenfell, 2000). Le bruit est d'autant plus important que la population est de petite taille. En somme, malgré une meilleure compréhension de la dynamique avec ces modèles, des anomalies persistent lors d'exercices de simulation. À titre d'exemple, un nombre important d'auteurs comme Aron (1990), Bolker (1993), Bolker et al. (1993) ont modélisé les effets de la variation saisonnière sur l'incidence de la rougeole à l'aide du modèle SEIR en intégrant la complexité due à la structure par âge. En dépit de leur capacité à produire des cycles oscillatoires réguliers et similaires aux séries chronologiques de la rougeole, les creux épidémiques sur la série sont souvent irréalistes et s'éloignent de la réalité observée. Selon Grenfell (2000), ce fait laisse place à

l'exploration et à l'approfondissement de la dynamique des maladies infectieuses à l'aide de données empiriques.

### 1.6.2 Simulations stochastiques et modélisation empirique

Il existe également des modèles stochastiques d'épidémie, c'est-à-dire des modèles qui tiennent compte des aspects probabilistes du processus de transmission. Il existe plusieurs modèles de simulation stochastique qui permettent de simuler les cycles épidémiques (Bailey, 1975). Ces modèles sont surtout utiles pour modéliser la dynamique des petites populations.

Il existe également des méthodes d'analyse stochastique des données empiriques comme la méthode de Box-Jenkins (connue aussi sous le nom de modèle ARIMA). Cette approche, encore peu utilisée dans l'analyse de la dynamique des maladies infectieuses, permet de caractériser la dynamique de base et de prévoir les futurs niveaux d'infections par le biais d'un ajustement utilitaire à partir de séries réelles. Il s'agit en effet d'une méthode essentiellement prédictive. La méthode peut également être utilisée pour analyser la relation entre l'émergence de deux phénomènes comme l'incidence de la maladie et la présence d'un facteur climatique (*cross-correlations*). La corrélation entre deux séries stationnaires permet de voir l'influence d'une série sur l'autre, tel l'effet de la température sur l'incidence des maladies cardiaques (Bowie et al., 1981). Choi et al. (1981a; 1981b) ont été les premiers auteurs à s'intéresser à la méthode dans le cadre des maladies infectieuses; leur but étant de mettre au point des modèles de prédiction de la mortalité relative à la pneumonie et à l'influenza. Allard (1998) utilise cette approche pour développer une méthode de détection et de projection à court terme

des épidémies, dans le but d'aider les départements de santé publique à réagir adéquatement en cas de recrudescence des maladies infectieuses.

Helfenstein (1986) utilise la méthode de Box-Jenkins pour identifier la dynamique de base de la varicelle et des oreillons. Pour ces deux maladies, la fonction d'auto-corrélation (ACF) des séries différenciées saisonnièrement indique un processus autorégressif (AR) et un processus de moyenne mobile (MA) saisonnier, ce qui signifie qu'au-delà de la composante saisonnière, le nombre de nouveaux cas qui apparaît chaque mois est déterminé, d'une part, par le nombre de cas présents au cours du mois précédent et, d'autre part, par des chocs aléatoires. En dépit du fait que ces deux maladies sont biologiquement différentes, elles montrent la même structure temporelle statistique. De plus, une analyse des séries chronologiques en fonction de différentes strates de temps (1928-1938, 1939-1949, 1950-1960) a montré le même type de modèle sans changement notable des estimations des paramètres. L'auteur conclut en mentionnant qu'il serait possible de définir la dynamique des maladies infectieuses non plus seulement en fonction de leur propriété biologique et épidémiologique, mais également en fonction de leur structure stochastique. Quoique la méthode connaisse des limites importantes - elle ne permet pas d'expliquer la dynamique; elle sert plutôt à la caractériser et à la prévoir à partir des données - elle pourrait être intéressante pour classer la structure stochastique des maladies infectieuses avant et après des programmes de vaccination de masse, par exemple.

### 1.6.3 Modèles stochastiques à implication déterministe

Plusieurs études de la rougeole, en particulier, ont suggéré que les maladies transmissibles peuvent être caractérisées par une dynamique complexe (Olsen et al.,

1990; Shaffer et al., 1990; Turchin, 1990; Grenfell et al., 1993; 1995; Bolker et al., 1993; Anderson, 1994; Engbert et al., 1994; Philippe et al., 1998; Earn et al., 2000). Au cours des dernières années, plusieurs discussions ont été soulevées concernant l'émergence possible d'une loi de puissance inverse au sein de ces systèmes. Déjà en 1961, Taylor décrivait la relation entre la variance temporelle/spatiale et l'abondance moyenne des populations en fonction de cette loi. Plusieurs phénomènes sociaux ont par la suite été décrits à l'aide de cette loi, comme la distribution des richesses dans la population (*Pareto's law*), la fréquence relative de l'utilisation des mots dans le langage (*Zipf's law*) ou encore la distribution de la taille des villes (*Auerbach's law*). Dans l'étude des tremblements de terre, les géophysiciens ont mis à jour ce type de loi pour en interpréter la dynamique (Bak et al., 1993). Depuis, on a tenté d'expliquer les patterns d'émergence de phénomènes biologiques en fonction de la loi stochastique de puissance inverse et ce, particulièrement en ce qui a trait à la nature dynamique du phénomène étudié, comme par exemple l'agrégation spatiale de la leucémie dans la ville de San Francisco (Philippe, 1999b), la relation entre la moyenne et la variance des cas de rougeole déclarés dans 366 communautés en Angleterre (Keeling et al., 1999), le réseau internet (Barabasi et al., 1999), ou encore le temps passé sur les listes d'attente à l'hôpital (Smethurst et al., 2001) ou à l'urgence (Philippe et al., 2004).

Rhodes et al. (1996a; 1996b; 1997; 1998) ont analysé la distribution de la taille et de la durée des épidémies de rougeole, d'oreillons et de coqueluche au sein des petites populations à l'aide de la loi stochastique de puissance inverse pour mettre en évidence des patterns de régularité. Selon eux, les changements temporels de l'incidence ont suscité de nombreuses discussions concernant les grandes populations, mais pour les petites populations, l'extinction temporaire des infections (*fade-out*), due à une influence

beaucoup plus considérable des facteurs aléatoires ou contingents, semble rendre les analyses traditionnelles (de type SEIR) inadéquates. Rhodes et al. (1996a; 1996b; 1997; 1998) montrent qu'une loi fondamentale gouverne les épidémies dans les petites populations, alors qu'on a longtemps pensé que leur dynamique, d'apparence stochastique, ne suivait pas de pattern régulier. Leurs analyses montrent, au contraire, que la loi de puissance inverse ajuste beaucoup mieux la distribution de la taille et de la durée des épidémies qu'une loi de distribution exponentielle. Le modèle permet d'estimer et de prédire les épidémies de taille et de durée spécifiques qui surviennent au cours d'un intervalle de temps donné. Le graphique log-log du nombre d'épidémies en fonction de leur taille fait apparaître une relation linéaire simple et permet de calculer la valeur de l'exposant critique (pente de la droite estimée). Il ressort de ces analyses que la loi est capable de prédire toute la série des épidémies, des plus petites aux plus grandes. Selon eux, cette loi est utile tout autant dans les grandes populations hautement vaccinées que dans les petites populations. En dépit du fait que le modèle est linéaire (après linéarisation) et stochastique, il permet d'interpréter les événements en fonction d'une loi déterministe d'organisation.

Incidentement, cette loi est typique des réseaux invariants d'échelle et fait référence aux notions de nœuds et de supernœuds des systèmes (réseaux) (Barabasi et al., 1999; 2003). Dans les réseaux invariants d'échelle quelques supernœuds sont reliés à presque tous les autres et déterminent les interactions. Ces réseaux n'ont pas d'échelle caractéristique. Plusieurs systèmes sociaux sont invariants d'échelle comme, par exemple, le réseau des relations sexuelles; la plupart des individus n'ont que quelques partenaires au cours de leur vie, alors que quelques-uns en ont des centaines (Liljeros et al., 2001; Schneeberger et al., 2004). Il en est de même des liens Internet reliés à Yahoo

qui dominent les échanges (Barabasi et al., 1999). En revanche, au sein des réseaux aléatoires, caractérisés par une loi exponentielle, le nombre d'interactions a plutôt la forme d'une cloche. Ces réseaux ont une échelle caractéristique. Cela signifie que la majorité des nœuds possède un nombre de liens proches d'une valeur moyenne : il n'y a quasiment pas de nœuds ayant peu de liens ou énormément de liens (Barabasi et al., 2003).

Selon Koopman et al. (1994; 1996; 1999, 2004), la présence d'une dynamique non linéaire à l'origine d'un phénomène épidémiologique peut rendre inadéquates les méthodes d'analyse traditionnelles, qui sont statiques, fixes et linéaires. Selon ces auteurs, le paradigme de la « boîte noire » (dans lequel on peut inclure la plupart des modèles d'ajustement stochastiques de régression) devrait être dépassé. Les bases théoriques des systèmes épidémiologiques devraient non seulement tenir compte des facteurs de risque à l'échelle individuelle, mais également de l'interaction entre les sujets, comme à l'intérieur d'un réseau. La notion de réseau est importante pour l'épidémiologie. En effet, la non-linéarité, associée à l'hétérogénéité des réseaux de contacts interindividuels, peut réduire le taux d'extinction des infections et être une cause de persistance de l'infection au sein de la population et ce, même en présence d'une couverture vaccinale de masse suffisante (Allen et al., 1993; Grenfell et al., 1993; Bolker et al., 1996; Barabasi et al., 2003). Contrairement aux réseaux aléatoires, les réseaux invariants d'échelle (réseaux caractérisés par une hétérogénéité des interactions telle que définie plus haut) résistent à la mise hors circuit de plusieurs nœuds, alors qu'une attaque ciblée de quelques supernœuds peut menacer l'ensemble du réseau (Barabasi et al., 2003). Une meilleure compréhension de cette dynamique pourrait aider à l'élimination des maladies infectieuses comme la rougeole en concentrant les efforts

sur l'organisation interindividuelle de la maladie. L'étude de la dynamique a donc une importance majeure pour les politiques de santé publique et plus précisément pour les programmes de vaccination (Association of Schools of Public Health, 1999). De plus, la fragmentation des séries en fonction des périodes de vaccination pourrait conduire à des pistes intéressantes concernant la dynamique non linéaire des maladies infectieuses (Grenfell, 2000).

### **1.7 Objectif de la thèse**

L'objectif général de cette thèse est d'analyser la dynamique de la rougeole, de la coqueluche, des oreillons et de la rubéole au Canada en tenant compte des périodes de prévacination et de postvacination. Ces maladies infectieuses transmissibles de l'enfance ont été sélectionnées en raison de la disponibilité des données et de leur intérêt pour la santé publique.

Les objectifs spécifiques de la recherche sont :

- 1) d'examiner la qualité des données du système de surveillance des maladies à déclaration obligatoire (MADO) analysées dans cette thèse (rougeole, coqueluche, oreillons, rubéole);
- 2) de caractériser (et classifier) la structure stochastique des séries chronologiques de maladies communes de l'enfance (rougeole, coqueluche, oreillons, rubéole), avant et après la vaccination de masse au Canada;
- 3) de vérifier l'hypothèse d'un ajustement convenable de la distribution des tailles d'épidémies en fonction de la loi de puissance inverse dans une population largement vaccinée.

La présente thèse est exploratoire. Nous sommes donc à la recherche de patterns et nous comptons sur la variable « temps » pour agir comme « proxy » des variables latentes d'intérêt. Les variables d'intérêt ne sont donc pas données, mais inférées à partir des patterns de la dynamique et des modalités d'organisation des réseaux de contacts. Le recours aux séries chronologiques canadiennes de la rougeole, de la coqueluche, des oreillons et de la rubéole permettra de mieux comprendre leur structure et leur dynamique.

### **1.8 Organisation de la thèse par articles**

Le chapitre 1 a passé en revue la littérature sur la dynamique des maladies infectieuses communes de l'enfance et a présenté les objectifs de la thèse.

Le chapitre 2, qui suit, discute les données et les méthodes utilisées dans la thèse.

Le chapitre 3 inclut les trois articles à soumettre.

Article #1. Le recours aux données d'incidence des maladies à déclaration obligatoire (MADO) amène un questionnement important sur la qualité de ces dernières. Ainsi, une analyse approfondie des biais pouvant affecter la qualité des MADO a été entreprise avec une attention particulière aux données de la rougeole, de la coqueluche, des oreillons et de la rubéole. Puisqu'il existe très peu de littérature sur le sujet, ce travail a fait l'objet de la rédaction d'un article.

Article #2. Un deuxième article a été rédigé qui couvre le deuxième objectif de la thèse, soit de caractériser, et de classifier si possible, la structure stochastique de la rougeole, de la coqueluche, des oreillons et de la rubéole, avant et après la période de vaccination de masse au Canada.

Article #3. On trouvera enfin un dernier article qui couvre le troisième objectif de la thèse, soit d'étudier les propriétés d'échelle des maladies infectieuses étudiées, avant et après la vaccination de masse au Canada.

Le chapitre 4 clôt la thèse avec une discussion et une conclusion générale.

## CHAPITRE 2

---

### DONNÉES ET MÉTHODES

Ce chapitre discute des données épidémiologiques auxquelles nous avons recours pour entreprendre les analyses (séries chronologiques de la rougeole, de la coqueluche, des oreillons et de la rubéole). On trouve également une ébauche des méthodes utilisées qui sont décrites, par ailleurs, en détail dans le corps de chacun des articles (chapitre 3).

#### **2.1 Données**

##### **2.1.1 Provenance des données**

Les données des maladies à déclaration obligatoire (MADO) proviennent du laboratoire de lutte contre la maladie (LLCM). Il s'agit des données canadiennes de surveillance de la rougeole, de la coqueluche, des oreillons et de la rubéole cumulées dans le temps à intervalle régulier. Les données sont disponibles à partir de 1924 jusqu'à aujourd'hui, à l'exception de quelques périodes où la rougeole (1959-1968), la rubéole (1959-1968) et les oreillons (1959-1986) n'étaient pas déclarés. Les séries chronologiques de ces infections sont disponibles sur une base de quatre semaines (treize observations annuelles) avant 1990 (Bureau fédéral de la statistique, 1924-1952; Bureau fédéral de la statistique, 1953-1970; Statistique Canada, 1971-1978; Santé et Bien-Être social Canada, 1975-1991) et sur une base mensuelle (douze observations annuelles) à partir de 1990 (Santé Canada, 1992-2004).

## 2.1.2 Épidémiologie des maladies transmissibles analysées

### 2.1.2.1 Rougeole

Depuis 1924, et jusqu'à ce que l'on introduise le vaccin antirougeoleux en 1963, on dénombrait au Canada des cycles épidémiques importants de la rougeole tous les 2 à 4 ans, allant environ de 13 000 à 83 000 cas déclarés par année (figure 1, annexe I). L'incidence de cette maladie a atteint un point culminant en 1935, avec 83 127 cas déclarés. La rougeole n'était pas une MADO de 1959 à 1968. Ainsi, dans la période qui a suivi immédiatement l'introduction du vaccin, on ne dispose pas des données d'incidence. Cependant, quand elle est redevenue une MADO en 1969, son incidence avait chuté considérablement (Santé Canada, 1997a). De 1969 à 1985, la rougeole a fluctué, environ, entre 1 000 et 25 000 cas par année et de 1986 à 1995, son incidence annuelle a diminué avec des fluctuations allant de 200 à 15 000 cas. En 1993, 203 cas de rougeole ont été signalés au Canada, ce qui représente une incidence presque quinze fois moindre qu'en 1992. Cependant, en 1994, 2,5 fois plus de cas de rougeole ont été déclarés qu'en 1993, et 4,4 fois plus l'ont été en 1995 par rapport à 1994. En 1998, seulement 12 cas de rougeole ont été signalés, il s'agit du plus petit nombre de cas annuels jamais déclarés depuis 1924. L'élimination de la rougeole dans tous les pays du continent américain constitue un sujet actuel de santé publique. Les campagnes massives de vaccination de rattrapage et l'adoption d'un programme d'immunisation systématique comportant deux doses du vaccin partout au Canada, en 1996, ont freiné la circulation du virus de la rougeole. Par contre, en 2000, l'incidence a connu à nouveau une remontée pour atteindre environ 200 cas déclarés.

### 2.1.2.2 Coqueluche

De 1924 jusqu'à ce que l'on introduise le vaccin à germes entiers en 1943, l'incidence de la coqueluche au Canada fluctuait autour de 6 000 à 19 000 cas chaque année en suivant des cycles épidémiques environ tous les 3 à 5 ans (figure 2, annexe I). Elle a atteint son sommet en 1934 avec 19 484 cas déclarés. Le vaccin contre la coqueluche était l'un des premiers vaccins utilisés au Canada. Dès son introduction en 1943, on a observé une nette réduction de l'incidence de la coqueluche et de la mortalité associée à cette dernière (Santé Canada, 1997a). Alors que l'on comptait en moyenne autour de 17 000 cas par an dans la période immédiatement antérieure à l'introduction de la vaccination, le nombre moyen de cas annuels est passé à environ 9 600 de 1945 à 1957, à 5 400 de 1958 à 1966 et à 2 350 de 1967 à 1985. Cependant, depuis quelques années, l'incidence de la coqueluche augmente partout au Canada (Santé Canada, 1993a) et aux États-Unis (Centers for Disease Control and Prevention, 1993; Farizo et al., 1992). Son incidence, d'après les cas déclarés au Canada en 1994 et 1995 (10 151 et 9 818 cas, respectivement), est la plus élevée au Canada depuis 1955. Un des aspects troublants de la tendance actuelle est la persistance de l'augmentation d'une année à l'autre depuis 1990. Il semble que l'incidence de la coqueluche soit passée brusquement de faible endémie à un état épidémique ou de très forte endémie. Aucune autre maladie à déclaration obligatoire n'a suivi une évolution analogue à celle de la coqueluche au cours des dernières années. Selon Santé Canada (1993a), la recrudescence de la coqueluche n'est vraisemblablement pas attribuable à des changements importants dans les méthodes de surveillance.

### 2.1.2.3 Oreillons

Avant l'usage généralisé du vaccin contre les oreillons en 1969, entre 10 000 et 50 000 cas d'oreillons étaient déclarés chaque année au Canada, suivant ainsi une recrudescence tous les 4 à 8 ans (figure 3, annexe I). Les oreillons n'étaient pas une maladie à déclaration obligatoire de 1959 à 1986. Ainsi, dans la période qui a suivi immédiatement l'introduction du vaccin, les données ne sont pas disponibles. Quand les oreillons sont redevenus une maladie à déclaration obligatoire en 1986, son incidence avait chuté considérablement et ses cycles épidémiques sont devenus moins perceptibles (Santé Canada, 1997a). Entre 1986 et 1995, la moyenne annuelle s'élevait à 509 cas déclarés. Au cours des dernières années (entre 1996 et 2000), il semble que son incidence ait chuté encore plus considérablement avec environ 200 cas en moyenne à chaque année. En 1996, l'adoption du programme d'immunisation systématique comportant deux doses du vaccin a freiné la circulation du virus. Cependant, certaines épidémies d'oreillons, comme l'écllosion récente de cas à Montréal (octobre 1998 à mars 1999), montrent que la population n'est pas à l'abri d'épidémies sporadiques (Caplan, 1999).

### 2.1.2.4 Rubéole

Avant la période de vaccination massive, le nombre de cas de rubéole déclarés allait environ de 800 à 70 000 cas par année (figure 4, annexe I). Elle se manifestait généralement par des cycles épidémiques irréguliers pouvant revenir tous les 3 à 10 ans. Les principaux sommets ont été observés en 1936, 1941, 1950 et 1956 (Santé Canada, 1997a). La fréquence la plus élevée a été enregistrée en 1936 avec 69 000 cas. Les données pour la rubéole ne sont pas disponibles de 1959 à 1968, période pour laquelle elle n'était pas une maladie à déclaration obligatoire au Canada. La vaccination contre la

rubéole a été instaurée en 1969, et depuis le milieu des années 1970, l'incidence de cette maladie est restée relativement faible. De 1986 à 1995, on a dénombré en moyenne 1 000 cas par année (de 237 à 2 450 cas). En 1997, on a déclaré 4 007 cas, niveau qui n'avait pas été atteint depuis 1983. Au cours des dernières années, plusieurs éclosions ont été signalées dans des collèges et des universités (Santé Canada, 1997a). Le problème le plus grave associé à la rubéole est la survenue de la rubéole congénitale chez le fœtus par suite de l'infection de la mère enceinte. C'est à partir de 1979 que la rubéole congénitale est devenue une MADO. De 1986 à 2000, 38 cas de rubéole congénitale ont été déclarés au Canada. Ils représentent environ 0,3 % du nombre total de cas de rubéole déclarés au cours de la même période (Santé Canada, 1997a; Santé Canada, 1992-2000). La rubéole congénitale n'est pas incluse dans les cas de rubéole étudiés dans cette thèse, mais il est possible que certains cas se retrouvent parmi les cas déclarés avant 1979.

### 2.1.3 Structure des séries chronologiques analysées

Seules les séries chronologiques de 1953-1959, 1970-1976 et 1991-2002 ont été analysées (figure 5, 6, 7 et 8; annexe I). Avant les années 1950, un nombre important de provinces et de territoires n'ayant pas intégré le système de surveillance canadien, ne déclaraient pas les données; aussi, les données antérieures à 1950 ont été exclues. Il faut également mentionner que le choix des séries de quatre semaines (jusqu'en 1990) a été influencé par les procédures de déclaration. Pour ajuster les données dans l'année civile correspondante, lorsqu'il restait au moins sept jours avant la fin d'une année civile, les données étaient enregistrées au fichier central sur la base d'une semaine et non de quatre

semaines. Par exemple, si la période de déclaration sur la base de quatre semaines en fin d'année commençait le 26 novembre et se terminait le 23 décembre (28 jours d'observation), les cas déclarés du 24 au 30 décembre (7 jours d'observation) étaient déclarés sur la base d'une semaine pour inclure ces derniers dans l'année civile correspondante. Cette procédure permettait d'ajuster les séries enregistrées, mais pose néanmoins un problème en termes d'analyse des données. Comme l'analyse des séries chronologiques doit se faire sur des périodes de temps constantes, la sélection des séries a dû se faire en tenant compte de ces ajustements.

Les premières séries analysées couvrent la période de 1953 à 1959. Il s'agit de séries chronologiques de quatre semaines. Cette période correspond à l'ère de prévaccination pour toutes les maladies étudiées, à l'exception de la coqueluche pour laquelle la vaccination a été implantée à partir de 1943. À partir de 1959, la rougeole, les oreillons et la rubéole ont été exclus des MADO, ce qui rend impossible leur analyse à partir de 1959, jusqu'au moment où ces maladies ont été réintroduites dans la liste des MADO. La série de coqueluche a été analysée de 1953 à 1962 ayant toujours été une maladie à déclaration obligatoire (la période couverte se termine à la fin de 1961 car, à ce moment, les données ont été enregistrées au fichier central sur la base d'une semaine et non de quatre semaines afin d'ajuster les données dans l'année civile correspondante).

Une deuxième période, de 1970 à 1976, a été analysée. Il s'agit de séries chronologiques de quatre semaines. L'année 1969 correspond au moment où la rougeole et la rubéole ont été réintégrées au sein des MADO. En 1976, les séries ont dû être sectionnées, les données ayant été enregistrées au fichier central sur la base d'une semaine et non de quatre semaines afin de les ajuster dans l'année civile correspondante. Les oreillons ne sont pas analysés au cours de cette période puisqu'ils ne sont redevenus

une maladie à déclaration obligatoire qu'en 1986. Les séries de 1970-1976 ont été ajustées pour les données déclarées non précisées (données sans date précise d'occurrence de la maladie au cours d'une année donnée). Par exemple, en 1973, trois cas de coqueluche sont survenus sans spécification de la période. Ces cas ont été répartis proportionnellement selon le pourcentage de cas survenus au cours de chaque période. Si 10 % des cas sont survenus dans la première partie de l'année, alors 10 % des données non précisées ont été réparties au sein de cette dernière. Il n'y avait que très peu de données sans date précise d'occurrence (moins de 1% chaque année).

Enfin, la période 1991-2002 a été analysée. L'année 1991 a été une année marquante pour le système de surveillance qui se base désormais sur des définitions spécifiques et uniformes et sur des cas confirmés à l'aide de tests de laboratoire. Les données sont également déclarées sur une base mensuelle et non plus de quatre semaines. Ces séries, comme celles de 1953-1959, n'avaient pas de données non précisées.

#### 2.1.4 Qualité des données

Tel que discuté dans l'article sur la qualité des données (chapitre 3), plusieurs limites des données se doivent d'être considérées. Il ne s'agit pas de biais relatifs aux données canadiennes uniquement, mais à l'ensemble des données de surveillance des pays industrialisés (Giesecke, 1994). La revue de la littérature (article 1, chapitre 3) a permis de faire ressortir des particularités très semblables concernant la qualité des données de surveillance du monde industrialisé que le Canada partage vraisemblablement avec ces pays.

D'abord, les données de surveillance sous-estiment considérablement l'incidence réelle des maladies (voir article 1, chapitre 3). Certains facteurs comme la symptomatologie, l'absence d'un médecin à proximité, le départ à l'extérieur du pays d'une personne infectée ou encore l'appartenance religieuse peuvent affecter la propension des gens à consulter pour une MADO. Par exemple, les consultations sont probablement moins importantes pour des maladies comme la coqueluche dont les symptômes peuvent être plus bénins chez l'adulte et l'adolescent ou encore pour des infections comme la rubéole, maladie pour laquelle 50 % des cas sont estimés asymptomatiques (article 1, chapitre 3). Par ailleurs, la grande faiblesse des systèmes de surveillance se situe souvent au niveau du diagnostic de la maladie. La sensibilité et la spécificité des diagnostics cliniques et/ou de laboratoires demeurent un défi de taille pour ces maladies et ce, surtout dans un contexte de faible prévalence. Finalement, une proportion importante de médecins ne déclare pas les cas aux autorités sanitaires. Ainsi, même si les individus infectés sont connus des services médicaux, bien souvent cette information ne parvient pas au fichier central.

Plusieurs études dans les pays industrialisés ont estimé la proportion des cas déclarés au système de surveillance. Les statistiques qui suivent sont relatives aux pays industrialisés, incluant le Canada. Pour la coqueluche, l'étendue des estimations de la proportion de cas déclarés par rapport au nombre de cas réels varie de 3 à 25 %. Pour la rougeole, l'étendue des estimations varie de 7 à 60 % (article 1, chapitre 3). Par ailleurs, la proportion de cas déclarés semble plus faible en période de postvaccination qu'en période de prévacination. Le tableau 2.1 résume le pourcentage de cas déclarés selon diverses études pour les différentes périodes analysées (voir aussi l'article 1, chapitre 3). Pour la coqueluche, les études montrent qu'après l'ère de vaccination (après 1950), la

proportion des cas déclarés s'étend de 9 à 11 %. Cette proportion semble avoir diminué à 6 % dans les années 1990. La rougeole semble mieux déclarée que la coqueluche, mais les estimations ne sont guère plus encourageantes. Autour des années 1950, on estime que la proportion de cas déclarés de rougeole se situe entre 13 et 63 %. On croit également qu'en période de post-vaccination (après 1970), cette proportion est plus faible, allant de 7 à 40 %. Par contre, au cours des années récentes, on peut présumer que sa déclaration est plus représentative de son incidence; en effet, il semble que la proportion des cas déclarés ait été de 33 à 41 %. Pour la rubéole et les oreillons, cette proportion semble se situer autour de 17 et 14 % respectivement dans les années qui suivent 1950. On peut supposer que ces maladies ont suivi les mêmes tendances que la rougeole et la coqueluche, c'est-à-dire qu'en période de postvaccination (années 1970), ces proportions ont été légèrement plus faibles. Dans les années récentes, on croit que la sous-estimation demeure également très importante pour la rubéole qui peut se présenter plus fréquemment sous forme asymptomatique (article 1, chapitre 3).

**Tableau 2.1**

Proportion de cas déclarés au système de surveillance, Canada et pays industrialisés.

Période	Coqueluche	Rougeole	Rubéole	Oreillons
Années 1950	9-11 %	13-63 %	17 %	14 %
Années 1970	9-10 %	7-40 %		
Années 1990	6 %	33-41 %		

Source : article 1, chapitre 3.

Il faut également ajouter que les données canadiennes sont incomplètes à certains moments puisque certaines provinces n'ont pas déclaré les cas des maladies concernées malgré leur inclusion à la liste des MADO. Ce phénomène peut contribuer à sous-estimer davantage le nombre de cas déclarés. Les données de la rougeole, de la rubéole

et des oreillons pour les séries de 1953-1959 ne sont pas exhaustives (tableau 2.2). La rougeole, les oreillons et la rubéole ont été exclus de la liste des MADO en 1959, mais certaines provinces avaient déjà arrêté le signalement des cas à partir de 1954 (Bureau fédéral de la statistique, 1954-1959). De façon générale, les provinces manquantes représentent au plus 13 % de la population canadienne, à l'exception des données sur les oreillons dont la proportion passe à près de 19 % en 1957 et 1958.

**Tableau 2.2**

Nombre de provinces qui ne déclaraient pas les cas de rougeole, de rubéole et d'oreillons de 1953 à 1959 et pourcentage de la population canadienne que ces provinces représentent.

Année	Rougeole		Rubéole		Oreillons	
	Province	% de la population canadienne		% de la population canadienne		% de la population canadienne
1953	0		0		0	
1954	0		2 (IPE, NB)	4,3	0	
1955	0		3 (IPE, NB, Sas)	10,0	1 (Alb)	6,8
1956	2 (IPE, NB)	4,1	3 (IPE, NB, Sas)	9,6	2 (Alb, NB)	10,4
1957	3 (IPE, NB, Sas)	9,3	3 (IPE, NB, Sas)	9,3	5 (TN, IPE, NB, Alb, Sas)	18,9
1958	3 (IPE, NB, Sas)	9,2	4 (IPE, NB, NE, Sas)	13,4	5 (TN, IPE, NB, Alb, Sas)	18,8

Alb=Alberta, CB=Colombie-Britannique, IPE=Île-du-Prince-Édouard, NB=Nouveau-Brunswick, Sas=Saskatchewan, TN=Terre-neuve.

En ce qui concerne la série de 1970-1976, il manque les données de certaines provinces qui n'ont pas réintroduit la rougeole et la rubéole dans leur liste des MADO et ce, malgré les recommandations de 1969 (tableau 2.3). En 1970 et 1971, l'Île-du-Prince-Édouard, le Nouveau-Brunswick et la Colombie-Britannique ne déclaraient pas la rougeole (13,5 % de la population canadienne); de même, l'Île-du-Prince-Édouard et la Colombie-Britannique ne déclaraient pas la rubéole (10,6 % de la population canadienne), tel que recommandé (Bureau fédéral de la statistique, 1970; 1971; Statistique Canada, 1972). De plus, il a fallu attendre jusqu'en 1988 pour que l'Île-du-Prince-Édouard déclare les cas de rubéole et la rougeole. Il faut donc garder à l'esprit

que certaines provinces ont mis quelques années avant de se conformer à l'ajout de la rougeole et de la rubéole sur leur liste de MADO, ce qui se reflète sur l'exhaustivité des données canadiennes. Toutes les provinces et les territoires déclarent cependant ces maladies pour la période 1991-2002.

**Tableau 2.3**

Nombre de provinces qui ne déclaraient pas les cas de rougeole et de rubéole de 1970 à 1976 et pourcentage de la population canadienne que ces provinces représentent.

Année	Rougeole		Rubéole	
	Province	% de la population canadienne		% de la population canadienne
1970	3 (IPE, NB, CB)	13,4	2 (IPE, CB)	10,5
1971	3 (IPE, NB, CB)	13,5	2 (IPE, CB)	10,6
1972	1 (IPE)	0,5	1 (IPE)	0,5
1973	1 (IPE)	0,5	1 (IPE)	0,5
1974	1 (IPE)	0,5	1 (IPE)	0,5
1975	1 (IPE)	0,5	1 (IPE)	0,5

CB=Colombie-Britannique, IPE=Île-du-Prince-Édouard, NB=Nouveau-Brunswick.

En somme, tous les cas ne sont pas portés à l'attention des médecins, et la déclaration des cas diagnostiqués est incomplète. Cependant, la sous-estimation des données ne causera pas de problème important si nous nous intéressons à la dynamique des maladies sur de courtes périodes. En effet, plusieurs études (article 1, chapitre 3) montrent que les données de surveillance sur de courtes périodes reflètent généralement bien les variations de l'incidence des maladies. Cependant, il ne serait pas adéquat de faire l'hypothèse que les fluctuations de la maladie sont fiables sur les 80 années de surveillance canadienne. Comme la surveillance a connu des changements significatifs au cours du temps (propension des gens à consulter les services médicaux, variation dans la façon de déclarer les cas, variations dans la définition des cas, introduction des tests de laboratoire pour confirmer les cas, etc.), ces facteurs peuvent avoir un effet sur

les données et en affecter l'analyse sur de longues périodes. Pour limiter l'impact des biais potentiels, il convient donc d'analyser les séries chronologiques sur de courtes périodes en considérant les changements au sein du système de surveillance (période de vaccination, modification des définitions des cas, etc.). Les séries chronologiques sélectionnées sont donc courtes et tiennent compte de ces changements. Ainsi, les variations dans un sens ou dans un autre des données au cours du temps seront le reflet d'une variation équivalente véritable dans la population.

Par ailleurs, il faut aussi noter qu'un nombre non négligeable de cas faussement positifs est sans doute aussi présent au sein des données de surveillance. Ce phénomène est d'autant plus important au cours des dernières années que la prévalence des maladies est faible. Cela touche particulièrement les oreillons qui ont souvent été déclarés sur la base d'une définition clinique (sans tests de laboratoire) (article 1, chapitre 3). En effet, même si les définitions de cas élaborés à partir de 1991 incluent les tests de laboratoire comme outil diagnostique, un nombre important de cas d'oreillons ont été déclarés sur la seule base des symptômes cliniques, ce qui cause un problème de spécificité important. Compte tenu de la faible incidence actuelle des oreillons, il est fort possible qu'un certain nombre de cas, déclarés à partir d'un diagnostic clinique, ne soient pas des cas d'oreillons étant donné que plusieurs autres virus peuvent aussi être mis en cause. Les fluctuations de l'incidence peuvent donc être influencées par d'autres maladies que celles des oreillons. D'autre part, il faut aussi noter que le manque d'exhaustivité des données, pour certaines années, peut causer des variations inopportunes des séries, dû à l'ajout ou au retrait graduel de certaines provinces dans le cumul des cas. Ces facteurs sont pris en compte dans l'interprétation des résultats. Le tableau 2.4 donne un résumé des caractéristiques et des biais relatifs aux séries d'infections analysées.

**Tableau 2.4**  
**Sommaire des caractéristiques des séries d'infections analysées**

1953-1959	1970-1976	1991-2002
<ul style="list-style-type: none"> <li>- Données par période de 4 semaines</li> <li>- Absence de définitions uniformes à l'échelle nationale</li> <li>- Manque de validité des diagnostics cliniques</li> <li>- Sous-estimation (tableau 2.1)</li> <li>- Manque d'exhaustivité des données de la rougeole, des oreillons et de la rubéole (tableau 2.2)</li> </ul>	<ul style="list-style-type: none"> <li>- Données par période de 4 semaines</li> <li>- Données des oreillons non disponibles</li> <li>- Absence de définitions uniformes à l'échelle nationale</li> <li>- Manque de validité des diagnostics cliniques</li> <li>- Sous-estimation (tableau 2.1)</li> <li>- Manque d'exhaustivité des données de la rougeole et de la rubéole (tableau 2.3)</li> </ul>	<ul style="list-style-type: none"> <li>- Données mensuelles</li> <li>- Déclaration généralement basée sur des définitions uniformes (incluant tests de laboratoire)</li> <li>- Faible valeur prédictive positive des diagnostics en raison de la faible prévalence des maladies</li> <li>- Sous-estimation (tableau 2.1)</li> <li>- Problème important de spécificité des cas d'oreillons en raison du fait que certaines provinces déclarent les cas d'oreillons sur la seule base des symptômes cliniques</li> </ul>

## 2.2 Méthodologie

La présente recherche analyse les séries chronologiques des infections sélectionnées à l'aide de :

- 1) la méthode de Box-Jenkins (ARIMA);
- 2) la méthode de loi de puissance inverse de Rhodes et al. (1996a; 1996b; 1997; 1998).

La première méthode identifie les dynamiques de base (objectif 2), la deuxième les interprète en termes de loi d'organisation (objectif 3).

### 2.2.1 Méthode de Box-Jenkins

La méthode de Box-Jenkins modélise la dépendance entre une observation et celle(s) qui la précède(nt) dans une suite d'observations temporelles. La description de la méthode est disponible dans plusieurs manuels (McCleary et al., 1980; McDowall et al., 1983) et

se trouve détaillée dans l'article 2 (Chapitre 3). Le logiciel « SPSS, version 11.5, incluant le module *Trend* », a été utilisé pour la modélisation.

Dans le cadre de l'analyse de Box-Jenkins uniquement, les données mensuelles (1991-2002) ont été rapportées sur une base de 28 jours. Cela permet d'obtenir des données davantage comparables avec les séries de 1953-1959 et de 1970-1976 qui sont déclarées sur une base de 28 jours. De plus, la base de 28 jours a été préférée, car elle correspond au mois le plus court. Choisir une correction à 30 jours aurait ajouté des cas aux mois plus courts tandis qu'une correction à 28 jours ne fait que retrancher ailleurs des cas effectivement survenus. Nous avons donc divisé l'effectif mensuel des cas par 29, 30 ou 31 jours, selon le mois, tout en tenant compte des années bissextiles, et nous avons multiplié le résultat par 28. Toutes les données de 1991 à 2002 tiennent compte de cette correction. Par ailleurs, pour les séries de 1991-2002 qui contenaient des valeurs nulles (rougeole et rubéole), la constante 1 a été ajoutée systématiquement à toutes les données de la série. Cette procédure permet de recourir à la transformation logarithmique des données. Pour la série de rougeole de 1991-2002 seulement 12 des 144 mois inclus dans l'analyse comportaient une valeur nulle. Pour la série de rubéole 1991-2002, il y en avait 6 sur 138. Aucune autre série ne comportait des valeurs nulles. L'ajout de la constante 1 n'a pas eu d'impact sur les résultats de l'analyse de Box-Jenkins puisqu'en ajoutant la constante 1 à toutes les données, on ne change pas la structure temporelle des séries. Par contre, cela permet d'en changer l'échelle d'analyse (avec la transformation logarithmique).

### 2.2.2 Méthode de loi de puissance inverse de Rhodes et al. (1996a; 1996b; 1997; 1998)

La méthode consiste d'abord à calculer le nombre des épidémies (d'une série chronologique) de tailles données par rapport au temps et, ensuite, à estimer la loi de puissance inverse sur les données. Cette loi est comparée à un modèle exponentiel pour en évaluer l'ajustement. La qualité de l'ajustement est donnée par le  $R^2$ . La méthode est détaillée dans le troisième article (chapitre 3). L'approche permet éventuellement d'identifier une loi d'organisation déterministe des épidémies. Aucune modification des données n'a été effectuée pour cette méthode (la valeur 1 n'a pas remplacé 0 comme dans l'analyse de Box-Jenkins). Le logiciel SPSS a été utilisé pour l'analyse.

CHAPITRE 3

---

ARTICLES

ARTICLE 1

## **Qualité des données de surveillance de la coqueluche, de la rougeole, de la rubéole et des oreillons au Canada**

### **Résumé**

**Contexte :** Depuis 1924, le système de surveillance canadien permet de compiler, dans une base de données centrale appelée MADO (maladies à déclaration obligatoire), les données relatives à l'émergence de certaines maladies comme la coqueluche, la rougeole, la rubéole et les oreillons. La surveillance des maladies implique une chaîne hiérarchique de liens (consultation médicale, diagnostic, déclaration des cas) susceptibles d'être influencés par une diversité de facteurs qui peuvent venir altérer la déclaration des cas au système de surveillance. Le but du présent article est d'évaluer la qualité des données de la coqueluche, de la rougeole, de la rubéole et des oreillons déclarées au fichier canadien des MADO en tenant compte des diverses étapes de la hiérarchie de saisie des données de surveillance.

**Méthode :** Une revue de littérature des articles publiés dans plusieurs pays industrialisés sur les données de surveillance de la coqueluche, de la rougeole, de la rubéole et des oreillons a été effectuée à l'aide de la ressource électronique PUBMED. Les différents rapports du gouvernement canadien sur la surveillance de ces maladies ont également été passés en revue. La revue a été limitée aux études portant sur la qualité des données.

**Résultats :** Les données de surveillance sous-estiment considérablement l'incidence réelle des maladies. Par exemple, en 1950, 11 % des cas de coqueluche étaient déclarés au Canada. Cette proportion semble avoir diminué à 6 % en 1991-1993. Les estimations

des pays industrialisés sont similaires, allant de 5 à 25 % pour les années 1940-1980 et de 3 à 12 % pour les années 1985-1988. Pour la rougeole, la proportion de cas déclarés au Canada était de 13 % en 1950. Cette proportion, apparemment faible par rapport à celle des pays industrialisés, ne s'écarte tout de même pas des estimés obtenus dans ces pays entre 1920 et 1980, qui vont de 7 à 63 %. Dans les années récentes, différentes études dans les pays industrialisés montrent que cette proportion est de 33 à 41 %. Pour la rubéole et les oreillons, la proportion de cas déclarés dans les années 1950 au Canada semblait être autour de 17 % et de 14 % respectivement. Par ailleurs, bien que le niveau réel de l'incidence soit sous-estimé, les études consultées montrent que les données donnent généralement une représentation juste des variations de l'incidence à court terme.

**Conclusion** : Les données posent un préjudice notable si l'on s'intéresse à l'incidence des maladies au cours du temps. Nous concluons que le système sert à suivre les tendances de l'incidence des maladies sous surveillance bien plus que leur incidence. Par ailleurs, les tendances sont difficilement comparables sur de longues périodes puisque les facteurs affectant la déclaration (propension des gens à consulter les services de santé, valeur prédictive positive des diagnostics, niveau de sous-estimation, etc.) ne peuvent être tenus constants, sinon sur de courtes périodes.

### **Mots-clés**

Surveillance, maladies à déclaration obligatoire (MADO), coqueluche, rougeole, oreillons, rubéole, qualité des données, proportion de cas déclarés, tendance.

## **Introduction**

Le Canada, comme bon nombre de pays, est doté depuis longtemps d'un système de surveillance des maladies transmissibles. Un tel système constitue une condition essentielle au contrôle et à la prévention des maladies transmissibles. Il permet de détecter les épidémies, d'approfondir les connaissances épidémiologiques (évolution et transmission de la maladie), de déceler les nouveaux problèmes de santé comme l'émergence d'agents infectieux inconnus, d'évaluer les programmes d'intervention comme la vaccination ou les programmes d'élimination de certaines maladies transmissibles évitables (Thacker et al., 1989; Santé et Bien-Être social Canada, 1991b; Doherty; 2000, Santé Canada, 2000). La compilation des données relatives aux maladies surveillées sur de longues périodes permet également de suivre l'émergence des maladies et d'analyser leurs patterns de variation au cours du temps. Grâce à un tel système de surveillance et à la compilation des données dans un fichier central des maladies à déclaration obligatoire (MADO) au Canada, il est possible de suivre l'émergence de certaines maladies transmissibles depuis 1924 (Bureau fédéral de la statistique, 1924-1952; Bureau fédéral de la statistique, 1953-1970; Statistique Canada, 1971-1978; Santé et Bien-Être social Canada, 1975-1991; Santé Canada, 1992-2004).

Cependant, lorsqu'on s'intéresse à la surveillance des maladies, plusieurs questions surgissent. L'incidence réelle est-elle sous-estimée ? Les données sont-elles fiables ? La qualité des données est-elle uniforme au cours du temps ? Le but du présent article est de poser un regard critique sur la qualité des données concernant les séries chronologiques de rougeole, de coqueluche, de rubéole et des oreillons, déclarés au fichier des MADO au Canada. Suite à un bref historique du système canadien des

MADO, nous analyserons les diverses étapes de la hiérarchie de saisie des données de surveillance afin de pouvoir évaluer la qualité des données canadiennes.

### **Bref historique du système de surveillance des MADO au Canada**

Depuis 1924, les données relatives aux maladies sous surveillance sont compilées dans un fichier central appelé MADO (maladies à déclaration obligatoire), suite à leur réception en provenance des provinces et des territoires. En 1924, c'était le Bureau fédéral de la statistique (BFS), devenu plus tard Statistique Canada, qui avait la charge de gérer le fichier national des MADO. Cette tâche a été transférée au Laboratoire de lutte contre la maladie (LLCM) en 1988. Les données de la coqueluche, de la rougeole, de la rubéole et des oreillons sont disponibles depuis 1924, à l'exception de la période 1959-1968 pour la rougeole et la rubéole et de la période 1959-1986 pour les oreillons (périodes où elles n'ont pas fait l'objet de déclaration obligatoire à l'échelle nationale)<sup>1</sup>.

Entre 1924 et 1991, aucun critère normalisé n'existait pour la déclaration des cas à l'échelle nationale. Les provinces et les territoires transmettaient les données agrégées selon leur propre réglementation sur la base des symptômes cliniques (dont il n'existait d'ailleurs pas de définition uniforme à l'échelle nationale). Les MADO devaient être signalées par les médecins et autres intervenants au médecin-hygiéniste, qui à son tour, en faisait rapport au ministère provincial de la santé. Celui-ci transmettait alors les renseignements de la province entière au BFS (devenu plus tard Statistique Canada). En 1991, de concert avec le LLCM et les épidémiologistes provinciaux et ceux des

---

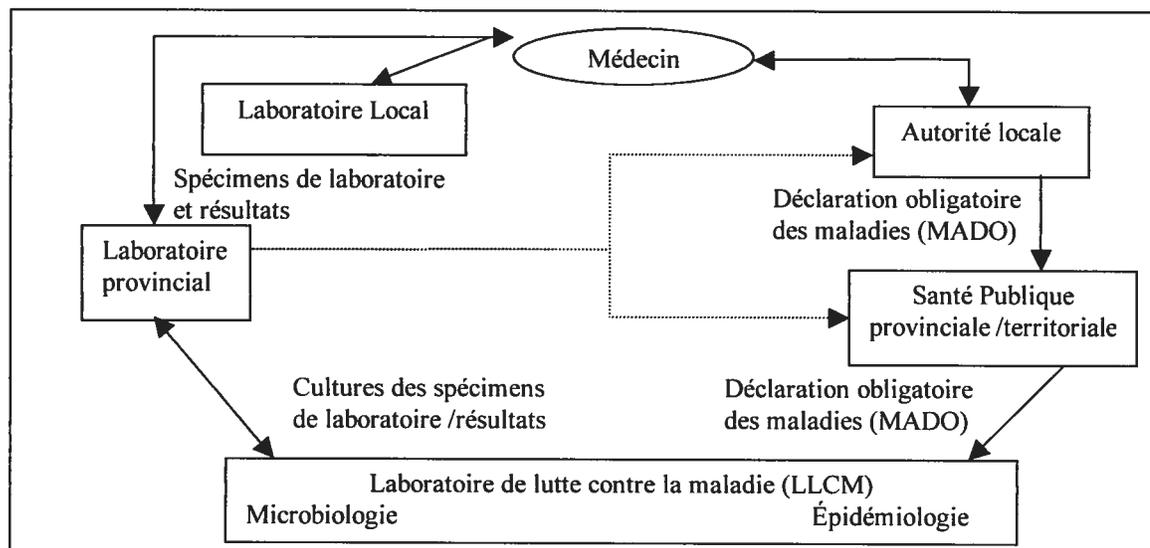
<sup>1</sup> A noter que l'Île-du-Prince-Édouard, Terre-Neuve, le Yukon et les Territoires du Nord-Ouest ne se sont joints au système de surveillance canadien qu'à partir de 1929, 1949, 1955 et 1959 respectivement (BFS, 1954a; BFS, 1955; BFS, 1959). De plus, en 1959 la rougeole, les oreillons et la rubéole ont été exclus de la liste des MADO, mais certaines provinces ont arrêté le signalement de ces dernières à partir de 1954. En 1969, la rougeole et la rubéole sont ré-introduites dans la liste, mais certaines provinces ont mis jusqu'en 1988 avant de les déclarer. Les données manquent donc d'exhaustivité pour certaines périodes.

territoires, le Comité consultatif de l'épidémiologie (CCE) a publié des définitions spécifiques pour chaque maladie faisant l'objet de surveillance à l'échelle nationale (Santé et Bien-Être social Canada, 1991b). Pour la première fois, ces définitions fournissaient des critères normalisés pour la déclaration des cas. De nouvelles réalités, comme l'utilisation de plus en plus fréquente des techniques de biologie moléculaire à titre d'outil diagnostique, ont été des éléments incitant à mettre à jour ces définitions. Par exemple, en 1997, le document sur la définition des cas a été révisé par le CCE (Santé Canada, 2000). La dernière révision des définitions de rougeole, de rubéole, de coqueluche et d'oreillons utilisée pour la déclaration de cas se trouve en annexe.

La surveillance épidémiologique est une activité complexe. La figure 1 montre le schéma général de la déclaration des cas au Canada (Sockett et al., 1996). Généralement, les médecins déclarent les cas diagnostiqués cliniquement, avec ou sans confirmation de laboratoire, à l'autorité sanitaire locale. Avant de déclarer officiellement le cas, celle-ci doit déterminer si le cas correspond à la définition établie aux fins de la surveillance. Le plus souvent, un cas est confirmé s'il est associé à un test de laboratoire, sans égard au tableau clinique. Par ailleurs, un cas peut aussi être confirmé en l'absence d'examen de laboratoire quand il présente un ensemble de symptômes et un lien épidémiologique avec un autre cas confirmé antérieurement par un test de laboratoire, le tout étant suffisamment spécifique pour que le degré de certitude entourant le diagnostic soit élevé. S'il existe des doutes, parce que les données sont incomplètes ou encore que les résultats ne sont pas concluants, le cas peut être déclaré comme cas probable, mais il devra être confirmé par la suite. S'il n'est pas confirmé, il est retiré de la compilation. Les cultures ou spécimens de laboratoire de certains cas peuvent être référés au laboratoire provincial pour confirmation du diagnostic. Les autorités locales et provinciales peuvent également

recevoir de l'information en provenance des laboratoires. Lorsque le cas est confirmé, il est déclaré au LLCM par l'autorité de santé publique provinciale ou territoriale, et il est compilé dans la base de données centrale canadienne des MADO.

**Figure 1.** Organisation du système de surveillance des MADO au Canada.



Source : Sockett et al., 1996.

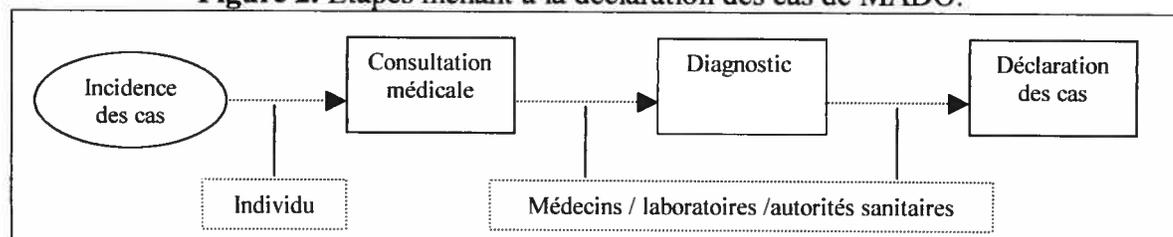
## Méthode

La revue de la qualité des données de surveillance canadiennes porte sur la coqueluche, la rougeole, la rubéole et les oreillons. Pour évaluer la qualité des données, il est nécessaire de confronter les études de qualité effectuées sur les données canadiennes aux études effectuées dans le monde industrialisé de pays comparables au cours des mêmes périodes. Une recherche a donc été entreprise avec PUBMED (regroupant la base de données MEDLINE; 1951-1965, 1966-2004) pour chaque maladie « measles », « rubella », « pertussis » et « mumps » en y ajoutant les termes suivants : « notification », « surveillance », « reporting ». Les références des articles retenus par la recherche ont permis d'obtenir des articles supplémentaires. De plus, une revue de

littérature des différents rapports du gouvernement canadien sur la surveillance de ces maladies depuis 1924 a été effectuée. Seuls les articles portant directement sur la qualité des données de surveillance ont été retenus. Ceux portant sur l'épidémiologie ont été exclus. Bien que le système de surveillance canadien ne soit pas directement comparable à celui des pays européens, des États-Unis, de l'Australie ou à celui d'autres pays industrialisés, les évaluations externes obtenues à partir des recherches effectuées ailleurs permettent néanmoins d'encadrer et de qualifier, dans une bonne mesure, la qualité des données canadiennes.

La déclaration des cas implique une chaîne hiérarchique complexe (figure 2). L'individu infecté doit d'abord accéder à la structure de soins. Le médecin doit ensuite faire le diagnostic approprié et voir à déclarer le cas à l'autorité sanitaire. Plusieurs facteurs peuvent intervenir à chacune des étapes de la chaîne et interférer avec la déclaration des cas au fichier central. Des estimés concernant la déclaration seront présentés en fonction de la hiérarchie du système de saisie des données (proportion des individus infectés qui consultent les services médicaux, proportion des cas diagnostiqués, proportion des cas déclarés par les intervenants). Par la suite, la proportion des cas déclarés (par rapport à incidence réelle) de chaque maladie sera estimée. Une dernière section présentera les études qui ont évalué les variations au cours du temps des données déclarées par rapport à l'incidence réelle.

**Figure 2.** Étapes menant à la déclaration des cas de MADO.



## Résultats

### Variation de la déclaration en fonction de la hiérarchie du système de surveillance

Le tableau 1 montre l'état de la déclaration des cas en fonction de chaque étape de la hiérarchie du système, selon diverses études.

#### 1) Consultation médicale

La première étape du processus de déclaration implique que l'individu infecté soit venu à l'attention des services de soins. Une étude au Québec sur la coqueluche entre 1991 et 1993 a montré que seulement 59 % des sujets qui avaient eu la coqueluche (selon la définition clinique utilisée au Canada) avaient entrepris une consultation médicale (Deeks et al., 1999). Ces derniers avaient généralement consulté en raison de symptômes plus graves. En ce qui concerne la rougeole, une étude à Los Angeles en 1990-1991 montre que 85 % des individus infectés ont eu une consultation médicale (Ewert et al., 1995). Cependant, en 1922-1923, Sydenstricker (1926) avait montré que cette proportion était de 64 % aux États-Unis. Aucune étude recensée n'a permis d'estimer la proportion des individus qui ont consulté après avoir été infecté par la rubéole et les oreillons.

#### 2) Diagnostic de la maladie

La deuxième étape du processus de déclaration concerne le diagnostic. Deeks et al. (1999) ont montré que seulement 67 (12 %) des 558 sujets qui avaient consulté et qui respectaient la définition clinique de la coqueluche ont eu un tel diagnostic par le médecin. Par ailleurs, Rivest et al. (2004) ont analysé les dossiers médicaux pour lesquels un diagnostic de coqueluche avait été posé en 1998 dans un centre hospitalier de Montréal. Parmi les 139 dossiers analysés, ils ont montré que 62 % des sujets avaient eu une culture ou un test par PCR (*Polymerase Chain Reaction*). Les tests diagnostiques de la coqueluche semblent donc sous-utilisés. Quoique que les cas déclarés ne

représentent qu'une mince proportion de ceux qui émergent plusieurs peuvent ne pas être des cas de coqueluche (Rivest et al., 2004).

Le diagnostic de la rougeole connaît aussi ses faiblesses. Une étude dans cinq villes aux États-Unis de 1995 à 1999 a montré que seulement 52 (23 %) des 253 sujets qui répondaient à la définition clinique de la rougeole, ont eu un tel diagnostic par le médecin (Kolasa et al., 2004). À Los Angeles entre 1990 et 1991, Ewert et al. (1995) ont montré que cette proportion était de 52 % (11 des 21 sujets qui répondaient à la définition clinique de la rougeole ont eu ce diagnostic). L'investigation des cas en laboratoire semble aussi problématique. Selon une étude en Alberta en 1988, 43 % des cas de rougeole déclarés étaient confirmés en laboratoire (Varughese et al., 1990). Au cours des dernières années, au Canada, les cas suspects ont été analysés par le Groupe de travail sur l'élimination de la rougeole qui a confirmé le diagnostic (avec test de laboratoire) avant que le cas ne soit ajouté à la base de données. Il suit de cette analyse qu'une proportion de cas ne semble toujours pas investiguée. Aux États-Unis, de 1995 à 1999, Kolasa et al. (2004) ont montré que 83 % des cas suspects de rougeole ont été investigués en laboratoire. Ewert et al. (1995) ont par ailleurs montré que seulement 6 (55 %) des 11 sujets de leur étude, qui ont eu un diagnostic de rougeole par le médecin, avaient eu une investigation en laboratoire. En Écosse, entre 1983 et 1991, seulement 2 % des cas déclarés ont été investigués (Brewster, 1993). Une étude jamaïcaine a montré que 37 % des cas suspects de rougeole ont été investigués et que la majorité s'est révélée être une rubéole (EPI newsl., 1992). Il est donc probable que les intervenants passent à côté de plusieurs cas de rougeole et que, parmi les cas déclarés, plusieurs n'en sont pas.

Le diagnostic de la rubéole n'est pas non plus chose facile. Plusieurs autres maladies peuvent causer une éruption cutanée similaire. De plus, se fier aux anticorps

spécifiques (IgM) comme moyen diagnostique peut causer des problèmes et entraîner des résultats faussement positifs comme, par exemple, chez les personnes dont les IgM restent longtemps présents dans le sang après une immunisation ou une infection (Santé Canada, 2002). Dans ce contexte, un nombre important de cas déclarés peuvent ne pas être des cas de rubéole. On peut également être perplexe concernant les cas déclarés d'oreillons. Une étude en 1997 a montré que plus de 90 % des cas d'oreillons déclarés au Québec sont des cas cliniques qui n'ont pas été investigués en laboratoire. Cela pose un problème de spécificité important (De Serres et al., 1997). Ainsi, dans une série de 20 cas de parotidite survenue chez des sujets vaccinés, Brunell (1972) avait montré que seulement 8 étaient réellement dus aux oreillons. Il est fort possible qu'un nombre important de cas déclarés ne soient pas des oreillons.

### 3) Déclaration des cas par les intervenants

Finalement, la troisième étape du processus de déclaration implique que les intervenants déclarent les cas diagnostiqués au système de surveillance. Deeks et al. (1999) ont montré que seulement 50 % des cas dûment diagnostiqués de coqueluche ont été déclarés par le médecin. En Australie, une étude a révélé que cette proportion est de 40 % pour la coqueluche et de 80 % pour la rougeole (Allen et al., 2000). À Sidney, 45 % des cas, qui avaient reçu un diagnostic de rougeole par le médecin lors de l'épidémie en 1993, ont été déclarés (McDonnell et al., 1994). Une analyse de la déclaration des cas de rougeole par 12 hôpitaux à New-York en 1991 a montré que 45 % des cas avaient été déclarés au Département de la santé (Davis et al., 1993). Cette proportion était de 67 % au sein de l'étude de Ewert et al. (1995) sur la rougeole. La déclaration des cas par les intervenants est donc loin d'être complète.

**Tableau 1. Évaluation de la déclaration des cas en fonction de la hiérarchie du système de surveillance selon différentes études**

Maladie	Lieu de l'étude	Période	Type d'étude	% de personnes infectées qui ont une consultation médicale	% de personnes qui respectent la définition clinique et qui ont un diagnostic par le médecin	% de cas diagnostiqués par le médecin qui ont fait l'objet de déclaration	Problème identifié	Référence
Coqueluche	Québec, Qc	1991-1993	Étude effectuée auprès de 8 235 enfants provenant d'écoles primaires et de centres de l'enfance (diagnostic parental)	59 %	67 / 558 (12 %)	33 / 67 (50 %)	- Faible consultation médicale - Cultures ou tests PCR sous-utilisés - Problème avec le diagnostic - Sous-déclaration des cas diagnostiqués par le médecin	Deeks et al. (1999)
	Sydney, Aus	2000	Enquête effectuée auprès de médecins			40 %	- Sous-déclaration des cas diagnostiqués par le médecin	Allen et al. (2000)
Rougeole	Hagerstown, É-U	1922-1923	Étude effectuée auprès de la population (diagnostic parental)	64 %				Sydenstricker (1926)
	Los Angeles, É-U	1990-1991	Étude effectuée auprès de 947 enfants (diagnostic parental)	85 %	52 %	67 %		Ewert et al. (1995)
	New York, É-U	1991	Étude effectuée auprès de 12 hôpitaux à partir du diagnostic clinique fait par le médecin			664 / 1487 (45 %)	-Sous-déclaration des cas diagnostiqués par les médecins	Davis et al. (1993)
	5 villes des États-Unis	1995-1999	Étude effectuée auprès de 5 départements de santé à partir du diagnostic clinique fait par le médecin		58 / 253 (23 %)			Kolasa et al. (2004)
	Sydney, Aus	1993	Étude effectuée auprès de 796 parents d'enfants fréquentant 5 écoles primaires (diagnostic parental)			37 / 83 (45 %)	-Sous-déclaration des cas diagnostiqués par les médecins	McDonnell et al. (1994)
	Sydney, Aus	2000	Enquête effectuée auprès de médecins			80 %	Sous-déclaration des cas diagnostiqués par les médecins	Allen et al. (2000)

### Évaluation de la proportion des cas déclarés au système de surveillance

Plusieurs études ont tenté d'évaluer l'incidence réelle des maladies afin d'estimer le niveau de cas déclarés au sein des systèmes de surveillance (tableau 2; elles ont surtout été effectuées au cours des dernières années, et très peu concernent la rubéole et les oreillons). Selon l'enquête nationale sur la maladie effectuée en 1950-1951 (ministère de la Santé nationale et du Bien-Être social, 1962), 11 % des cas de coqueluche ont été déclarés au système de surveillance canadien. Au Royaume-Unis, Clarkson et al. (1985) montrent que 5 à 25 % des cas de coqueluche ont été déclarés durant la période 1941-1981. De plus, les auteurs montrent que la proportion de cas déclarés est plus complète en période de prévacination (16-17 %) qu'en période de postvacination (9-10 %). Une étude américaine sur les données entre 1985 et 1988 a montré que seulement 3 à 12 % des cas de coqueluche chez les jeunes enfants ont été déclarés au CDC (Sutter et al., 1992). Au Québec, entre 1991 et 1993, l'étude de Deeks et al. (1999) montre que 6 % des cas ont été déclarés. Une étude récente à Sydney montre que le nombre des hospitalisations relatives à la coqueluche, entre 1993 et 1999, excède le nombre des cas déclarés de 32 % (Torvaldsen et al., 2003) Cette situation est la même pour l'ensemble de l'Australie (McIntyre et al., 2000) et la Nouvelle-Zélande (Blakely et al., 1999). Comme seulement une mince proportion des cas est hospitalisée (les cas les plus sévères), les auteurs estiment que la sous-déclaration est très importante et ce, surtout chez les adultes et les adolescents qui peuvent présenter des symptômes plus atténués. À Sydney, au cours des dernières années, on a estimé que le nombre réel de cas est douze fois supérieur à celui des cas déclarés (8 % des cas incidents)(Cagney et al., 2001).

La déclaration des cas de rougeole semble plus complète que celle de la coqueluche, mais les estimations ne sont guère plus encourageantes. Avant la période de

vaccination de masse, Sydenstricker (1926) a montré que 26 % des cas de rougeole étaient déclarés en 1922-1923 aux États-Unis. Au Canada, l'enquête nationale sur la maladie effectuée en 1950-1951 (Ministère de la Santé nationale et du Bien-Être social, 1962) permet de constater que 469 000 cas de rougeole ont été déclarés pour 1950-1951 (sur une période de douze mois), alors que seulement 61 300 cas de rougeole avaient été enregistrés au système de surveillance au cours de l'année 1951 (Bureau fédéral de la statistique, 1954a). Selon cette enquête, il y aurait donc environ 13 % des cas qui auraient été déclarés. Au Royaume-Uni, entre les années 1941 et 1981, Clarkson et al. (1985) montrent que la proportion des cas déclarés pour la rougeole atteint des niveaux de 40 à 60 %. Les auteurs montrent également que la proportion du nombre de cas déclarés semble plus complète en période de prévacination (56-63 %) qu'en période de postvacination (33-40 %). Harpaz (2004) a publié une revue de littérature concernant la validité des données de surveillance de la rougeole dans différentes villes des États-Unis pour différentes périodes; l'ensemble de leurs études montre que la déclaration des cas de rougeole se situe entre 3 % et 29 %. Au cours des dernières années, Ewert et al. (1995) a montré que cette proportion se situe autour de 18-33 % aux États-Unis. En Australie, McDonnell et al. (1994) ont montré que 41 % des cas incidents avaient été déclarés lors de l'épidémie survenue dans la ville de Sydney en 1993.

En ce qui concerne la rubéole et les oreillons, très peu d'études se sont penchées sur la qualité des données déclarées. Cependant, la comparaison des données de l'enquête nationale sur la maladie effectuée au Canada en 1950-1951 avec celles des cas déclarés dans le fichier des MADO, montre que les pourcentages de cas déclarés auraient été de 14 % pour les oreillons et de 17 % pour la rubéole (Ministère de la Santé nationale et du Bien-Être social, 1962).

**Tableau 2. Variations de la proportion des cas déclarés selon différentes études**

Lieu de l'étude	Période	Mode d'estimation du nombre de cas réels	% de cas déclarés	Problème identifié	Référence
Coqueluche Canada	1950-1951	Diagnostic parental avec l'enquête nationale sur la maladie	11 %		Ministère de la Santé nationale et du Bien-Être social (1962)
Grande-Bretagne	1941-1981	Comparaison des cas déclarés avec différents fichiers (naissance, hospitalisation, autres systèmes de déclaration)	5-25 %	- Sous estimation très importante - Proportion de cas déclarés varie en fonction des périodes de vaccination : prévaccination: 16-17 % postvaccination: 9-10 %	Clarkson et al. (1985)
États-unis	1985-1988	Comparaison des cas déclarés avec différents fichiers (données d'hospitalisation par âge)	3-12 %		Sutter et al. (1992)
Québec, Qc.	1992-1993	Diagnostic parental avec une étude de cohorte rétrospective dans 14 écoles primaires et centre de l'enfance de la ville de Québec (8 235 enfants)	6 %	- Cultures sous-utilisées - Problème avec le diagnostic	Deeks et al. (1999)
Rougeole Hagerstown, É-U	1922-1923	Diagnostic parental avec une enquête dans la communauté	26 %		Sydenstricker (1926)
Canada	1950-1951	Diagnostic parental avec l'enquête nationale sur la maladie	13 %		Ministère de la Santé nationale et du Bien-Être social (1962)
Grande-Bretagne	1941-1981	Comparaison des cas déclarés avec différents fichiers (naissance, hospitalisation, autres systèmes de déclaration)	40-60 %	- Sous estimation importante - Proportion de cas déclarés varie en fonction des périodes de vaccination : prévaccination: 56-63 % postvaccination: 33-40 %	Clarkson et al. (1985)
St-Louis, É-U	1970-1971	Diagnostic parental avec une enquête par la poste (10 000 enfants)	7 %		Cherry et al. (1972)
Los Angeles, É-U	1987-1989	Diagnostic parental avec une enquête auprès de familles (947 enfants)	18 %	- Forte sous-estimation de la part des médecins en clinique privé	Ewert et al. (1995)
Sydney, Aus	1990-1991	Diagnostic parental avec une enquête auprès de 796 parents d'enfants fréquentant 5 écoles primaires	33 %		McDonnell et al. (1994)
Rubéole Canada	1950-1951	Diagnostic parental avec l'enquête nationale sur la maladie	41 %	-Sous-déclaration très importante des cas diagnostiqués par les médecins	Ministère de la Santé nationale et du Bien-Être social (1962)
Oreillons Canada	1950-1951	Diagnostic parental avec l'enquête nationale sur la maladie	17 %		Ministère de la Santé nationale et du Bien-Être social (1962)
			14 %		Ministère de la Santé nationale et du Bien-Être social (1962)

### Évaluation des variations de l'incidence des maladies au cours du temps

Quoique les données semblent sous-estimer considérablement l'incidence réelle des maladies, on s'entend généralement pour dire que le système est utile pour établir les tendances de la maladie, c'est-à-dire les fluctuations de l'incidence au cours du temps. Par exemple, les patterns épidémiques de la coqueluche dans le fichier des cas déclarés et le fichier des hospitalisations, entre 1993 et 1999, à Sydney, sont similaires à travers le temps laissant suggérer qu'ils sont le reflet d'une réelle variation de l'incidence (Torvaldsen et al., 2003). Les cas de coqueluche et de rougeole rapportés par les laboratoires suivent les mêmes tendances que celles des cas déclarés au Royaume-Uni (Clarkson et al., 1985). En Écosse, les cas déclarés de rougeole entre 1983 et 1991 suivent les mêmes tendances que celles rapportées par les laboratoires (Brewster, 1993). Au Canada, à partir de 1981, les laboratoires de diagnostics virologiques à travers le Canada ont envoyé leurs rapports au LLCM qui ont pu être comparés avec les cas déclarés (Santé et Bien-Être social Canada, 1983); quoique les deux types fichiers ne reflètent pas l'incidence de la maladie, les tendances (variations) qui y sont notées sont similaires. Toute variation dans un sens ou dans un autre dans les données, au cours du temps, serait donc le reflet d'une variation équivalente véritable dans la population.

### **Discussion**

La revue de la littérature sur les données de surveillance de la coqueluche, de la rougeole, de la rubéole et des oreillons fait ressortir des lacunes importantes du système de surveillance, tant pour le Canada que pour l'ensemble des pays industrialisés. Il est néanmoins nécessaire de prendre note des limites du mode de saisie des articles (PUBMED) qui ont conduit à ces résultats. En effet, les ressources électroniques

permettent de détecter les seuls articles publiés; le système ne permet pas d'inclure des références comme les rapports internes des gouvernements sur la surveillance. Par conséquent, des références utiles portant sur la qualité des données n'ont pas pu être recensées. L'analyse des articles trouvés a néanmoins permis de faire ressortir des particularités très semblables concernant la qualité des données de surveillance du monde industrialisé, particularités que le Canada partage avec ces pays. Par exemple, en 1950, 11 % des cas de coqueluche étaient déclarés au Canada. Cette proportion semble avoir diminué à 6 % en 1991-1993. Les estimations des pays industrialisés vont de 5 à 25 % pour les années 1940-1980 et de 3 à 12 % en 1985-1988. Pour la rougeole, en 1950, la proportion de cas déclarés au Canada était de 13 %. Cette proportion semble faible par rapport à celle des autres pays, mais elle ne s'écarte pas non plus des estimés obtenus pour la période 1920 à 1980, qui s'étendent de 7 à 63 %. La déclaration des cas pour ces deux maladies au Canada semble donc comparable à celle des autres pays.

La cohérence des résultats mise à part, il reste que la sous-déclaration des cas est considérable pour tous les pays considérés. D'abord, tous les cas ne sont pas portés à l'attention des services médicaux. Des facteurs comme la symptomatologie, l'absence d'un médecin à proximité, le départ à l'extérieur du pays d'une personne infectée ou encore l'appartenance religieuse peuvent affecter la propension des gens à consulter pour une MADO. Ainsi, les consultations sont probablement moins importantes pour des maladies comme la coqueluche dont les symptômes peuvent être plus bénins chez l'adulte et l'adolescent (Halperin et al., 1989; Deeks et al., 1999; Rivest et al., 2004) ou encore pour des infections comme la rubéole pour laquelle 50 % des cas sont estimés asymptomatiques (Santé Canada, 2002). Par ailleurs, la grande faiblesse des systèmes de surveillance se situe souvent au niveau du diagnostic de la maladie. La sensibilité et la

spécificité des diagnostics cliniques et/ou de laboratoires demeurent un défi de taille pour ces maladies et ce, surtout dans un contexte de faible prévalence. Finalement, une proportion importante de médecins ne déclare pas les cas aux autorités sanitaires. Ainsi, même si les individus infectés sont connus des services médicaux, bien souvent l'information ne parvient pas au fichier central.

Évidemment le nombre absolu de cas non-déclarés a été moins important au cours des années récentes puisque l'incidence des maladies est beaucoup plus faible, mais il reste que les données déclarées semblent ne représenter qu'une mince proportion du nombre total des cas survenus. Il appert que cette constatation est d'autant plus grave pour la coqueluche. Les difficultés associées à la déclaration de cette dernière sont nombreuses : une proportion importante de cas n'est pas portée à l'attention des soins médicaux; il y a un manque de spécificité de la définition du cas; il semble également y avoir une réticence de la part des médecins à poser un diagnostic de coqueluche même en présence de symptômes classiques; d'autres micro-organismes peuvent être responsables d'une toux sévère; une faible sensibilité de la culture; et une forte proportion de médecins qui ne déclarent pas les cas diagnostiqués (Deeks et al., 1999, Rivest et al., 2004). Il est donc probable que parmi les cas déclarés, plusieurs ne sont pas des cas de coqueluche et que ceux qui le sont ne représentent qu'une mince proportion de ceux qui émergent (Rivest et al.; 2004). Selon les études recensées, de 3 à 25 % des cas de coqueluche semblent déclarés. De plus, il semble que cette proportion soit encore plus faible après une période de vaccination de masse, passant de 16-17 % à 9-10 % (Clarkson et al., 1985). Il est difficile d'identifier les facteurs responsables de la sous-déclaration plus grande dans un contexte de postvaccination. La plus faible incidence de la maladie et l'idée générale selon laquelle un certain contrôle de la maladie est atteint

avec la vaccination amènent peut-être une plus grande proportion de médecins à ne pas en faire le diagnostic et à ne pas déclarer les cas (Deeks et al., 1999).

La rougeole semble conduire à plus de consultation médicale et semble mieux déclarée que la coqueluche. Cependant, le diagnostic n'est pas simple et une proportion non négligeable de médecins semble ne pas déclarer les cas. Les estimations de la proportion des cas déclarés s'étendent de 7 à 63 %. Il semble également que cette proportion est plus faible en période de postvaccination (Clarkson et al. 1985). Cependant, à l'heure actuelle, la transmission indigène de la rougeole semble interrompue en Amérique, la majorité des cas étant importés ou reliés à des cas importés par immigration (King et al; 2004, Kolasa et al., 2004). Le nombre absolu de cas non-déclarés dans ce contexte peut donc être faible. Les études sur la qualité des données de la rubéole et des oreillons sont rares, mais l'enquête canadienne de 1950 révèle que 17 % et 14 % des cas sont déclarés pour la rubéole et les oreillons respectivement. On peut penser que la déclaration des cas a suivi la tendance des autres maladies, et qu'elle reste très faible. Ce phénomène est sans doute d'autant plus important pour la rubéole qui peut se présenter fréquemment de manière asymptomatique (Santé Canada; 2002).

Il faut, par ailleurs, noter des limites importantes concernant le calcul des estimations de la proportion de cas déclarés au système de surveillance. L'incidence réelle des maladies est souvent estimée à partir d'une évaluation parentale du diagnostic clinique sur la base d'une enquête dans la communauté. Ce type de diagnostic ne permet pas d'établir le nombre réel des cas survenus dans la population. La valeur prédictive positive du diagnostic clinique des données ainsi obtenues, sur plusieurs années d'intervalle, est questionnable et peut par conséquent surestimer le nombre réel de cas et sous-estimer, partant, la proportion des cas déclarés. Il en est de même pour les études

qui estiment la proportion de cas qui sont diagnostiqués par les médecins. Pour estimer le nombre de cas infectés qui se sont présentés chez le médecin, on utilise souvent la définition clinique de la maladie alors que cette dernière manque de spécificité et de sensibilité. La proportion des cas diagnostiqués par les médecins est alors, sans doute, considérablement sous-estimée puisqu'un nombre important d'autres maladies peuvent causer des symptômes cliniques similaires aux maladies étudiées. Par exemple, des études effectuées au cours des dernières années ont estimé la valeur prédictive positive de la définition clinique de la coqueluche à 27,5 % (Rivest et al., 2004) et celle des oreillons à 28 % (Gaulin et al., 1997). Les estimations obtenues comportent donc des limites importantes.

D'une part, en somme, nous pensons que les données sous-estiment considérablement le niveau réel de l'incidence mais, d'autre part, un nombre important de cas faussement positifs est sans doute aussi présent au sein des données de surveillance. Par exemple, à partir de 1991, en dépit des définitions de cas incluant les tests de laboratoire comme moyen de confirmation diagnostique, un nombre important de cas d'oreillons semble avoir été déclaré sur la seule base des symptômes cliniques (De Serres et al, 1997). Compte tenu de la faible incidence actuelle des oreillons, il est fort possible que le nombre de cas déclarés en fonction d'un diagnostic clinique ne soit pas des cas d'oreillons étant donné que plusieurs autres virus comme l'influenza et les entérovirus peuvent aussi causer des parotidites (Gaulin et al. 1997).

Finalement, plusieurs études ont montré que bien que le niveau réel de l'incidence des maladies est sous-estimé, les données fournissent tout de même l'heure juste quant aux variations de l'incidence dans le temps (Santé et Bien-Être social Canada, 1983; Clarkson et al., 1985; Brewster, 1993; Torvaldsen et al., 2003).

Cependant, on ne saurait espérer que les biais relatifs à la déclaration des cas soient constants dans le temps. Par exemple, la propension des gens à recourir aux soins médicaux semble plus importante au cours des périodes plus récentes. Aux États-Unis, 64 % des gens consultaient pour la rougeole dans les années 1920 alors que cette proportion semble avoir augmenté à 85 % dans les années 1990. De plus, une augmentation du nombre de cas peut également être due à la mise en place de programmes de détection des malades asymptomatiques, telle une surveillance plus ciblée de la rubéole chez les femmes enceintes. Par ailleurs, les facteurs affectant la validité des diagnostics ne sont pas constants dans le temps. Les différences de critères utilisés pour déclarer un cas (importance accordée aux critères cliniques par rapport aux confirmations de laboratoires) peuvent influencer la sensibilité et la spécificité des cas déclarés au cours du temps. Les variations dans les pratiques diagnostiques, de même que l'interprétation des définitions des maladies dans chaque province et territoire ont inévitablement entraîné des biais qualitatifs et quantitatifs qui ont, à leur tour, une incidence sur la qualité des données. En 1991, la surveillance au Canada a été marquée par l'élaboration de définitions des cas incluant généralement les tests de laboratoire comme outils diagnostiques. En présence de tests de laboratoire plus valides, la sensibilité et la spécificité des cas déclarés au système national ont pu s'améliorer dans les périodes plus récentes, mais étant donné la faible prévalence des maladies infectieuses au cours des dernières décennies, il est possible que la fréquence des cas déclarés faussement positifs soit proportionnellement plus élevée. Dans ce contexte, les variations de l'incidence notées au cours des dernières années peuvent donc être davantage contaminées par la présence d'autres virus.

## **Conclusion**

Les données de surveillance canadiennes comme celles des autres pays industrialisés souffrent de biais importants. Les évaluations externes permettent d'encadrer et de qualifier, dans une bonne mesure, la qualité des données canadiennes. Par exemple, il est clair que les données de surveillance sous-estiment considérablement l'incidence des maladies au Canada. Cependant, les difficultés associées à la surveillance ne tiennent pas à des particularités canadiennes; ces difficultés sont rencontrées dans tous les pays industrialisés. La déclaration des cas au Canada semble donc comparable à celle des autres pays.

En somme, le système de surveillance sert à suivre la tendance des maladies bien plus que leur émergence. En effet, plusieurs études ont montré que les données, bien qu'elles sous-estiment le niveau réel de l'incidence, fournissent néanmoins une représentation adéquate des variations de l'incidence dans le temps. Les données ne semblent donc pas poser de problèmes majeurs pour l'étude de la dynamique temporelle des maladies. En revanche, on ne saurait espérer que les facteurs affectant la déclaration soient constants sur de longue période. Les changements dans les pratiques de déclaration, dans les définitions des maladies, dans la prévalence des maladies, etc., entraînent inévitablement des biais qualitatifs et quantitatifs qui, à leur tour, ont une incidence sur la représentativité et l'intégrité des séries temporelles d'infections. Pour limiter l'impact de ces changements, il est donc préférable, si l'on s'intéresse à la dynamique temporelle des MADO, de faire l'analyse des données sur de courtes périodes. Un découpage des séries devrait idéalement respecter les changements comme les périodes de prévacination, de postvacination, et tenir compte des changements de définition des cas.

**Remerciements**

Les auteurs souhaitent remercier Carole Scott du LLCM et D<sup>r</sup> Yves Robert pour avoir fourni des renseignements concernant les MADO, et le Conseil de la recherche en sciences humaines du Canada (CRSH) pour le support financier de HT.

## Bibliographie

Allen C.J., Ferson M.J. (2000). Notification of infectious diseases by general practitioners: a quantitative and qualitative study. *The Medical journal of Australia*, 172(7), 325-328.

Blakely T., Mansoor O., Baker M. (1999). The 1996 pertussis epidemic in New Zealand: descriptive epidemiology. *The New Zealand medical journal*, 112(1081), 30-33.

Bureau fédéral de la statistique. (1924-1952). Division de la santé et du bien-être. Section de la santé publique. *Sommaire de maladies à déclaration obligatoire*. Annuel.

Bureau fédéral de la Statistique. (1953 à 1970). Division de la santé et du bien-être. Section de la santé publique. *Rapport annuel sur les maladies à déclaration obligatoire*. Catalogue 82-201, annuel.

Bureau fédéral de la statistique. (1954a). Division de la santé et du bien-être. Section de la santé publique. *Sommaire de maladies à déclaration obligatoire : 1924-1952*.

Brewster D. (1993). Surveillance of measles in Scotland 1983-1991; a comparison of statutory notification and laboratory reporting. *Journal of Infection*, 27(2), 193-201.

Brunell P.A., Brickman A., Steinberg S., Allen E. (1972). Parotitis in children who had previously received mumps vaccine. *Pediatrics*, 50(3), 441-444.

Cagney M., MacIntyre C.R., McIntyre P.B., Torvaldsen S., Hanlon M. (2001). *Estimating the burden of pertussis infection in Children aged 6-14 years in Western Sydney Area Health Service, New South Wales, Australia*. Westmead, NSW: National center for immunization Research and Surveillance.

Cherry J.D., Feigin R.D., Lobes L.A. Jr, Hinthorn D.R., Shackelford P.G., Shirley R.H., Lins R.D., Choi S.C. (1972). Urban measles in the vaccine era: a clinical, epidemiologic, and serologic study. *The Journal of pediatrics*, 81(2), 217-230.

Clarkson J.A., Fine P.E. (1985). The efficiency of measles and pertussis notification in England and Wales. *International journal of epidemiology*, 14(1), 153-168.

Davis S.F., Strebel P.M., Atkinson W.L., Markowitz L.E., Sutter R.W., Scanlon K.S., Friedman S., Hadler S.C. (1993). Reporting efficiency during a measles outbreak in New York City, 1991. *American journal of public health*, 83(7), 1011-1015.

Deeks S., De Serres G., Boulianne N., Duval B., Rochette L., Dery P., Halperin S. (1999). Failure of physicians to consider the diagnosis of pertussis in children. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 28(4), 840-846.

- De Serres G., Boulianne N., Bussi eres N., Pouliot B., Marin-Lira A. (1997).  pidemiologie des oreillons au Qu ebec, 1970-1995. *Relev  des maladies transmissibles au Canada*, 23(2), 9-13.
- Doherty J.A. (2000). Establishing priorities for national communicable disease surveillance. *The Canadian journal of infectious diseases and medical microbiology*, 11(1), 21-24.
- EPI Newsl. (1992). Jamaican measles surveillance system. *EPI newsletter*, 14(1), 6.
- Ewert D.P., Westman S., Frederick P.D., Waterman S.H. (1995). Measles reporting completeness during a community-wide epidemic in inner-city Los Angeles. *Public health reports*, 110(2), 161-165.
- Gaulin C., De Serres G. (1997). N cessit  d'utiliser une d finition sp cifique d'oreillons dans une population hautement vaccin e. *Relev  des maladies transmissibles au Canada*, 23(2), 14-16.
- Halperin S.A., Bortolussi R., MacLean D., Chisholm N. (1989). Persistence of pertussis in an immunized population: results of the Nova Scotia Enhanced Pertussis Surveillance Program. *The Journal of Pediatrics*, 115 (5 pt 1), 686-693.
- Harpaz R. (2004). Completeness of measles case reporting: review of estimates for the United States. *The Journal of infectious diseases*, 189 (Suppl.1), S185-S190.
- King A., Varughese P., De Serres G., Tipples G.A., Waters J. (2004). Measles elimination in Canada. *The Journal of infectious diseases*, 189 (Suppl.1), S236-S242.
- Kolasa M., Alexopoulos N., Diaz P., Kellachan J., Lowrey M.J., Shelton B., Harpaz R., Papania M.J. (2004). Measles surveillance in five major US cities: Chicago, Houston, Los Angeles, Miami, and New York. *The Journal of infectious diseases*, 189 (Suppl.1), S216-S221.
- McDonnell I.F., Jorm L.R. (1994). Measles epidemic in western Sydney. *New South Wales Public Health Bulletin*, 5(6), 59-60.
- McIntyre P., Amin J., Gidding H., Hull B., Torvaldsen S., Tucker A., Turnbull F., Burgess M. (2000). Vaccine preventable diseases and vaccination coverage in Australia, 1993-1998. *Communicable diseases intelligence*, 24 (supp.), S1-S83.
- Minist re de la Sant  nationale et du Bien- tre social. (1962). *La maladie et les soins de sant  au Canada. Enqu te sur la maladie : 1950-1951*. Imprimeur de la Reine et contr leur de la papeterie. Ottawa.
- Rivest P., Richer F., Bedard L. (2004). Difficult s associ es   la surveillance de la coqueluche. *Relev  de maladies transmissibles au Canada*, 30(4), 29-33, 36.

- Santé Canada. (1992-2004). *Relevé des maladies transmissibles au Canada*. Bimensuel.
- Santé Canada. (2000). Comité consultatif de l'Épidémiologie et la Division de la surveillance des maladies, Bureau des maladies infectieuses, Laboratoire de lutte contre la maladie, Direction générale de la protection de la santé. Définitions de cas de maladies faisant l'objet d'une surveillance nationale. *Relevé des maladies transmissibles au Canada*, 26(S3).
- Santé Canada. (2002). Rapport d'une réunion du comité consultatif d'experts sur la rubéole au Canada. *Relevé des maladies transmissibles au Canada*, 28 (S4).
- Santé et Bien-Être social Canada. (1975-1991). *Rapport hebdomadaire des maladies au Canada*.
- Santé et Bien-Être social Canada. (1983). Résultats de laboratoire concernant les infections virales au Canada. *Rapport hebdomadaire des maladies au Canada*, 9(34), 133-136.
- Santé et Bien-Être social Canada. (1991b). Comité consultatif de l'épidémiologie, Bureau d'épidémiologie des maladies transmissibles. Programme canadien de surveillance des maladies transmissibles. Définitions de cas et méthodes de surveillance particulières à chaque maladie. *Rapport hebdomadaire des maladies au Canada*, 17(S3).
- Sockett P.N., Garnett M.J., Scott C. (1996). Communicable disease surveillance: Notification of infectious diseases in Canada. *Canadian journal of Infectious Diseases*, 7(5), 293-295.
- Statistique Canada. (1971 à 1978). Division de la santé et du bien-être. Section de la santé publique. *Rapport annuel sur les maladies à déclaration obligatoire*. Catalogue 82-201, annuel.
- Sutter R.W., Cochi S.L. (1992). Pertussis hospitalizations and mortality in the United States, 1985-1988. Evaluation of the completeness of national reporting. *JAMA : the journal of the American Medical Association*, 267(3), 386-391.
- Sydenstricker E. (1926). The reporting of notifiable diseases in a typical small city. Hagerstown morbidity studies no.II. *Public Health Reports*, 41, 2186-2191.
- Thacker S.B., Berkelman R.L., Stroup D.F. (1989). The science of public health surveillance. *Journal of Public Health Policy*, 10(2), 187-203.
- Torvaldsen S., McIntyre P. (2003). Do variations in pertussis notifications reflect incidence or surveillance practices? A comparison of infant notification rates and hospitalisation data in NSW. *New South Wales public health bulletin*, 14(4-5), 81-84.
- Varughese P.V., Carter A.O., Ellis E. (1990). La rougeole au Canada –1988. *Rapport hebdomadaire des maladies au Canada*, 16(1), 1-6.

## ANNEXE

### Définitions de cas de rougeole, d'oreillons, de coqueluche et de rubéole faisant l'objet d'une surveillance nationale au Canada

#### ROUGEOLE

**Cas confirmé** : Présence d'une des trois conditions suivantes en l'absence d'immunisation récente avec un vaccin contenant l'antigène rougeoleux :

- 1) isolement du virus de la rougeole d'un échantillon clinique approprié; **ou**
- 2) détection d'une augmentation significative du taux d'anticorps contre le virus de la rougeole entre le sérum prélevé en phase aiguë et celui prélevé en phase de convalescence; **ou**
- 3) détection sérologique d'IgM contre le virus de la rougeole à l'aide d'une méthode recommandée.

**OU** : Manifestations cliniques\* compatibles **et** lien épidémiologique avec un cas confirmé.

**Cas probable** : Maladies cliniques\* en l'absence des épreuves de laboratoire appropriées et d'un lien épidémiologique avec un cas confirmé.

**\*Cas clinique** : Caractérisé par tous les symptômes suivants: 1) fièvre  $\geq 38,3^{\circ}$  C; 2) toux ou coryza ou conjonctivite et 3) éruption maculopapulaire généralisée d'une durée minimale de trois jours.

#### OREILLONS

**Cas confirmé** : Présence d'une des trois conditions suivantes en l'absence d'immunisation récente contre les oreillons :

- 1) isolement du virus ourlien dans un échantillon clinique approprié; **ou**
- 2) détection sérologique d'IgG spécifiques dans le sérum ou séroconversion démontrée par une épreuve sérologique standard; **ou**
- 3) sérologie positive pour l'anticorps IgM dirigé contre le virus ourlien.

**OU** : Manifestations cliniques\* compatibles **et** lien épidémiologique avec un cas confirmé.

**Cas probable** : Maladies cliniques\* en l'absence des épreuves de laboratoire appropriées et d'un lien épidémiologique avec un cas confirmé.

**\*Cas clinique** : Caractérisé par l'apparition aiguë d'un œdème unilatéral ou bilatéral douloureux, spontanément résolutif, des parotides ou d'autres glandes salivaires durant deux jours, sans autres causes apparentes.

#### COQUELUCHE

**Cas confirmé** : présence d'une des conditions suivantes :

- 1) Isolement de *bordetella pertussis* dans un échantillon clinique approprié; **ou**
- 2) Détection de *bordetella pertussis* par PCR (*Polymerase chain reaction*).

**OU** : Manifestations cliniques\* compatibles **et** lien épidémiologique avec un cas confirmé.

**Cas probable** : Toux durant au moins deux semaines en l'absence de test de laboratoire ou d'un lien épidémiologique avec un cas confirmé et une ou les deux manifestations suivantes : toux paroxystique et sifflement inspiratoire (chant du coq).

**\*Cas clinique** : Définit par la présence des symptômes suivants pour lesquels il n'y a aucune autre cause connue : 1) toux paroxystique (quintes de toux); **et** 2) sifflement inspiratoire (chant du coq); **et** 3) toux se terminant par des vomissements ou par de l'apnée.

#### RUBÉOLE

**Cas confirmé** : Présence d'une des trois conditions suivantes en l'absence d'immunisation récente contre la rubéole :

- 1) isolement du virus de la rubéole d'un échantillon clinique approprié; **ou**
- 2) détection d'une augmentation significative du taux d'anticorps IgG dans le sérum mise en évidence par une épreuve sérologique standard; **ou**
- 3) détection sérologique d'IgM spécifique de la rubéole.

**OU** : Manifestations cliniques\* compatibles **et** lien épidémiologique avec un cas confirmé

**Cas probable** : Maladie clinique\* en l'absence de tests de laboratoire appropriés ou de lien épidémiologique avec un cas confirmé.

**\*Cas clinique** : inclut l'ensemble des symptômes suivants : 1) fièvre; 2) éruption cutanée maculopapulaire généralisée; 3) arthralgie ou arthrite ou lymphadénopathie ou conjonctivite.

ARTICLE 2

## **Time series analysis of some childhood infectious diseases before and after mass vaccination in Canada: Box-Jenkins modelling**

### **Abstract**

The goal of this paper is to analyze the stochastic dynamics of childhood infectious disease time series. We present an univariate time series analysis of pertussis, mumps, measles and rubella based on Box-Jenkins or ARIMA (autoregressive, integrated, moving average) modelling. The method, which allows to model the dependent structure embedded in the time series data, has potential research applications in a study of infectious disease dynamics. Canadian chronological series of pertussis, mumps, measles and rubella, before and after mass vaccination, are analyzed to characterize the statistical structure of these diseases. Despite the fact that these infectious diseases are biologically different, it is found that they all may be represented by simple models which have basically the same statistical structure. Aside from seasonal effects, the number of new cases is given by the incidence in the previous period and by periodically recurrent random factors. It is also shown that mass vaccination does not change this stochastic dependency. We conclude that the Box-Jenkins methodology does identify the collective pattern of the dynamics beyond the specifics of the diseases at the biological individual level.

### **Key words**

Time series, childhood infectious diseases, stochastic dynamics, ARIMA modelling.

## **Introduction**

Childhood infectious diseases that induce lasting immunity, such as measles, mumps, and rubella are particularly well documented dynamically. Theoreticians have focused their attention on some of these infections because of the comparative simplicity of the processes that influence their transmission dynamics, the existence of long runs of data over many decades and also as a consequence of the remarkable periodicity of their incidences (Anderson, 1994).

Modern theoretical epidemiology has been significantly influenced by applying mathematics to the study of infectious diseases and by the development of the theoretical framework of the threshold theorem, according to which the introduction of a few infectious individuals into a community of susceptibles will not give rise to an epidemic outbreak unless the number of susceptibles is above a critical value (Kermack et al., 1927). An example of the expansion of this theorem was the development of the concept of a basic reproductive number (which defines the average number of secondary cases of infection generated by one primary case in a susceptible population). The above-mentioned assumptions allowed the development of systems of non linear equations whose solutions describe the transmission dynamics in a population. The application of these explanatory models have provided insight into the transmission dynamics such as the discovery of a critical community size for measles (Bartlett, 1960) or the critical level of mass vaccination required to block transmission within a defined community (Dietz, 1976). Essential insights into the persistence and stability of transmitted viral and bacterial infections that induce lasting immunity within large human communities was also gained by simple deterministic models (simulation) (Anderson et al., 1991). These mechanistic models, such as SEIR, predict the spread of infectious diseases, exhibit

damped oscillations and mirror well the observed trends, such as for measles, which generally exhibit biennial cycles (with a seasonal pattern) before mass vaccination. The cycles are generated by exhaustion by infection and replenishment by new births (of the supply of susceptibles), and may be perpetuated by seasonal changes in the rate of transmission, resulting from the aggregation and disaggregation of children for school terms and holiday periods. Cycles are more pronounced in infections with short infectious and latent periods such as measles.

These simple mechanistic models (especially the seasonally forced SEIR model) generate a rich array of dynamical behaviours including chaotic patterns. However, some authors have shown that the incorporation of more biological realism into the model (i.e., age structure), of realism in the seasonal forcing function, or of other spatial factors, can suppress the complex dynamic (Bolker, 1993; Bolker et al., 1993; Grenfell et al., 1995). Adding biological complexity and heterogeneity generates a buffering effect that tends to reduce or eliminate the chaotic behaviour. The variety of dynamics seen in real populations with different demographic and geographic patterns suggests caution in the construction of models of complicated systems. Moreover, the existence of multiple stable limit cycles implies model sensitivity on initial conditions (Ferguson et al., 1996). Thus, and even if these “explanatory” models are very useful to understand the dynamics, it is impossible to mimic the details of an epidemic on more than a qualitative level.

By contrast to simulation, empirical data on infectious diseases can be studied with a stochastic approach that models the dependent structure embedded in the time series. Box-Jenkins modelling, also called the ARIMA method (AutoRegressive, Integrated, Moving Average), seems promising to complement infectious disease

theories by describing the component structure (“non explanatory” analysis) of statistical time series. This method widely used in the biological sciences has rarely been applied to infectious disease modelling despite that the method seems promising for many common infections that show seasonal behaviour and random changes or trends (non-stationary time series). Helfenstein (1986) was the first to show that mumps and chicken pox time series from New York City (1928 to 1960) both have the same simple statistical structure. He concluded from his endeavour that apart from their clinical and epidemiological properties, it could be possible to classify the infectious diseases by the structure of their corresponding time series and that this might provide a better understanding of the disease. For example, he suggested that if the statistical structure of the time series for infectious diseases is different from that of non-infectious diseases then this might provide insight into diseases such as multiple sclerosis or certain forms of leukemia for which an infectious agent is suspected but not established.

ARIMA models may also be relevant to investigate the relationship between different time series, to evaluate the impact of intervention analysis or to make forecasts. A few authors have used Box-Jenkins modelling in epidemiology to forecast the mortality incidence of pneumonia and influenza in the United States (Choi et al., 1981a; 1981b) or infectious diseases in Canada (Allard, 1998). However, these authors were not really concerned with the possibility of evaluating the statistical structure of their disease time series.

In line with Helfenstein (1986)’s contention, this paper will attempt to model the time series of four childhood infectious diseases (pertussis, mumps, measles and rubella) in order to compare their specific statistical structure before and after mass vaccination in Canada. More specifically, we expect to correlate specific models with either

vaccination phase, disease dynamics, or data quality. In other words, we attempted to characterize the basic statistical dynamics of these infectious diseases to obtain a better understanding of the different disease patterns, and we explored the possibility of classifying the infections by their corresponding time series structure.

### **Data**

Data were obtained from Health Canada (Center for Infectious Diseases Prevention Control, Division of Disease Surveillance). Time series of pertussis, mumps, measles and rubella in Canada were available on a 4-weekly basis (13 observations per year) before 1990 and on a monthly basis (12 observations per year) from 1991 to 2002 except for some periods where measles and rubella (1959-1968) and mumps (1959-1986) were not declared. The 4-weekly and monthly time series were analyzed from the 1950s. In Canada, mass vaccination began in 1943 for pertussis, in 1963 for measles, and in 1969 for mumps and rubella (this could have varied for some provinces and territories). The study covered the pre- and post-vaccination era (except for pertussis where only the post-vaccination period is analyzed).

Notifications of infectious diseases on a 4-weekly basis can be found in the *Annual Report of Notifiable Diseases* (Dominion Bureau of Statistics, 1953-1970; Statistics Canada, 1971-1978). For the pre-vaccination era, 4-weekly notification data from 1953 to 1959 for measles, mumps and rubella were analyzed. In the post-vaccination period, 4-weekly notification data from 1970 to 1976 for measles and rubella were used (mumps was not declared from 1959 to 1986; it is therefore impossible to analyze data in the early post-vaccination era). Also, 4-weekly notifications of pertussis from 1953 to 1962 and from 1970 to 1976 were studied.

Finally, to cover the current period, monthly notifications of measles, rubella, pertussis were used from 1991 to 2002 and mumps from 1991 to 1998 (Health Canada, 1992-2004). The crude time series are presented in Appendix A (figure A1, A2 and A3).

Some 4-weekly time series from 1970 to 1976 had unclassified annual data (i.e., without specifying in which of the 4-weekly period of the year the cases appeared). For example, the pertussis time series had 3 unclassified data in 1973. These data were proportionally distributed with respect to the percentage of cases that appeared in each 4-weekly period. Since only a few cases were unclassified (<1% each year), we assumed that there is no effect on the dynamics. No other series analyzed (1950s and 1990s) had unclassified data. Further, monthly basis observations (1991-2002 data) do not have constant time intervals (they vary between 28 and 31 days). To control for this, we transformed the reported time series on a basis of 28 days. For example, the total number of cases that occurred in January were divided by 31 and multiplied by 28. This procedure allowed to equally space the cases over a period of 28 days and allowed to obtain time series more comparable. For time series that contained null values we added the constant 1 to the data. Monthly time series for measles (1991-2002) and rubella (1991-2002) were adjusted by adding the constant 1 to all the data. Only 12 months (out of 144) for measles and 6 (out of 138) for rubella contained null values. This procedure did not change the temporal structure of the series since the constant 1 is added to all the data and allowed us to use the logarithmic transformation.

Certain limitations must be considered about the data. They are inherent in all notification data (Giesecke, 1994). There are strong indications that not all cases come to the attention of physicians, which leads to underreporting of cases; further, the reporting of diagnosed cases may also be considered incomplete (Clarkson et al., 1985; Davis et

al., 1993; McDonnell et al., 1994; Ewert et al., 1995; Deeks et al., 1999; Allen et al., 2000; Harpaz, 2004; Trottier et al., 2004). In spite of these shortcomings, the reporting system still reflects the dynamics quite accurately (i.e., relative changes) (Dominion Bureau of Statistics, 1953-1970; Statistics Canada, 1971-1978, Clarkson et al., 1985; Brewster, 1993; Torvaldsen et al., 2003). This type of underreporting does not create a problem since our objective is to analyze the dynamics. Furthermore, the notification system in Canada has changed since 1950 (variation in the ways cases are reported, variation in case definition, consideration of laboratory confirmation, completeness of reporting) (Dominion Bureau of Statistics, 1953-1970; Statistics Canada, 1971-1978; Health and Welfare Canada, 1975-1991; Clarkson et al., 1985; Health and Welfare Canada, 1991b; Health Canada, 1992-2004; Harpaz, 2004; Trottier et al., 2004). Had these phenomena been of some extent, they would be expected to cause problems by entailing inappropriate variations in the course of time. As stated above, the time series have been broken down into short sub-periods. This limited the extent of inappropriate variations in the data.

The choice of the series was conditioned by the changes in the surveillance system. The choice of the series was also limited by the fact that 4-weekly time series were sometimes broken by a 1-week notification period. For example, if the end-of-the-year notification period began on November 26 and ended on December 23 (28 days), the data between December 24 and 30 (7 days) were notified on a 1-week basis in order to include the cases in the corresponding year. This procedure that allowed to adjust the data nevertheless caused problems by entailing non constant time intervals in the series. The series have been selected with respect to these adjustments. Firstly, data before 1950 were excluded because some provinces and territories did not report the infectious

diseases. The first time series thus covers the period from 1953 to 1959. In 1959, measles, rubella and mumps were excluded from the list of notifiable diseases in Canada. Pertussis were analyzed from 1953 to 1962. The time series were then truncated at the end of 1961 since the data were notified on a 1-week period (to adjust for the data in the corresponding year). Measles and rubella were reintroduced in 1969. The second time series was then selected from 1970 to 1976. Mumps were not notified from 1959 to 1986; it was therefore impossible to analyze the data from 1970 to 1976. In 1976, the series were truncated because the data were notified on a 1-week rather than a 4-week basis. The last series covered the period from 1991 to 2002. In 1991, the surveillance system was modified to introduce disease-specific case definitions by laboratory tests. The data were then declared on a month rather than 4-week period.

## Methods

Box-Jenkins modelling requires a series of observations, repeated at equal time intervals on the same population. Also, time series should be stationary with respect to mean and variance. The following introduction is necessarily brief, but a more extensive introduction to this stochastic approach is available in many textbooks (Box et al., 1994; McCleary et al., 1980; Abraham et al., 1983; McDowall et al., 1983).

Denote the values of a series at equally spaced time  $t, t-1, t-2, \dots$  by  $Z_t, Z_{t-1}, Z_{t-2}, \dots$ . The Box-Jenkins method distinguishes three possible processes to extract the dependent structure between successive observations:

- 1) Autoregressive process (AR):  $Z_t = a_t + \phi_1 Z_{t-1}$

Here, the current time series observation ( $Z_t$ ) is determined by a portion of the preceding observation ( $Z_{t-1}$ ), and a current random shock ( $a_t$ ) (order 1 process).

The basic model above is a special case of the following general model (an autoregressive model of order  $p$ ):

$$Z_t = a_t + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p}$$

2) Moving average process (MA):  $Z_t = a_t - \theta_1 a_{t-1}$

In this model, the current time series observation,  $Z_t$ , is composed of a current random shock ( $a_t$ ) and portion of the preceding random shock ( $a_{t-1}$ ) (order 1 process).

The simple model above is a special case of the following general model (a moving average model of order  $q$ ):

$$Z_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

3) Integration (I):  $Z_t = a_t + Z_{t-1}$

In this process,  $Z_t$  is the sum of the current random shock ( $a_t$ ) and of the preceding observation.

Empirical time series can be adjusted by any one of these models. Alternatively, more than one component model can be used to fit the data like an autoregressive moving average model (ARMA( $p,q$ ) model) such as:

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} + U \quad (\text{where } U \text{ is the model constant})$$

For adequate ARIMA modelling, the time series should be stationary with respect to mean and variance. If a series has to be differenced to stabilize the mean, then the model corresponding to the original series is called integrated. Usually, time series show secular trends or other dependencies that require regular differencing (integration), and AR or MA processes co-exist. Moreover, most of the monthly, 4-weekly or quarterly time series have a strong seasonal component (cyclical or periodic fluctuations that recur or repeat itself at the same phase of the cycle period). Indeed, some series could display seasonal non-stationarity which would require seasonal differencing. A seasonal moving average process (SMA) and/or a seasonal autoregressive process (SAR) could also co-exist. It is also possible but rare to find a complex series that embeds more than one term of the same process. Any successful modelling transforms time series into white noise (random residuals series). On the other hand, a time series that does not require modelling (no trend and no dependent structure embedded) is white noise, i.e., a series of random shocks.

One of the first steps in ARIMA modelling consists in stabilizing the variance of the series (stabilization of the mean is done by differencing). Various transformations like the square root or the logarithmic transformations can be appropriate. A practical tool for the choice of the appropriate transformation, based on the Box-Cox (1964) power transformation, is the mean-range plot (the range of the data is plotted against the mean for each seasonal period) (Helfenstein, 1986; Jenkins, 1979). If the range is independent of the mean, no transformation is required. If the plot displays random scatter about a straight line, a logarithmic transformation is suggested. Other transformations can be used if a satisfactory model is not obtained by usual transformations.

ARIMA modelling is best done while following a protocol (see Appendix B for details). The main steps required by Box-Jenkins modelling are model identification, parameter estimation and diagnosis. Identification is the key of model building. It allows to specify the model, i.e., ARIMA(p,d,q) or SARIMA(p,d,q)(P,D,Q)<sub>s</sub> model where:

p	is the order of the AR process,
q	is the order of the MA process,
d	is the order of the regular difference process,
P	is the order of the seasonal AR process,
Q	is the order of the seasonal MA process,
D	is the order of the seasonal difference process,
s	is the length of the seasonal period.

The equation for the seasonal univariate ARIMA model written using the backward shift operator (B)  $Bz_t = z_{t-1}$  is given by (see Appendix B for details):

$$\Phi_P(B^s) \phi_p(B) [\nabla^d \nabla_s^D Z_t - U] = \theta_q(B) \Theta_Q(B^s) a_t, \text{ where:}$$

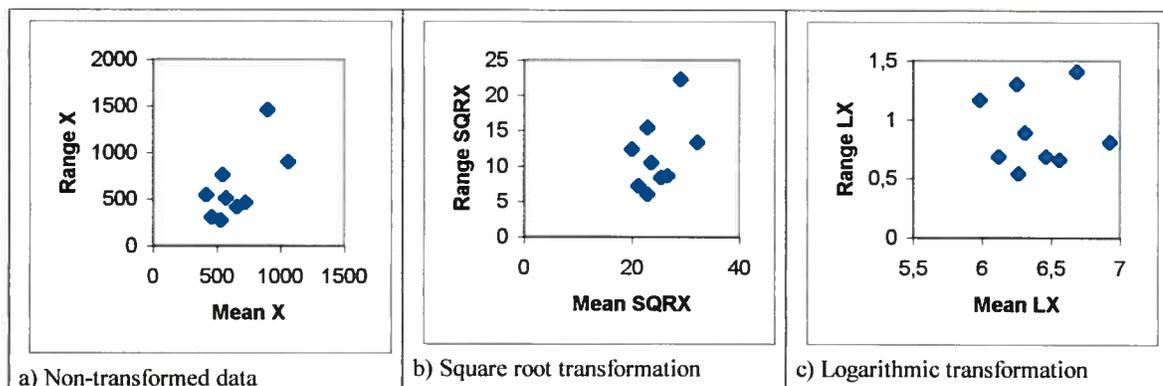
$\phi_p(B)$	is the autoregressive operator,
$\theta_q(B)$	is the moving average operator,
$\Phi_P(B^s)$	is the seasonal autoregressive operator,
$\Theta_Q(B^s)$	is the seasonal moving average operator,
$\nabla$	is the non-seasonal differencing operator $\nabla = 1 - B$ ,
$\nabla_s$	is the seasonal differencing operator with seasonality s, $\nabla_s = 1 - B^s$ ,
$a_t$	is the white noise series normally distributed with mean zero and variance $\sigma_a^2$ ,
B	is the backward shift operator ( $Bz_t = z_{t-1}$ ),
U	is the model constant.

The dependent structure of a stationary time series may be ascertained by the autocorrelation function (ACF). This function measures the correlation between  $Z_t$  and  $Z_{t+k}$ , or in other words, it is the correlation between a series and the same series lagged one, two, three... units. To help identify the process order we also use the partial autocorrelation function (PACF). The PACF is the same as ACF except that the effect of the intervening observation(s) is removed. ACF and PACF patterns are generally recognized as belonging to certain types of ARIMA(p,d,q) or SARIMA(p,d,q)(P,D,Q)<sub>s</sub> models. Minimization of either residual variance, Akaike's information criterion (AIC) (Akaike, 1969) or Schwarz's Bayesian Criterion (SBC) (Schwarz, 1978) help identify the appropriate model (see Appendix C for the meaning of these tools). Once the model has been specified, parameters of the model are estimated. This iterative process is rather straightforward. Diagnosis of the model consists of validating the model by examining its residuals. Residuals of a tentative model must be equivalent to white noise. Only random variations should be left in the residuals once main trends, dependencies, and cyclic variations have been removed from the series by modelling. ACF of the residuals and Ljung-Box statistics (global test for hypothesis of no correlation across a specified number of time lags) (Ljung et al., 1978) are useful to test the randomness of the residuals. In the case where ACF shows significant autocorrelations in the residuals, the protocol implies that one returns to the step of identification to explore the possibility of having missed an important component of the model or because of instable variance. If different models represent the data equally well, one chooses the simplest. An essential concept consists in finding a representation of the series that includes as few parameters as possible (this is the parsimony principle). SPSS version 11.5 (including Trend) was used to perform the analyses.

## Results

This section provides the results ARIMA modelling for one of the time series. This will include transformation of the series to stabilize the variance, ACF and PACF, and the whole whitening process. A complete example of the Box-Jenkins protocol (identification, estimation, diagnostic) application to the pertussis time series (1953-1962) is presented. Parameters estimates of the whole set of series are shown in table 2.

Mean-range plots were prepared for each infection time series. For example, one sees the mean-range plot (figure 1) for the pertussis time series (1953-1962). It is seen that the square root transformation does not suffice to make the range independent from the mean, whereas the logarithmic transformation appears appropriate. Mean-range plots have suggested the logarithmic transformation for every infection time series except rubella (1953-1959) where the inverse of the square root transformation would perhaps be more appropriate (not presented). Since the analyzed time series are short, it was not always clear which transformation was best; in this case we preferred to keep the same comparative base and to apply the logarithmic transformation for every time series, including the rubella case (1953-1959).



**Figure 1.** Mean-Range Plot for Pertussis (1953-1962). (a) Mean and range for non-transformed data (X). (b) Mean and range applying the square root transformation (SRX). (c) Mean and range applying the logarithmic transformation (LX).

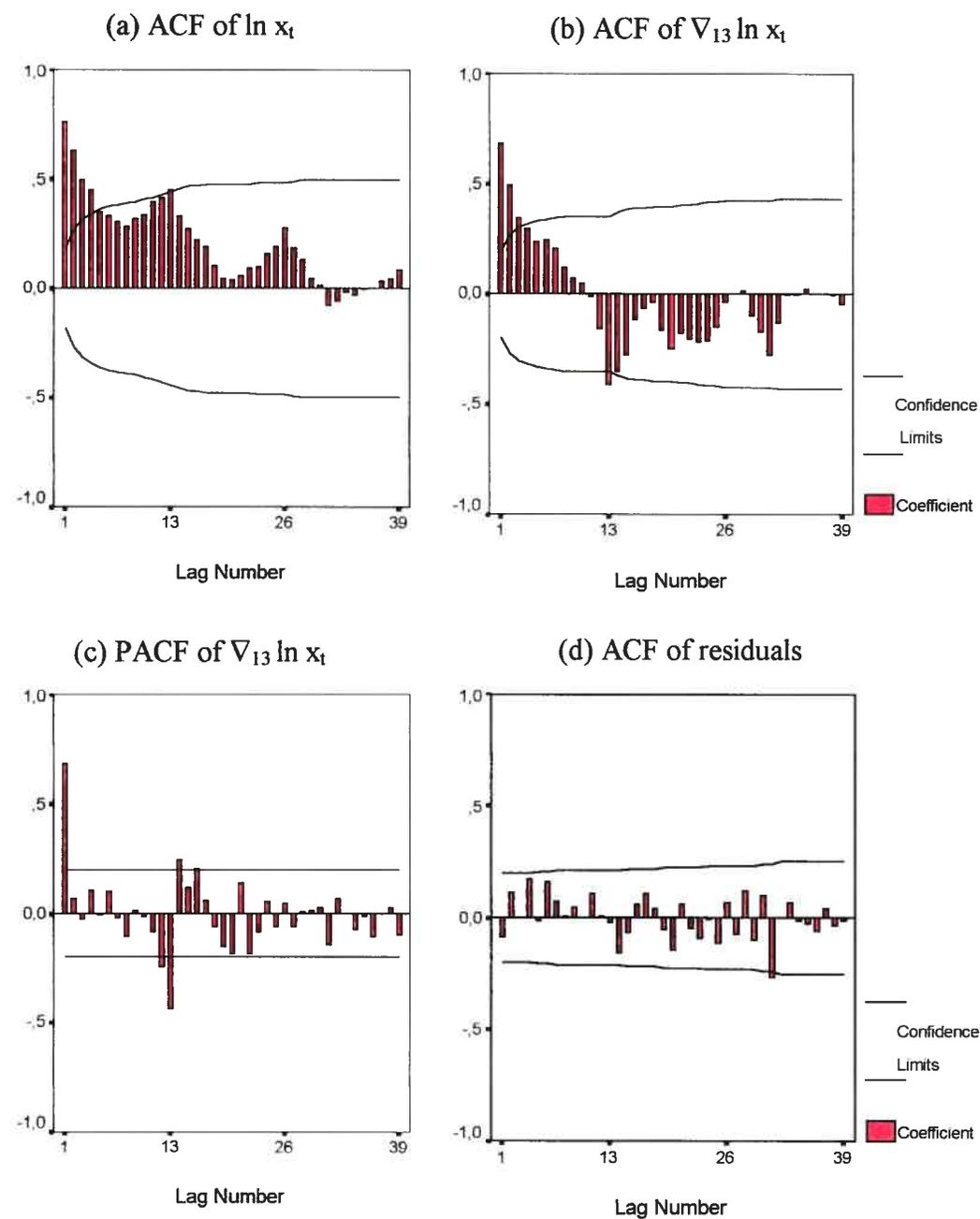
The following is a complete description of the analysis of the pertussis time series (1953-1962). The ACF of the logarithmically transformed series (figure 2a) exhibits periodicity of length 13. This was expected since this childhood disease shows a seasonal cycle. Such ACF pattern requires seasonal differencing ( $\nabla_{13} \ln x_t = \ln x_t - \ln x_{t-13}$ ) to induce seasonal stationarity. The periodicity disappears after seasonal differencing. The ACF and the PACF of  $\nabla_{13} \ln x_t$  in figures 2(b) and 2(c) allow one to identify a model with an AR (1) term and an MA (13) term. The decaying ACF and the single peak at lag 1 in PACF suggest one AR (1) term. PACF also gives indications that an MA(13) term is needed since a sole peak at lag 13 is detected. Diagnostic checking shows that the series of residuals behaves as white noise. Residuals ACF (figure 2(d)) exhibits no specific pattern and no marked peaks. This is confirmed by the Box-Ljung statistics (statistics Q), which gives lags 1, 13 and 26 P-values of 0.366, 0.643, and 0.579 respectively. It is therefore concluded that the residuals have no remaining structure and that a SARIMA model  $(1,0,0)(0,1,1)_{13}$  best fits the data series. The estimated parameters and the residuals autocorrelations of each lag (with Ljung-Box statistics) of the pertussis time series (1953-1962) are presented in table 1a and 1b.

The following model has completely characterized the pertussis (1953-1962) time series:

$$(1-\phi B)\nabla_{13} \ln x_t = U + (1-\Theta B^{13}) a_t$$

with parameter estimates (and standard errors) and residual variance:

$$\hat{\phi} = 0.65 (0.07), \hat{\Theta} = 0.84 (0.17), \hat{U} = -0.09 (0.02), \hat{\sigma}_a^2 = 0.047.$$



**Figure 2.** Pertussis (1953-1962): (a) autocorrelation function (ACF) of  $\ln x_t$ ; (b) autocorrelation function (ACF) of  $\nabla_{13} \ln x_t$ ; (c) partial autocorrelation function (PACF) of  $\nabla_{13} \ln x_t$ ; (d) autocorrelation function of the residuals.

Number of residuals	104			
Standard error	0.2159537			
Log likelihood	4,8975731			
AIC	-3,7951462			
SBC	4.1380265			
Analysis of Variance:				
	DF	Adj. Sum of Squares	Residual Variance	
Residuals	101	5,4967722	0.04663602	
Variables in the Model:				
	B	SEB	T-RATIO	APPROX. PROB.
AR1	0.65236437	0,07159567	9,1117856	0.00000000
SMA1	0.84313398	0,17157796	4.9139994	0.00000345
CONSTANT	-0,08709594	0,02178515	-3.9979503	0.00012172

**Table 1a.** Estimates of parameters for the pertussis series (1953-1962). Arima output SPSS Trend 11.5

Lag	Corr.	Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	-,087	,098					**					,816	,366
2	,112	,099						**				2,167	,338
3	-,002	,100					*					2,167	,538
4	,175	,100						****				5,546	,236
5	-,016	,103					*					5,575	,350
6	,158	,103						***				8,380	,212
7	,073	,105					*					8,980	,254
8	,004	,106					*					8,982	,344
9	,045	,106					*					9,212	,418
10	,000	,106					*					9,212	,512
11	,106	,106						**				10,550	,482
12	,009	,107					*					10,559	,567
13	-,022	,107					*					10,615	,643
14	-,160	,107						***				13,737	,469
15	-,069	,109					*					14,331	,501
16	,059	,110					*					14,762	,542
17	,104	,110						**				16,145	,514
18	,041	,111					*					16,361	,567
19	-,056	,111					*					16,763	,606
20	-,144	,111						***				19,486	,490
21	,058	,113					*					19,938	,525
22	-,048	,113					*					20,246	,568
23	-,092	,114					**					21,401	,557
24	-,005	,114					*					21,404	,615
25	-,116	,114					**					23,271	,562
26	,069	,115					*					23,945	,579
27	-,076	,116					**					24,770	,587
28	,121	,116						**				26,880	,525
29	-,099	,117					**					28,312	,501
30	,096	,118					*	**				29,688	,482
31	-,264	,119					*****					40,189	,125
32	,000	,125					*					40,189	,152
33	,067	,125					*					40,896	,162
34	-,016	,125					*					40,935	,192
35	-,025	,125					*					41,031	,223
36	-,058	,125					*					41,571	,241
37	,039	,125					*					41,822	,269
38	-,036	,125					*					42,035	,300
39	-,010	,125					*					42,053	,340

Plot Symbols: Autocorrelations \* Two Standard Error Limits .  
Standard errors are based on the Bartlett (MA) approximation.

**Table 1b.** Diagnosis checking for the pertussis series (1953-1962). Residual ACF and Box-Ljung statistics for each lag. Arima output SPSS Trend 11.5

The steps described for pertussis (1953-1962) analysis are not repeated here for the other time series since they lead to the same type of model (the complete description of the analysis of each time series can however be found in the thesis II and III appendices). The rubella time series (1953-1959) and the measles time series (1970-1976) do not follow this model consistently, however. The major epidemic of rubella in 1956 (Figure A1; Appendix A) introduces significant variance in the time series. ACF of the seasonally differenced time series of rubella (1953-1959) suggests that it might be better to also apply regular differencing. The model that best fits the regularly and seasonally differenced series has an AR (1) and an MA (13) term. Regular and seasonal differencing were also applied to the measles time series (1970-1976) which exhibits a significant epidemic peak in 1970. A model with an AR(1) term and an MA (13) term was also found. A better model was searched for but the residuals proved to not comply with white noise. Since it was not clear that regular differencing was appropriate, these time series should be interpreted with caution.

A summary of the results is presented in table 2. For each time series, we provide  $N$ , the number of observations, the estimated parameters with their corresponding estimated standard error (in parenthesis), the estimated residual variance ( $\hat{\sigma}_a^2$ ), the variance of the transformed and differenced initial observed time series ( $\hat{\sigma}_w^2$ ), the coefficient of determination ( $R^2$ ), the Akaike's information criterion (AIC), the Schwartz Bayesian Criterion (SBC) and the statistic  $Q$  (Box-Ljung statistics for lag 1, 13 and 26 for the 4-weekly notifications and lag 1, 12 and 24 for the monthly notifications) (Appendix C). The comparison of  $\hat{\sigma}_a^2$  and  $\hat{\sigma}_w^2$  ( $R^2$ ) indicates that except for mumps

(1991-1998), the models explain between 45 to 89% of the variance of the differenced series. This is quite respectable in a time series context.

**Table 2 : Estimates of parameters (and standard errors)**

**a) Pertussis**

Period	N	AR ( $\phi$ )	SMA ( $\Theta$ )	Constant	$\hat{\sigma}_w^2$	$\hat{\sigma}_a^2$	$R^2$	SBC	AIC	Box-Ljung (stat Q)		
										Lag 1	Lag 12 / 13	Lag 24 / 26
1953-1962 <sup>1</sup>	117	0,65* (0,07)	0,84* (0,17)	-0,08* (0,02)	0,144	0,046	0,68	4,1	-3,8	0,816 p=0,37	10,615 p=0,64	23,945 p=0,58
1970-1976 <sup>1</sup>	78	0,91* (0,05)	0,86* (0,44)	0,10 (0,14)	0,487	0,082	0,83	49,6	43,1	0,164 p=0,69	17,689 p=0,17	29,532 p=0,29
1991-2002 <sup>2</sup>	132	0,93* (0,03)	0,57* (0,11)	-0,05 (0,13)	0,326	0,045	0,86	-12,7	-21,0	0,316 p=0,57	4,173 p=0,98	12,244 p=0,99

**b) Mumps**

Period	N	AR ( $\phi$ )	SMA ( $\Theta$ )	Constant	$\hat{\sigma}_w^2$	$\hat{\sigma}_a^2$	$R^2$	SBC	AIC	Box-Ljung (stat Q)		
										Lag 1	Lag 12 / 13	Lag 24 / 26
1953-1959 <sup>1</sup>	78	0,79* (0,06)	0,91 (0,59)	-0,16* (0,04)	0,114	0,024	0,79	-28,3	-34,8	1,072 p=0,30	8,117 p=0,84	15,718 p=0,94
1991-1998 <sup>2</sup>	84	0,47* (0,11)	0,54* (0,22)	-0,08 (0,05)	0,167	0,133	0,20	73,5	66,7	0,005 p=0,95	14,719 p=0,26	23,017 p=0,52

**c) Measles**

Period	N	AR ( $\phi$ )	SMA ( $\Theta$ )	Constant	$\hat{\sigma}_w^2$	$\hat{\sigma}_a^2$	$R^2$	SBC	AIC	Box-Ljung (stat Q)		
										Lag 1	Lag 12 / 13	Lag 24 / 26
1953-1959 <sup>1</sup>	78	0,88* (0,06)	0,78* (0,24)	-0,04 (0,09)	0,320	0,042	0,87	1,7	-5,35	0,001 p=0,98	11,468 p=0,57	16,442 p=0,93
1970-1976 <sup>1,3</sup>	78	-0,28* (0,11)	0,90 (0,55)	0,01 (0,01)	0,129	0,069	0,45	39,7	33,2	0,139 p=0,71	17,053 p=0,20	32,449 p=0,18
1991-2002 <sup>2</sup>	132	0,78* (0,05)	0,90* (0,17)	-0,41* (0,10)	2,833	0,696	0,75	329,4	321,0	0,159 p=0,69	8,940 p=0,71	17,761 p=0,81

**d) Rubella**

Period	N	AR ( $\phi$ )	SMA ( $\Theta$ )	Constant	$\hat{\sigma}_w^2$	$\hat{\sigma}_a^2$	$R^2$	SBC	AIC	Box-Ljung (stat Q)		
										Lag 1	Lag 12 / 13	Lag 24 / 26
1953-1959 <sup>1,3</sup>	78	0,25* (0,11)	0,95 (1,28)	-0,02 (0,02)	0,153	0,075	0,51	48,3	41,8	0,091 p=0,76	12,150 p=0,52	20,771 p=0,75
1970-1976 <sup>1</sup>	78	0,93* (0,04)	0,81* (0,26)	-0,07 (0,18)	0,676	0,071	0,89	36,9	30,5	0,267 p=0,61	12,385 p=0,50	20,826 p=0,75
1991-2002 <sup>2</sup>	126	0,82* (0,05)	0,94* (0,34)	-0,34 (0,10)	2,795	0,492	0,82	279,0	270,8	0,812 p=0,37	8,316 p=0,76	26,461 p=0,33

\* p<0.05

<sup>1</sup>4-weekly notifications

<sup>2</sup>Monthly notifications

<sup>3</sup>The rubella (1953-1959) and the measles (1970-1976) time series do not follow the SARIMA model (1,0,0)(0,1,1)<sub>13</sub> consistently. Regular differencing was also applied but it was not clear that it was required. The results for these time series should be analysed with caution.

## Discussion

The results show that time series of pertussis, mumps, measles and rubella have the same stochastic dependence in their consecutive data. Despite the fact that these childhood infectious diseases are biologically different, the number of new cases in one period is given by the number of cases in the previous period and by periodically recurrent random shocks. An AR (1) term assumes that each observation is directly dependent (autocorrelated) on the past observation (incidence in April is, for example, dependent of incidence in March). This process shows that the dynamics of infectious diseases are not only related by random or exogenous factors but also by factors that keep endogenous memory of the past. An MA (12 or 13) term process assumes that each value is dependent upon the current and past random shocks (random factors). Dependence on the previous-year random shock would mean that random factors have a carryover effect on this-year infectious diseases occurrence. Periodically recurrent random factors act in the same way over incidence variations. Temperature would be an example of a random factor that has this effect on infectious diseases.

Box-Jenkins modelling is based only on the mathematical properties of the series and not on the transmission dynamics of infectious diseases as alluded to in the Introduction. Describing the dependent structure of the time series does not necessarily allow to explain the pattern. But it is nevertheless possible that some factors account for the emergence of these processes. The potential for an epidemic lies not only with the mean biological parameters of a disease, such as transmission risk or the duration of the infectivity or latent period, but just as much as with the way society is organized, how often we travel, the size of our family, division of the school year, density of the population, etc. In the case of measles, for example, it is well known that the

transmission potential is higher in autumn with school entry (Anderson et al., 1991). Seasonal patterns of births are also well documented (Huntington, 1938; Lam et al., 1991a; 1991b; Kallas, 1994). Moreover, existence of different viral strains magnifies the problem to elucidate the complex dynamics of infectious diseases. Future research, particularly in molecular epidemiology, could help to understand these dependencies. For example, molecular typing of rubella isolates in the Americas during the 1990s identified three statistically distinct genotypic groups (Reef et al., 2002). Phylogenetic analysis has shown that dissimilar viruses from all three genotypic groups were involved in outbreaks at different times.

It appears that childhood infections share some basic factors and that mass vaccination does not seem to have changed the stochastic dependency. In Canada, the pertussis vaccine was introduced in 1943, the measles vaccine in 1963, and the mumps and rubella vaccines in 1969. Unfortunately, national estimates of childhood vaccination coverage were not collected in a standardized manner or on a regular basis prior to 1994. In 1994 and 1998, annual surveys were conducted by the Division of Immunization, LCDC, to obtain population-based estimates of vaccination coverage of 2-year-olds in Canada (LCDC, 1998). Results of the 1998 survey show stable coverage levels compared to the 1994 previous surveys. The coverage levels for one dose of measles (96.0%), mumps (95.4%) and rubella (95.3%) vaccines are the highest for any of the routinely recommended vaccines and the closest to national targets for 97% coverage by the second birthday. Coverage levels for pertussis in the 1994 and 1998 surveys were reported as approximating 80% by age 2. The Canadian population of the 1990s is therefore considered as a highly vaccinated population. Notwithstanding the large proportion of susceptibles vaccinated, vaccine failure can occur so that a proportion

remains susceptible. Higher immunity is reached for viral infection (measles, rubella, mumps). Thus measles vaccine efficacy varies between 85 to 95% with a single dose of the vaccine given at 12 or 15 months and may reach 100% with a second dose (De Serres et al., 1995; Vitek et al., 1999; Health Canada, 2002b; Redd et al., 2004). A single dose of the mumps and rubella vaccines produces an antibody response in over 95% and 97% of susceptible individuals, respectively (Health Canada, 2002b; Redd et al., 2004). For pertussis (bacterial infection), immunity might be lower. Efficacy of the pertussis vaccines was estimated to be no more than 85% and will not induce lasting immunity (Onorato et al., 1992; Bentsi-Enchill et al., 1997, Halperin et al. 2000; Health Canada, 2002b). We therefore conclude that vaccination coverage and vaccine efficacy are relatively high for measles, rubella and mumps, but could be less for pertussis.

Vaccination is the crucial public-health intervention that limits epidemics by reducing the net case reproductive number ( $R$ ) (number of transmissions / number of infected sources). The higher  $R$  is, the more explosive the disease is. If  $R$  reduces to 1, the disease turns from an epidemic to an endemic state. On the other hand, if  $R$  gets less than 1, the disease eventually disappears from the population. Among other interventions that have a bearing upon epidemics (to lower  $R$ ) is isolating infected cases from susceptibles during the infectivity period, or administering antibiotics, which can reduce the infectivity period of some diseases. Notwithstanding, our analyses show that changes in these parameters do not affect the general statistical pattern of the epidemic.

Beyond this crude feature (same statistical structure), the infection series differ in the magnitude of their AR and seasonal MA coefficients. However, interpretation of these coefficients has to be made by keeping in mind that many biases inherent in the notification data may affect the results (Giesecke, 1994; Trottier et al., 2004). Pertussis

in the 1950s obtained a lower AR (0.65) than in the 1970s and 1990s (around 0.90). That means that both series after 1970 are closer to the limit of non-stationarity and more dependent on preceding incidence. Comparing the seasonally differenced ( $\nabla_{13} \ln x_t$ ) series of pertussis for the three periods (figure A4; Appendix A), one sees that pertussis, in the 1970s and 1990s, shows more pronounced slow fluctuations than in the 1950s. Moreover, a greater impact of periodically recurrent random factors seems to explain the dynamics of pertussis in the 1950s and 1970s. The seasonal MA process changes from about 0.85 in the 1950s and 1970s to 0.57 in the 1990s. It seems that variability in the 1950s and 1970s is more random or subject to a few more random events. Finally, the deterministic linear trend in the 1950s, which is shown by the constant ( $p < 0.05$ ) of the model, could be explained by the mass immunization (started in 1943 for pertussis).

Results for the mumps series should be analyzed with caution. For example, in some provinces, such as Quebec, most of the cases recorded ( $> 90\%$ ) are clinical cases, which cause problems in terms of disease specificity (De Serres et al., 1997). Given today's high vaccination coverage and the low incidence of mumps, the positive predictive value of a diagnosis based solely on the presence of parotitis is probably low. Therefore, the reported number of clinically diagnosed cases probably over-reports true mumps occurrence (Gaulin et al., 1997). The low positive predictive value is due to the low prevalence of mumps in an immunized population and to the fact that other viruses, including enteroviruses and influenza viruses, may also cause parotitis. Any overestimation of the number of mumps cases on the basis of the surveillance definition may create problems for modelling, specifically for the 1990s. The AR process for mumps was expected to be higher in the 1950s (AR=0.79) and the 1990s (AR=0.47).

The New York mumps time series (from 1928 to 1960) exhibits  $AR = 0.93$  and  $MA(12) = 0.87$  (Helfenstein, 1986). Despite that, a SARIMA  $(1,0,0)(0,1,1)_{12}$  model best fits the data, but the estimated parameters must be interpreted with caution.

The estimated parameters of the measles series exhibit relatively constant results. However, the regular trend of the series (1970-1976) (figure A4: Appendix A) has been removed by regular differencing. Replacement of the AR operator by a differencing operator  $\nabla = 1-B$  (which corresponds to setting  $\phi = 1$ ) leads to the model presented in table 2. A significant epidemic peak of measles, which occurred in 1970 (figure A2; Appendix A), could explain part of the non-stationarity. From 1990 to 2002, the measles time series shows a significant deterministic linear trend that could be induced by the effort to eliminate the disease.

The estimated parameters of the rubella series also exhibit relatively constant results. However, from 1953 to 1959, the rubella time series exhibits both regular and seasonal non-stationarity. The estimated parameters are thus based on both seasonally and regularly differenced series. But they should be analyzed with caution since a significant variance is introduced in the time series by the major epidemic of rubella in 1956 (figure A1; Appendix A). Finally, from 1991 to 2002, the seasonally differenced series of rubella (and measles) exhibit pronounced fluctuations (figure A4; Appendix A). Childhood infectious disease time series generally exhibit highly significant seasonal and long term cycles in the pre-vaccination era. Mass immunization can reduce the relative importance of the cycles (Anderson et al., 1991). Seasonal oscillations of measles and rubella are less evident in the 1990s and seasonal differencing may be less appropriate for the series.

## **Conclusion**

Despite the fact that these infectious diseases are biologically different, it was found that they may be represented by simple models which have basically the same statistical structure. Basically, the number of new cases in one month is given by the number of cases in the previous month and by periodically recurrent random factors. Further, all series displayed seasonal drift that was taken care of by seasonal differencing. It was also shown that mass vaccination does not change this structural dependency. Even though mass vaccination was expected to have a major impact on the transmission dynamics (i.e., disease incidence, average age at infection, long-term periodicity, seasonal cycles), it does not seem to affect the stochastic dynamics. We therefore conclude that Box-Jenkins stochastic modelling is most robust to factors that might have been expected to entail significant effects. It follows that our initial objective of attempting to classify the series according to their dynamics has failed. It indeed came as a surprise that all the series proved to be similar and independent of the particular circumstances characterizing the dynamics of these infections. Notwithstanding, it is highly significant that our models did not differ from those obtained by Helfenstein (1986) despite our dealing with much shorter series and different diseases. We conclude two things from the above: (1) that the Box-Jenkins methodology is not sensitive enough to rescue the finely-tuned biological dynamics of the various types of infections; we therefore expect that future analyses of most common childhood infections will behave similarly; (2) that the Box-Jenkins methodology has rather been sensitive to the generic organizational principles of the diseases population dynamics, i.e., it identified the collective pattern of the dynamics beyond the specifics of the diseases at the biological individual level. The next paper will deal with these aspects more in depth.

**Acknowledgment**

The authors wish to thank Carole Scott from Health Canada (Center for Infectious Disease Prevention and Control / Division of Disease Surveillance) for providing the infectious diseases time series and Social Sciences and Humanities Research Council of Canada for financial support to HT.

## References

Abraham B., Ledolter J. (1983). *Statistical methods for forecasting*. New York, John Wiley and Sons, 281-321.

Akaike H. (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics*, 21, 243-247.

Allard R. (1998). Use of time-series analysis in infectious disease surveillance. *Bulletin of the World Health Organization*, 76(4), 327-333.

Allen C.J., Ferson M.J. (2000). Notification of infectious diseases by general practitioners: a quantitative and qualitative study. *The Medical journal of Australia*, 172(7), 325-328.

Anderson R.M., May R.M. (1991). *Infectious disease of human. Dynamics and control*. New-York, Oxford University Press, 757p.

Anderson R.M. (1994). The Croonian lecture, 1994. Populations, infectious disease and immunity: a very non linear world. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, 346, 457-505.

Bartlett M.S. (1960). The critical community size for measles in the united states. *Journal of the Royal Statistical Society. Series A (general)*, 123(1), 37-44.

Bentsi-Enchill A.D., Halperin S.A., Scott J., MacIsaac K., Duclos P. (1997). Estimates of the effectiveness of a whole-cell pertussis vaccine from an outbreak in an immunized population. *Vaccine*, 15(3), 301-306.

Bolker B. (1993). Chaos and complexity in measles models: a comparative numerical study. *IMA Journal of Mathematics Applied in Medicine and Biology*, 10(2), 83-95.

Bolker B.M., Grenfell B.T. (1993). Chaos and biological complexity in measles dynamics. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 251(1330), 75-81.

Box G.E.P., Cox D.R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B*, 26, 211-243.

Box G.E.P., Jenkins G.M., Reinsel G.C. (1994). *Time series analysis: forecasting and control*. Third edition. New Jersey, Prentice Hall, 598p.

Brewster D. (1993). Surveillance of measles in Scotland 1983-1991; a comparison of statutory notification and laboratory reporting. *Journal of Infection*, 27(2), 193-201.

Choi K., Thacker S.B. (1981a). An evaluation of influenza mortality surveillance, 1962-1979. I. Time series forecasts of expected pneumonia and influenza death. *American Journal of Epidemiology*, 113, 215-226.

Choi K., Thacker S.B. (1981b). An evaluation of influenza mortality surveillance 1962-1979. II. Percentage of pneumonia and influenza deaths as an indicator of influenza activity. *American Journal of Epidemiology*, 113, 227-235.

Clarkson J.A., Fine P.E. (1985). The efficiency of measles and pertussis notification in England and Wales. *International journal of epidemiology*, 14(1), 153-168.

Davis S.F., Strebel P.M., Atkinson W.L., Markowitz L.E., Sutter R.W., Scanlon K.S., Friedman S., Hadler S.C. (1993). Reporting efficiency during a measles outbreak in New York City, 1991. *American journal of public health*, 83(7), 1011-1015.

Deeks S., De Serres G., Boulianne N., Duval B., Rochette L., Dery P., Halperin S. (1999). Failure of physicians to consider the diagnosis of pertussis in children. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 28(4), 840-846.

De Serres G., Boulianne N., Bussi eres N., Pouliot B., Marin-Lira A. (1997). Epidemiology of mumps in Quebec, 1970-1995. *Canada Communicable Disease Report*, 23(2), 9-13.

De Serres G., Boulianne N., Meyer F., Ward B.J. (1995). Measles vaccine efficacy during an outbreak in a highly vaccinated population: incremental increase in protection with age at vaccination up to 18 months. *Epidemiology and infection*, 115(2), 315-323.

Dietz K. (1976). The incidence of infectious diseases under the influence of seasonal fluctuations. Lecture notes in *Biomathematics*, 11, 1-15.

Dominion Bureau of Statistics. (1953-1970). Health and Welfare Division, Public Health Section. *Annual Report of Notifiable Diseases*. Catalogue number 82-201.

Ewert D.P., Westman S., Frederick P.D., Waterman S.H. (1995). Measles reporting completeness during a community-wide epidemic in inner-city Los Angeles. *Public health reports*, 110(2), 161-165.

Ferguson N.M., Nokes D.J., Anderson R.M. (1996). Dynamical complexity in age-structured models of the transmission of the measles virus: epidemiological implications at high levels of vaccine uptake. *Mathematical Biosciences*, 138(2), 101-130.

Gaulin C., De Serres G. (1997). Need for a specific definition of mumps in a highly immunized population. *Canada Communicable Disease Report*, 23(2), 14-16.

Giesecke J. (1994). *Modern infectious disease epidemiology*. London, Edward Arnold, 256p.

Grenfell B.T., Kleczkowski A., Gilligan C.A., Bolker B.M. (1995). Spatial heterogeneity, nonlinear dynamics and chaos in infectious diseases. *Statistical Methods in Medical Research*, 4(2), 160-183.

Halperin S.A., Smith B., Russell M., Hasselback P., Guasparini R., Skowronski D., Meekison W., Parker R., Lavigne P., Barreto L. (2000). An adult formulation of a five-component acellular pertussis vaccine combined with diphtheria and tetanus toxoids is safe and immunogenic in adolescents and adults. *Vaccine*, 18(14), 1312-1319.

Harpaz R. (2004). Completeness of measles case reporting: review of estimates for the United States. *The Journal of infectious diseases*, 189 (Suppl.1), S185-S190.

Health and Welfare Canada. (1975-1991). Bureau of Epidemiology, Laboratory Centre for Disease Control. *Canada Diseases Weekly Report*. ISBN 0382-232X

Health and Welfare Canada. (1991b). Canadian Communicable Disease Surveillance System. Diseases-Specific Case definitions and surveillance methods. *Canada Diseases Weekly Report*, 17(S3).

Health Canada. (1992-2004). Laboratory Centre for Disease Control. *Canada Communicable Disease Report*. ISBN 1188-4169.

Health Canada. (2002b). *Canadian Immunization Guide*. Sixth edition. 292p.

Helfenstein U. (1986). Box-Jenkins modelling of some viral infectious diseases. *Statistics in Medicine*, 5(1), 37-47.

Huntington E. (1938). *Season of birth. Its relation to human abilities*. New-York, John Wiley and Sons.

Jenkins G.M. (1979). *Practical experiments with modelling and forecasting time series*. St. Helier, Gwilym Jenkins and partners (Overseas) Ltd.

Kallas G. (1994). *Variations saisonnières des naissances au Québec*. Mémoire, Université de Montréal.

Kermack W.O., McKendrick A.G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A*, 115, 700-721.

Lam D.A., Miron J.A. (1991a). Temperature and seasonality of birth. *Advances in Experimental Medicine and Biology*, 286, 73-88.

Lam D.A., Miron J.A. (1991b). Seasonality of births. *Social Biology*, 38, 51-78.

LCDC. (1998). Canadian National Report on Immunization. *Paediatrics and Child Health*, 4 (Suppl C).

Ljung G.M., Box G.E.P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65, 297-303.

McCleary R., Hay R.A. (1980). *Applied time series analysis for the social sciences*. Beverly Hills, Sage Publications, 331p.

McDonnell I.F., Jorm L.R. (1994). Measles epidemic in western Sydney. *New South Wales Public Health Bulletin*, 5(6), 59-60.

McDowall D., McCleary R., Meidinger E.E., Hay R.A. (1983). *Interrupted time series analysis*. Third printing, Sage Publication, 93p.

Onorato I.M., Wassilak S.G., Meade B. (1992). Efficacy of whole-cell pertussis vaccine in preschool children in the United States. *JAMA : the journal of the American Medical Association*, 267(20), 2745-2749.

Redd S.C., King G.E., Heath J.L., Forghani B., Bellini W.J., Markowitz L.E. (2004). Comparison of vaccination with measles-mumps-rubella vaccine at 9, 12, and 15 months of age. *The Journal of infectious diseases*, 189 (Suppl.1), S116-122.

Reef S.E., Frey T.K., Theall K., Abernathy E., Burnett C.L., Icenogle J., McCauley M.M., Wharton M. (2002). The changing epidemiology of rubella in the 1990s: on the verge of elimination and new challenges for control and prevention. *JAMA : the Journal of the American Medical Association*, 287(4), 464-472.

Schwarz G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461-464.

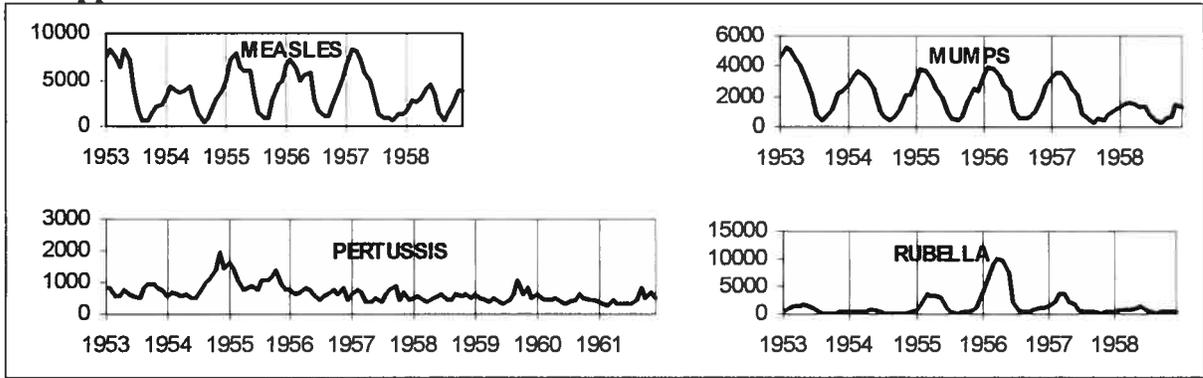
Statistics Canada. (1971-1978). Health and Welfare Division, Public Health Section. *Annual Report of Notifiable Diseases*. Catalogue number 82-201.

Torvaldsen S., McIntyre P. (2003). Do variations in pertussis notifications reflect incidence or surveillance practices? A comparison of infant notification rates and hospitalisation data in NSW. *New South Wales public health bulletin*, 14(4-5), 81-84.

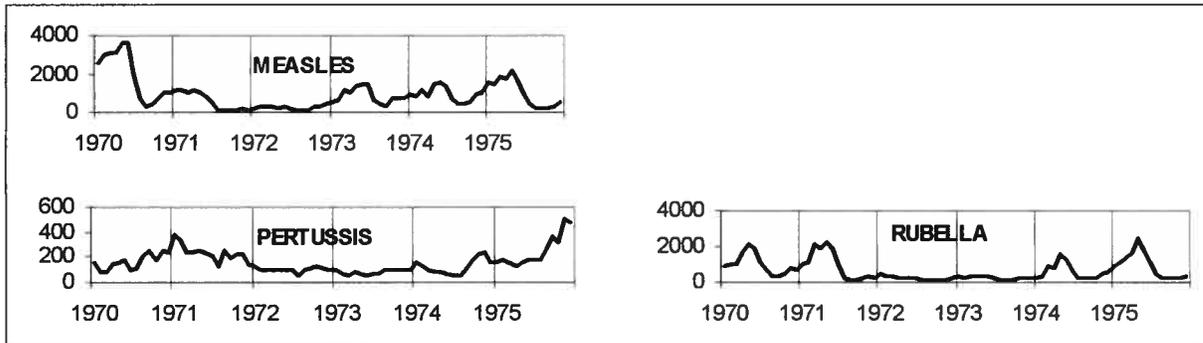
Trottier H., Philippe P. (2004). *Qualité des données de surveillance de la coqueluche, de la rougeole, de la rubéole et des oreillons au Canada*. (In preparation).

Vitek C.R., Aduddell M., Brinton M.J., Hoffman R.E., Redd S.C. (1999). Increased protections during a measles outbreak of children previously vaccinated with a second dose of measles-mumps-rubella vaccine. *The Pediatric infectious disease journal*, 18(7), 620-623.

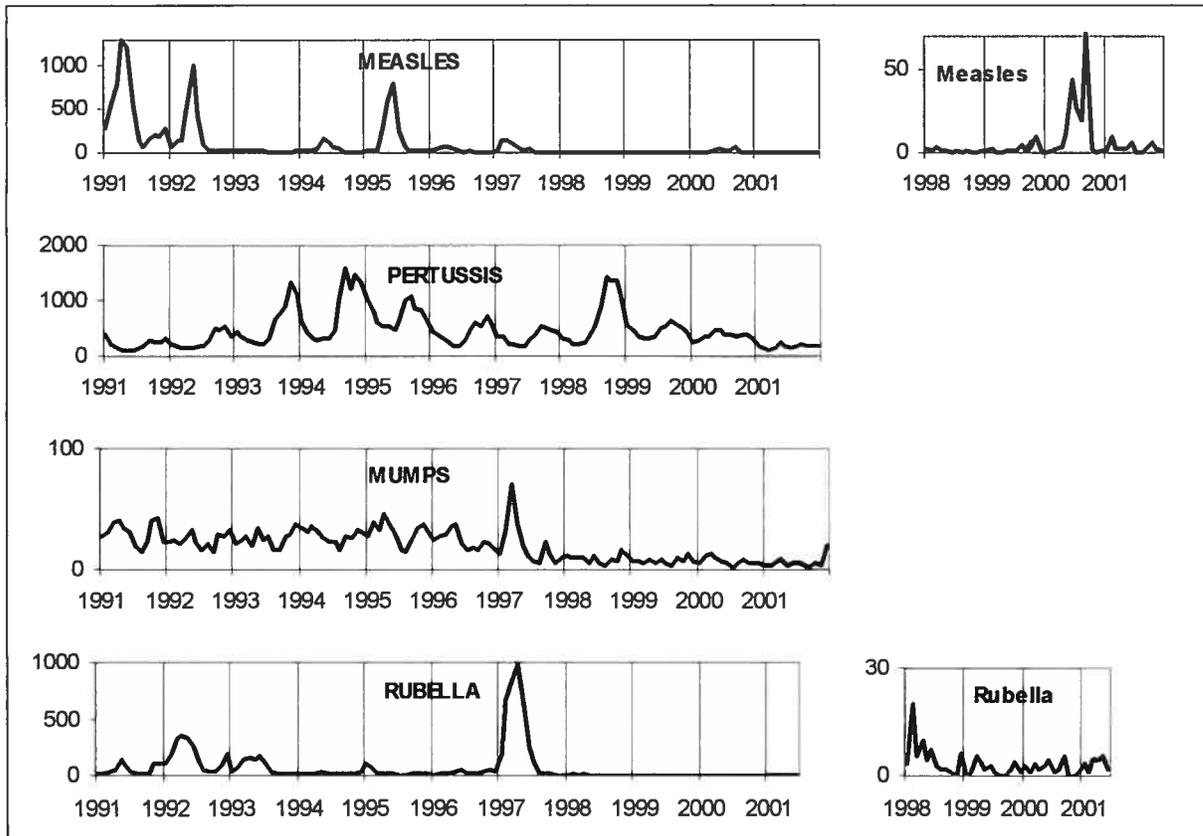
**Appendix A : Time series data**



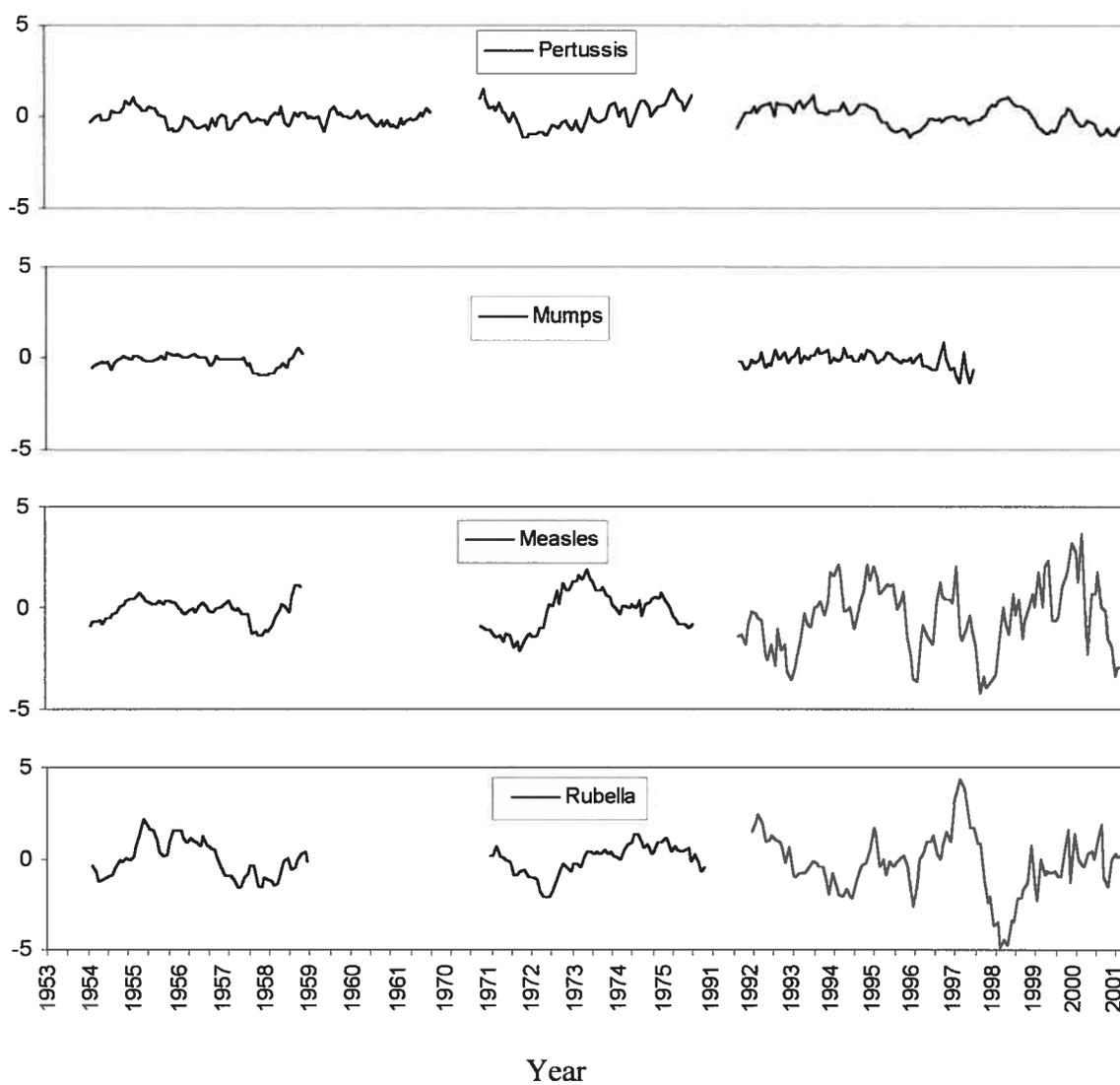
**Figure A1.** 4-weekly notifications of measles, mumps, rubella (1953-1959) and pertussis (1953-1962) in Canada.



**Figure A2.** 4-weekly notifications of measles, rubella and pertussis (1970-1976) in Canada. Adjusted for annual non-stated data (if for a given year, 10% of the cases occurred in period I, then 10% of the annual non-stated data were replaced in period I)



**Figure A3.** Monthly notifications of measles, mumps, rubella and pertussis (1991-2002) in Canada adjusted on a 28-day basis.



**Figure A4.** Seasonal differenced and logarithm transformed time series of 4-weekly (1953-1959, 1970-1976) and monthly (1991-2002) notifications of pertussis, mumps, measles and rubella in Canada.

## Appendix B: Box-Jenkins (ARIMA) modelling

Beginning with an introduction to Box-Jenkins modelling, this Appendix is structured to lead the reader through the logic of time series analysis. The ARIMA protocol is also presented. The objective of ARIMA is to model the dependent structure embedded in the time series data. Denote the values of a series at equally spaced time  $t, t-1, t-2, \dots$  by  $Z_t, Z_{t-1}, Z_{t-2}, \dots$ . This method distinguishes three possible processes to extract the dependent structure between successive observations:

- 1) Autoregressive processes (AR):  $Z_t = a_t + \phi_1 Z_{t-1}$   
 General model (autoregressive model of order  $p$ ):  

$$Z_t = a_t + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p}$$
- 2) Moving Average processes (MA):  $Z_t = a_t - \theta_1 a_{t-1}$ , or  
 General model (moving average model of order  $q$ )  

$$Z_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$
- 3) Integration (I):  $Z_t = a_t + Z_{t-1}$ .

Empirical time series can be adjusted by any one of these models. Alternatively, more than one model can be used to fit the data like an autoregressive moving average model (ARMA( $p, q$ ) model):

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}.$$

Equation 1) and 2) are usually written more concisely using the backward shift operator "B" such that  $Bz_t = z_{t-1}$ :

$$\begin{aligned} \text{AR (1) model: } (1 - \phi B) Z_t = a_t \quad \text{and} \quad \text{AR (p) model: } (1 - \phi_1 B - \dots - \phi_p B^p) Z_t = a_t \\ \text{MA (1) model: } Z_t = (1 - \theta B) a_t \quad \text{and} \quad \text{MA (q) model: } Z_t = (1 - \theta_1 B - \dots - \theta_q B^q) a_t \end{aligned}$$

Combining the AR ( $p$ ) model and the MA ( $q$ ) model, we obtain the ARMA( $p, q$ ) model:

$$\begin{aligned} (1 - \phi_1 B - \dots - \phi_p B^p) Z_t = ((1 - \theta_1 B - \dots - \theta_q B^q) a_t, \text{ or} \\ \phi(B) Z_t = \theta(B) a_t, \quad \text{where} \\ \phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p \text{ and } \theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q \end{aligned}$$

Box and Jenkins (Box et al., 1994) have extended the above concepts to cope with seasonal time series. Monthly notifications can be related by:

$$\begin{aligned} Z_t - \Phi_1 Z_{t-12} - \Phi_2 Z_{t-24} - \dots - \Phi_p Z_{t-12p} = \alpha_t - \Theta_1 \alpha_{t-12} - \Theta_2 \alpha_{t-24} - \dots - \Theta_Q \alpha_{t-12}, \quad (\alpha = \text{error component}) \text{ or} \\ \Phi(B^{12}) Z_t = \Theta(B^{12}) \alpha_t \text{ where,} \\ \Phi(B^{12}) = 1 - \Phi_1 B^{12} - \dots - \Phi_p B^{12p} \\ \Theta(B^{12}) = 1 - \Theta_1 B^{12} - \dots - \Theta_Q B^{12Q} \\ \text{and } B^{12} \alpha_t = \alpha_{t-12} \end{aligned}$$

Joining the seasonal and the non-seasonal parts gives:  $\phi(B) \Phi(B^{12}) Z_t = \theta(B) \Theta(B^{12}) a_t$ ,

Extending the concept of ordinary ( $\nabla = 1 - B$ ) and seasonally ( $\nabla_s = 1 - B^s$ ) differencing, one obtains the seasonal ARIMA model abbreviated: ARIMA ( $p, d, q$ )( $P, D, Q$ ) $_s$ :

$\Phi_p(B^s) \phi_p(B) [\nabla^d \nabla_s^D Z_t - U] = \theta_q(B) \Theta_Q(B^s) a_t$ , where  $U$  is an optional model constant.

Box-Jenkins modelling consists in adjusting the time series with the above-mentioned equation by stepping through several steps of the ARIMA protocol:

### 1) Identification:

Identification is the key to model building. A time series must be stationary with respect to the mean and variance ( $E(Z_t)=0$ ). Once the variance is stabilized, the next step consists in identifying the model embedded in the series. In the presence of a secular trend in the series, regular differencing is indicated (each observation is replaced by the difference between it and the previous observation). In the presence of seasonal trend, seasonal differencing is indicated (each observation is replaced by the difference between it and the previous-year observation). For want of a better word, the series might also exhibit a drift (stochastic behaviour) that can be taken care of by differencing. Both differencing (regular or seasonal) can be required for the same series. The dependent structure of a stationary time series is described by the autocorrelation function (ACF). To help precise the process order, we also use the partial autocorrelation function (PACF). Autoregressive (AR) processes are characterized by decaying ACF and spiking PACF. Moving average (MA) processes are characterized by spiking ACF and decaying PACF. A model with both processes (AR and MA) is expected to have both decaying ACF and PACF. The PACF is also helpful to determine the number of terms that can be included in the model. For example, decaying ACF and spiking PACF at only one lag will lead to an AR(1) term; PACF that shows two spiking lags will lead to an AR(2) term, and so on.

### 2) Estimation

Some computer softwares like SPSS (including *Trend*) have routines for carrying out ARIMA time series analysis. Estimation is an iterative process by the least square method that gives estimates of the parameters (with standard error) (SPSS). The concept of parsimony is mandatory; it consists in finding a representation of the series which includes as few parameters as possible.

### 3) Diagnosis

The adequacy of the model selected must be assessed to determine if model assumptions are met. One basic assumption is that the residuals  $a_t$  form a white-noise series. Different tests are possible to analyze residuals correlation. A common test is the Box-Ljung (Statistic Q) (Ljung et al., 1978) that verifies the null hypothesis that residuals are white noise.

It is also common in time series analysis that several models be adequate in the sense that model residuals are reduced to white noise. Some criteria can be helpful in model selection, such as, the residual variance  $\hat{\sigma}_a^2$ ; the coefficient of determination  $R^2$ , which is the amount of variance "explained" by the model; the Akaike's Information Criterion (AIC; Akaike 1969); and the Schwarz's Bayesian Criterion (SBC; Schwarz 1978). The AIC and SBC are performance statistics that balance statistical fit with model parsimony. SBC utilizes a larger penalty function than AIC, thereby often suggesting a model with fewer parameters (Appendix C).

## Appendix C: Formulas

### Estimated residual variance

$\hat{\sigma}_a^2 = SSQ / df$ , where  $df = N - N_p$  and  $SSQ$  is the residual sum of square,  $N_p$  is the number of parameters and  $N$  is the total number of observations

### Log-Likelihood

$L = -N \ln(\hat{\sigma}_a^2) - SSQ' / 2(\hat{\sigma}_a^2) - N \ln(2\pi) / 2$ , where  $SSQ'$  is the adjusted residual sum of square

### Residuals adjusted sum of square

$SSQ' = (SSQ) |\Omega|^{1/N}$  where  $\Omega$  is the theoretical covariance matrix of the observation vector computed at MLE (Maximum likelihood estimate)

### Coefficient of determination ( $R^2$ )

$R^2 = 1 - (\hat{\sigma}_a^2 / \hat{\sigma}_w^2)$  where  $\hat{\sigma}_w^2$  represents the variance of the transformed and differenced initial observed values

### Aikaike information criterion (AIC)

$AIC = -2 \ln(L) + 2 N_p$  where  $L$  is the likelihood function

### Schwartz Bayesian Criterion (SBC)

$SBC = -2L + \ln(N)N_p$ ,

### Ljung-Box (Statistic Q)

The chi-square statistics used in the test for lack of fit are computed using the Ljung-Box formula. Each chi-square statistic is computed for all lags up to the indicated lag value and is not independent of the preceding chi-square values. The null hypothesis tested is that the current set of autocorrelations is white noise. The Ljung-Box statistic with  $k$  degree of freedom (lags):

$Q_{(k)} = N(N+2) \sum_{j=1}^k ACF(j)^2 / (N-j)$  where  $ACF(j)$  is the autocorrelation coefficient at lag  $j$ .

ARTICLE 3

# **Scaling properties of childhood infectious diseases before and after mass vaccination in Canada**

## **Abstract**

The goal of this paper is to analyze the scaling properties of childhood infectious disease time series. We present a scaling analysis of the distribution of epidemic sizes of measles, rubella, pertussis and mumps outbreaks in Canada. This application provides a new approach in assessing infectious disease dynamics in large populations. The findings suggest that there is relevant information hidden in the heterogeneity of the common communicable disease time series. We find that the rubella epidemic size distribution follows an inverse power law. This behaviour also governs the propagation of measles for highly vaccinated populations. The IPL suggests the presence of a scale-invariant network in the propagation of measles and rubella as a result of the heterogeneity of the distribution of the contact rate. The topology network of these diseases seems to be up to the social determinants of behaviour at the individual scale. By contrast, it is found that pertussis and mumps are characterized by a network of transmission of the exponential type, which suggests homogeneity in the contact rate. The application of scaling analysis is helpful to interpret the dynamics of childhood infection in a real population and allows to suggest more strategic vaccination programs.

## **Key words**

Childhood infectious diseases, time series, complex dynamics, scaling properties, power-law, scale-invariant network, exponential network, large population, mass vaccination.

## Introduction

There are basically three methods to extract the structure of infectious disease time series data. Firstly, the deterministic models, such as SEIR, which allows to identify the main actors of the transmission of infections in populations (Anderson et al., 1991). This well-known method mirrors qualitatively the observed trends of the infectious disease time series in large populations insofar as the biological characteristics of the modelled disease are known. The SEIR model reflects a collective dynamic from the interactions of the mean individual biological parameters of the disease. SEIR is an *explanatory* model of the collective dynamics.

Secondly, there is the stochastic modelling technique of Box-Jenkins. The method models an empirical series by seeking the parameters that will reproduce the variations in the series as accurately as possible. This method is also well known, but was rarely applied to communicable infectious disease time series. The method allows one to describe the collective pattern of the dynamics while ignoring the biological specifics of the disease. This has been confirmed by our latest results (Trottier et al., 2004a). In fact, the Box-Jenkins methodology is not sensitive to the finely tuned biological characteristics of the various types of infections. It is a *predictive* model of the collective dynamics.

Finally, a new approach, the inverse power-law (IPL) modelling technique, has a promising research application for the study of infectious disease dynamics in small populations and in large, mass vaccinated, populations (Rhodes et al., 1996a; 1996b; 1997; 1998). This method is stochastic. However, it allows one to recognize the existence of a deterministic structure embedded in the series. This deterministic structure can determine the frequency of epidemic sizes over time and the topological structure of the contact network that caused the epidemics. In other words, the structure of the interactive

network of contacts, when described by an IPL distribution function, explains the type of propagation and the size of epidemics in the population over time (Liljeros et al., 2001; Barabasi et al., 2003).

In a previous article (Trottier et al., 2004a), we have shown that Box-Jenkins modelling does rescue the population dynamics of infections, but ignores both the biological characteristics of the diseases and the mass immunization impact. As far as our results are concerned, this type of modelling proves to be strictly collective. Box-Jenkins modelling, therefore, comes close to the SEIR assumptions. Although the SEIR model assumes knowledge of the biological characteristics of the disease, it supposes a law of mass action at the origin of the inter-individual contact rate and average value of the disease biological parameters (i.e. period of infectivity, latent period). Therefore, SEIR simulates infectious diseases on the collective level by positing an homogenous disease process. The method is therefore free from the individual level properties of the transmission process. This is confirmed by Rhodes et al. (1996a; 1996b; 1997) who have suggested that in small populations, where the IPL properly adjusts the data, the SEIR model is not applicable. These observations, therefore, show the limits of the Box-Jenkins and the SEIR modelling techniques from the standpoint of individual transmission properties: both techniques model the population dynamics and say nothing about the individual behaviour of the susceptibles and infectious. By the same token, these methods underline the gap between the population and the individual scales of investigation.

The Box-Jenkins and SEIR modelling techniques are thus in stark contrast with the IPL method that can extract the contact network topology at the individual level. Thus, scale-invariance of epidemic sizes in time (the absence of a typical scale of epidemics), such as the one modelled by the IPL, supposes heterogeneity of the links

between individuals. On the contrary, a network of contacts with only one characteristic scale, modelled for example by an exponential distribution of contacts, rather corresponds to a homogenous network of contacts, uniform for each individual. We, therefore, suggest that a uniform network of contacts corresponds well to the mass action principle of the SEIR model and is also akin to the Box-Jenkins methodology.

Our objective in this paper is to test the hypothesis of Rhodes et al. (1996a; 1996b; 1997; 1998) of an adequate adjustment of the IPL to epidemic sizes in a large and highly vaccinated population. In contradistinction with the Box-Jenkins method, the IPL model might give information on the individual structure of contacts that leads to epidemics in the population. If this were the case, it might bear upon vaccination strategy. More specifically, an IPL and an exponential function will be adjusted, trying to correlate the results with the disease, the vaccination period, and the contact network topology. Suggestions will then be made about strategic vaccination campaigns.

### **Data**

Data were obtained from Health Canada (Center for Infectious Diseases Prevention Control, Division of Disease Surveillance). Time series of pertussis, mumps, measles and rubella in Canada were available on a 4-weekly basis (13 observations per year) before 1990 and on a monthly basis (12 observations per year) from 1991 to 2002 except for some periods where measles and rubella (1959-1968) and mumps (1959-1986) were not declared. The 4-weekly and monthly time series were analyzed from the 1950s. In Canada, the pertussis vaccine was introduced in 1943, the measles vaccine in 1963, and the mumps and rubella vaccines in 1969. The study covered the pre- and post-vaccination era (except for pertussis where only the post-vaccination period was analyzed).

Notifications of infectious diseases on a 4-weekly basis can be found in the *Annual Report of Notifiable Diseases* (Dominion Bureau of Statistics, 1953-1970; Statistics Canada, 1971-1978). For the pre-vaccination era, 4-weekly notification data from 1953 to 1959 for measles, mumps and rubella were analyzed. In the post-vaccination period, 4-weekly notification data from 1970 to 1976 for measles and rubella were used (mumps was not declared from 1959 to 1986; it is therefore impossible to analyze data in the early post-vaccination era). Also, 4-weekly notifications of pertussis from 1953 to 1962 and from 1970 to 1976 were studied. Finally, to cover the current period, monthly notifications of measles, rubella, pertussis were used from 1991 to 2002 and mumps from 1991 to 1998 (Health Canada, 1992-2004). The crude time series are presented in the thesis I Appendix.

## Method

### Epidemic size

Rhodes et al. (1996a; 1996b; 1997; 1998) determine the size of epidemics in small and isolated populations in the following way. An epidemic event is ascertained by the presence of a finite number of cases recorded in a sequence of consecutive time series data bounded by an absence of cases. An epidemic has size  $s = \sum_{t_{\text{start}}}^{t_{\text{end}}} C(t)$ , where  $t_{\text{start}}$  is the time point when cases in an epidemic event first appear and  $t_{\text{end}}$  is the following time point when there are no more cases present.  $C(t)$  is the number of recorded cases in the period  $t$ .

Due to frequent fade-outs of infections in small communities it is rather easy to distinguish epidemics and calculate the distribution of epidemic sizes (Rhodes et al., 1996a; 1996b; 1997). In large populations, things go otherwise. Therefore, there may be

different ways to estimate epidemic sizes. Since the number of cases usually never reaches 0 in large populations such as Canada, our classification schema has consisted of estimating the epidemic size from the cumulated number of cases bounded by troughs. In other words, an epidemic event has been ascertained by the presence of a number of cases recorded in a sequence of consecutive time series data bounded by a diminution of cases.  $t_{\text{start}}$  is the time point when cases in an epidemic event begins to increase and  $t_{\text{end}}$  is the following time point when the lowest number of cases appear before a newly increasing number of cases is notified.  $C(t)$  is the number of recorded cases in the  $t$  period.

It goes without saying that this way of assessing the size of epidemics may bear classification errors. For example, one epidemic for which the number of cases slightly decreases between two different time periods of observation (month or 4-week periods) might have been classified as distinct epidemics. By contrast, distinct epidemics occurring in one period of observation might have been identified as only one overestimated epidemic (i.e. the monthly or 4-weekly notifications could hide several epidemics). Moreover, we had to take an arbitrary decision to avoid overlapping between epidemics ( $t_{\text{start}}$  was defined as the time point when cases in an epidemic event begins to increase right after a trough and  $t_{\text{end}}$  is the following time point when a new trough appears). We used this classification schema because we believe it involves the least biases. But to control somewhat for any shortcomings of this classification, we also analyzed the size of epidemics in a different manner: we used the cumulative distribution of the monthly (or 4-weekly) notifications (not presented). This type of classification is similar to the aggregation of the weekly data into monthly figures for the Carlisle data used by Rhodes et al. (1998). This type of classification may also be affected by

classification errors in that it ignores the peaks and troughs of the epidemics. We were nevertheless confident that two classification schemes would be safer to ascertain the size of epidemics than a single one. In the results section only the first classification of the data will be presented formally.

### **Modelling**

Scaling analysis consists in adjusting epidemic size distribution to an inverse power-law (IPL). The IPL function can be written on an arithmetic scale as:

$$N(>s)=as^{-b}$$

or on a logarithmic scale as:

$$\log N(>s) = \log a - b \log s$$

where  $s$  is the epidemics size and  $N(>s)$ , the frequency of epidemics of at least  $s$  size. (Rhodes et al., 1996a; 1996b; 1997). A log-log plot of the number of epidemics greater than a given size  $s$ , ( $N(>s)$ ), against  $s$ , allows to estimate the scaling exponent  $b$  and an estimate of  $a$  from the intercept. A good fit to an IPL function should give a straight line with slope  $b$  on a log-log plot.

The particularity of this function is the absence of any particular scale (West et al., 1990; Bak et al., 1991; Barabasi et al., 1999). West et al. (1990) explain how the IPL can be generated from a one-characteristic-scale distribution; they suggest that when there are interactions between factors at the origin of an outcome, amplification of the outcome may occur by way of positive feedback and lead to the loss of the characteristic scale. When this occurs, new properties of the studied phenomenon arise at a higher scale of observation that are rescued by the IPL. For example, new population disease properties (resistance of disease prevalence to vaccination) can come out from the nonlinear effects of interactions setting up at the individual level (the transmission network organization).

This population effect has been well explained by Koopman et al. (1994; 1999) in the case of dengue and AIDS. When a probability density has an IPL form there is a high probability that the extreme values of the curve dominate and account, therefore, for most of the complex dynamics. The IPL involves a hierarchy of scales of all sizes embedded into one another that can explain why very small epidemics do co-exist with very large ones. This is the meaning of the expression «scale-invariance». Large events of very low probability (generally ascribed to randomness) are expected from a scale-invariant network of interactions. The IPL predicts that an infinity of observations is theoretically possible depending on the number and size of the correlations linking the individual units of the complex system (Philippe, 2000). If applied to the probability of contacts among individuals, an IPL assumes that contacts are not random or uniform. In a scale-invariant network of interactions a few nodes have many connections (supernodes) while most have only a few ones (Barabasi et al., 2003).

As a benchmark for the IPL, the exponential function,  $N(>s) = a e^{-bs}$ , has been used. Although other plausible probability models could be used, such as the lognormal (West et al., 1990; Philippe, 2000), the exponential remains the model of choice for the study of epidemic size distribution (Rhodes et al., 1996a; 1997). Unlike the IPL that has no typical scale (the distribution is unbounded to the right), the exponential distribution has only one characteristic scale (the distribution has finite variance) (West et al., 1990). Since the exponential has a shorter right tail than the IPL, it is less likely to explain very large epidemics (large events of very low probability). A good fit to the exponential function means that the disease has only one typical scale (involving a short right tail) that dominates the dynamics. An exponential law has constant random hazard through time and, thereby, describes memoryless phenomena (Lee, 1992). Therefore, if applied to

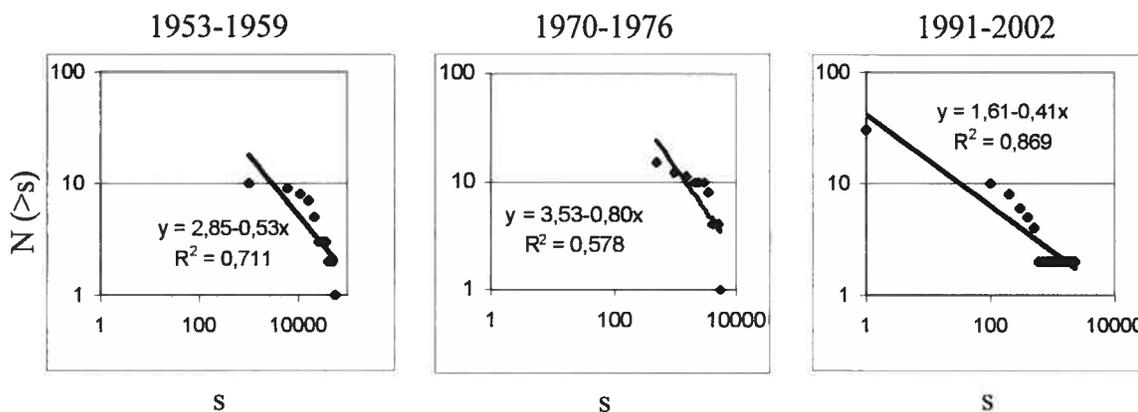
the probability of contacts among individuals, the exponential assumes that contacts are random events whose probability is uniform in the population (Barabasi et al., 2003). For the IPL, the probability of contacts is rather variable among individuals, with a few individuals having very numerous and most, rare contacts. Goodness-of-fit ( $R^2$ ) to the exponential and IPL will be assessed graphically on the arithmetic and doubly-logarithmic scales in this paper. In line with previous publications, no statistical tests will be used to test curve fitting.

## Results

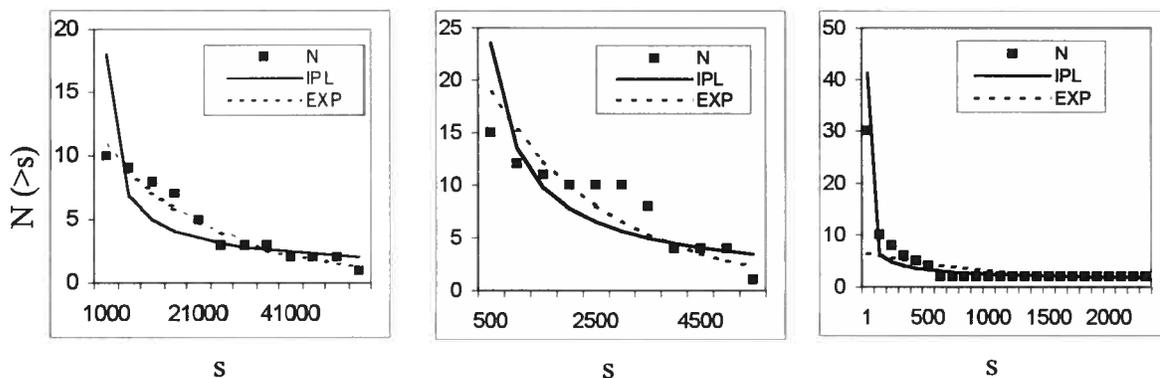
This section provides the results of the scaling analysis for each time series. Figure 1 depicts the measles epidemic size distribution analysis. The upper diagrams (figure 1a) represent IPL fits on a log-log plot for each period: 1953-1959, 1970-1976 and 1991-2002 (original data on logarithmic scales). The linear regression equation of the logarithms of the data estimates the IPL slope,  $b$ . The lower diagrams (figure 1b) represent the IPL and exponential fits on the arithmetic scale. Goodness-of-fit ( $R^2$ ) to IPL and exponential functions are given in table 1. The presentation of the diagrams is identical for rubella (figure 2), pertussis (figure 3) and mumps (figure 4).

From 1953 to 1959, the time series of measles is made up of 10 distinct epidemics ranging in size from 1 000 cases up to 56 000 cases. From 1970 to 1976 and 1991 to 2002, 15 distinct epidemics (from 500 cases up to 5 900 cases) and 30 distinct epidemics (from 1 case up to 2 400 cases) occurred, respectively. An example of computation of the epidemic size distribution is presented in Appendix. For the 1950's and 1970's, the exponential function provides a better fit (0.95 and 0.78 respectively) of the distribution than the IPL (0.71 and 0.58 respectively). From the scaling plot of the 1990's in figure 1b

one sees that the IPL provides a better fit (0.87) than the exponential (0.52). The measles time series seems to reflect scale-invariant behaviour only for the highly vaccinated population. The IPL slope,  $b$ , is estimated at 0.41 in the 1990s.



**Figure 1a.** Measles epidemic size distribution for Canada data (log-log plot).



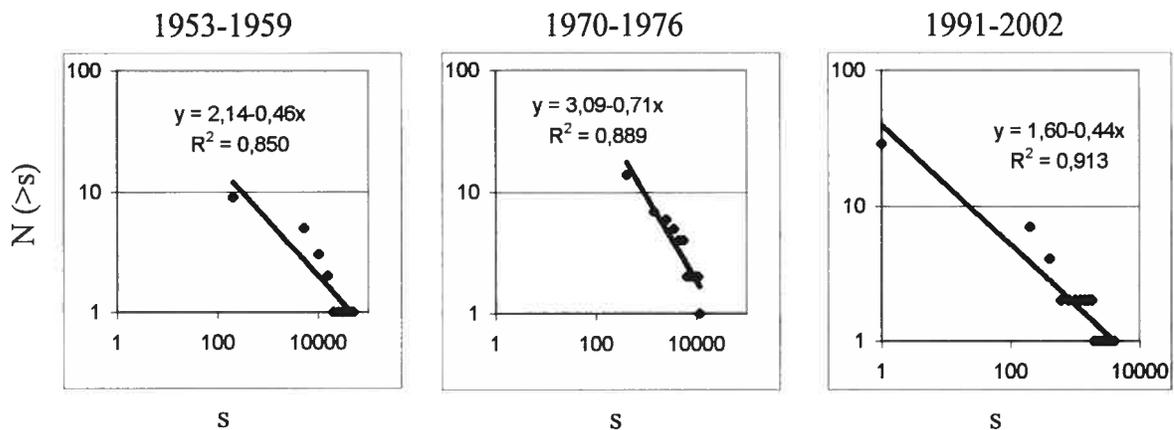
**Figure 1b.** Measles epidemic size distribution (linear plot) fitted with power-law (solid line) and exponential (dashed line) functions.

Period	IPL	EXP
1953-1959	0.71	0.95
1970-1976	0.58	0.78
1991-2002	0.87	0.52

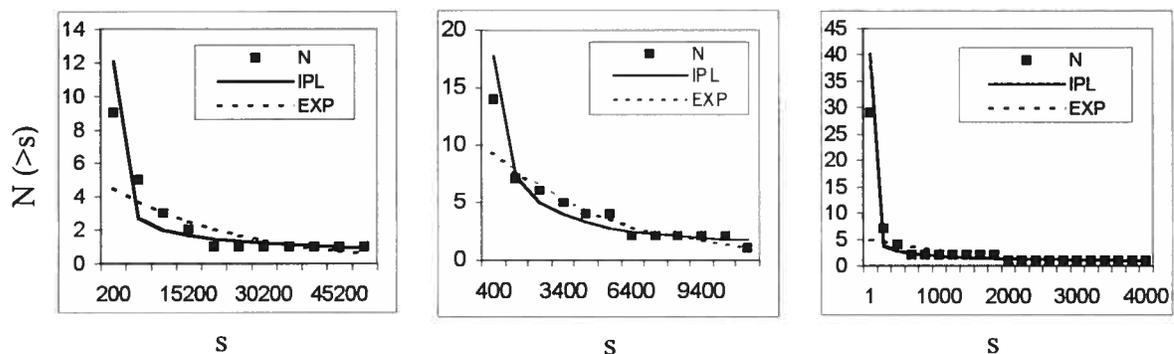
**Table 1.** Measles. Goodness-of-fit ( $R^2$ ) to IPL and exponential distribution functions

From 1953 to 1959, the time series of rubella is made up of 9 distinct epidemics ranging in size from 200 cases up to 53 000 cases. From 1970 to 1976 there are 14 distinct epidemics (from 400 cases up to 13 000 cases). Finally, from 1990 to 2002, 29

distinct epidemics (from 1 case up to 4 000 cases) occurred. The IPL generally provides a better fit for the 1950s, 1970s and 1990s series (0.85, 0.89 and 0.91 respectively) than the exponential (0.71, 0.90 and 0.62 respectively) (table 2). Whether there is mass vaccination or not, rubella seems to be associated with scale-invariance. The IPL slopes are estimated at 0.46, 0.71 and 0.44 for the 1950s, 1970s and 1990s respectively.



**Figure 2a.** Rubella epidemic size distribution for Canada data (log-log plot).



**Figure 2b.** Rubella epidemic size distribution (linear plot) fitted with power-law (solid line) and exponential (dashed line) functions.

Period	IPL	EXP
1953-1959	0.85	0.71
1970-1976	0.89	0.90
1991-2002	0.91	0.62

**Table 2.** Rubella. Goodness-of-fit ( $R^2$ ) to IPL and exponential distribution functions.

From 1953 to 1959, the time series of pertussis is made up of 28 distinct epidemics ranging in size from 1 000 cases up to 8 000 cases. From 1970 to 1976 and

1991 to 2002, 18 distinct epidemics (from 100 cases up to 1 100 cases ) and 23 distinct epidemics (from 200 cases up to 10 000 cases) occurred respectively. The exponential function better fits the pertussis series for the 1970s and the 1990s but not for the 1950s (table 3). However, a similar analysis from the monthly incidence cumulative distribution classification schema has shown that all the series follow an exponential function including the 1950s series (not shown). We will discuss this point in the next section.

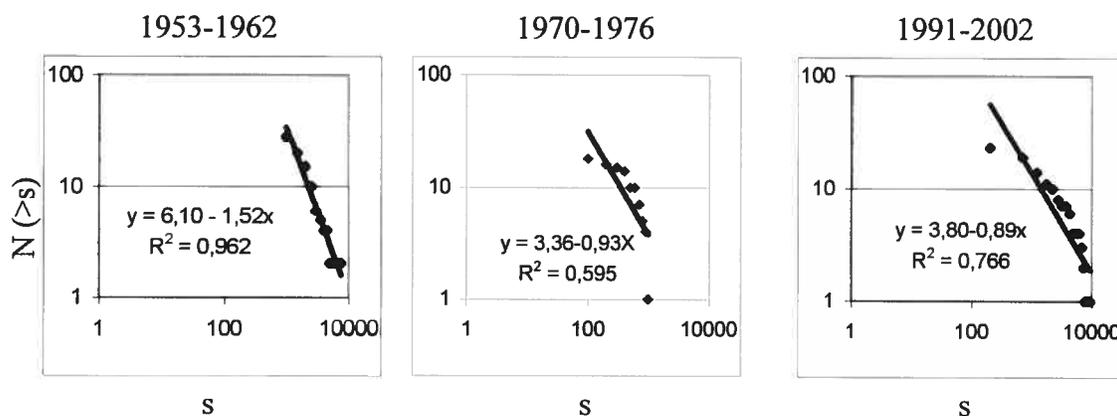


Figure 3a. Pertussis epidemic size distribution for Canada data (log-log plot).

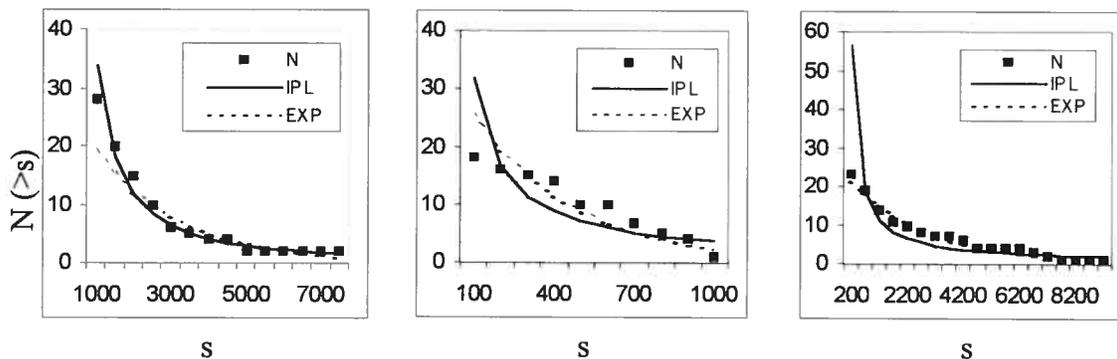
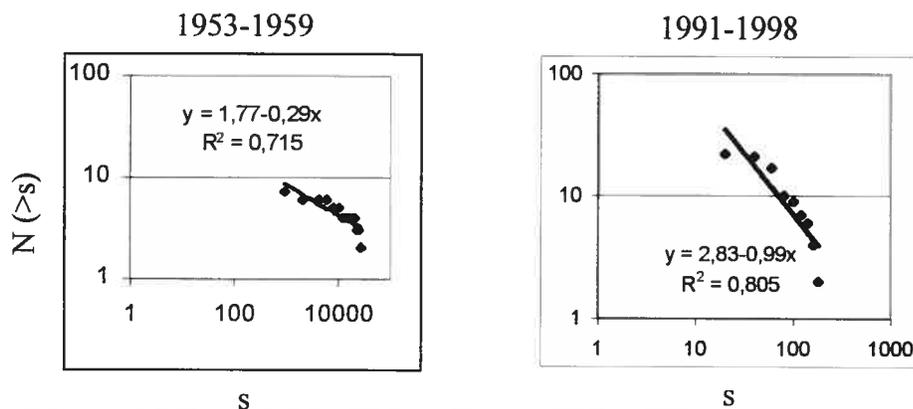


Figure 3b. Pertussis epidemic size distribution (linear plot) fitted with power-law (solid line) and exponential (dashed line) functions.

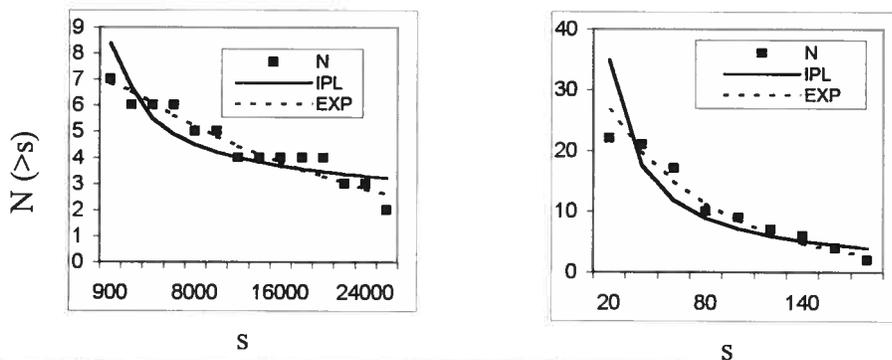
Period	IPL	EXP
1953-1962	0.96	0.89
1970-1976	0.60	0.81
1991-2002	0.77	0.96

Table 3. Pertussis. Goodness-of-fit ( $R^2$ ) to IPL and exponential distribution functions

Finally, from 1953 to 1959, the time series of mumps is made up of 7 distinct epidemics ranging in size from 900 cases up to 28 000 cases. From 1991 to 2002, 22 distinct epidemics occurred ranging in size from 20 cases up to 200 cases. The exponential function seems to provide a better fit (0.89 and 0.95) for the mumps series than the IPL (0.72 and 0.81) in the 1950s and 1990s respectively (table 4).



**Figure 4a.** Mumps epidemic size distribution for Canada data (log-log plot).



**Figure 4b.** Mumps epidemic size distribution (linear plot) fitted with power-law (solid line) and exponential (dashed line) functions.

Period	IPL	EXP
1953-1959	0.72	0.89
1991-1998	0.81	0.95

**Table 4.** Mumps. Goodness-of-fit ( $R^2$ ) to power law and exponential distribution functions

## Discussion

Complex systems and non-linear effects in population dynamics have stimulated much research since a few decades (May, 1976; Olsen et al., 1988; Bolker et al., 1993; Anderson, 1994; Earn et al., 1998; Philippe et al., 1998; Billings et al., 2002). A variety of mathematical models, such as SEIR, have shown that a well-organized, non-linear, system may describe the transmission dynamics of several childhood infectious diseases at the population scale (Anderson et al., 1991). In fact, the incidence of infectious diseases over time is viewed as a moderately complex system whose dynamics are determined by a few actors involving interaction and feedback processes.

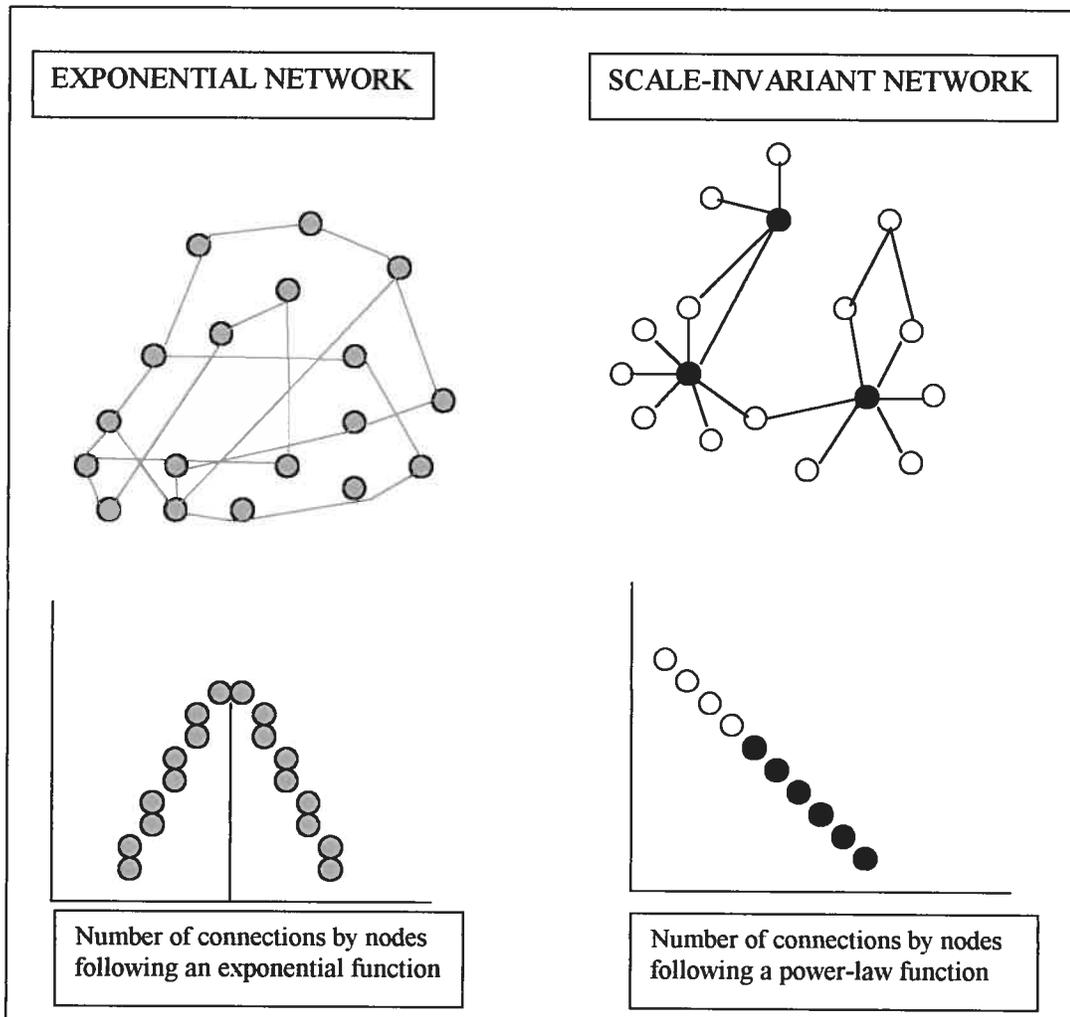
There has been much discussion concerning the possible emergence of an inverse power-law in some of these complex systems. Thus, as early as 1961, Taylor describes one of the few generalities in ecology, i.e., the species-specific relationship between the temporal or spatial variance of populations and their mean abundances (Taylor, 1961). Many social phenomena were later described by an IPL such as the distribution of wealth (Pareto's law), the relative frequency in the use of words in language (Zipf's law) or the distribution in the sizes of cities (Auerbach's law). Other examples of dynamics that comply with an IPL are the distribution of the intensity of earthquakes (Bak et al., 1991), the relationship of the variance to the mean for measles cases reported in 366 communities in England and Wales (Keeling et al., 1999), the emergency-ward length of stay (Philippe et al., 2004), the spatial clustering of leukemia in the city of San Francisco (Philippe, 1999), the distribution of reported number of sexual partners (Liljeros et al., 2001; Schneeberger et al., 2004), and many more.

Moreover, it has been shown that the pattern of measles, mumps and pertussis, which are characterized by frequent fade-outs in small populations, can also be

represented by an IPL (Rhodes et al., 1996a; 1996b; 1997; 1998). For measles, the critical community size above which chains of transmission persist has been estimated at 250 000-300 000 individuals (Bartlett, 1957; 1960). Populations smaller than this critical size are characterized by frequent fade-outs and irregular epidemics. According to Rhodes et al. (1998), there are three types of measles epidemics. Type I epidemics occur in large populations well above the critical community size. There is no fade-out of infections between epidemics in these populations. Type II epidemics occur in populations near the critical size. These populations are characterized by regular epidemic outbreaks and occasional fade-outs. Finally, type III epidemics occur in populations smaller than the critical size and are subject to frequent fade-outs of infections resulting in long intervals with no disease. The IPL describes well type III epidemic size; in contradistinction, the exponential function shows overt lack of fit to epidemic size in these populations (Rhodes et al., 1996a; 1997). The previous authors also advocated that the pattern of infectious diseases in highly vaccinated populations might well be described by an IPL (therefore ruling out the exponential) since such populations can give rise to type III epidemics.

A good fit to an exponential distribution of epidemic size means that the disease dynamics has only one typical scale. An exponential law fits a homogenous network of contacts, uniform for each individual (figure 5), since no constraint is at work in an exponential process. The contacts occur at random. In a single-scale network, each node has approximately the same number of links that varies randomly among individuals. This contrasts with a power-law distribution, which has no typical scale. In a scale-invariant or scale-free network, the links are concentrated within a few hubs (figure 5). An IPL fit means that the randomness of the individual contacts has been curbed by a constraint to growth. In the scale-free network a few nodes, therefore, have many

connections while most have only a few ones. In the scale-free network the distribution of contacts is heterogeneous. This may be explained by contingent events selected for by positive feedback processes and/or preferential attachment of the new links to the highly connected nodes (supernodes) (Barabasi et al., 1999; 2003; Liljeros et al., 2001).



**Figure 5.** In the exponential network, the number of connections is randomly distributed. In the scale-invariant network a few nodes have many connections while most have few ones.

Our analysis shows that the rubella epidemics distribution may be scale-invariant since it is strongly associated with the IPL distribution before and after mass vaccination.

The measles series also reflects a scale-invariant behaviour for the highly vaccinated population 1990 sub-period. By contrast, pertussis and mumps (and measles before mass vaccination) rather comply with an exponential function. We have also analyzed the cumulative distribution of the monthly (or 4-weekly) notifications for every series (not shown, see the Methods section). We obtained the same results for all our series with this different classification scheme except for the pertussis 1950s sub-period, which was best fitted by the exponential.

It is relatively clear that mumps (vaccine introduced in 1969) is a single-scale infectious process. The goodness-of-fit to the exponential function ranges from 0.89 to 0.95. It also seems to be the case for pertussis (vaccine introduced in 1943). If vaccination is expected to contribute to the loss of the characteristic-scale epidemic size distribution in a large mass-vaccinated population as Rhodes et al. (1997; 1998) contend, it does not seem to be the case for mumps and pertussis. Either data quality prevents rescue of the scale-invariant distributional pattern or these diseases are fundamentally different from rubella and measles. If the latter, then it would mean that the propagation of pertussis and mumps is related to a more homogenous system of contacts, i.e., an unconstrained network. Another hypothesis is that we did not reach a high level of immunization in the population due to vaccine failure or inadequate vaccine coverage. In the 1990s, there has been a resurgence of pertussis, and adolescents and adults are currently considered the major reservoir playing a predominant role in the transmission to young children in whom the disease is most severe (Tam et al., 1998). National estimates of childhood vaccination coverage were not collected in a standardized manner or on a regular basis prior to 1994. In 1994 and 1998, the vaccination coverage was 96.0% for measles, 95.4% for mumps, and 95.3% for rubella (LCDC, 1998). Coverage levels for pertussis were reported as

approximating 80%. Notwithstanding the large proportion of susceptibles vaccinated, vaccine failure could have occurred. Higher immunity is reached for viral infection (measles, rubella, mumps). Thus measles vaccine efficacy varies between 85 to 95% with a single dose of the vaccine given at 12 or 15 months and may reach 100% with a second dose (De Serres et al., 1995; Vitek et al., 1999; Health Canada, 2002b; Redd et al., 2004). A single dose of the mumps and rubella vaccines produces an antibody response in over 95% and 97% of susceptible individuals, respectively (Health Canada, 2002b; Redd et al., 2004). For pertussis (bacterial infection), immunity might be lower. Efficacy of the pertussis vaccines was estimated to be no more than 85% and will not induce lasting immunity (Onorato et al., 1992; Bentsi-Enchill et al., 1997; Halperin et al., 2000; Health Canada, 2002b). We therefore conclude that, in the 1990s, vaccination coverage and vaccine efficacy are relatively high for measles, rubella and mumps, but could be less for pertussis. Given the relatively high transmissibility of the disease and a significant number of susceptibles in the population, pertussis could propagate randomly with no need of supernodes. However, mumps appear to have a high level of vaccination coverage and high vaccine efficacy. It is possible that the propagation of mumps is related to a system of contacts that is more homogenous, but the quality of data should not be ruled out when considering this interpretation. Incidentally, in the 1990s, some provinces, such as Quebec, used to notify cases of mumps on a clinical basis ( $> 90\%$ ), an event that might have overestimated the number of cases as other viruses may also cause parotitis, a key mumps symptom (De Serres et al., 1997).

Rubella is different from the other infections because it exhibits a scale-invariant network both before and after vaccination. Of all the infectious diseases studied, rubella has the lowest transmissibility (Anderson et al., 1991). This might explain why IPL best

fits rubella even in the pre-vaccination era. Even for an unvaccinated population, rubella is not transmissible enough to randomly infect individuals. The lower the transmissibility, the higher the impact of social interaction according to Rhodes et al. (1997). The impact of stochasticity or heterogeneity on the network topology can, therefore, become significant and drive the epidemics even in an unvaccinated population. For example, the average age at infection for rubella is usually high (Anderson et al., 1991); the social organization of individual contacts in this disease might then show a different pattern of organization.

Measles in the 1950s and 1970s is related to an exponential network. This might be explained by the very high transmissibility of measles. Within an unvaccinated population, the transmissibility is so high that all susceptibles contract the disease anyhow. In this context, the contingent factors liable to curb the network topology can hardly affect the system dynamics. Mass vaccination began in 1963 in Canada and seems to have contributed to the loss of the characteristic scale thereafter. There is a relationship between the type of measles epidemic observed in small populations and that observed in a large highly vaccinated population. When an infection is introduced in a small community, an epidemic outbreak is triggered and long intervals of time elapse before a newly epidemic outbreak occurs. Successful vaccination reduces the number of available susceptibles in the same way.

In the 1990s, the measles and rubella series do not really differ in their IPL slope, (summarized in table 5). The IPL exponent may reflect the transmissibility of the disease and can determine the basic reproduction number of the disease (Rhodes et al., 1997). Rhodes et al. (1996a; 1996b; 1997) have shown that the epidemic size distribution of measles, mumps and pertussis in small isolated populations generally break down for

very large epidemics. They have estimated the slope of the IPL for measles ( $\approx 0.27$ ), pertussis ( $\approx 0.26$ ), and mumps ( $\approx 0.45$ ). Unfortunately, these values cannot be compared with ours because the authors have excluded the very large epidemics before fitting IPL. However, the estimated IPL exponent for Carlisle's measles, where all the data points were used to fit the model, was found to be 0.46 (Rhodes et al., 1998), a result quite similar to ours. Presently, the size of the exponent can hardly be interpreted any further, as it is influenced by the transmissibility of the disease, the social structure, and the individual mixing pattern (Rhodes et al., 1998).

Infectious diseases	1953-1959	1970-1976	1991-2002
Measles	-	-	0.41 (0.03)
Rubella	0.46 (0.06)	0.71 (0.08)	0.44 (0.03)

**Table 5.** The IPL slope,  $b$ , for series represented by IPL with their corresponding estimated standard error (in parenthesis).

Studying the distribution of the epidemics size is important to understand the law of transmission (exponential or IPL) among individuals that are involved in a network of social interactions. IPL arises in a network driven by preferential attachment; the exponential, from a randomly interconnected network. Clearly, control and prevention of the disease will be more effective if one understands the network topology. Where supernodes are at work, a classical random vaccination program is expected to fail because infection is transmitted by the supernodes that are not zeroed in by the standard vaccination program. Since supernodes are related to a multitude of other nodes, one of the supernodes may always be infected by an infectious. When a supernode is infected, it can contaminate a multitude of other nodes that can propagate the virus through the whole network in a short time lapse. Incidentally, Pastor-Satorras et al. (2001) found in a

simulated scale-invariant network that there is no contacts threshold (defined by the cases reproduction rate) that can limit epidemic outbreaks since there is always a supernode that can be infected. Scale-free networks are therefore resilient to random intervention (Barabasi et al., 2003). Infections propagate much faster in scale-free networks than in a single-scale network. In fact, in a scale-invariant network almost all the individuals should be immunized to avoid the spread of infections (Barabasi et al., 2003). The classical approach, which consists of randomly immunizing individuals, may therefore be inadequate since it neglects supernodes.

These results should, however, be interpreted with caution. There are strong indications that not all cases of disease come to the attention of physicians, which can lead to underreporting; also, the reporting of diagnosed cases may also be incomplete (Clarkson et al., 1985; Davis et al., 1993; McDonnell et al., 1994; Giesecke, 1994; Ewert et al., 1995; Deeks et al., 1999; Allen et al., 2000; Health Canada, 2002b; Harpaz, 2004; Trottier et al., 2004b). In spite of these shortcomings, the reporting system still reflects the dynamics quite accurately (i.e., variation of the incidence in the course of time) (Dominion Bureau of Statistics, 1953-1970; Statistics Canada, 1971-1978, Clarkson et al., 1985; Brewster, 1993; Torvaldsen et al., 2003). However, the notification system in Canada has changed since 1950 (variation in the way cases are reported, variation in case definition, laboratory confirmation, and improvement of the quality of data) (Dominion Bureau of Statistics, 1953-1970; Statistics Canada, 1971-1978; Health Canada, 1975-1991; 1992-2004; Trottier et al., 2004b). If these changes were of significant extent, they could have entailed biased variations in the data across time. As stated above, however, the time series have been broken down into three short sub-periods. This can help limit the extent of inappropriate variations in the data. There is a price to pay for more

homogeneity, i.e., series of small size, it is that they can increase the variability of the parameters estimates of the models. All these shortcomings can have an impact on the interpretation of results in principle. Be this as it may, it is a relief that our results show coherence across series and diseases.

In this paper, the spatial dimension of the data was not considered. Disease spatial dynamics if left aside may introduce confounding in the time dimension as space and time are related in disease transmission. Even though the space dynamics can modulate the time dynamics of diseases, to adequately manage space and time interaction the spatial data base must concern populations that are relatively close in space. Clearly, coupling of populations over space decreases with distance (decoupling also occurs with time for time correlations). We, therefore, hypothesize that our results are independent of space as they pertain to data from one very large spatial unit. However, even for a country as large as Canada, time-space interaction occurs at a regional level. Therefore, forthcoming studies should look into the role of space in the dynamics rescued by the IPL: nearby towns disease notifications would be a strategic asset for this endeavour.

### **Conclusion**

Our analyses of the dynamics of common infectious diseases in Canada suggest that relevant information in the form of individual interaction is hidden in the heterogeneity of the common communicable diseases time series. Pertussis and mumps were related to the exponential network model, which assumes that individuals are connected randomly or uniformly. These diseases' dynamics are close to the assumptions of the SEIR model, which suppose homogeneity in inter-individual contacts.

Epidemiology usually treats population as a collection of independent individuals rather than as a system of interacting individuals (Koopman et al., 1999). The development of the IPL scaling method indicates that supernodes play an important role in network development (Barabasi et al., 1999). The impact of stochasticity in the network topology is therefore significant and drives the epidemics. This phenomenon gives useful information about the type of connectivities in the network. Social interactions might also explain this phenomenon. Rubella before and after mass vaccination and measles after mass vaccination were associated with an IPL. For example, the average age at infection for rubella is usually high; the social organization of individual contacts in this disease might then prove to be different.

There are some aspects of real networks that are not incorporated in theoretical models such as SEIR. Because epidemiological analysis usually ignores network connections between individuals, it misses important determinants of these diseases. For infectious diseases that conform to IPL, this could explain why attempts to reduce them by vaccination have only a temporary effect. Complex systems are composed of coupled subsystems. Random intervention (vaccination) at the population scale may entail less-than-optimal effect because the most important subsystems for disease transmission are not directly zeroed in. Effects are sometimes counter-intuitive because the main active factors of disease occurrence are embedded in simple linear cause-effect relationships. Interconnectivity is an important ingredient of complexity and imposes the potential for unexpected large-scale events (large epidemics) to co-exist with small occurrences. Small variations in the magnitude of events occur frequently, whereas large catastrophic changes are rare. The scale-free distribution of rubella and measles indicates that their development is governed by robust self-organizing individual interactions that go beyond

the smoothing observed at the population scale (the time series of observations). Understanding the tenets of open-systems dynamics in the population transmission of disease would certainly help to control infectious diseases more efficiently (Association of Schools of Public Health, 1999). In fact, the IPL distribution suggests that some infectious diseases will not be eliminated without either immunization of entire populations (a black-box approach) or, more strategically, an understanding of the network topology of disease transmission.

### **Acknowledgements**

The authors wish to thank Carole Scott from Health Canada (Center for Infectious Disease Prevention and Control / Division of Disease Surveillance) for providing the infectious diseases time series and Social Sciences and Humanities Research Council of Canada for financial support to HT.

## References

- Allen C.J., Ferson M.J. (2000). Notification of infectious diseases by general practitioners: a quantitative and qualitative study. *The Medical journal of Australia*, 172(7), 325-328.
- Anderson R.M. (1994). The Croonian Lecture, 1994. Populations, infectious disease and immunity: a very nonlinear world. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 346(1318), 457-505.
- Anderson R.M., May R.M. (1991). *Infectious Diseases of Human. Dynamics and Control*. New York, Oxford University Press, 757p.
- Association of Schools of Public Health. (1999). *Demonstrating excellence in academic public health practice*. Prepared by The Association of Schools of Public Health Council of Public Health Practice Coordinators, Washington, DC.
- Bak P., Chen K. (1991). Self-organized critically. *Scientific American*, 264, 46-53.
- Barabasi A.L., Albert R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509-512.
- Barabasi A.L., Bonabeau E. (2003). Réseaux invariants d'échelle. *Pour la Science*, 314, 58-63.
- Bartlett M.S. (1957). Measles periodicity and community size. *Journal of the Royal Statistical Society. Series A (general)*, 120(1), 48-70.
- Bartlett M.S. (1960). The critical community size for measles in the United States. *Journal of the Royal Statistical Society. Series A (general)*, 123(1), 37-44.
- Bentsi-Enchill A.D., Halperin S.A., Scott J., MacIsaac K., Duclos P. (1997). Estimates of the effectiveness of a whole-cell pertussis vaccine from an outbreak in an immunized population. *Vaccine*, 15(3), 301-306.
- Billings L., Schwartz I.B. (2002). Exciting chaos with noise: unexpected dynamics in epidemic outbreaks. *Journal of Mathematical Biology*, 44(1), 31-48.
- Bolker B.M., Grenfell B.T. (1993). Chaos and biological complexity in measles dynamics. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 251(1330), 75-81.
- Brewster D. (1993). Surveillance of measles in Scotland 1983-1991; a comparison of statutory notification and laboratory reporting. *Journal of Infection*, 27(2), 193-201.
- Clarkson J.A., Fine P.E. (1985). The efficiency of measles and pertussis notification in England and Wales. *International journal of epidemiology*, 14(1), 153-168.

Davis S.F., Strebel P.M., Atkinson W.L., Markowitz L.E., Sutter R.W., Scanlon K.S., Friedman S., Hadler S.C. (1993). Reporting efficiency during a measles outbreak in New York City, 1991. *American journal of public health*, 83(7), 1011-1015.

Deeks S., De Serres G., Boulianne N., Duval B., Rochette L., Dery P., Halperin S. (1999). Failure of physicians to consider the diagnosis of pertussis in children. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 28(4), 840-846.

De Serres G., Boulianne N., Bussi eres N., Pouliot B., Marin-Lira A. (1997). Epidemiology of mumps in Quebec, 1970-1995. *Canada Communicable Disease Report*, 23(2), 9-13.

De Serres G., Boulianne N., Meyer F., Ward B.J. (1995). Measles vaccine efficacy during an outbreak in a highly vaccinated population: incremental increase in protection with age at vaccination up to 18 months. *Epidemiology and infection*, 115(2), 315-323.

Dominion Bureau of Statistics. (1953-1970). Health and Welfare Division, Public Health Section. *Annual Report of Notifiable Diseases*. Catalogue number 82-201.

Earn D.J., Rohani P., Grenfell B.T. (1998). Persistence, chaos and synchrony in ecology and epidemiology. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 265(1390), 7-10.

Ewert D.P., Westman S., Frederick P.D., Waterman S.H. (1995). Measles reporting completeness during a community-wide epidemic in inner-city Los Angeles. *Public health reports*, 110(2), 161-165.

Giesecke J. (1994). *Modern Infectious Disease Epidemiology*. London, Edward Arnold, 256p.

Halperin S.A., Smith B., Russell M., Hasselback P., Guasparini R., Skowronski D., Meekison W., Parker R., Lavigne P., Barreto L. (2000). An adult formulation of a five-component acellular pertussis vaccine combined with diphtheria and tetanus toxoids is safe and immunogenic in adolescents and adults. *Vaccine*, 18(14), 1312-1319.

Harpaz R. (2004). Completeness of measles case reporting: review of estimates for the United States. *The Journal of infectious diseases*, 189 (Suppl.1), S185-S190.

Health Canada. (1975-1991). Bureau of Epidemiology, Laboratory Centre for Disease Control. *Canada Diseases Weekly Report*. ISBN 0382-232X

Health Canada. (1992-2004). Laboratory Centre for Disease Control. *Canada Communicable Disease Report*. ISBN 1188-4169.

Health Canada. (2002b). *Canadian Immunization Guide*. 6<sup>th</sup> edition. 292p.

- Keeling M., Grenfell B. (1999). Stochastic dynamics and a power law for measles variability. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 354(1384), 769-776.
- Koopman J.S., Longini I.M. (1994). The ecological effects of individual exposures and nonlinear disease dynamics in populations. *American Journal of Public Health*, 84, 836-842.
- Koopman J.S., Lynch J.W. (1999). Individual causal models and population system models in epidemiology. *American Journal of Public Health*, 89(8), 1170-1174.
- LCDC. (1998). Canadian National Report on Immunization. *Paediatrics and Child Health*, 4 (Suppl C).
- Lee E.T. (1992). *Statistical Methods for Survival Data Analysis*. Wiley Interscience, 2nd edition, New York.
- Liljeros F., Edling C.R., Amaral L.A., Stanley H.E., Aberg Y. (2001). The web of human sexual contacts. *Nature*. 411, 907-908.
- May R.M. (1976). Simple mathematical models with very complicated dynamics. *Nature*. Jun 10, 261(5560), 459-467.
- McDonnell I.F., Jorm L.R. (1994). Measles epidemic in western Sydney. *New South Wales Public Health Bulletin*, 5(6), 59-60.
- Olsen L.F., Truty G.L., Schaffer W.M. (1988). Oscillations and chaos in epidemics: a nonlinear dynamic study of six childhood diseases in Copenhagen, Denmark. *Theoretical Population Biology*, 33(3), 344-370.
- Onorato I.M., Wassilak S.G., Meade B. (1992). Efficacy of whole-cell pertussis vaccine in preschool children in the United States. *JAMA : the journal of the American Medical Association*, 267(20), 2745-2749.
- Pastor-Satorras R., Vespignani A. (2001). Epidemic spreading in scale-free networks. *Physical Review Letters*, 86(14), 3200-3203.
- Philippe P., Mansi O. (1998). Nonlinearity in the epidemiology of complex health and diseases process. *Theoretical Medicine and Bioethics*, 19, 591-607.
- Philippe P. (1999). The scale-invariant spatial clustering of leukemia in San Francisco. *Journal of Theoretical Biology*, 199, 371-381.
- Philippe P. (2000). Epidemiology and self-organized critical systems: An analysis in waiting times and diseases heterogeneity. *Nonlinear Dynamics, Psychology, and Life Sciences*, 4(4), 275-295.

Philippe P., Garcia M.L., West B.J. (2004). Evidence of "essential uncertainty" in emergency-ward length of stay (EWLS). *Fractals*, 12(2), 197-209.

Redd S.C., King G.E., Heath J.L., Forghani B., Bellini W.J., Markowitz L.E. (2004). Comparison of vaccination with measles-mumps-rubella vaccine at 9, 12, and 15 months of age. *The Journal of infectious diseases*, 189 (Suppl.1), S116-122.

Rhodes C.J., Anderson R.M. (1996a). A scaling analysis of measles epidemics in a small population. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 351(1348), 1679-88.

Rhodes C.J., Anderson R.M. (1996b). Power laws governing epidemics in isolated populations. *Nature*, 381(6583), 600-602.

Rhodes C.J., Jensen H.J., Anderson R.M. (1997). On the critical behaviour of simple epidemics. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 264(1388), 1639-1646.

Rhodes C.J., Butler A.R., Anderson R.M. (1998). Epidemiology of communicable diseases in small population. *Journal of Molecular Medicine*, 76, 111-116.

Schneeberger A., Mercer C.H., Gregson S.A., Ferguson N.M., Nyamukapa C.A., Anderson R.M., Johnson A.M., Garnett G.P. (2004). Scale-free networks and sexually transmitted diseases: a description of observed patterns of sexual contacts in Britain and Zimbabwe. *Sexually transmitted diseases*, 31(6), 380-387.

Statistics Canada. (1971-1978). Health and Welfare Division, Public Health Section. *Annual Report of Notifiable Diseases*. Catalogue number 82-201.

Tam T.W.T., Bentsi-Enchill A. (1998). The return of the 100-day cough: resurgence of pertussis in the 1990s. *Canadian Medical Association journal*, 159(6), 695-696.

Taylor L.R. (1961). Aggregation, variance and the mean. *Nature*, 189, 732-735.

Torvaldsen S., McIntyre P. (2003). Do variations in pertussis notifications reflect incidence or surveillance practices? A comparison of infant notification rates and hospitalisation data in NSW. *New South Wales public health bulletin*, 14(4-5), 81-84.

Trottier H., Philippe P. (2004a). Time series analysis of some childhood infectious diseases before and after mass vaccination in Canada: Box-Jenkins modelling. (In preparation).

Trottier H., Philippe P. (2004b). Qualité des données de surveillance de la coqueluche, de la rougeole, de la rubéole et des oreillons au Canada. (In preparation).

Vitek C.R., Aduddell M., Brinton M.J., Hoffman R.E., Redd S.C. (1999). Increased protections during a measles outbreak of children previously vaccinated with a second dose of measles-mumps-rubella vaccine. *The Pediatric infectious disease journal*, 18(7), 620-623.

West B.J., Shlesinger M. (1990). The noise in natural phenomena. *American Scientist*, 78, 40-45.

## APPENDIX

### Estimation of the epidemic size distribution: measles (1990-2002)

The table below shows the sizes  $s$  of epidemics and the epidemic size distributions (two are shown, the strict distribution according to  $s$  and a cumulative display) used to analyze the measles time series data (1991-2002). The table shows that 30 distinct epidemics (from 1 case up to 2 400 cases) occurred during the whole time period from 1991 to 2002. More specifically, there were exactly 2 large epidemics of 2 300-2 399 cases, 2 others of 500-599 cases, 1 epidemic of 400-499 case, etc. The epidemic size distribution used to fit the models is the cumulated number of epidemics of size  $s$ . The latter is the frequency of epidemics of at least size  $s$  (the large epidemics are cumulated with the smaller ones). The plot of figure 1b (Results section) shows the number of epidemics of at least size  $s$ . In this figure, one can clearly see that there are 2 epidemics of at least 2 300 cases, 4 epidemics of at least 500 cases (including epidemics of sizes 2 300 to 2 399 and those of sizes 500 to 599). The cumulating procedure has been generalized to the other values of the distribution.

**Table A. Epidemic size distribution, measles (1991-2002)**

Size ( $s$ )	Number of epidemics of size $s$	Epidemic size distribution ( $N \geq s$ ) (frequency of epidemics of at least size $s$ )
2300-2399	2	2
500-599	2	4
400-499	1	5
300-399	1	6
200-299	2	8
100-199	2	10
1-99	20	30

## CHAPITRE 4

---

### DISCUSSION ET CONCLUSION

Le projet de recherche a consisté à analyser les séries chronologiques des maladies infectieuses transmissibles communes (rougeole, oreillons, rubéole, coqueluche) au Canada afin de caractériser leur dynamique à l'échelle populationnelle et d'examiner les bases de leur « comportement » temporel en termes de loi d'organisation. De plus, pour permettre d'interpréter adéquatement les résultats de la recherche, une analyse de la qualité des MADO a été entreprise. Les séries chronologiques ont été analysées pour les périodes de 1953-1959, 1970-1976 et 1991-2002 en vue de couvrir les périodes de prévacination et de postvacination.

L'analyse de la qualité des données a d'abord permis de faire ressortir des lacunes importantes (article 1, chapitre 3). En effet, les données sont fortement sous-estimées. Malgré ce biais important, il n'y a pas de raison de croire, puisque l'on s'intéresse à la dynamique, que les données posent un problème important dans notre étude, dans la mesure où elles font ressortir les tendances, c'est-à-dire les variations de l'incidence des maladies au cours du temps (article, chapitre 3). En conséquence, toute variation dans un sens ou dans un autre dans les données, au cours du temps, serait le reflet d'une variation équivalente véritable dans la population. L'hypothèse est ainsi posée que les fluctuations au sein des données, sur de courtes périodes, reflètent les fluctuations réelles de l'incidence des maladies.

Étant donné que le niveau de sous-enregistrement et les facteurs affectant la déclaration des cas ne peuvent être considérés constants sur de longues périodes, il semblait approprié de faire l'analyse de la dynamique des maladies sur des périodes courtes. Par exemple, entre 1953 et 1959, on peut supposer que les facteurs affectant le niveau de sous-déclaration des données sont restés relativement stables, permettant ainsi d'avancer l'hypothèse que les fluctuations des sous-périodes reflètent les fluctuations réelles. De plus, l'analyse sur de courtes périodes nous semblait d'autant plus appropriée que le système de surveillance a connu des changements importants depuis 1924. En effet, les variations dans les pratiques de déclaration, le manque d'intégrité du système entre les provinces, les changements dans les définitions des maladies, etc., ont entraîné, inévitablement, des biais qualitatifs et quantitatifs qui, à leur tour, ont eu une incidence sur la représentativité et sur l'intégrité des données. Les séries chronologiques ont été choisies en tenant compte de ces changements afin qu'elles soient les plus homogènes possibles. Notons, cependant, que la surestimation des cas, tels les oreillons au cours des dernières années, est sans doute plus problématique dans la mesure où les tendances que l'on trouve au sein des données sont également issues de celles d'autres maladies se présentant avec une symptomatologie similaire. Malgré la segmentation dont les séries ont fait l'objet pour des motifs d'homogénéité, nous nous sommes assurés que la taille des séries respectait néanmoins les exigences des méthodes d'analyses auxquelles elles se sont prêtées. D'ailleurs, la cohérence et la plausibilité des résultats obtenus militent largement en faveur de la validité des données.

La modélisation stochastique de Box-Jenkins a permis de modéliser les séries en cherchant les paramètres qui reproduisent le plus fidèlement possible les variations dans le temps (chapitre 3, article 2). Le modèle général qui ajuste le mieux les séries (en log)

est le modèle SARIMA  $(1,0,0)(0,1,1)_{12\text{ou}13}$ . Une différenciation saisonnière est généralement requise pour rendre les séries stationnaires. Le paramètre autorégressif (AR) de premier degré et le paramètre de moyenne mobile (MA) saisonnier (12 ou 13 degrés) sont ressortis comme les paramètres qui, de manière générale, ajustent le mieux les séries. La modélisation obtenue signifie que le nombre de cas de maladies infectieuses qui apparaît au cours d'un mois est donné par le nombre de cas qui apparaît le mois précédent et par des facteurs aléatoires saisonniers.

Quoi qu'il en soit, certaines procédures méritent d'être discutées concernant la modélisation ARIMA des séries. D'abord, la transformation logarithmique des données a été appliquée pour rendre les séries homogènes (indépendance de la moyenne et de la variance). Puisque l'intérêt de notre analyse est de comparer les modèles entre eux, la transformation logarithmique a été appliquée à toutes les séries en dépit du fait que cette transformation n'était peut-être pas la plus adéquate pour quelques séries (Annexe II). Quoique le type de transformation (racine carrée, log, réciproque) ne change généralement pas le modèle identifié, le cas de la série de rubéole de 1953-1959 s'est avéré distinct. Comme nos séries sont courtes, il nous a semblé approprié d'appliquer la transformation logarithmique à l'ensemble des séries. Par ailleurs, la différenciation saisonnière des séries a semblé moins appropriée pour les séries de coqueluche de 1970-1976, et de rubéole et de rougeole de 1991-2002 (Annexe III). En effet, dans un contexte de vaccination de masse, les cycles saisonniers peuvent être moins perceptibles (Anderson et al., 1991). La différenciation saisonnière a tout de même été appliquée à l'ensemble des séries, étant donné la présence à peu près générale d'une dérive (*drift*) ou tendance saisonnière (moins perceptible pour certaines séries).

Les séries de rubéole de 1953-1959 et de rougeole de 1970-1976 s'écartent quelque peu des modèles trouvés, dans la mesure où une différenciation régulière a aussi été appliquée. Au Canada, un des plus grands pics épidémiques de rubéole jamais apparus a émergé en 1956. C'est aussi le cas de la série de rougeole qui montre un pic épidémique considérable en 1970. Ce phénomène a semblé générer une non-stationnarité des séries. On a donc également appliqué une différenciation régulière à ces séries. En dépit de cette opération, le modèle de base est resté le même. Cependant, étant donné que la nécessité d'appliquer une différenciation régulière n'était pas claire, ces résultats doivent être interprétés avec prudence.

L'application d'une composante de moyenne mobile (MA) régulière de premier degré a aussi été considérée pour ajuster les modèles, car la série de rubéole de 1953-1959, sans différenciation régulière, semblait montrer un bon ajustement avec les paramètres AR, MA régulier, et MA saisonnier. Comme les autres séries de rubéole ne se pliaient pas convenablement à ce type de modèle, et que les données pour la période en question sont de moins bonne qualité parce que plusieurs provinces ont graduellement arrêté le signalement de la rubéole à partir de 1956 jusqu'en 1959 (chapitre 2), le paramètre MA régulier a été exclu des modèles. Il faut également mentionner que l'introduction de la composante MA régulière de premier degré conduisait souvent, pour les autres séries d'infections, à une augmentation de la variance résiduelle.

De façon générale, les estimations des paramètres du modèle sont relativement semblables, mais il semble difficile de comparer ces derniers en raison de la qualité des données. De plus, les données des séries de 1953-1959 et de 1970-1976 portent sur treize observations annuelles, alors que les données de 1991-2002 portent sur douze

observations annuelles. Ce facteur peut aussi introduire une variation au niveau des paramètres estimés.

Au terme des analyses, les séries de rougeole, de coqueluche, d'oreillons et de rubéole partagent la même structure stochastique. En effet, de manière générale, le modèle qui ajuste le mieux l'ensemble des séries est le modèle SARIMA  $(1,0,0)(0,1,1)_{12}$  ou  $13$ . La modélisation de Box-Jenkins s'est avérée intéressante dans la mesure où les maladies infectieuses étudiées partagent toutes la même dynamique stochastique de base, malgré le fait qu'elles soient biologiquement différentes et en dépit de l'impact de facteurs présumés importants comme la vaccination de masse. Cette constatation, à elle seule, est remarquable. Elle constitue un progrès notable sur la modélisation de séries d'infections faites par Helfenstein (1986). Il suit cependant que l'objectif de classer les maladies en fonction de leur propriété stochastique, en tenant compte des périodes de prévacination et de postvacination, s'est avéré inatteignable. Il apparaît, en effet, que la modélisation de Box-Jenkins identifie la dynamique populationnelle des infections tout en restant insensible aux caractéristiques biologiques spécifiques des maladies et aux effets de la vaccination. Cette modélisation est donc strictement collective et s'apparente, de ce fait, aux postulats de la modélisation SEIR. Ainsi, et bien que le modèle SEIR assume une connaissance des caractéristiques biologiques de la maladie, il simule une série chronologique d'infections au niveau collectif sur la base des valeurs moyennes de ses paramètres principaux. La méthode de Box-Jenkins en fait autant, mais à rebours, c'est-à-dire à partir d'une série réelle. L'une et l'autre profilent néanmoins une série moyenne à l'échelle populationnelle.

Comme la méthode de Box-Jenkins profilait une dynamique strictement collective sans pouvoir en caractériser plus amplement les assises, il devenait

particulièrement heuristique de soumettre les séries à une modélisation par la loi de puissance inverse capable, en principe, de saisir les bases organisationnelles des différentes épidémies. Aussi, l'analyse de la distribution des tailles d'épidémies a-t-elle permis d'étudier les propriétés d'échelle des maladies infectieuses concernées (chapitre 3, article 3). Les résultats montrent que la distribution des tailles d'épidémies, pour l'ensemble des séries de coqueluche et d'oreillons, et pour les séries de rougeole dans un contexte de population peu ou pas vaccinée (séries 1953-1959, 1970-1976), adhère le plus souvent à une loi exponentielle. *A contrario*, la rubéole, dans un contexte avec ou sans vaccination de masse, est caractérisée par une loi de puissance inverse (séries 1953-1959, 1970-1976, 1991-2002). La rougeole semble également associée à une loi de puissance inverse quand la population est largement vaccinée (série 1991-2002).

Il convient de discuter des biais pouvant affecter ces résultats. Il y a d'abord la possibilité d'une mauvaise classification des épidémies. En effet, dans les grandes populations, contrairement à ce qui se passe dans les petites, il n'y a généralement pas d'extinction de la maladie; ce phénomène rend donc plus difficile de différencier les épidémies les unes des autres. Aussi, pour faire une analyse plus approfondie des résultats, la distribution des tailles d'épidémie a-t-elle été analysée en fonction de deux schémas de classification. Le premier schéma a consisté à estimer le nombre et la taille des épidémies à partir du nombre cumulé des cas au sein de chaque pic épidémique, limité aux deux extrémités par des creux épidémiques (voir article 3, chapitre 3). Dans un deuxième temps, l'analyse a été reprise à partir de la distribution du nombre de cas mensuels (ou par période de 4 semaines). Ce faisant, les deux schémas ont permis de confirmer les résultats pour toutes les séries analysées. En effet, les résultats ont été exactement les mêmes à l'exception de la première série de coqueluche (1953-1962).

Selon le premier schéma de classification, la coqueluche adhère à une loi de puissance inverse alors qu'avec le second, elle adhère plutôt à une distribution exponentielle. Il semble donc difficile de conclure pour cette série. Cependant, puisque les autres séries de coqueluche se distribuent en fonction d'une loi exponentielle, il semble plus probable que la série des années 1950 adhère également à une loi exponentielle.

De plus, il faut mentionner que la dimension spatiale des données n'a pas été considérée. Il suit que des foyers infectieux indépendants ont pu être regroupés. Par exemple, si cinq cas de rougeole sont survenus au Canada dans cinq provinces différentes, ces derniers ont été classés comme foyer infectieux unique, et non comme foyers indépendants. Aussi, des interactions spatiales, tant aux niveaux provincial que régional devraient être prises en compte si l'on veut tenir compte de l'interférence de l'espace avec le temps. Ce n'était pas le but de cette thèse. Aussi, des recherches futures sur le rôle de la dynamique spatiale dans l'émergence d'IPL devraient-elles permettre de clarifier la signification de l'ajustement de la loi de puissance inverse à la distribution des épidémies.

Il convient également de dire un mot sur la longueur des séries. Des séries aussi courtes rendent difficile la discrimination entre une loi exponentielle et une loi de puissance inverse. Des séries plus longues auraient, en principe, donné plus d'assurance à la reconnaissance d'une loi de puissance inverse dont les valeurs extrêmes (épidémies de très grandes tailles) ne peuvent être reconnues qu'en présence d'un temps d'attente assez long qui permet l'apparition de telles épidémies. On doit néanmoins tempérer cette limite par le fait que Rhodes et al. (1996a; 1996b; 1997; 1998), avec des séries plus longues que les nôtres, a reconnu un déficit d'épidémies de très grandes tailles. Autrement dit, la durée de nos séries ne semblerait pas constituer un handicap important.

L'ajustement des séries des maladies contagieuses à une loi exponentielle ou de puissance inverse permet d'approfondir notre connaissance de leur dynamique de transmission. L'ajustement de la taille des épidémies à une loi exponentielle montre que leur distribution est le reflet d'une seule échelle typique. La loi de distribution exponentielle suggère que les taux (en angl : «hazard») de contacts qui génèrent les épidémies ont une valeur moyenne qui varie peu d'un individu à l'autre (ci-après nommés, les nœuds); autrement dit, il y a peu de nœuds avec peu ou énormément de contacts (aux extrémités de la distribution des contacts), de sorte que la distribution des contacts est homogène. Par conséquent, cette observation fait état de l'existence d'un réseau homogène sous-jacent. L'ensemble des interactions (contacts) entre les susceptibles et les infectieux dans une population se ferait donc principalement au hasard. Entre 1953-1959, toutes les séries s'ajustent à une loi exponentielle à l'exception de la série de rubéole. La grande contagiosité de ces maladies pourrait expliquer la distribution exponentielle des tailles d'épidémie. La rougeole, la coqueluche et les oreillons sont des maladies plus contagieuses que la rubéole. Leur force de transmissibilité semble assez élevée pour prévenir une éventuelle hétérogénéité du réseau de contacts. En d'autres mots, dans le cas de la rougeole, de la coqueluche et des oreillons, la transmissibilité est assez forte pour contrecarrer un éventuel effet de topologie hétérogène du réseau. Par contre, pour la rubéole, maladie relativement peu contagieuse, l'hétérogénéité du réseau peut devenir importante pour expliquer une distribution différente des patterns épidémiques. En effet, plus faible est la transmissibilité de la maladie, plus grand est l'impact des interactions sociales (Rhodes et al., 1997). La topologie hétérogène des réseaux de contacts que nous avons trouvée pourrait donc être en mesure d'expliquer la distribution des épidémies de rubéole qui se

trouve bien ajustée par une loi de puissance inverse et ce, même dans un contexte de population non vaccinée. Soulignons, incidemment, que l'âge moyen à l'infection d'une rubéole est généralement plus élevé que pour les autres maladies étudiées, ce qui peut induire une organisation sociale différente des contacts individuels.

Les épidémies d'oreillons et de coqueluche sont ajustées par une loi exponentielle dans un contexte de vaccination de masse. Or, la coqueluche connaît une recrudescence depuis quelques années. Tel que souligné à l'article 3 (chapitre 3), les adultes et les adolescents constituent des réservoirs importants de transmission chez les jeunes enfants. Ce phénomène pourrait expliquer une structure organisationnelle des taux de contacts plus homogène dans la population. Dans le cas des oreillons, la mauvaise qualité des données (problème de surestimation) a pu altérer la dynamique de masse et occulter une loi de puissance, même dans un contexte de forte vaccination. Soit que la mauvaise qualité des données dans les années 1990 empêche d'ajuster convenablement une loi de puissance inverse aux épidémies, soit que cette maladie est fondamentalement différente de la rougeole et de la rubéole.

La loi de puissance inverse, celle qui semble dominer la distribution des épidémies de rougeole dans un contexte de population hautement vaccinée ou encore de rubéole, s'applique à des réseaux hétérogènes où quelques nœuds seulement sont reliés à presque tous les nœuds, et dominent les contacts. Il ne s'agit pas là de réseaux aléatoires, comme ceux de la coqueluche et les oreillons. La distribution des épidémies semble plutôt court-circuiter les échelles (pas d'échelle typique). Dans un réseau hétérogène, dit aussi «invariant d'échelle», même si une majorité de nœuds est mis hors circuit, par la vaccination par exemple, les supernœuds (les individus avec un grand nombre de contacts) continuent à assurer la connexité du réseau (Barabasi et al., 2003). La

robustesse du réseau tient donc à sa topologie. La suppression aléatoire de nœuds dans ce cas perturbe surtout les petits nœuds, qui sont beaucoup plus nombreux que les supernœuds.

L'approche classique de santé publique consiste à immuniser les personnes de façon aléatoire; elle risque d'échouer par conséquent si elle néglige les supernœuds dans le cas des maladies dont les épidémies sont invariantes d'échelle. Cela signifie que la quasi-totalité de la population devrait être vaccinée pour une efficacité vaccinale optimale avec ce type d'épidémies. Par exemple, on peut se demander pourquoi on n'a pas réussi à éliminer la rubéole qui, selon la théorie conventionnelle, devrait l'avoir été avec un seuil de vaccination de 82-87 % (voir Anderson et al., 1991, p.88, pour le seuil critique de vaccination requis pour éradiquer certaines maladies infectieuses). Ce seuil est pourtant dépassé depuis plusieurs années au Canada (article 3, chapitre 3). La théorie conventionnelle du seuil de vaccination est basée sur l'homogénéité des contacts; elle néglige donc des facteurs importants de transmission de la maladie, comme l'organisation du réseau. Il est intéressant de faire l'hypothèse que deux populations semblables ayant des réseaux organisés différemment, l'un homogène et l'autre hétérogène, auront des comportements distincts en regard de l'efficacité de la vaccination.

Il convient, enfin, de mentionner que l'immigration constitue un facteur contemporain important qui contraint la lutte à l'élimination des maladies infectieuses communes de l'enfance. Une proportion non négligeable d'immigrants qui entre au pays n'est pas immunisée contre ces maladies (Barnett et al., 2002; Sheridan et al., 2002; Grosman et al., 1999). Depuis 1990, plus de trois millions d'immigrants sont entrés au Canada (plus de 200 000 chaque année) (Statistique Canada, 2004). Certains immigrants

peuvent également être infectés et la source d'épidémies importantes. Par exemple, Rangel et al. (1999) et Coughlan et al. (2002) ont mis en évidence l'émergence d'épidémies importantes de rougeole et de rubéole à la suite de l'entrée d'immigrants infectés aux États-Unis et en Irlande. En somme, l'introduction des nouveaux immigrants peut non seulement générer des épidémies nouvelles, mais également «bousculer» les patterns épidémiques existant; par ailleurs, une immigration de nœuds infectés, d'origine ethnique, pourrait modifier l'organisation collective des épidémies.

Nous concluons cette thèse en affirmant que l'élimination rationnelle des maladies contagieuses de l'enfance ne pourra passer que par une meilleure compréhension de la topologie des réseaux de contacts propres à chacune. Les réseaux invariants d'échelles ont une propriété commune : ils dépendent de quelques supernœuds; ils sont donc vulnérables aux attaques ciblées (Barabasi et al., 2003). Des recherches futures sur la topologie des réseaux de contacts ne pourront être que bénéfiques pour éliminer plus efficacement les maladies transmissibles.

Plus généralement, nous pouvons conclure que le paradigme linéaire traditionnel de l'analyse épidémiologique semble comporter des limites qui ne peuvent plus être ignorées. Une alternative fructueuse semble être la mise en contexte qui peut apporter un éclairage neuf et offrir des solutions plus appropriées aux problèmes complexes de l'épidémiologie contemporaine. A l'opposé des méthodes traditionnelles, l'approche « syndémique » (Centers for Disease Control and Prevention, 2002) semble une alternative intéressante mettant l'accent sur l'étude des liens entre les problèmes de santé et des interconnexions au niveau de la communauté.

## BIBLIOGRAPHIE

---

- Allard R. (1998). Use of time-series analysis in infectious disease surveillance. *Bulletin of the World Health Organization*, 76(4), 327-333.
- Allen J., Shaffer W., Rosko D. (1993). Chaos reduces species extinction by amplifying local population noise. *Nature*, 364, 229-232.
- Anderson R.M., Grenfell B.T., May R.M. (1984). Oscillatory fluctuations in the incidence of infectious disease and the impact of vaccination: time series analysis. *The Journal of Hygiene*, 93(3), 587-608.
- Anderson R.M., May R.M. (1991). *Infectious diseases of humans. Dynamics and control*. Oxford University Press. New York.
- Anderson R.M. (1994). The Croonian lecture, 1994. Populations, infectious disease and immunity: a very non linear world. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, 346, 457-505.
- Aron J.L. (1990). Multiple attractors in the response to a vaccination program. *Theoretical Population Biology*, 38, 58-67.
- Association of Schools of Public Health. (1999). *Demonstrating excellence in academic public health practice*. Prepared by The Association of Schools of Public Health Council of Public Health Practice Coordinators, Washington, DC.
- Bailey N.T.J. (1975). *The mathematical theory of infectious diseases and its applications*, London, Griffin.
- Bak P., Sneppen K. (1993). Punctuated equilibrium and criticality in a simple model of evolution. *Physical Review Letters*, 71, 4083-4086.
- Barabasi A.L., Albert R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509-512.
- Barabasi A.L., Bonabeau E. (2003). Réseaux invariants d'échelle. *Pour la Science*, 314, 58-63.
- Barnett E.D., Christiansen D., Figueira M. (2002). Seroprevalence of measles, rubella, and varicella in refugees. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 35(4), 403-408.

Bartlett M.S. (1957). Measles Periodicity and Community Size. *Journal of the Royal Statistical Society. Series A (general)*, 20(1), 48-70.

Bartlett M.S. (1960). The critical community size for measles in the U.S. *Journal of the Royal Statistical Society. Series A (general)*, 123(1), 37-44.

Bolker B. (1993). Chaos and complexity in measles models: a comparative numerical study. *IMA Journal of Mathematics Applied in Medicine and Biology*, 10(2), 83-95.

Bolker B.M., Grenfell B.T. (1993). Chaos and biological complexity in measles dynamics. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 251(1330), 75-81.

Bolker B.M., Grenfell B.T. (1996). Impact of vaccination on the spatial correlation and persistence of measles dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 12648-12653.

Bowie C., Prothero D. (1981). Finding cause of seasonal diseases using time series Analysis. *International Journal of Epidemiology*, 10(1), 87-92.

Bureau fédéral de la statistique. (1924-1952). Division de la santé et du bien-être. Section de la santé publique. *Sommaire de maladies à déclaration obligatoire*. Annuel.

Bureau fédéral de la Statistique. (1953 à 1970). Division de la santé et du bien-être. Section de la santé publique. *Rapport annuel sur les maladies à déclaration obligatoire*. Catalogue 82-201, annuel.

Caplan C.E. (1999). Mumps in the era of vaccines. *Canadian Medical Association Journal*, 160(6), 865-6.

Centers for Disease Control and Prevention. (1991). Cholera associated with imported frozen coconut milk –Maryland, 1991. *MMWR. Morbidity and Mortality Weekly Report*, 39, 266-273.

Centers for Disease Control and Prevention. (1993). Resurgence of pertussis - United States. *MMWR. Mortality and Morbidity Weekly Report*, 42, 952-953.

Centers for Disease Control and Prevention. (2002). *Introduction to the Syndemics Prevention Network*. Working Draft. <http://www.cdc.gov/syndemics>.

Choi K., Thacker S.B. (1981a). An evaluation of influenza mortality surveillance, 1962-1979. I. Time series forecasts of expected pneumonia and influenza death. *American Journal of Epidemiology*, 113, 215-226.

Choi K., Thacker S.B. (1981b). An evaluation of influenza mortality surveillance 1962-1979. II. Percentage of pneumonia and influenza deaths as an indicator of influenza activity. *American Journal of Epidemiology*, 113, 227-235.

Coughlan S., Connell J., Cohen B., Jin L., Hall W.W. (2002). Suboptimal measles-mumps-rubella vaccination coverage facilitates an imported measles outbreak in Ireland. *Clinical Infectious Diseases : an official publication of the Infectious Diseases Society of America*, 35(1), 84-86.

Dacunha-Castelle D. (1996). *Chemins de L'aléatoire. Le hasard et le risque dans la société moderne*. Champs, Flammarion.

Dietz K. (1976). The incidence of infectious diseases under the influence of seasonal fluctuations. *Lecture notes in Biomathematics*, 11, 1-15.

Earn D.J., Rohani P., Bolker B.M., Grenfell B.T. (2000). A simple model for complex dynamical transitions in epidemics. *Science*, 287(5453), 667-670.

Engbert R., Drepper F. (1994). Chance and chaos in population biology – models of recurrent epidemics and food chain dynamics. *Chaos, Solitons, and Fractals*, 4, 1147-1169.

Farizo K.M., Cochi S.L., Zell E.R. (1992). Epidemiological features of pertussis in United State. *Clinical Infectious Diseases : an Official Publication of the Infectious Diseases Society of America*, 14, 708-719.

Feldmann H., Czub M., Jones S., Dick D., Garbutt M., Grolla A., Artsob H. (2002). Emerging and re-emerging infectious diseases. *Medical Microbiology and Immunology*, 191(2), 63-74.

Garrett L. (1994). *The Coming plague*. Penguin Books.

Giesecke J. (1994). *Modern infectious disease epidemiology*. Edward Arnold. London.

Gomes M.C., Gomes J.J., Paulo A.C. (1999). Diphtheria, pertussis, and measles in Portugal before and after mass vaccination: a time series analysis. *European Journal of Epidemiology*, 15, 791-798.

Gottman J.M. (1981). *Time series analysis: a comprehensive introduction for social scientists*. New York, Cambridge University Press.

Grenfell B.T, Bolker B.M., Kleczkowski A. (1993). *In epidemic models: their structure and relation to data*. Ed. Mollison, D. Cambridge University Press, Cambridge, U.K. pp.248-268.

Grenfell B.T., Bolker B.M., Kleczkowski A. (1995). Seasonality and extinction in chaotic metapopulations. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 259, 97-103.

Grenfell B.T. (2000). Measles as a testbed for characterising nonlinear behaviour in Ecology. In *Chaos in real data: The analysis of non-linear dynamics from short ecological time Series*. Edited by JN Perry, RH Smith, IP Woivod and DR Morse. Kluwer Academic Publishers, 49-72.

Grmek M.D. (1990). *Histoire du sida : début et origine d'une pandémie actuelle*. Paris, Payot.

Grossman D.W., Hans L.M., Glazier R. (1999). Geographic origin and risk for congenital infection in a Canadian inner city: findings and implications for policy. *Canadian journal of public health*, 90(6), 385-388.

Hamer W.H. (1906). Epidemic disease in England. *The Lancet*, i, 733-739.

Helfenstein U. (1986). Box-Jenkins modelling of some viral infectious diseases. *Statistics in Medicine*, 5(1), 37-47.

Keeling M., Grenfell B. (1999). Stochastic dynamics and a power law for measles variability. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 354(1384), 769-776.

Kermack W.O., McKendrick A.G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society*, A 115, 700-721.

Koopman J.S., Longini I.M. (1994). The ecological effects of individual exposures and nonlinear disease dynamics in populations. *American Journal of Public Health*, 84, 836-842.

Koopman J.S. (1996). Emerging objectives and methods in epidemiology. *American Journal of Public Health*, 86(5), 630-632.

Koopman J.S., Lynch J.W. (1999). Individual causal models and population system models in epidemiology. *American Journal of Public Health*, 89(8), 1170-1174.

Koopman J.S. (2004). Modeling Infection Transmission. *Annual Review of Public Health*, 25, 303-326.

Lashley F.R. (2003). Factors contributing to the occurrence of emerging infectious diseases. *Biological Research for Nursing*, 4(4), 258-267.

Liljeros F., Edling C.R., Amaral L.A., Stanley H.E., Aberg Y. (2001). The web of human sexual contacts. *Nature*, 411, 907-908.

London W.P., Yorke J.A. (1973). Recurrent outbreaks of measles, chickenpox and mumps. I. Seasonal variation in contact rates. *American Journal of Epidemiology*, 98(6), 453-468.

McCleary R., Hay R.A. (1980). *Applied time series analysis for the social sciences*. Beverly Hills, Sage Publications.

McDowall D., McCleary R., Meidinger E.E., Hay R.A. (1983). *Interrupted time series analysis*. Third printing, Sage Publication.

Morse S.S. (1995). Factors in the emergence of infectious diseases. *Emerging Infectious Diseases*, 1(1), 7-15.

Olsen L.F., Shaffer W.M. (1990). Chaos versus noisy periodicity: alternative hypotheses for childhood epidemics. *Science*, 249, 499-504.

Philippe P., Mansi O. (1998). Nonlinearity in the epidemiology of complex health and disease processes. *Theoretical Medicine*, 19, 591-607.

Philippe P. (1999a). Épistémologie de la recherche causale et épidémiologie des systèmes complexes. *Revue d'Épidémiologie et de Santé Publique*, 47(5), 474-477.

Philippe P. (1999b). The scale-invariant spatial clustering of leukemia in San Francisco. *Journal of Theoretical Biology*, 199(4), 371-381.

Philippe P., Garcia M.L., West B.J. (2004). Evidence of "essential uncertainty" in emergency-ward length of stay (EWLS). *Fractals*, 12(2), 197-209.

Rangel M.C., Sales R.M., Valeriano E.N. (1999). Rubella outbreaks among Hispanics in North Carolina: lessons learned from a field investigation. *Ethnicity & disease*, 9(2), 230-236.

Rhodes C.J., Anderson R.M. (1996a). A scaling analysis of measles epidemics in a small population. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 351(1348), 1679-88.

Rhodes C.J., Anderson R.M. (1996b). Power laws governing epidemics in isolated populations. *Nature*, 381(6583), 600-602.

Rhodes C.J., Jensen H.J., Anderson R.M. (1997). On the critical behaviour of simple epidemics. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 264(1388), 1639-1646.

Rhodes C.J., Butler A.R., Anderson R.M. (1998). Epidemiology of communicable diseases in small population. *Journal of Molecular Medicine*, 76, 111-116.

Ross R. (1908). *Report on the prevention of the malaria in Mauritius*. London.

Santé Canada. (1992-2004). *Relevé des maladies transmissibles au Canada*. Bimensuel.

Santé Canada. (1993a). Conférence de concertation sur la coqueluche. *Relevé des maladies transmissibles au Canada*, 19, 124-135.

Santé Canada. (1997a). Direction générale de la protection de la santé - Laboratoire de lutte contre la maladie. Rapport sur l'immunisation au Canada, 1996. *Relevé des maladies transmissibles au Canada* – Supplément, 23 (23S4).

Santé et Bien-Être social Canada. (1975-1991). *Rapport hebdomadaire des maladies au Canada*.

Schenzle D. (1984). An age-structured model of pre- and post-vaccination measles transmission. *IMA Journal of Mathematics Applied in Medicine and Biology*, 1(2), 169-191.

Schneeberger A., Mercer C.H., Gregson S.A., Ferguson N.M., Nyamukapa C.A., Anderson R.M., Johnson A.M., Garnett G.P. (2004). Scale-free networks and sexually transmitted diseases: a description of observed patterns of sexual contacts in Britain and Zimbabwe. *Sexually transmitted diseases*, 31(6), 380-387.

Shaffer W.M., Olsen L.F., Truty G.L., Fulmer S.L. (1990). The case of chaos in childhood epidemics. In *The ubiquity of chaos*. Edited by Saul Krasner. American Association for the Advancement of Science. Washington. pp138-166.

Sheridan E., Aitken C., Jeffries D., Hird M., Thayalasekaran P. (2002). Congenital rubella syndrome: a risk in immigrant populations. *Lancet*, 359(9307), 674-675.

Smethurst D.P., Willimas H.C. (2001). Are hospital waiting lists self-regulating? *Nature*, 410, 652-653.

Soper M.A. (1929). The interpretation of periodicity in disease prevalence. *Journal of the Royal Statistical Society*, A92, 34-61.

Statistique Canada. (1971 à 1978). Division de la santé et du bien-être. Section de la santé publique. *Rapport annuel sur les maladies à déclaration obligatoire*. Catalogue 82-201, annuel.

Statistique Canada. (2004). *Site Internet de Statistique Canada, données d'immigration*, [http://estat.statcan.ca/content/francais/overview/pop-o\\_f.shtml](http://estat.statcan.ca/content/francais/overview/pop-o_f.shtml).

Taylor L.R. (1961). Aggregation, variance and the mean. *Nature*, 189, 732-735.

Trottier H., Philippe P. (2003). Deterministic Modeling of Infectious Diseases: Measles Cycles and the Role of Births and Vaccination. *The Internet Journal of Infectious Diseases*. Vol. 2 (2). <http://www.ispub.com>.

Turchin P. (1990). Non-linear modeling of time series data: limit cycles and chaos in forest insects, voles, and epidemics. Papers presented at the symposium: Does chaos exist in ecological systems. IUFRO XIX World Congress. Montréal, Canada.

WHO (2002). *The world health report 2001*.

## ANNEXE I

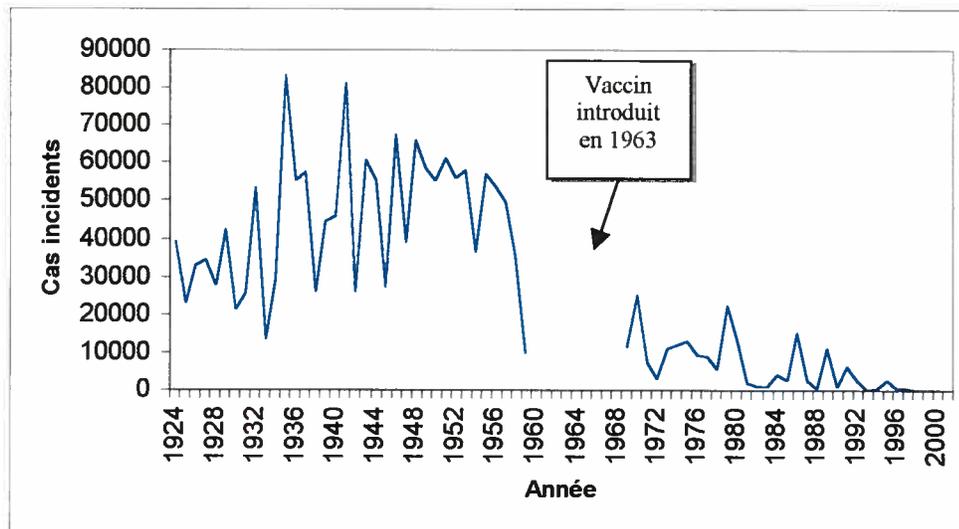
---

Séries chronologiques de rougeole, de coqueluche, d'oreillons et de rubéole au Canada :

- 1) Séries chronologiques annuelles, 1924-2002
- 2) Séries chronologiques analysées

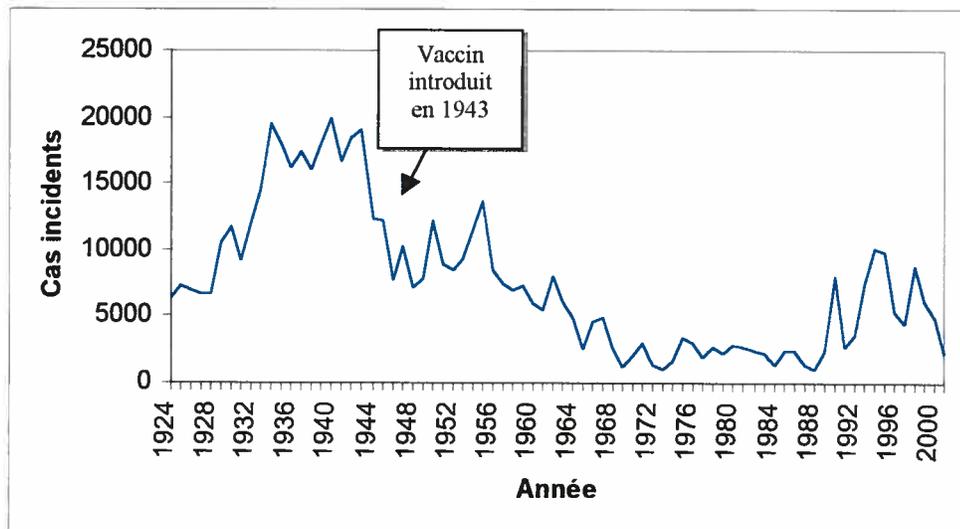
## 1. Séries chronologiques d'infections annuelles, 1924-2002

**Figure 1**  
**Nombre déclaré de cas de rougeole par année au Canada, 1924-2002\***



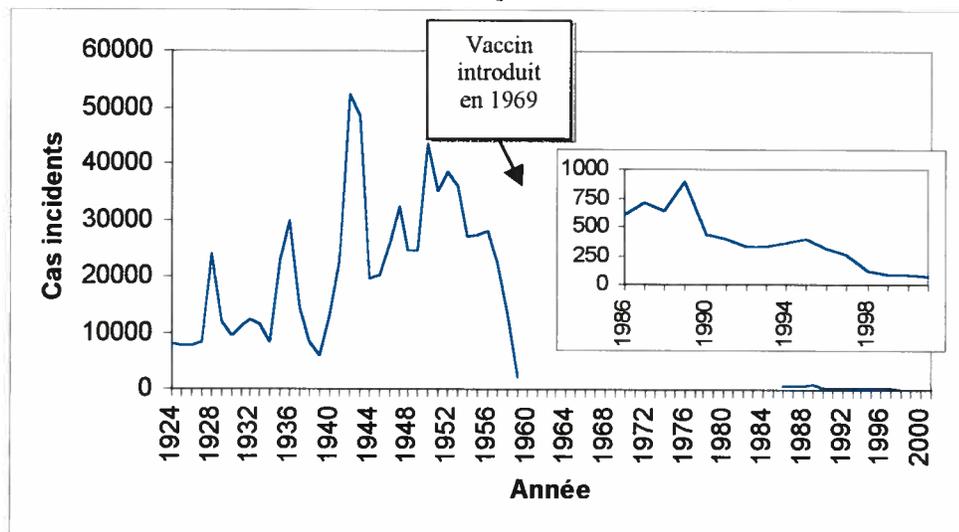
\* La déclaration des cas de rougeole n'était pas obligatoire de 1959 à 1968.  
 Source: Laboratoire de lutte contre la maladie, Santé Canada.

**Figure 2**  
**Nombre déclaré de cas de coqueluche par année au Canada, 1924-2002**



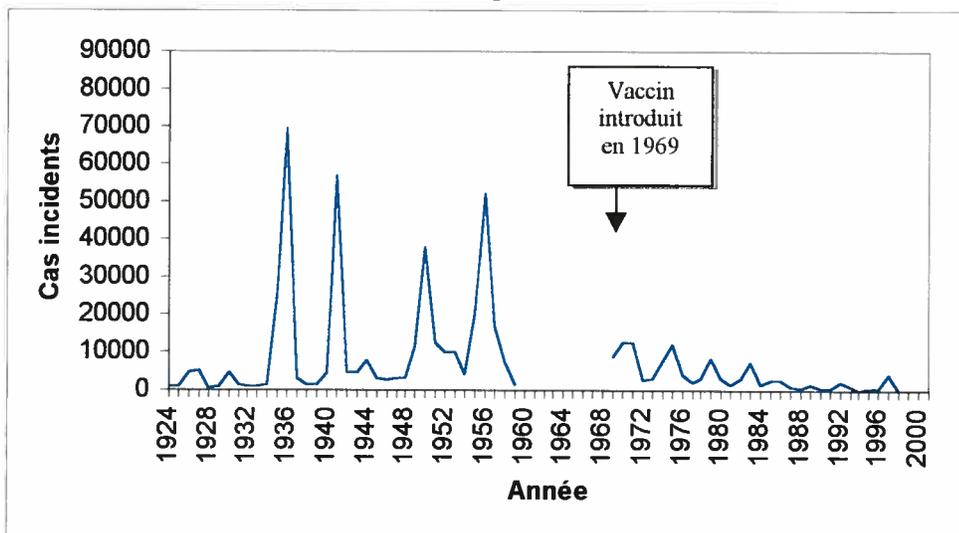
Source: Laboratoire de lutte contre la maladie, Santé Canada.

**Figure 3**  
**Nombre déclaré de cas d'oreillons par année au Canada, 1924-2002\***



\*La déclaration des cas d'oreillons n'était pas obligatoire de 1959 à 1986.  
 Source: Laboratoire de lutte contre la maladie, Santé Canada.

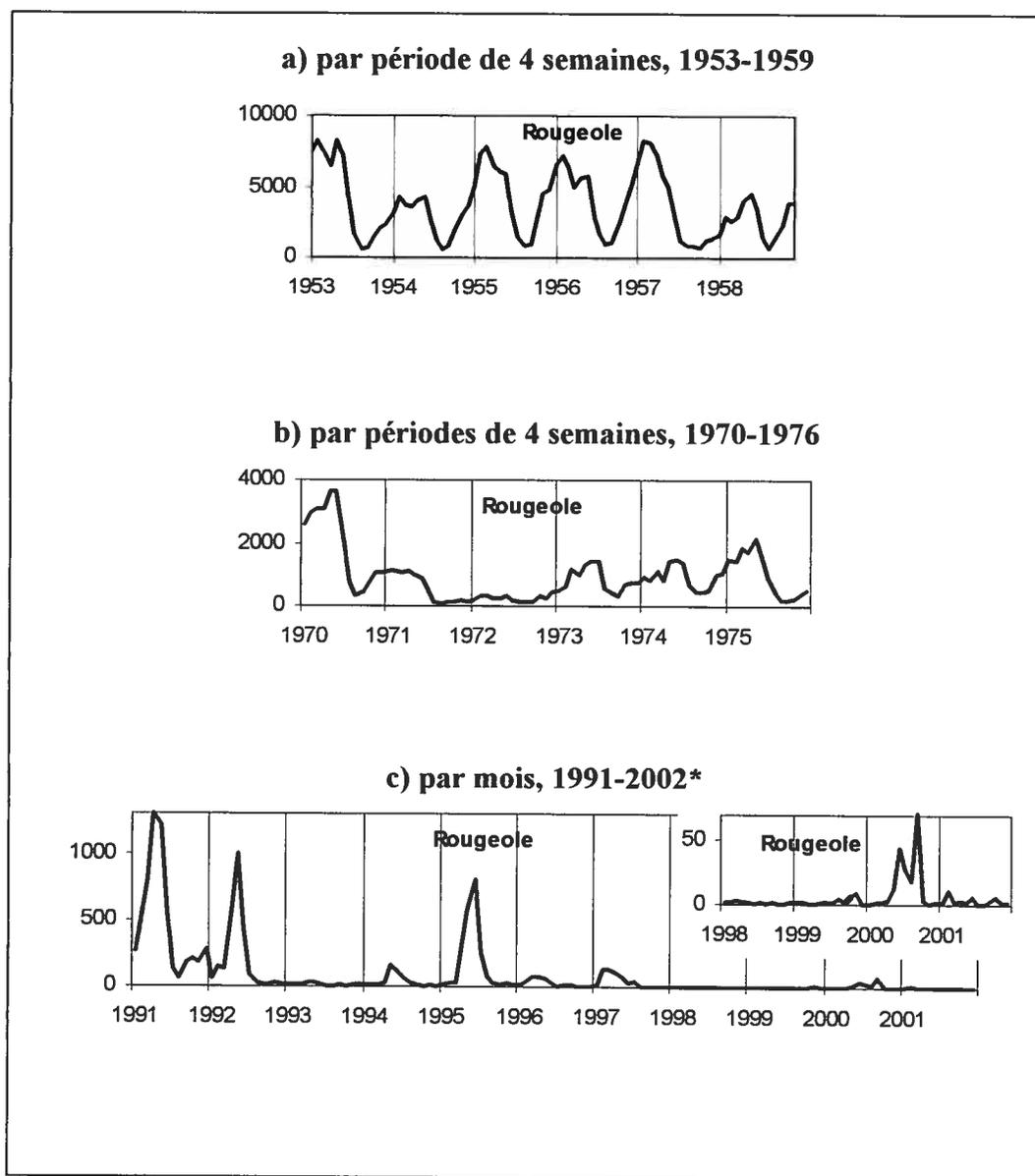
**Figure 4**  
**Nombre déclaré de cas de rubéole par année au Canada, 1924-2002\***



\*La déclaration des cas de rubéole n'était pas obligatoire de 1959 à 1968.  
 Source: Laboratoire de lutte contre la maladie, Santé Canada.

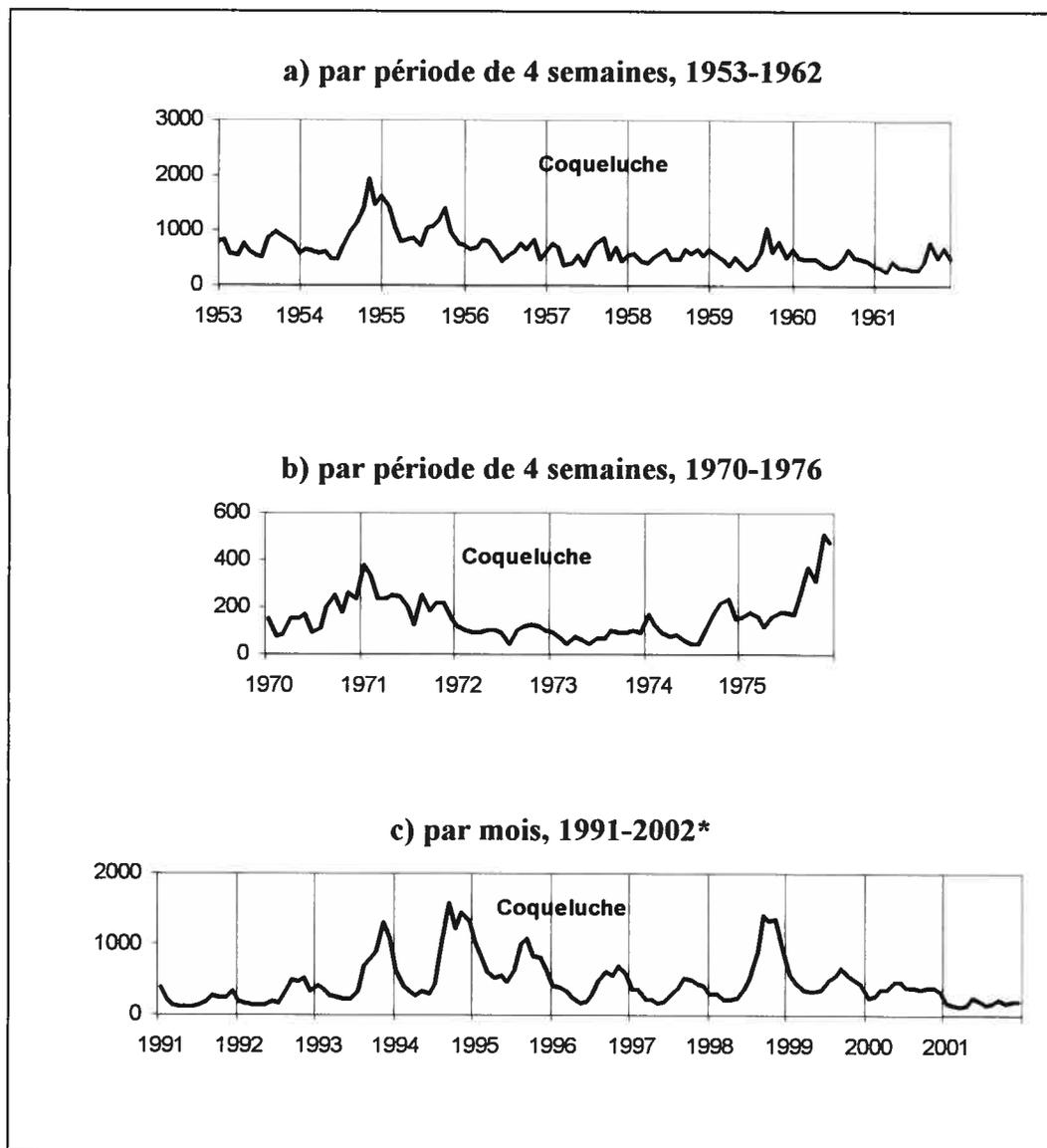
## 2. Séries chronologiques analysées

**Figure 5**  
**Nombre déclaré de cas de rougeole au Canada**



\*À partir de 1991, des définitions spécifiques pour chaque maladie sous surveillance ont servi de base à la déclaration des cas à l'échelle nationale.  
Source: Laboratoire de lutte contre la maladie, Santé Canada.

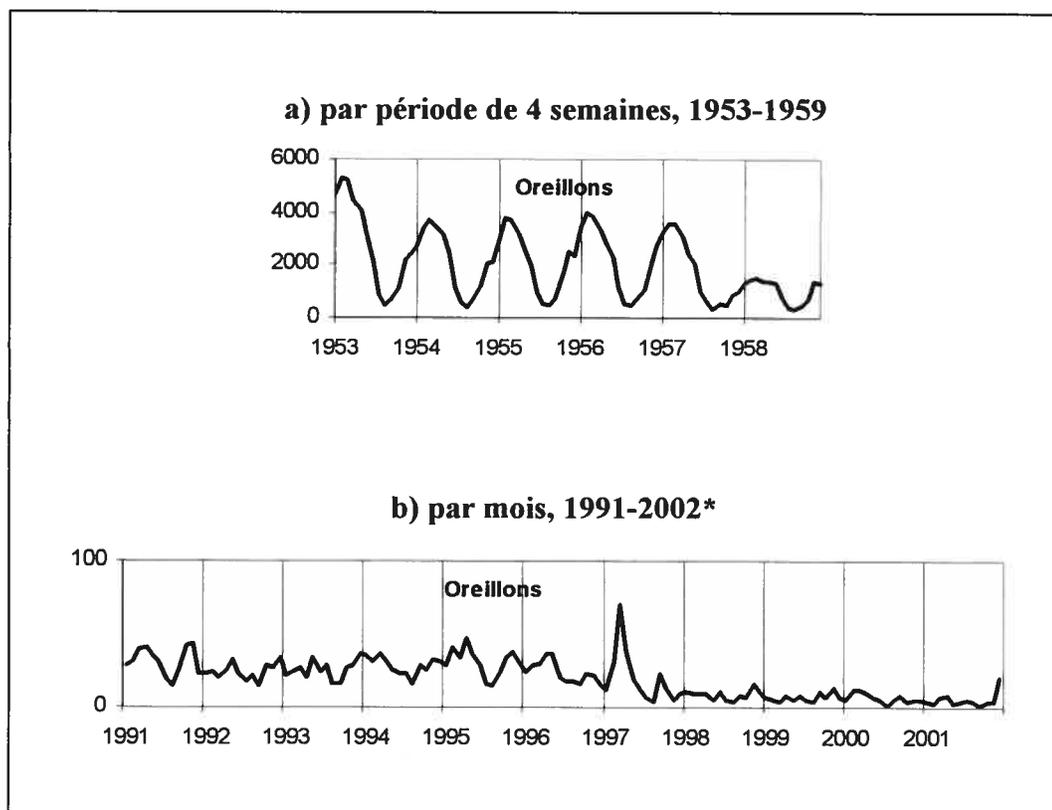
**Figure 6**  
**Nombre déclaré de cas de coqueluche au Canada**



\*À partir de 1991, des définitions spécifiques pour chaque maladie sous surveillance ont servi de base à la déclaration des cas à l'échelle nationale.

Source: Laboratoire de lutte contre la maladie. Santé Canada.

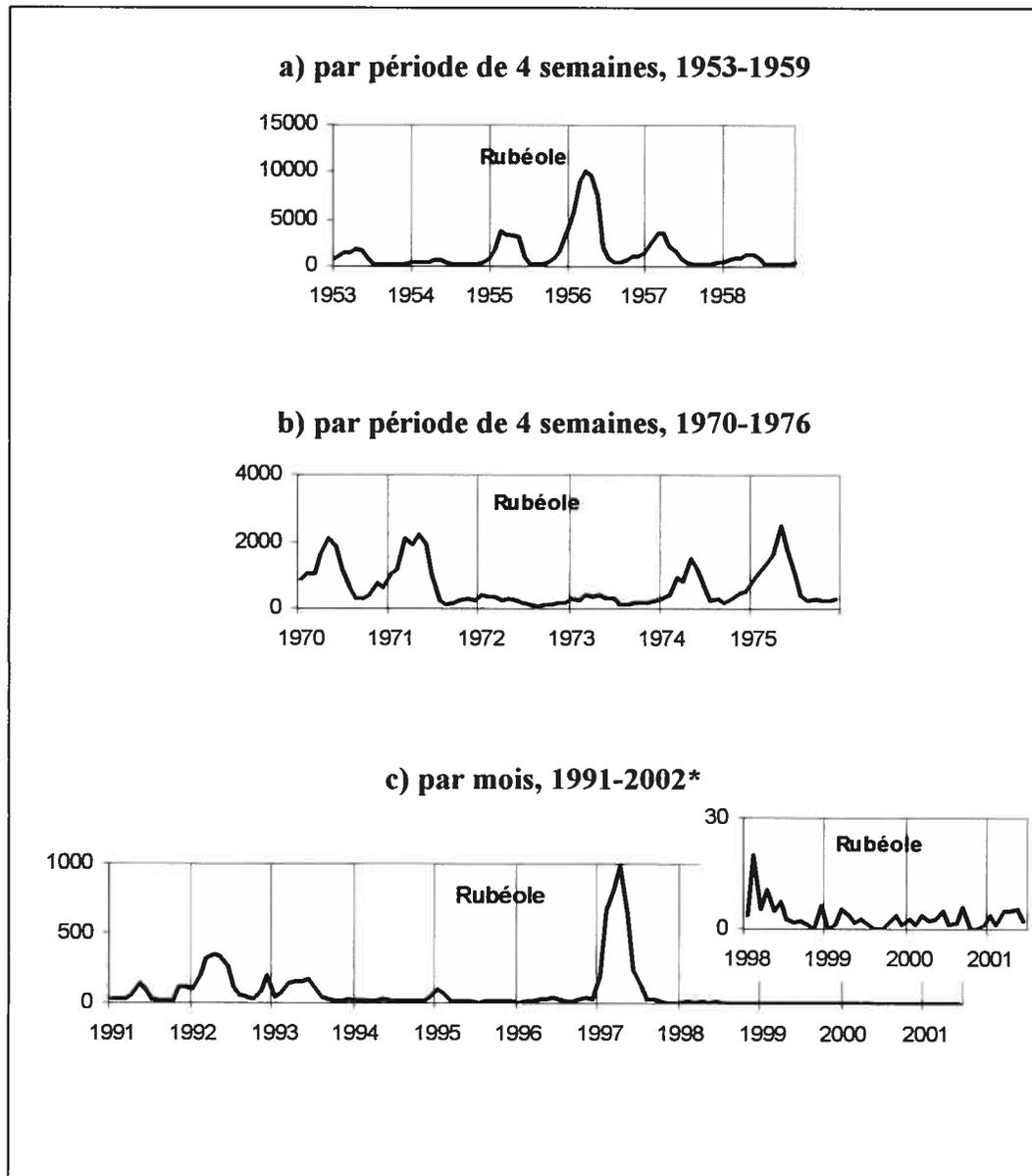
**Figure 7**  
**Nombre déclaré de cas d'oreillons au Canada**



\*A partir de 1991, des définitions spécifiques pour chaque maladie sous surveillance ont servi de base à la déclaration des cas à l'échelle nationale.

Source: Laboratoire de lutte contre la maladie, Santé Canada.

**Figure 8**  
**Nombre déclaré de cas de rubéole au Canada**



\*À partir de 1991, des définitions spécifiques pour chaque maladie sous surveillance ont servi de base à la déclaration des cas à l'échelle nationale.

Source: Laboratoire de lutte contre la maladie, Santé Canada.

## ANNEXE II

---

### **Graphique « moyenne-étendue » pour chaque série chronologique :**

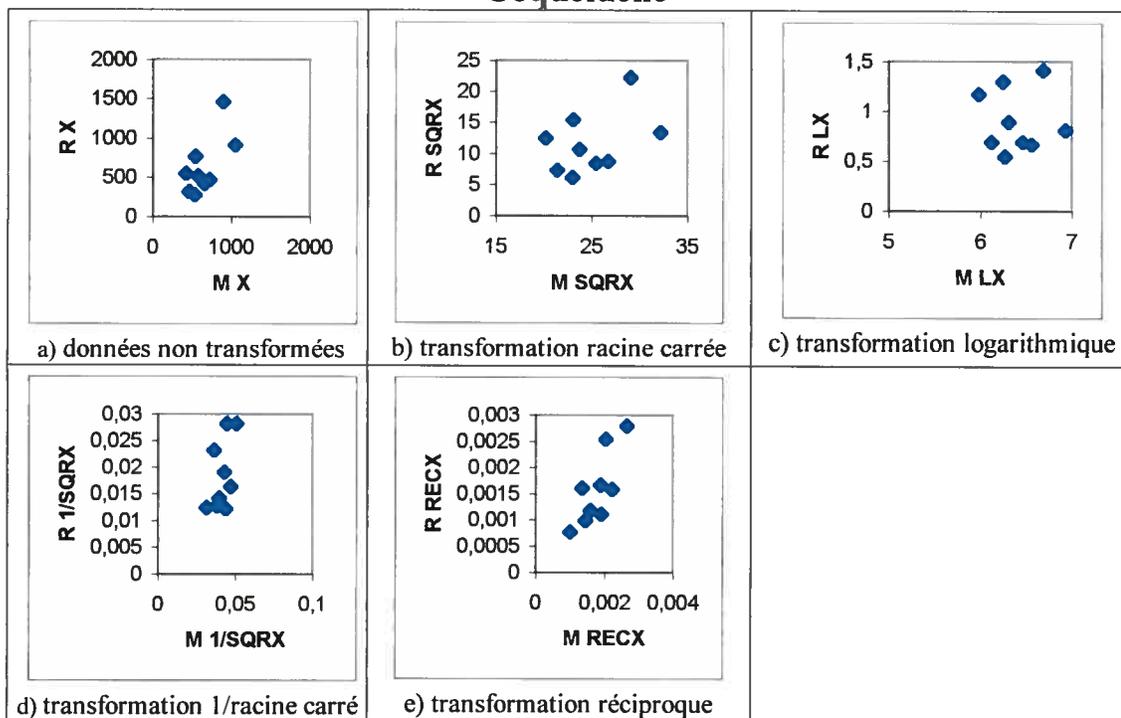
1) Coqueluche (1953-1962, 1970-1976, 1991-2002)

2) Oreillons (1953-1959, 1991-1998)

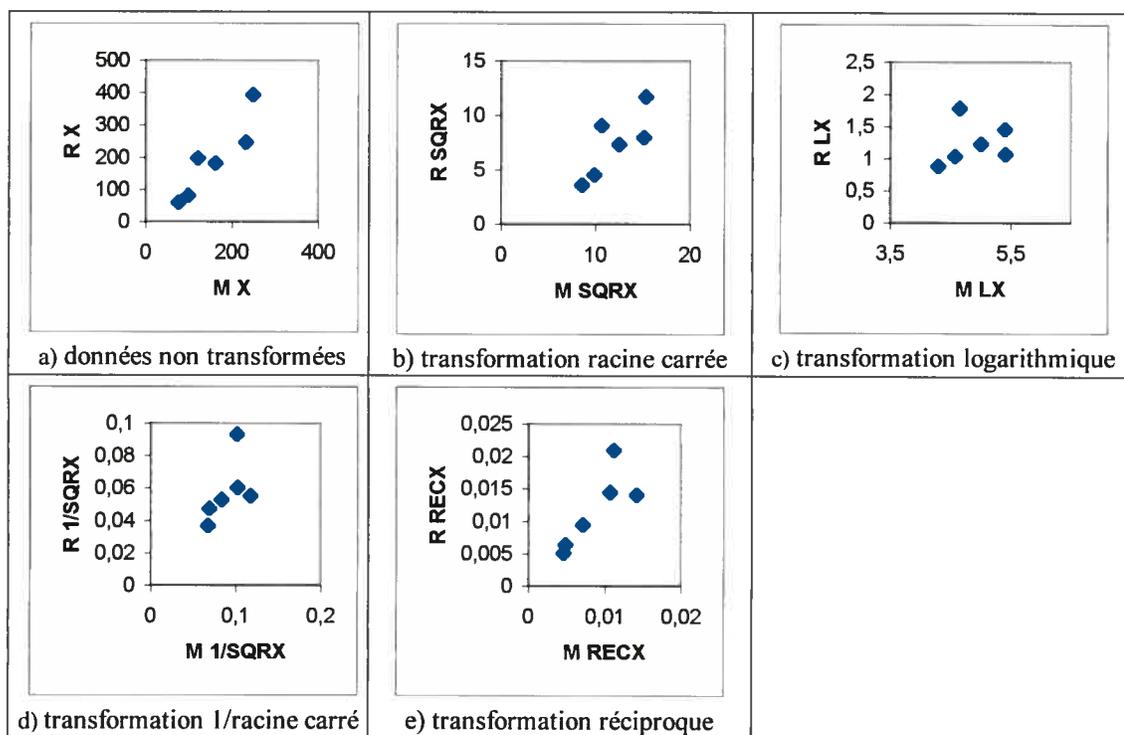
3) Rougeole (1953-1959, 1970-1976, 1991-2002)

4) Rubéole (1953-1959, 1970-1976, 1991-2002)

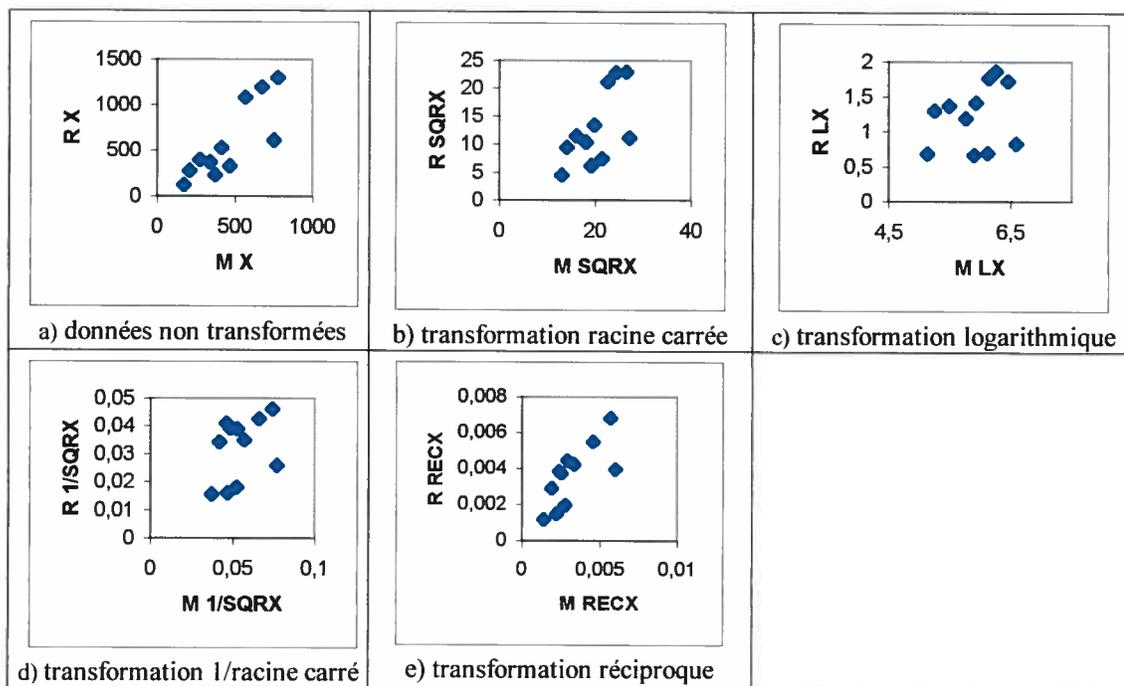
## Coqueluche



Graphique “moyenne-étendue” pour la **coqueluche** (1953-1962). Moyenne (M) et étendue (R) pour (a) les données non transformées (X), et moyenne et étendue pour les données transformées avec (b) la transformation racine carrée (SQRX), (c) la transformation logarithmique (LX), (d) la transformation inverse de la racine carrée (1/SQRX) et (e) la transformation réciproque (RECX).

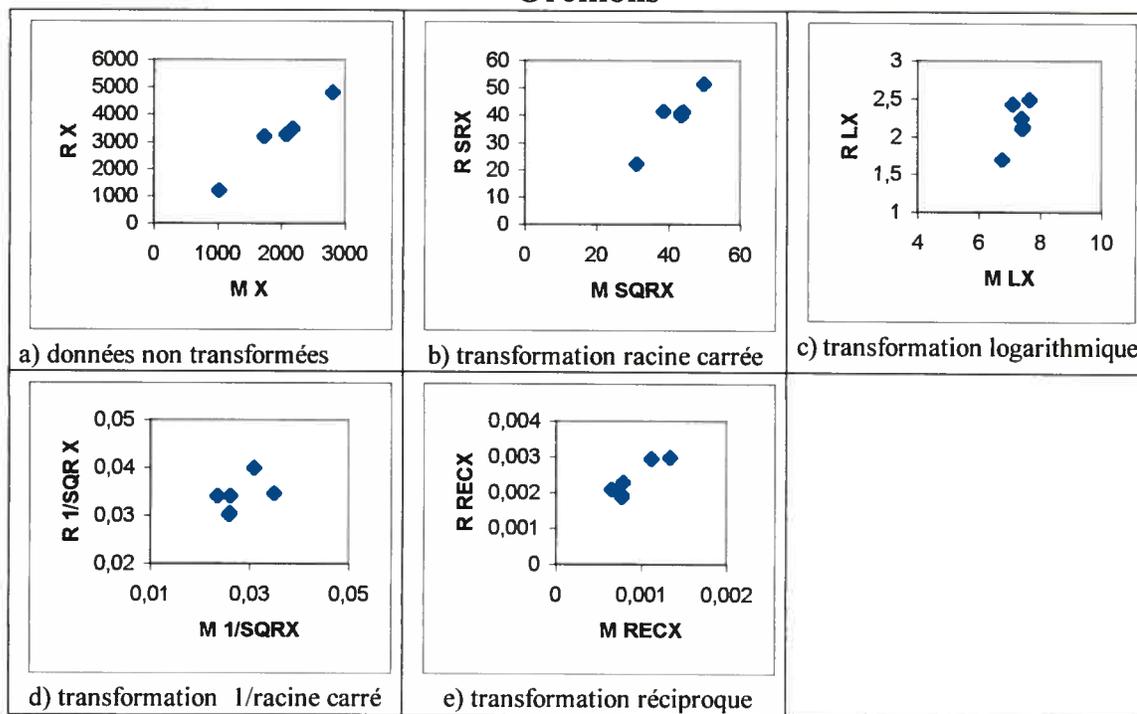


Graphique “moyenne-étendue” pour la **coqueluche** (1970-1976). Moyenne (M) et étendue (R) pour (a) les données non transformées (X), et moyenne et étendue pour les données transformées avec (b) la transformation racine carrée (SQRX), (c) la transformation logarithmique (LX), (d) la transformation inverse de la racine carrée (1/SQRX) et (e) la transformation réciproque (RECX).

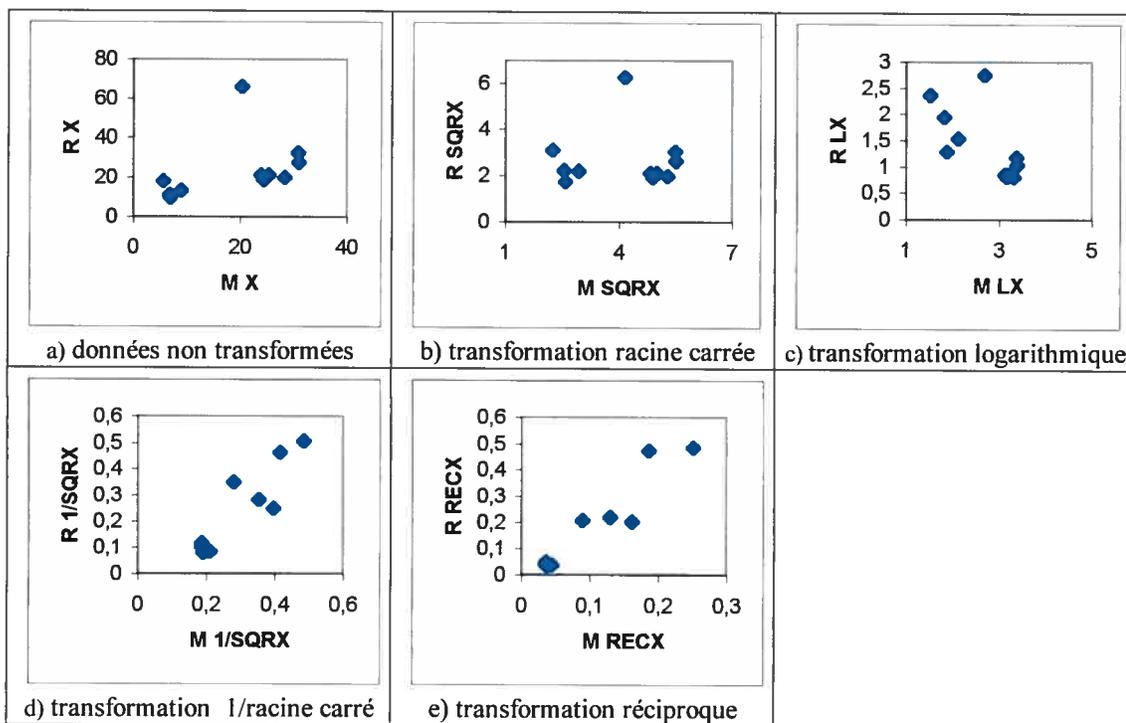


Graphique “moyenne-étendue” pour la **coqueluche (1991-2002)**. Moyenne (M) et étendue (R) pour (a) les données non transformées (X), et moyenne et étendue pour les données transformées avec (b) la transformation racine carrée (SQRX), (c) la transformation logarithmique (LX), (d) la transformation inverse de la racine carrée (1/SQRX) et (e) la transformation réciproque (RECX).

### Oreillons

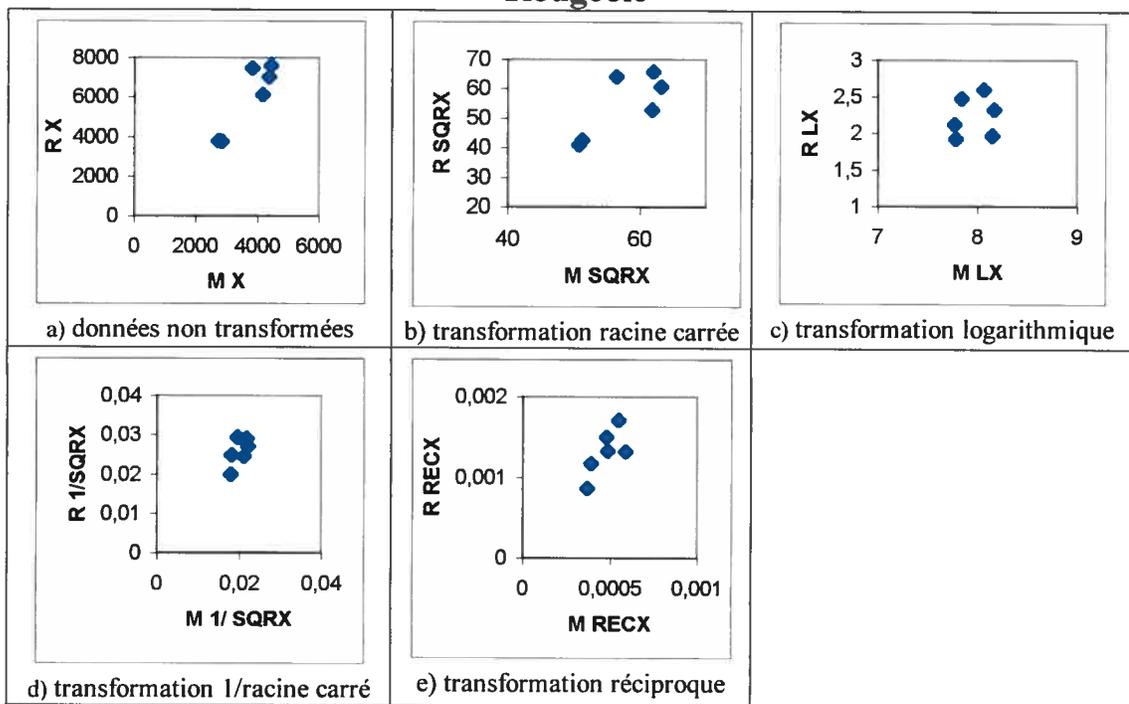


Graphique “moyenne-étendue” pour les **oreillons (1953-1959)**. Moyenne (M) et étendue (R) pour (a) les données non transformées (X), et moyenne et étendue pour les données transformées avec (b) la transformation racine carrée (SQRX), (c) la transformation logarithmique (LX), (d) la transformation inverse de la racine carrée (1/SQRX) et (e) la transformation réciproque (RECX).

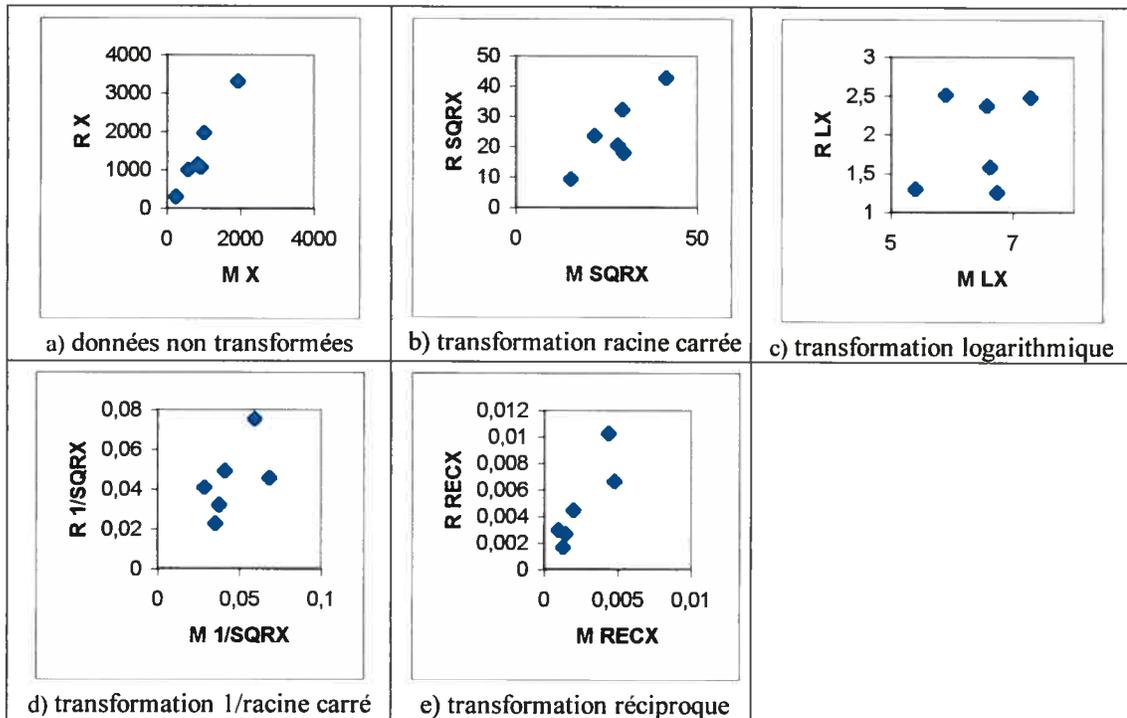


Graphique “moyenne-étendue” pour les **oreillons (1991-1998)**. Moyenne (M) et étendue (R) pour (a) les données non transformées (X), et moyenne et étendue pour les données transformées avec (b) la transformation racine carrée (SQRX), (c) la transformation logarithmique (LX), (d) la transformation inverse de la racine carrée (1/SQRX) et (e) la transformation réciproque (RECX).

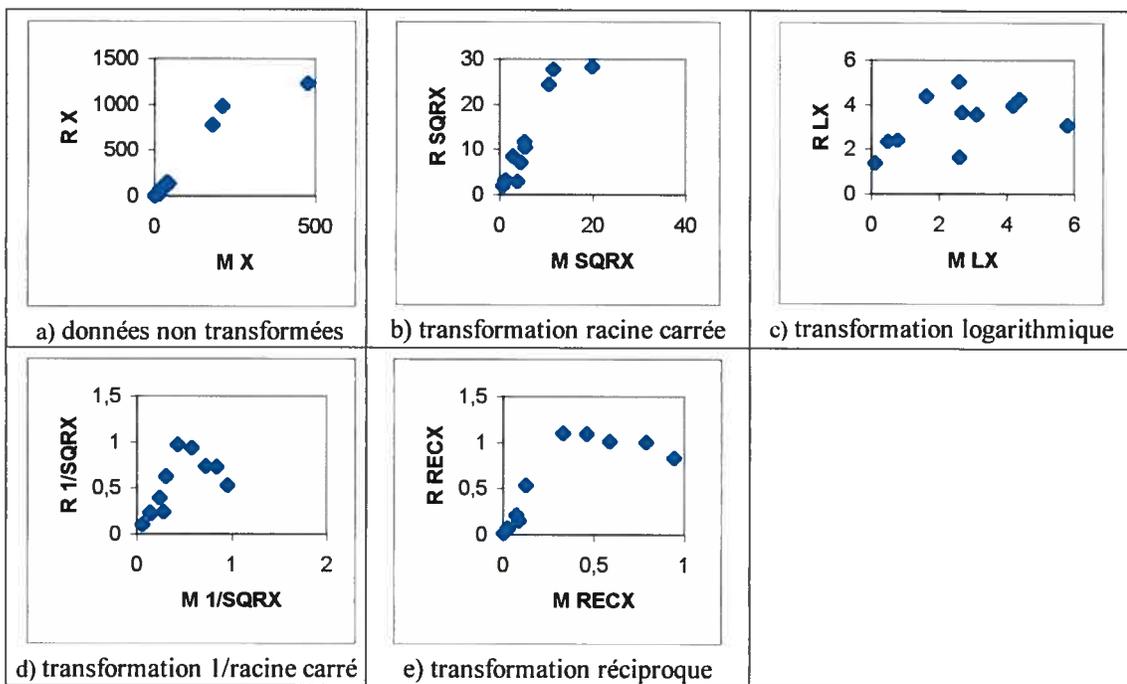
### Rougeole



Graphique “moyenne-étendue” pour la **rougeole (1953-1959)**. Moyenne (M) et étendue (R) pour (a) les données non transformées (X), et moyenne et étendue pour les données transformées avec (b) la transformation racine carrée (SQRX), (c) la transformation logarithmique (LX), (d) la transformation inverse de la racine carrée (1/SQRX) et (e) la transformation réciproque (RECX).

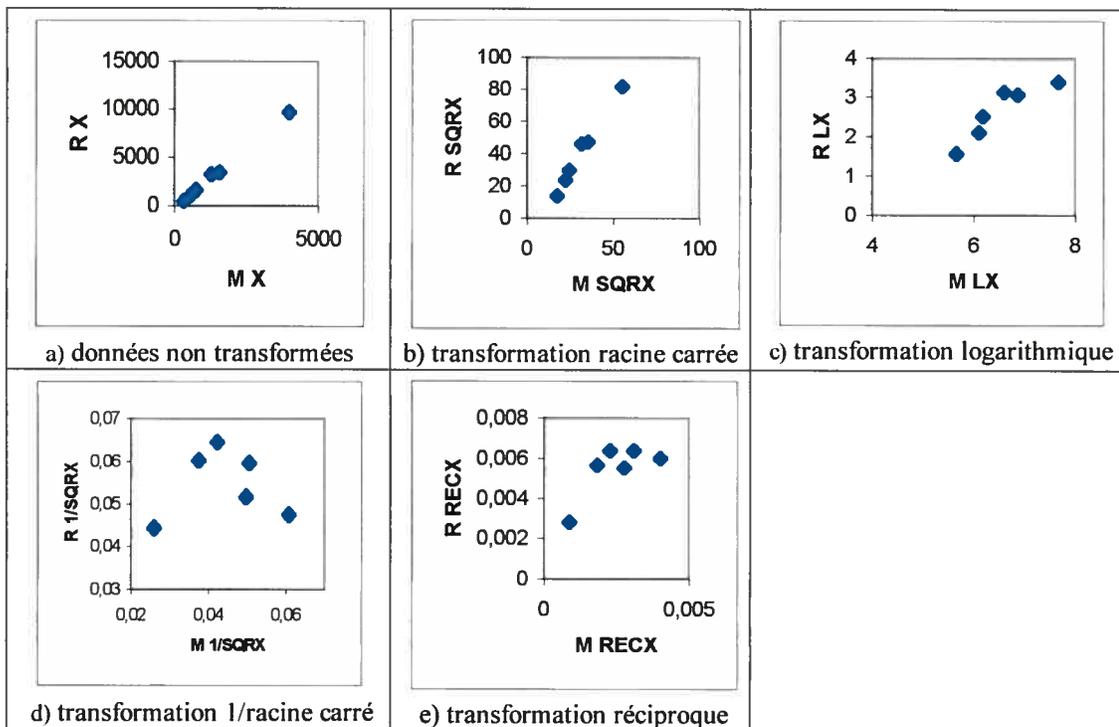


Graphique “moyenne-étendue” pour la **rougeole (1970-1976)**. Moyenne (M) et étendue (R) pour (a) les données non transformées (X), et moyenne et étendue pour les données transformées avec (b) la transformation racine carrée (SQRX), (c) la transformation logarithmique (LX), (d) la transformation inverse de la racine carrée (1/SQRX) et (e) la transformation réciproque (RECX).

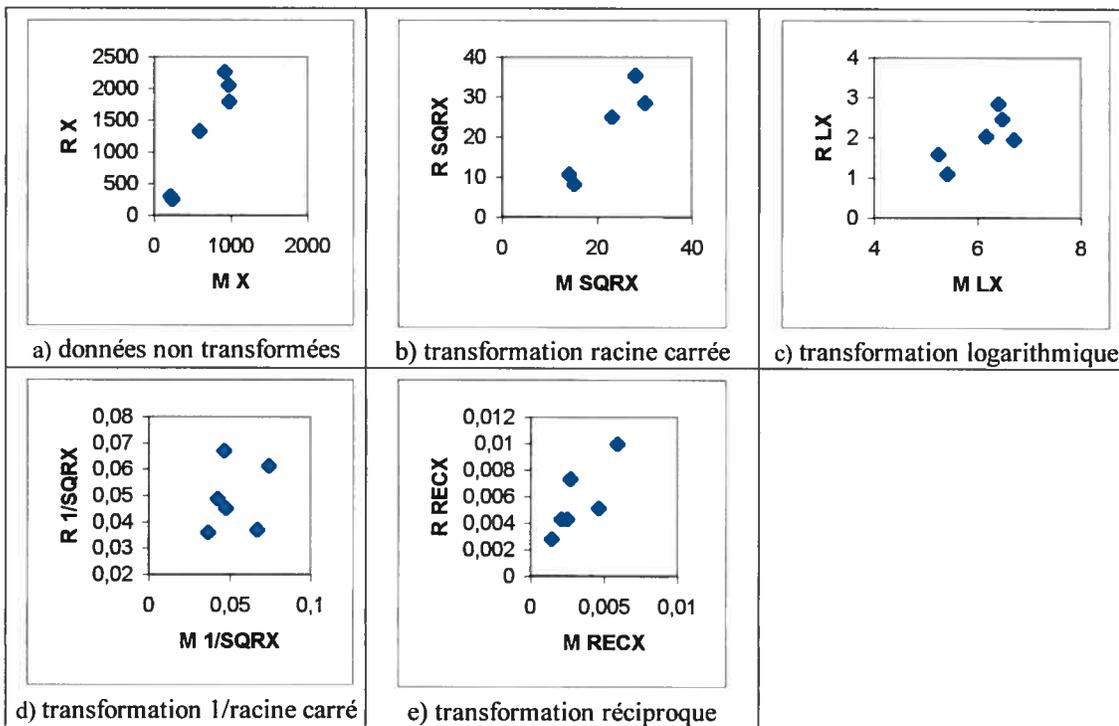


Graphique “moyenne-étendue” pour la **rougeole (1991-2002)**. Moyenne (M) et étendue (R) pour (a) les données non transformées (X), et moyenne et étendue pour les données transformées avec (b) la transformation racine carrée (SQRX), (c) la transformation logarithmique (LX), (d) la transformation inverse de la racine carrée (1/SQRX) et (e) la transformation réciproque (RECX).

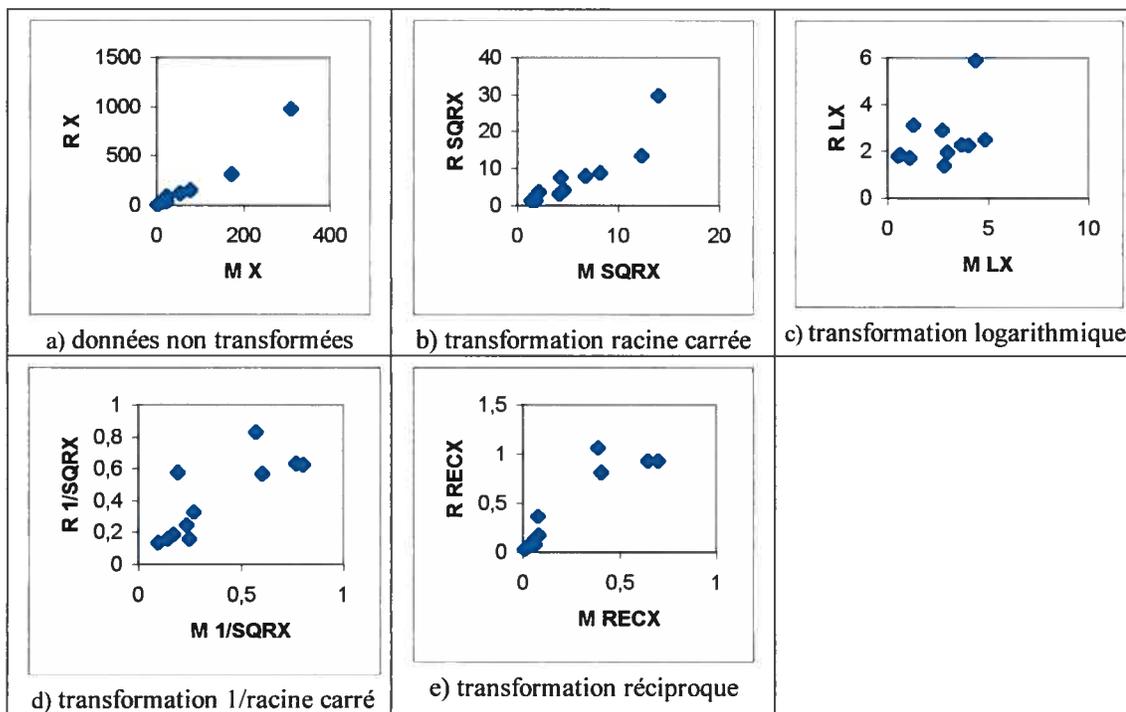
## Rubéole



Graphique “moyenne-étendue” pour la **rubéole (1953-1959)**. Moyenne (M) et étendue (R) pour (a) les données non transformées (X), et moyenne et étendue pour les données transformées avec (b) la transformation racine carrée (SQRX), (c) la transformation logarithmique (LX), (d) la transformation inverse de la racine carrée (1/SQRX) et (e) la transformation réciproque (RECX).



Graphique “moyenne-étendue” pour la **rubéole (1970-1976)**. Moyenne (M) et étendue (R) pour (a) les données non transformées (X), et moyenne et étendue pour les données transformées avec (b) la transformation racine carrée (SQRX), (c) la transformation logarithmique (LX), (d) la transformation inverse de la racine carrée (1/SQRX) et (e) la transformation réciproque (RECX).



Graphique “moyenne-étendue” pour la **rubéole (1991-2002)**. Moyenne (M) et étendue (R) pour (a) les données non transformées (X), et moyenne et étendue pour les données transformées avec (b) la transformation racine carrée (SQRX), (c) la transformation logarithmique (LX), (d) la transformation inverse de la racine carrée (1/SQRX) et (e) la transformation réciproque (RECX).

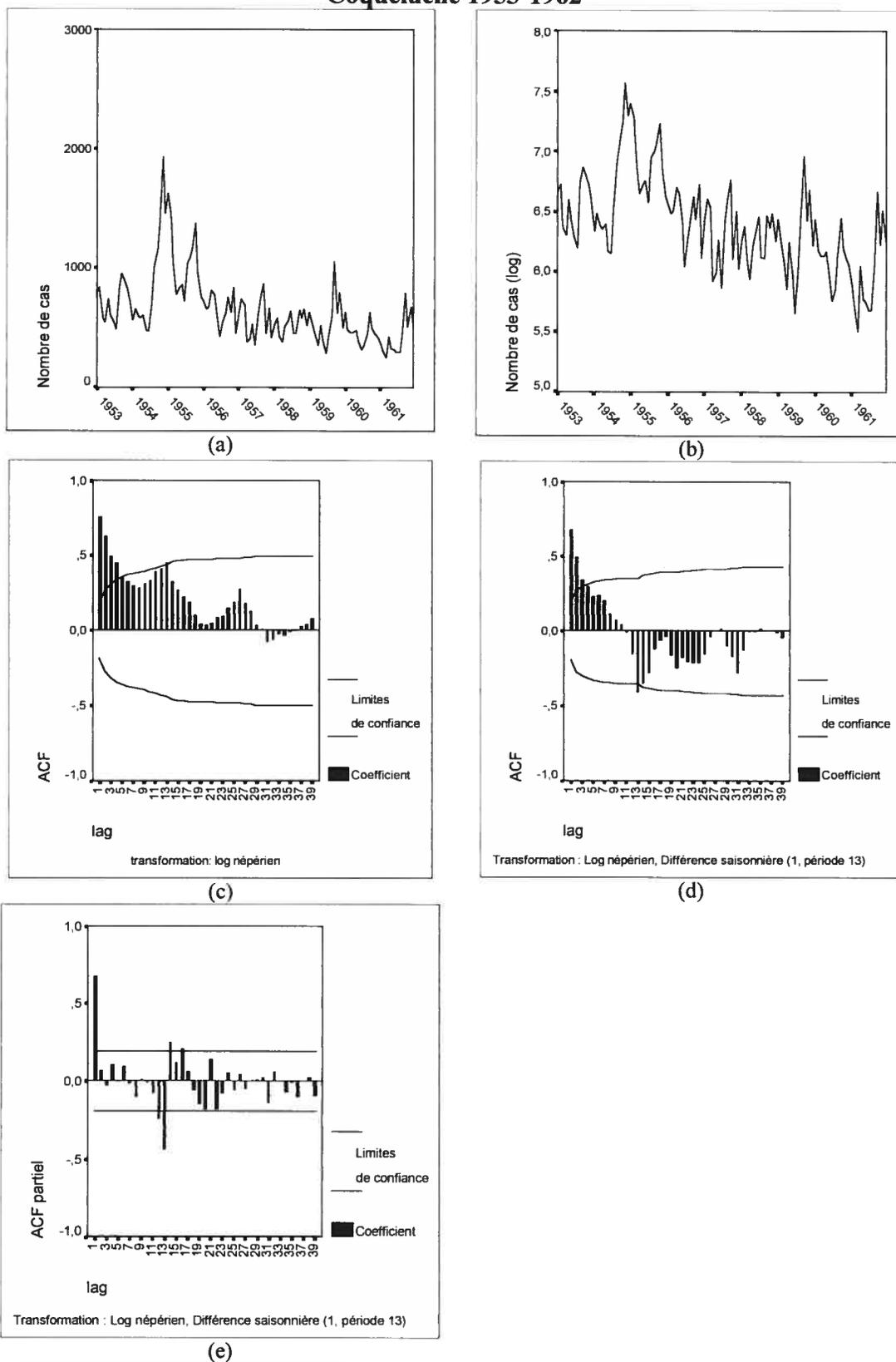
## ANNEXE III

---

Analyse Box-Jenkins détaillée (identification, estimation des paramètres et diagnostic)  
pour chaque série chronologique:

- 1) Coqueluche (1953-1962, 1970-1976, 1991-2002)
- 2) Oreillons (1953-1959, 1991-1998)
- 3) Rougeole (1953-1959, 1970-1976, 1991-2002)
- 4) Rubéole (1953-1959, 1970-1976, 1991-2002)

## Coqueluche 1953-1962



Incidence de la coqueluche par période de 4 semaines au Canada (1953-1962). (a) série chronologique, (b) série transformée en log, (c) ACF de la série (log), (d) ACF de la série (log) différenciée saisonnièrement (e) PACF de la série (log) différenciée saisonnièrement.

## Coqueluche (1953-1962) (suite)

## Estimation des paramètres

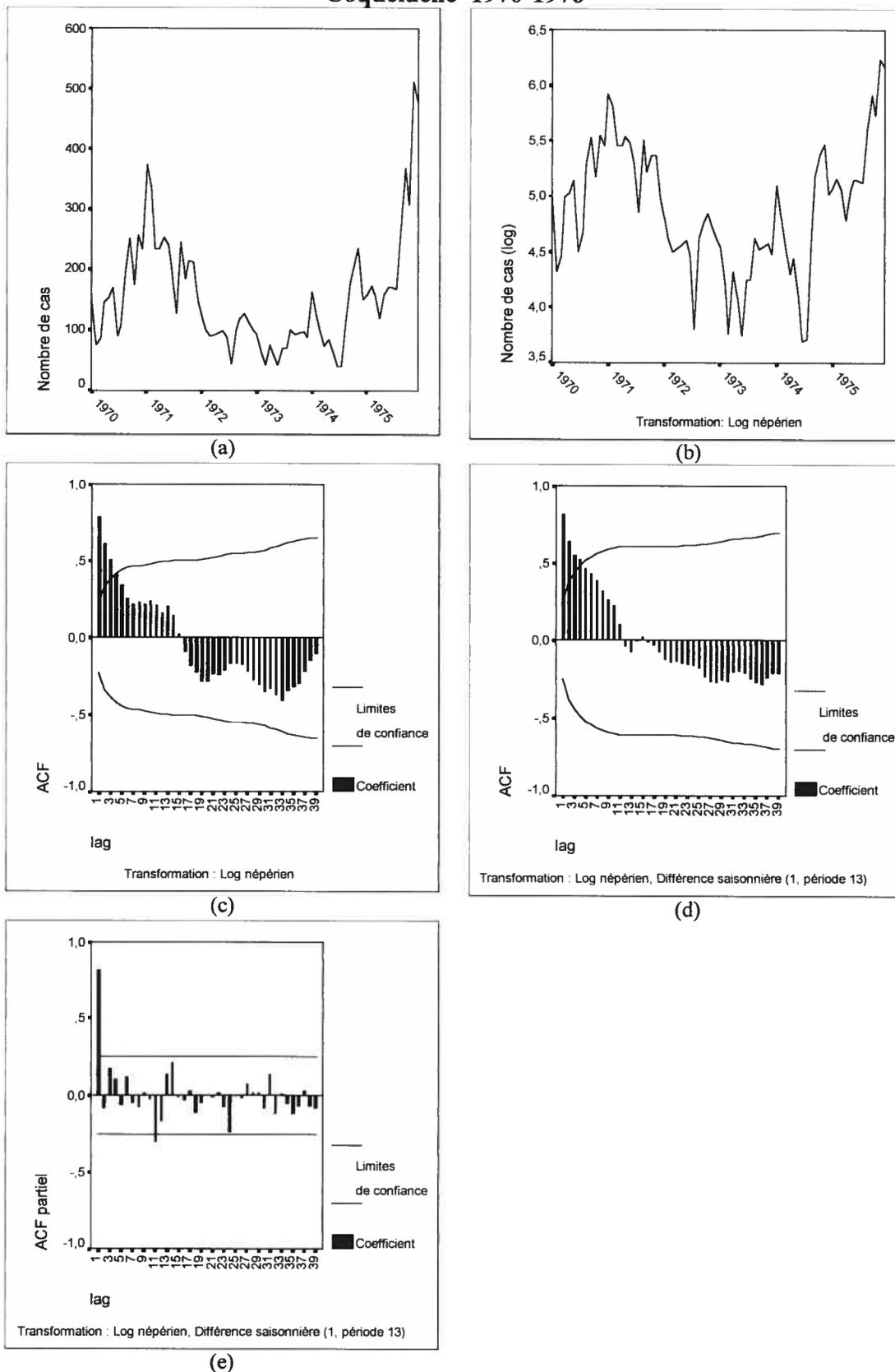
Number of residuals	104			
Standard error	,21595374			
Log likelihood	4,8975731			
AIC	-3,7951462			
SBC	4,1380265			
Analysis of Variance:				
	DF	Adj. Sum of Squares	Residual Variance	
Residuals	101	5,4967722	,04663602	
Variables in the Model:				
	B	SEB	T-RATIO	APPROX. PROB.
AR1	,65236437	,07159567	9,1117856	,00000000
SMA1	,84313398	,17157796	4,9139994	,00000345
CONSTANT	-,08709594	,02178515	-3,9979503	,00012172

## Diagnostic (ACF des résidus)

Lag	Auto-Corr.	Stand. Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	-,087	,098					**					,816	,366
2	,112	,099					**					2,167	,338
3	-,002	,100					*					2,167	,538
4	,175	,100					****					5,546	,236
5	-,016	,103					*					5,575	,350
6	,158	,103					***					8,380	,212
7	,073	,105					*					8,980	,254
8	,004	,106					*					8,982	,344
9	,045	,106					*					9,212	,418
10	,000	,106					*					9,212	,512
11	,106	,106					**					10,550	,482
12	,009	,107					*					10,559	,567
13	-,022	,107					*					10,615	,643
14	-,160	,107					***					13,737	,469
15	-,069	,109					*					14,331	,501
16	,059	,110					*					14,762	,542
17	,104	,110					**					16,145	,514
18	,041	,111					*					16,361	,567
19	-,056	,111					*					16,763	,606
20	-,144	,111					***					19,486	,490
21	,058	,113					*					19,938	,525
22	-,048	,113					*					20,246	,568
23	-,092	,114					**					21,401	,557
24	-,005	,114					*					21,404	,615
25	-,116	,114					**					23,271	,562
26	,069	,115					*					23,945	,579
27	-,076	,116					**					24,770	,587
28	,121	,116					**					26,880	,525
29	-,099	,117					**					28,312	,501
30	,096	,118					**					29,688	,482
31	-,264	,119					*****					40,189	,125
32	,000	,125					*					40,189	,152
33	,067	,125					*					40,896	,162
34	-,016	,125					*					40,935	,192
35	-,025	,125					*					41,031	,223
36	-,058	,125					*					41,571	,241
37	,039	,125					*					41,822	,269
38	-,036	,125					*					42,035	,300
39	-,010	,125					*					42,053	,340

Plot Symbols:            Autocorrelations \*            Two Standard Error Limits .  
Standard errors are based on the Bartlett (MA) approximation.  
Total cases: 117            Computable first lags: 103

## Coqueluche 1970-1976



Incidence de la coqueluche par période de 4 semaines au Canada (1970-1976) ajustée pour les données non-précisées. (a) série chronologique, (b) série transformée en log, (c) ACF de la série (log), (d) ACF de la série (log) différenciée saisonnièrement (e) PACF de la série (log) différenciée saisonnièrement.

## Coqueluche 1970-1976 (suite)

## Estimation des paramètres

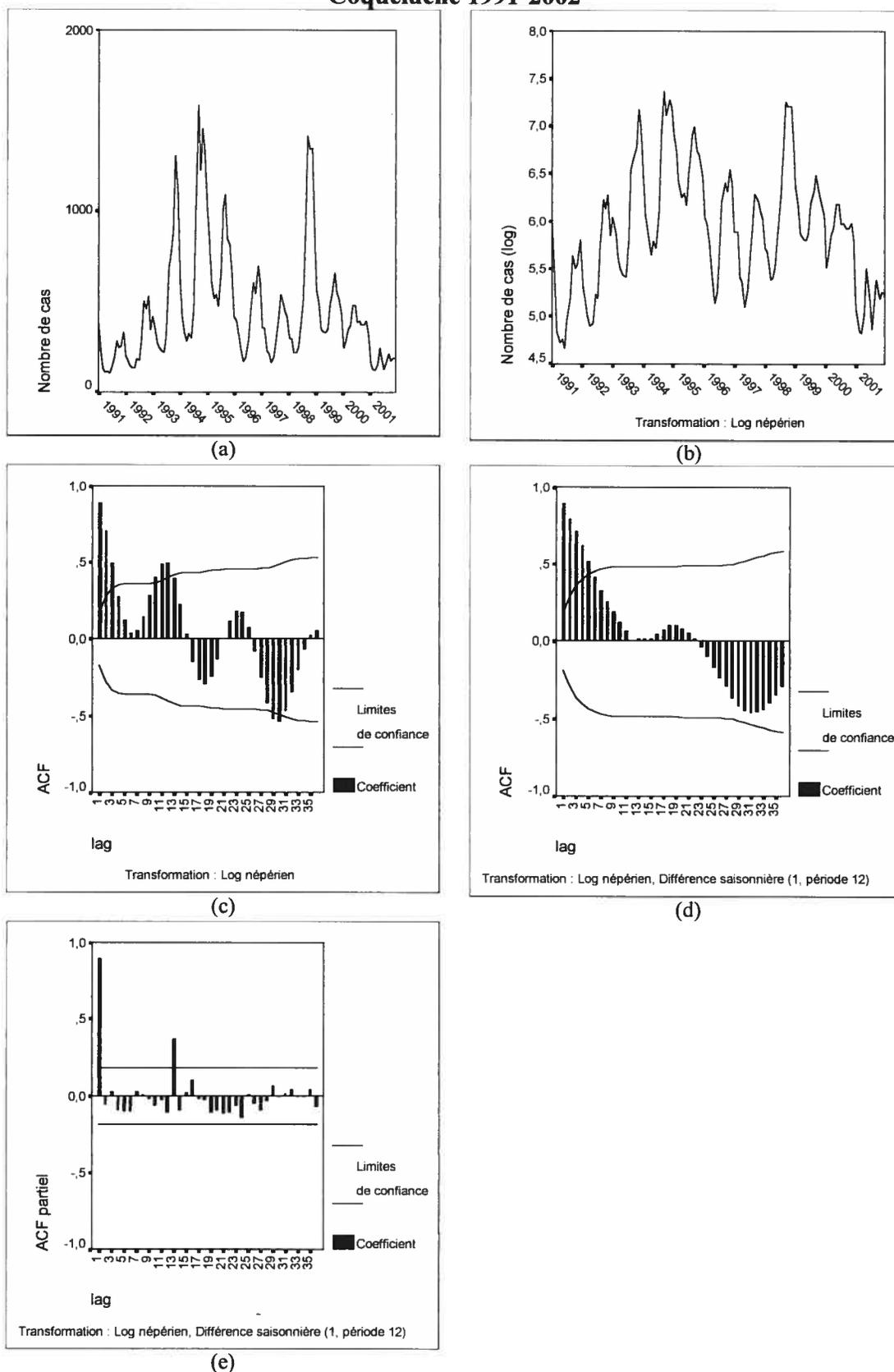
Number of residuals	65
Standard error	,2859994
Log likelihood	-18,535359
AIC	43,070717
SBC	49,593879
Analysis of Variance:	
	DF Adj. Sum of Squares Residual Variance
Residuals	62 6,5713187 ,08179565
Variables in the Model:	
	B SEB T-RATIO APPROX. PROB.
AR1	,91078020 ,05952344 15,301202 ,00000000
SMA1	,86948657 ,44568034 1,950920 ,05558952
CONSTANT	,10562932 ,14685442 ,719279 ,47467143

## Diagnostic (ACF des résidus)

Lag	Auto-Corr.	Stand. Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	-,049	,124					*					,164	,686
2	-,097	,124					**					,818	,664
3	-,164	,125					***					2,705	,439
4	,191	,129						****				5,306	,257
5	-,047	,133					*					5,469	,361
6	,029	,133						*				5,531	,478
7	,072	,133						*				5,918	,549
8	-,012	,134						*				5,930	,655
9	,013	,134						*				5,942	,746
10	,168	,134						***				8,177	,612
11	,120	,137						**				9,328	,592
12	-,318	,139					*****					17,640	,127
13	-,024	,150					*					17,689	,170
14	,088	,150						**				18,357	,191
15	,123	,150						**				19,680	,185
16	-,108	,152					**					20,711	,190
17	,125	,153						**				22,127	,180
18	,074	,155						*				22,638	,205
19	-,137	,155					***					24,428	,180
20	-,147	,157					***					26,532	,149
21	,070	,159						*				27,013	,170
22	,015	,160						*				27,036	,210
23	-,117	,160					**					28,456	,199
24	,069	,161						*				28,960	,222
25	,038	,162						*				29,116	,259
26	-,061	,162					*					29,532	,287
27	-,045	,162					*					29,766	,325
28	-,042	,162					*					29,976	,364
29	,039	,162						*				30,161	,406
30	-,199	,163					****					35,093	,239
31	,094	,166						**				36,232	,238
32	,090	,167						**				37,292	,239
33	-,027	,168					*					37,388	,275
34	-,077	,168					**					38,229	,283
35	-,021	,168					*					38,294	,322
36	-,124	,168					**					40,593	,275
37	-,086	,170					**					41,742	,272
38	,068	,170					*					42,484	,284
39	,059	,171					*					43,062	,302

Plot Symbols: Autocorrelations \* Two Standard Error Limits .  
Standard errors are based on the Bartlett (MA) approximation.  
Total cases: 78 Computable first lags: 64

## Coqueluche 1991-2002



Incidence mensuelle de la coqueluche au Canada (1991-2002) ajustée sur 28 jours. (a) série chronologique, (b) série transformée en log, (c) ACF de la série (log), (d) ACF de la série (log) différenciée saisonnièrement (e) PACF de la série (log) différenciée saisonnièrement.

## Coqueluche 1991-2002 (suite)

## Estimation des paramètres

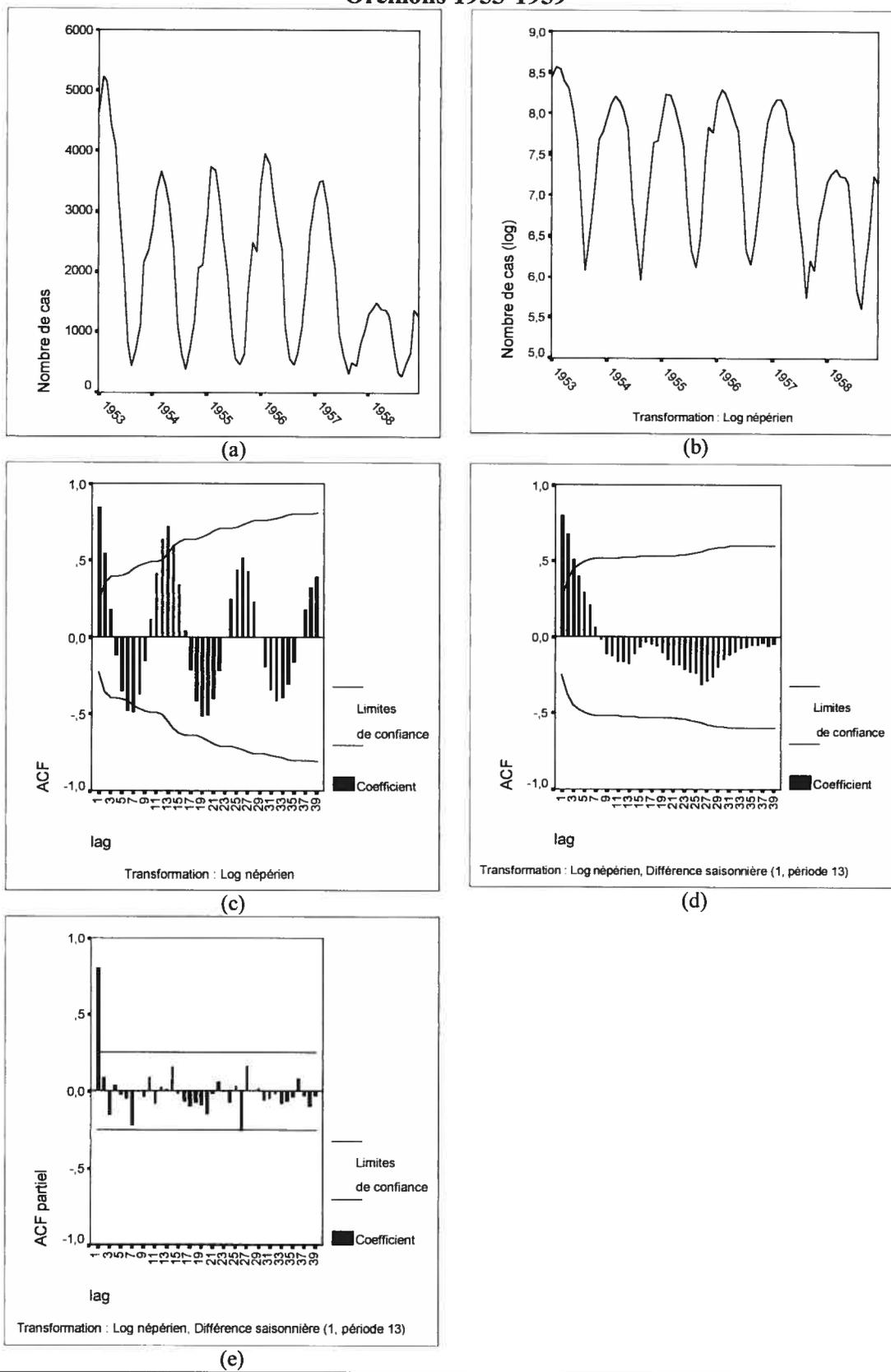
Number of residuals	120
Standard error	,21318972
Log likelihood	13,520229
AIC	-21,040459
SBC	-12,677983
Analysis of Variance:	
	DF Adj. Sum of Squares Residual Variance
Residuals	117 5,6063205 ,04544986
Variables in the Model:	
	B SEB T-RATIO APPROX. PROB.
AR1	,93916579 ,03432157 27,363720 ,00000000
SMA1	,57609697 ,10505469 5,483782 ,00000024
CONSTANT	-,05311859 ,13658424 -,388907 ,69805167

## Diagnostic (ACF des résidus)

Lag	Auto-Corr.	Stand. Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	,051	,091					*					,316	,574
2	-,075	,092					*					1,009	,604
3	,002	,092					*					1,010	,799
4	-,009	,092					*					1,020	,907
5	,029	,092					*					1,129	,952
6	-,031	,092					*					1,250	,974
7	,012	,092					*					1,268	,989
8	-,053	,092					*					1,631	,990
9	,029	,092					*					1,740	,995
10	-,004	,093					*					1,742	,998
11	,134	,093					***					4,136	,966
12	-,017	,094					*					4,173	,980
13	,037	,094					*					4,356	,987
14	,016	,094					*					4,393	,993
15	-,142	,094					***					7,204	,952
16	-,066	,096					*					7,820	,954
17	-,010	,096					*					7,833	,970
18	,120	,096					**					9,900	,935
19	,041	,098					*					10,147	,949
20	,032	,098					*					10,295	,963
21	,021	,098					*					10,362	,974
22	,009	,098					*					10,374	,983
23	,076	,098					**					11,243	,981
24	,034	,098					*					11,417	,986
25	-,062	,099					*					12,006	,987
26	-,039	,099					*					12,244	,990
27	,009	,099					*					12,256	,993
28	-,105	,099					**					14,008	,987
29	-,047	,100					*					14,356	,989
30	-,058	,100					*					14,912	,990
31	-,063	,100					*					15,571	,990
32	-,031	,101					*					15,734	,993
33	-,091	,101					**					17,135	,990
34	-,075	,101					**					18,102	,988
35	,052	,102					*					18,570	,990
36	-,055	,102					*					19,100	,991

Plot Symbols: Autocorrelations \* Two Standard Error Limits .  
Standard errors are based on the Bartlett (MA) approximation.  
Total cases: 132 Computable first lags: 119

## Oreillons 1953-1959



Incidence des oreillons par période de 4 semaines au Canada (1953-1959). (a) série chronologique, (b) série transformée en log, (c) ACF de la série (log), (d) ACF de la série (log) différenciée saisonnièrement (e) PACF de la série (log) différenciée saisonnièrement.

## Oreillons 1953-1959 ( suite)

## Estimation des paramètres

Number of residuals	65
Standard error	,15426708
Log likelihood	20,394143
AIC	-34,788286
SBC	-28,265125
Analysis of Variance:	
	DF Adj. Sum of Squares Residual Variance
Residuals	62 1,9688073 ,02379833
Variables in the Model:	
	B SEB T-RATIO APPROX. PROB.
AR1	,79275688 ,06728658 11,781798 ,00000000
SMA1	,91048470 ,59094011 1,540739 ,12846898
CONSTANT	-,15540408 ,04143331 -3,750704 ,00039025

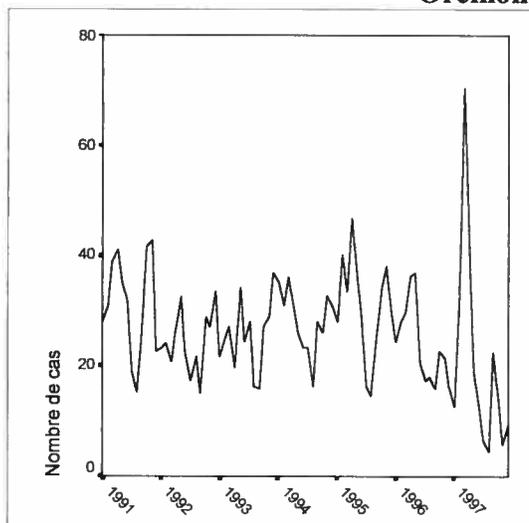
## Diagnostic (ACF des résidus)

Lag	Auto-Corr.	Stand. Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	-,125	,124					***					1,072	,301
2	,079	,126					**					1,506	,471
3	,011	,127					*					1,515	,679
4	,004	,127					*					1,516	,824
5	,062	,127					*					1,794	,877
6	,140	,127					***					3,232	,779
7	-,082	,130					**					3,741	,809
8	-,024	,130					*					3,785	,876
9	-,167	,130					***					5,962	,744
10	,040	,134					*					6,092	,807
11	-,094	,134					**					6,812	,814
12	-,017	,135					*					6,835	,868
13	,124	,135					**					8,117	,836
14	-,052	,137					*					8,345	,871
15	-,062	,137					*					8,679	,894
16	-,005	,137					*					8,681	,926
17	,022	,137					*					8,726	,948
18	,003	,137					*					8,727	,966
19	,040	,137					*					8,875	,975
20	-,052	,138					*					9,132	,981
21	-,081	,138					**					9,789	,982
22	-,016	,139					*					9,813	,988
23	-,071	,139					*					10,334	,989
24	,008	,139					*					10,341	,993
25	,075	,139					*					10,946	,993
26	-,207	,140					****					15,718	,943
27	,010	,144					*					15,729	,958
28	-,180	,144					****					19,558	,880
29	-,048	,148					*					19,840	,898
30	,059	,148					*					20,275	,909
31	,004	,149					*					20,276	,930
32	-,025	,149					*					20,362	,945
33	-,020	,149					*					20,417	,957
34	-,018	,149					*					20,464	,968
35	,039	,149					*					20,683	,974
36	-,042	,149					*					20,950	,979
37	,128	,149					***					23,518	,958
38	-,055	,151					*					24,014	,962
39	-,026	,151					*					24,129	,970

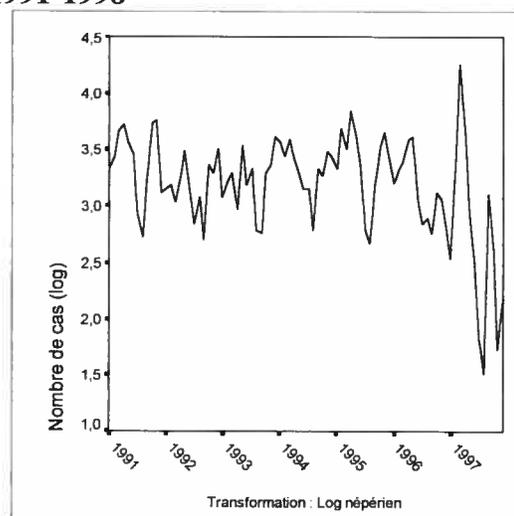
Plot Symbols: Autocorrelations \* Two Standard Error Limits .Standard errors are based on the Bartlett (MA) approximation.

Total cases: 78 Computable first lags: 64

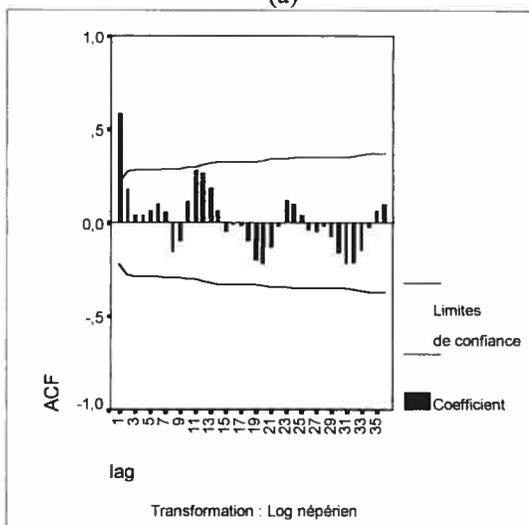
## Oreillons 1991-1998



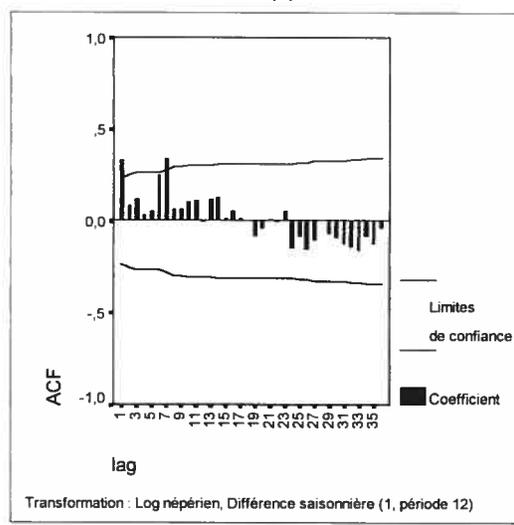
(a)



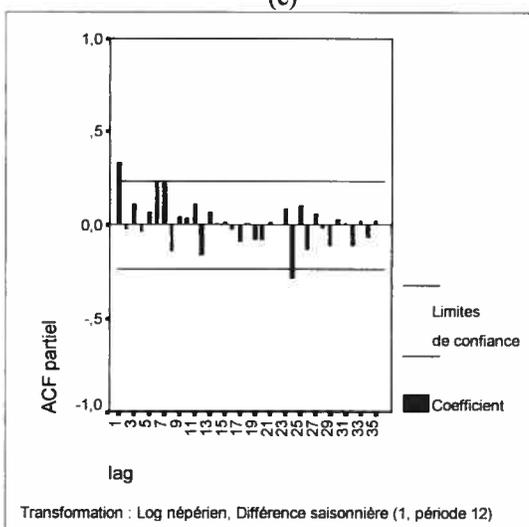
(b)



(c)



(d)



(e)

Incidence mensuelle des oreillons au Canada (1991-1998) ajustée sur 28 jours. (a) série chronologique, (b) série transformée en log, (c) ACF de la série (log), (d) ACF de la série (log) différenciée saisonnièrement (e) PACF de la série (log) différenciée saisonnièrement.

## Oreillons 1991-1998 ( suite)

## Estimation des paramètres

Number of residuals	72
Standard error	,3653533
Log likelihood	-30,346936
AIC	66,693872
SBC	73,523871
Analysis of Variance:	
	DF Adj. Sum of Squares Residual Variance
Residuals	69 9,7921756 ,13348304
Variables in the Model:	
	B SEB T-RATIO APPROX. PROB.
AR1	,47033277 ,10926739 4,3044202 ,00005419
SMA1	,54141955 ,21766233 2,4874288 ,01528810
CONSTANT	-,08665004 ,04641896 -1,8666951 ,06619353

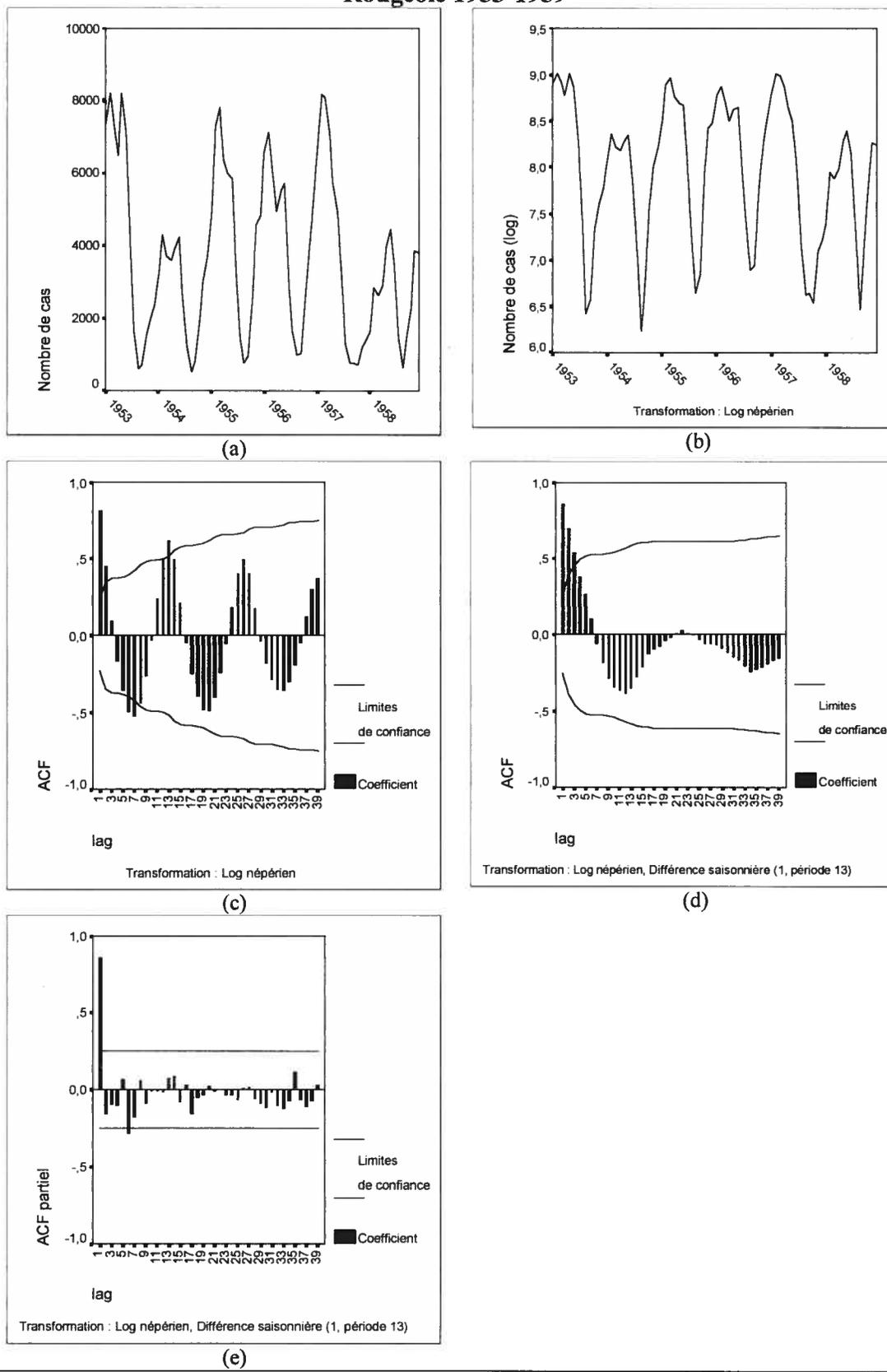
## Diagnostic (ACF des résidus)

Lag	Auto-Corr.	Stand. Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	,008	,118					*					,005	,945
2	-,070	,118					*					,375	,829
3	,063	,118					*					,677	,879
4	-,017	,119					*					,700	,951
5	,050	,119					*					,900	,970
6	,183	,119					****					3,608	,730
7	,297	,123					*****					10,815	,147
8	-,165	,133					***					13,078	,109
9	-,038	,135					*					13,202	,154
10	,050	,136					*					13,417	,201
11	,121	,136					**					14,698	,197
12	,015	,137					*					14,719	,257
13	,090	,137					**					15,456	,280
14	,089	,138					**					16,183	,302
15	-,119	,139					**					17,510	,289
16	,058	,140					*					17,831	,334
17	,004	,141					*					17,832	,399
18	,019	,141					*					17,867	,464
19	-,054	,141					*					18,165	,511
20	-,065	,141					*					18,594	,548
21	,004	,141					*					18,596	,611
22	-,071	,141					*					19,128	,637
23	,115	,142					**					20,563	,608
24	-,149	,143					***					23,017	,519
25	,040	,145					*					23,196	,566
26	-,103	,146					**					24,418	,552
27	-,069	,147					*					24,984	,575
28	,038	,147					*					25,161	,619
29	-,081	,147					**					25,975	,627
30	-,043	,148					*					26,207	,665
31	-,100	,148					**					27,499	,647
32	-,031	,149					*					27,626	,688
33	-,086	,149					**					28,641	,684
34	-,053	,150					*					29,027	,710
35	-,125	,150					***					31,287	,648
36	,002	,151					*					31,287	,692

Plot Symbols: Autocorrelations \* Two Standard Error Limits .Standard errors are based on the Bartlett (MA) approximation.

Total cases: 84 Computable first lags: 71

## Rougeole 1953-1959



Incidence de la rougeole par période de 4 semaines au Canada (1953-1959). (a) série chronologique, (b) série transformée en log, (c) ACF de la série (log), (d) ACF de la série (log) différenciée saisonnièrement (e) PACF de la série (log) différenciée saisonnièrement.

## Rougeole 1953-1959 (suite)

## Estimation des paramètres

Number of residuals	65			
Standard error	,20469096			
Log likelihood	5,6785521			
AIC	-5,3571042			
SBC	1,1660576			
Analysis of Variance:				
	DF	Adj. Sum of Squares	Residual Variance	
Residuals	62	3,1588831	,04189839	
Variables in the Model:				
	B	SEB	T-RATIO	APPROX. PROB.
AR1	,88377613	,05885627	15,015836	,00000000
SMA1	,78065426	,24436536	3,194619	,00220224
CONSTANT	-,04498166	,09030385	-,498114	,62016572

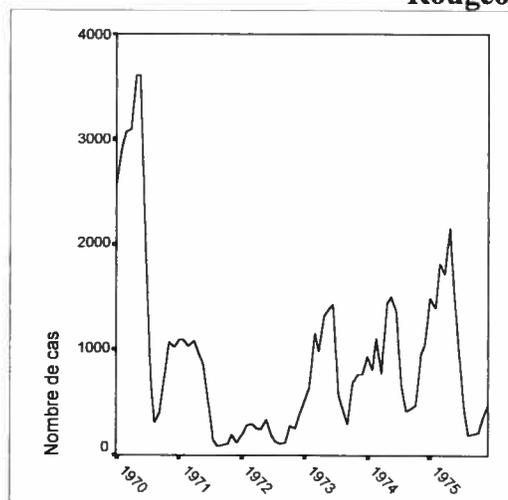
## Diagnostic (ACF des résidus)

Lag	Auto-Corr.	Stand. Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	,003	,124					*					,001	,978
2	,040	,124					*					,112	,946
3	,158	,124					***					1,859	,602
4	-,054	,127					*					2,069	,723
5	,065	,128					*					2,375	,795
6	,003	,128					*					2,375	,882
7	-,171	,128					***					4,577	,711
8	,006	,132					*					4,580	,801
9	-,184	,132					****					7,200	,616
10	-,039	,135					*					7,320	,695
11	-,034	,136					*					7,415	,765
12	-,220	,136					****					11,387	,496
13	,031	,141					*					11,468	,572
14	,002	,141					*					11,468	,649
15	-,125	,141					**					12,826	,616
16	,008	,143					*					12,831	,685
17	,041	,143					*					12,985	,737
18	-,079	,143					**					13,556	,758
19	,044	,144					*					13,739	,799
20	,020	,144					*					13,777	,842
21	-,059	,144					*					14,117	,865
22	,065	,144					*					14,551	,881
23	,007	,145					*					14,556	,910
24	,053	,145					*					14,850	,925
25	-,001	,145					*					14,850	,945
26	-,119	,145					**					16,442	,925
27	,112	,147					**					17,877	,907
28	-,057	,148					*					18,256	,920
29	-,041	,148					*					18,456	,934
30	-,015	,149					*					18,485	,950
31	-,050	,149					*					18,806	,958
32	,140	,149					***					21,388	,923
33	-,034	,151					*					21,548	,937
34	-,168	,151					***					25,505	,853
35	,045	,154					*					25,795	,871
36	-,030	,154					*					25,932	,892
37	,007	,154					*					25,940	,913
38	-,058	,154					*					26,475	,920
39	-,058	,154					*					27,030	,926

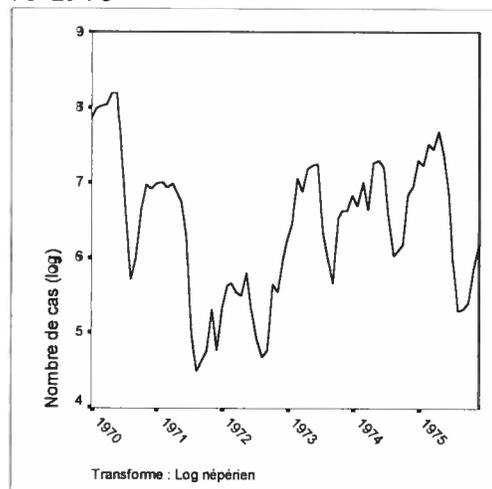
Plot Symbols: Autocorrelations \* Two Standard Error Limits .Standard errors are based on the Bartlett (MA) approximation.

Total cases: 78 Computable first lags: 64

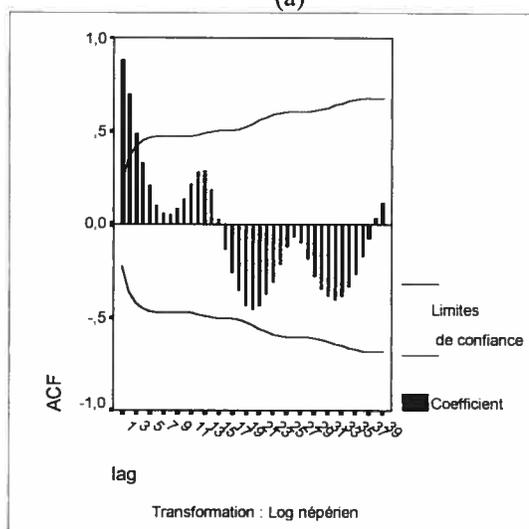
## Rougeole 1970-1976



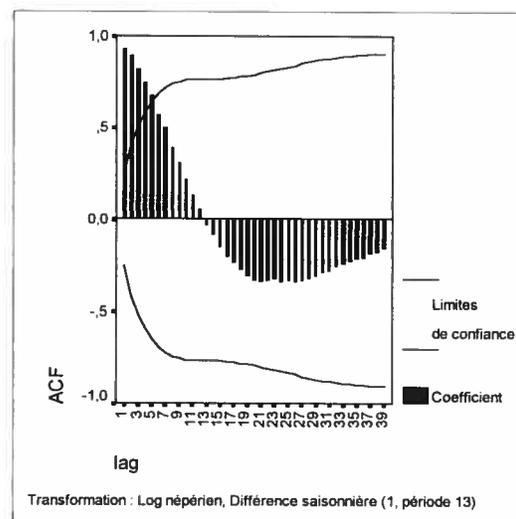
(a)



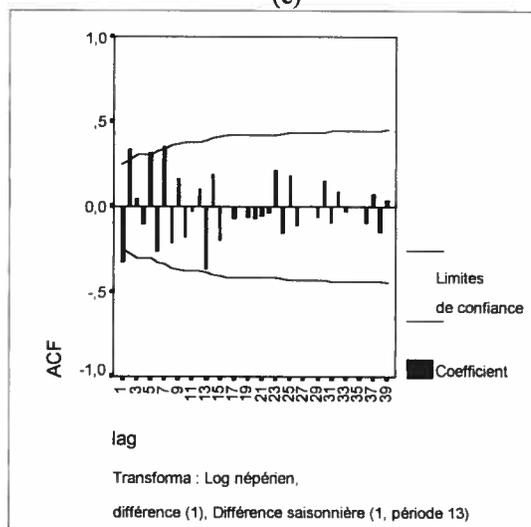
(b)



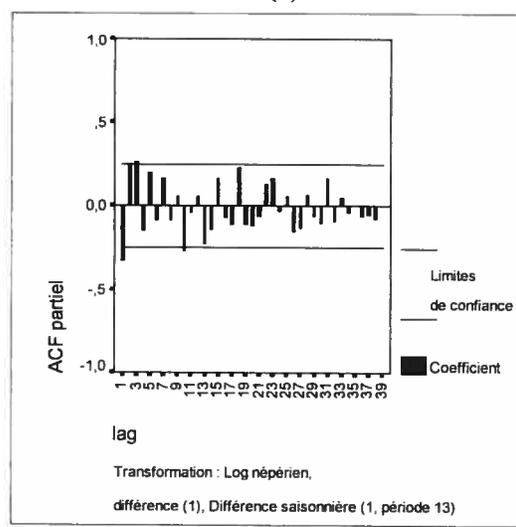
(c)



(d)



(e)



(f)

Incidence de la rougeole par période de 4 semaines au Canada (1970-1976) ajustée pour les données non précisées. (a) série chronologique, (b) série transformée en log, (c) ACF de la série (log), (d) ACF de la série (log) différenciée saisonnièrement (e) ACF de la série (log) différenciée régulièrement et saisonnièrement, (f) PACF de la série (log) différenciée régulièrement et saisonnièrement.

## Rougeole 1970-1976 (suite)

## Estimation des paramètres

Number of residuals	64			
Standard error	,26382132			
Log likelihood	-13,605907			
AIC	33,211815			
SBC	39,688464			
Analysis of Variance:				
	DF	Adj. Sum of Squares	Residual Variance	
Residuals	61	5,5782649	,06960169	
Variables in the Model:				
	B	SEB	T-RATIO	APPROX. PROB.
AR1	-,28080238	,10909603	-2,5739011	,01250086
SMA1	,90385683	,55027164	1,6425648	,10561870
CONSTANT	,01337913	,01354370	,9878489	,32712893

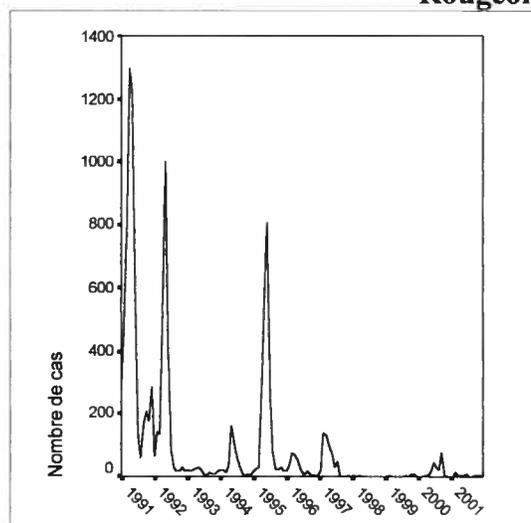
## Diagnostic (ACF des résidus)

Lag	Auto-Corr.	Stand. Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	,046	,125					*					,139	,709
2	,256	,125					*****					4,615	,099
3	,009	,133					*					4,621	,202
4	-,028	,133					*					4,676	,322
5	,296	,133					*****	*				10,967	,052
6	-,049	,143					*					11,141	,084
7	,232	,143					*****					15,126	,034
8	-,043	,149					*					15,268	,054
9	,052	,149					*					15,479	,079
10	-,132	,150					***					16,845	,078
11	-,017	,152					*					16,868	,112
12	,045	,152					*					17,029	,149
13	,017	,152					*					17,053	,197
14	,005	,152					*					17,054	,253
15	-,136	,152					***					18,643	,230
16	-,177	,154					*****					21,387	,164
17	-,120	,157					**					22,690	,160
18	,050	,158					*					22,920	,194
19	-,109	,159					**					24,025	,195
20	,022	,160					*					24,071	,239
21	-,167	,160					***					26,798	,178
22	-,013	,162					*					26,816	,218
23	,116	,162					**					28,210	,208
24	-,085	,164					**					28,970	,221
25	,136	,164					***					30,975	,190
26	-,115	,166					**					32,449	,179
27	-,059	,167					*					32,844	,202
28	-,106	,168					**					34,153	,196
29	-,155	,169					***					37,046	,145
30	,041	,171					*					37,251	,170
31	-,040	,171					*					37,453	,197
32	-,004	,171					*					37,455	,233
33	-,034	,171					*					37,617	,266
34	-,115	,171					**					39,469	,239
35	-,069	,173					*					40,164	,252
36	-,061	,173					*					40,723	,270
37	-,039	,173					*					40,966	,301
38	-,034	,173					*					41,155	,334
39	-,064	,174					*					41,845	,348

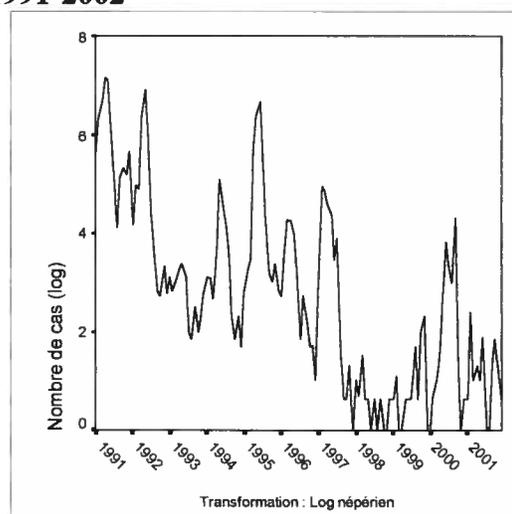
Plot Symbols: Autocorrelations \* Two Standard Error Limits .Standard errors are based on the Bartlett (MA) approximation.

Total cases: 78 Computable first lags: 63

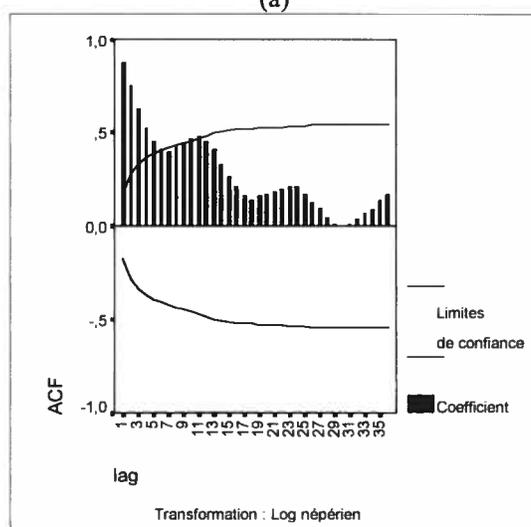
## Rougeole 1991-2002



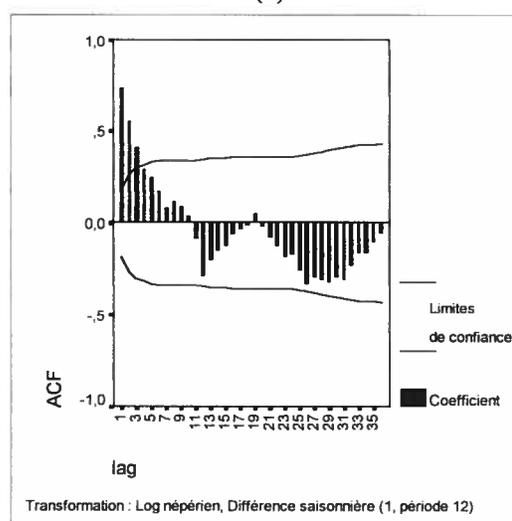
(a)



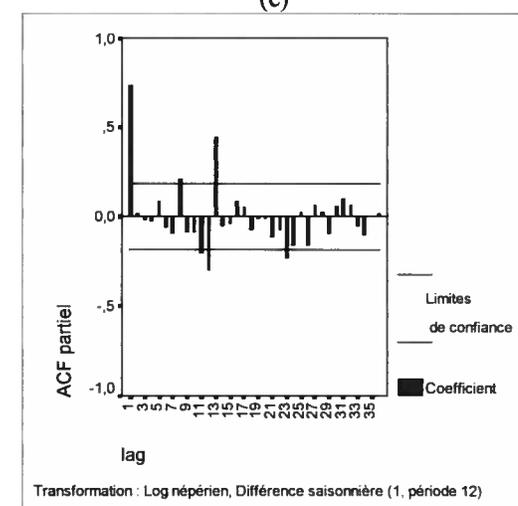
(b)



(c)



(d)



(e)

Incidence mensuelle de la rougeole au Canada (1991-2002) ajustée sur 28 jours et ajustée en ajoutant la constante 1 à toutes les données. (a) série chronologique, (b) série transformée en log, (c) ACF de la série (log), (d) ACF de la série (log) différenciée saisonnièrement (e) PACF de la série (log) différenciée saisonnièrement.

## Rougeole 1991-2002 (suite)

## Estimation des paramètres

Number of residuals	120
Standard error	,83443892
Log likelihood	-157,50484
AIC	321,00967
SBC	329,37215
Analysis of Variance:	
	DF Adj. Sum of Squares Residual Variance
Residuals	117 96,020530 ,69628832
Variables in the Model:	
	B SEB T-RATIO APPROX. PROB.
AR1	,78049482 ,05335135 14,629335 ,00000000
SMA1	,90239822 ,17475386 5,163824 ,00000100
CONSTANT	-,41763770 ,09804947 -4,259459 ,00004163

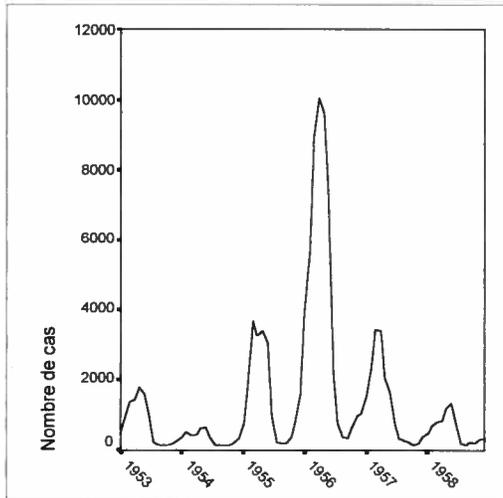
## Diagnostic (ACF des résidus)

Lag	Auto-Corr.	Stand. Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	,036	,091					*					,159	,690
2	,046	,091					*					,424	,809
3	-,035	,092					*					,576	,902
4	-,060	,092					*					1,036	,904
5	-,008	,092					*					1,043	,959
6	,074	,092					*					1,751	,941
7	-,114	,093					**					3,424	,843
8	,093	,094					**					4,543	,805
9	,033	,094					*					4,685	,861
10	,101	,095					**					6,041	,812
11	,136	,095					***					8,529	,665
12	-,055	,097					*					8,940	,708
13	,070	,097					*					9,614	,725
14	-,046	,098					*					9,907	,769
15	-,089	,098					**					11,015	,752
16	,002	,099					*					11,015	,809
17	-,052	,099					*					11,398	,835
18	-,032	,099					*					11,547	,870
19	,176	,099					****					16,018	,656
20	-,023	,101					*					16,098	,711
21	-,025	,102					*					16,190	,759
22	-,019	,102					*					16,243	,804
23	-,042	,102					*					16,513	,832
24	,090	,102					**					17,761	,814
25	-,032	,102					*					17,915	,846
26	-,149	,102					***					21,369	,723
27	-,049	,104					*					21,748	,750
28	-,134	,104					***					24,597	,650
29	-,105	,106					**					26,360	,606
30	-,004	,107					*					26,363	,656
31	-,063	,107					*					27,015	,671
32	-,021	,107					*					27,091	,713
33	,038	,107					*					27,331	,745
34	-,194	,107					****					33,714	,482
35	,001	,110					*					33,714	,530
36	-,053	,110					*					34,203	,554

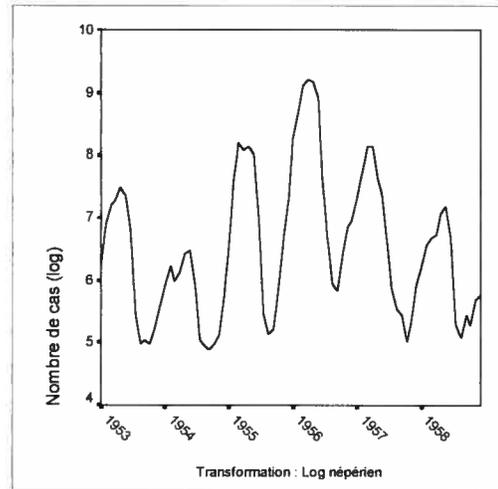
Plot Symbols: Autocorrelations \* Two Standard Error Limits .Standard errors are based on the Bartlett (MA) approximation.

Total cases: 132 Computable first lags: 119

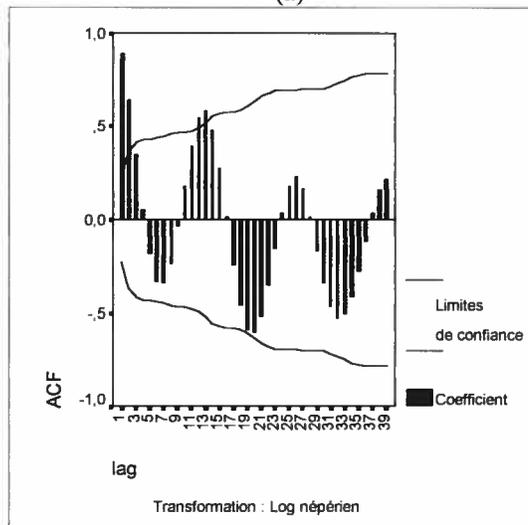
Rubéole 1953-1959



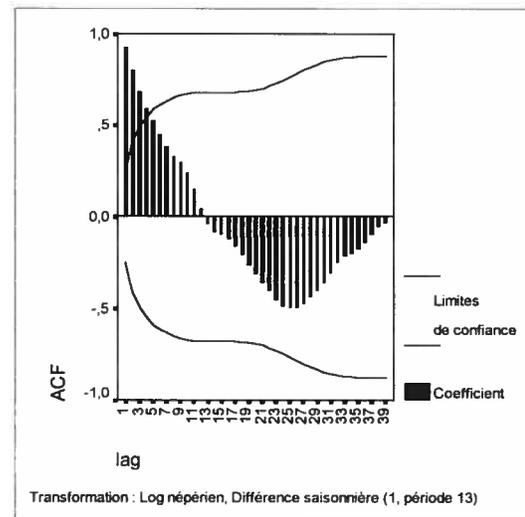
(a)



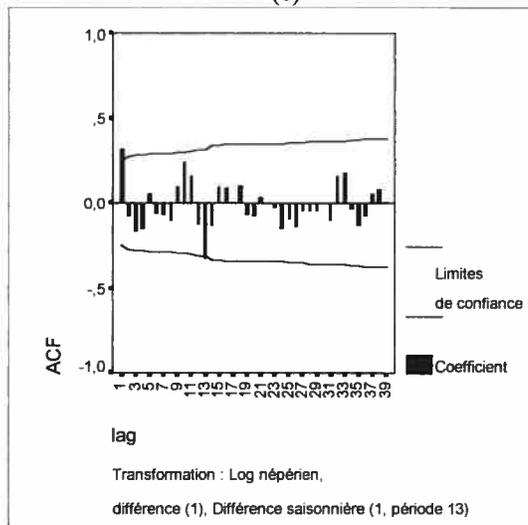
(b)



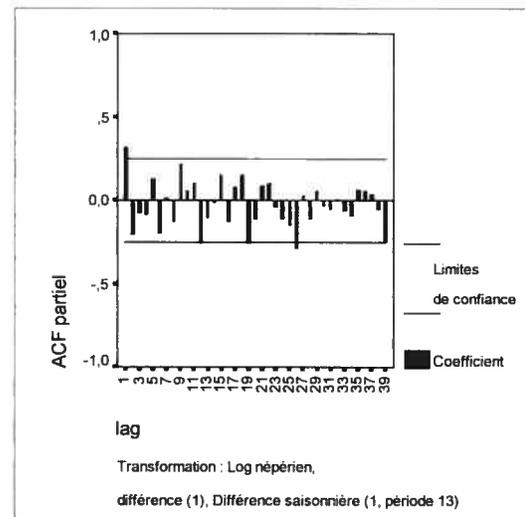
(c)



(d)



(e)



(f)

Incidence de la rubéole par période de 4 semaines au Canada (1953-1959). (a) série chronologique, (b) série transformée en log, (c) ACF de la série (log), (d) ACF de la série (log) différenciée saisonnièrement (e) ACF de la série (log) différenciée régulièrement et saisonnièrement, (f) PACF de la série (log) différenciée régulièrement et saisonnièrement.

Rubéole 1953-1959 (suite)

Estimation des paramètres

Number of residuals	64			
Standard error	,27446545			
Log likelihood	-17,888765			
AIC	41,77753			
SBC	48,254179			
Analysis of Variance:				
	DF	Adj. Sum of Squares	Residual Variance	
Residuals	61	6,3013434	,07533128	
Variables in the Model:				
	B	SEB	T-RATIO	APPROX. PROB.
AR1	,25422879	,1076407	2,3618275	,02139314
SMA1	,95200520	1,2833833	,7417933	,46105916
CONSTANT	-,01509604	,0240253	-,6283403	,53212668

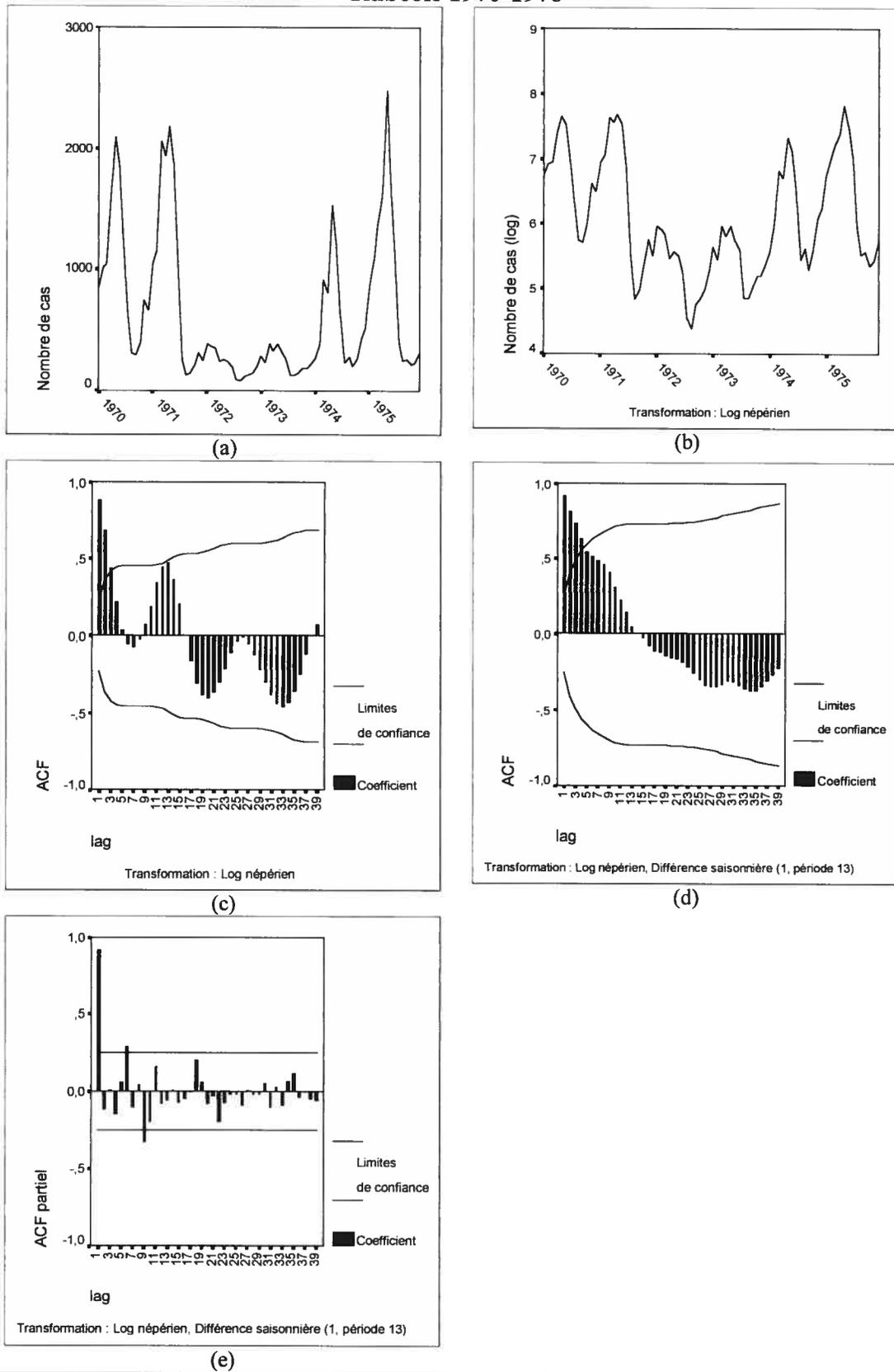
Diagnostic (ACF des résidus)

Lag	Auto-Corr.	Stand. Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	,037	,125					*					,091	,762
2	-,029	,125					*					,148	,929
3	-,023	,125					*					,184	,980
4	-,178	,125				****						2,413	,660
5	,124	,129					**					3,510	,622
6	-,173	,131				***						5,685	,459
7	-,109	,135				**						6,574	,475
8	-,070	,136				*						6,938	,543
9	,012	,137				*						6,949	,642
10	,223	,137				****						10,850	,369
11	,079	,142					**					11,352	,414
12	-,098	,143				**						12,126	,436
13	-,017	,144				*						12,150	,515
14	-,070	,144				*						12,566	,561
15	,115	,144					**					13,712	,547
16	,014	,146				*						13,729	,619
17	-,071	,146				*						14,184	,654
18	,089	,146					**					14,914	,668
19	-,041	,147				*						15,074	,718
20	-,044	,147				*						15,258	,761
21	,058	,148				*						15,589	,792
22	-,021	,148				*						15,633	,834
23	,008	,148				*						15,640	,870
24	-,072	,148				*						16,186	,881
25	-,043	,149				*						16,390	,903
26	-,199	,149				****						20,771	,754
27	-,049	,153				*						21,048	,784
28	-,008	,153				*						21,056	,823
29	-,085	,153				**						21,932	,823
30	,096	,154					**					23,069	,812
31	-,123	,155				**						24,992	,768
32	,174	,156					***					29,007	,619
33	,167	,159					***					32,828	,476
34	-,023	,162				*						32,904	,521
35	-,067	,162				*						33,555	,538
36	-,082	,163				**						34,581	,536
37	,045	,163				*						34,894	,568
38	,015	,163				*						34,928	,612
39	-,096	,163				**						36,483	,585

Plot Symbols: Autocorrelations \* Two Standard Error Limits .Standard errors are based on the Bartlett (MA) approximation.

Total cases: 78 Computable first lags: 63

## Rubéole 1970-1976



Incidence de la rubéole par période de 4 semaines au Canada (1970-1976) ajustée pour les données non précisées. (a) série chronologique, (b) série transformée en log, (c) ACF de la série (log), (d) ACF de la série (log) différenciée saisonnièrement (e) PACF de la série (log) différenciée saisonnièrement.

### Rubéole 1970-1976 (suite)

#### Estimation des paramètres

Number of residuals	65
Standard error	,26580619
Log likelihood	-12,228632
AIC	30,457264
SBC	36,980425
Analysis of Variance:	
	DF Adj. Sum of Squares Residual Variance
Residuals	62 5,4574922 ,07065293
Variables in the Model:	
	B SEB T-RATIO APPROX. PROB.
AR1	,93906909 ,04431429 21,191112 ,00000000
SMA1	,81521217 ,26192424 3,112397 ,00280609
CONSTANT	-,07882060 ,18033437 -,437080 ,66357172

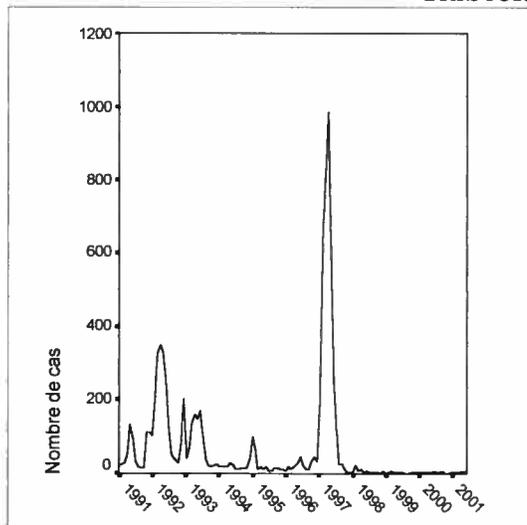
#### Diagnostic (ACF des résidus)

Lag	Auto-Corr.	Stand. Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	,063	,124					*					,267	,605
2	,056	,125					*					,483	,786
3	,057	,125					*					,708	,871
4	-,006	,125					*					,710	,950
5	-,242	,125				*****						4,953	,422
6	,043	,132					*					5,087	,533
7	-,051	,132					*					5,284	,625
8	,115	,133					**					6,289	,615
9	,236	,134					*****					10,629	,302
10	-,053	,141					*					10,853	,369
11	,041	,141					*					10,991	,444
12	,076	,141					**					11,469	,489
13	,105	,142					**					12,385	,496
14	-,161	,143					***					14,590	,407
15	,038	,146					*					14,714	,472
16	-,134	,146					***					16,307	,432
17	-,054	,148					*					16,569	,484
18	-,105	,148					**					17,588	,483
19	,020	,149					*					17,627	,547
20	,036	,149					*					17,750	,604
21	,104	,149					**					18,818	,597
22	,047	,150					*					19,042	,643
23	,036	,151					*					19,175	,691
24	-,030	,151					*					19,270	,737
25	-,091	,151					**					20,167	,738
26	-,077	,152					**					20,826	,751
27	-,219	,152					****					26,325	,501
28	-,066	,157					*					26,836	,527
29	-,133	,157					***					28,986	,466
30	,099	,159					**					30,217	,455
31	-,014	,160					*					30,243	,505
32	,084	,160					**					31,167	,509
33	,010	,161					*					31,181	,558
34	-,020	,161					*					31,239	,604
35	-,115	,161					**					33,153	,557
36	-,050	,162					*					33,529	,587
37	-,147	,162					***					36,886	,474
38	-,008	,164					*					36,897	,520
39	,035	,164					*					37,101	,557

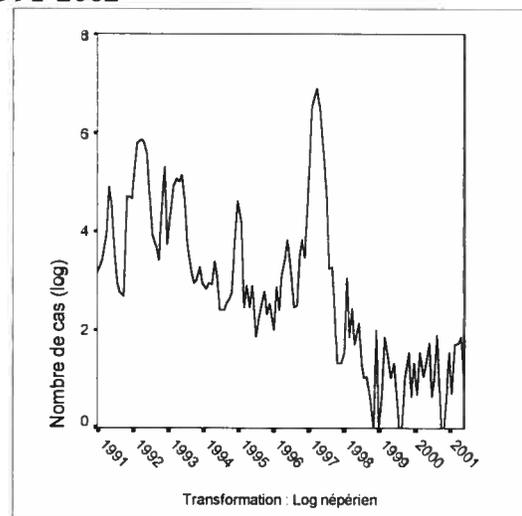
Plot Symbols: Autocorrelations \* Two Standard Error Limits .Standard errors are based on the Bartlett (MA) approximation.

Total cases: 78 Computable first lags: 64

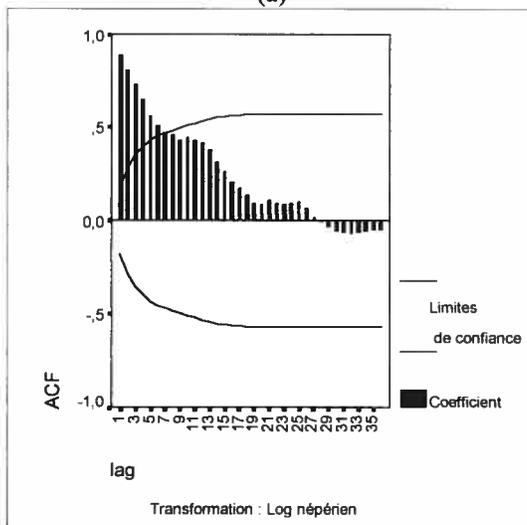
## Rubéole 1991-2002



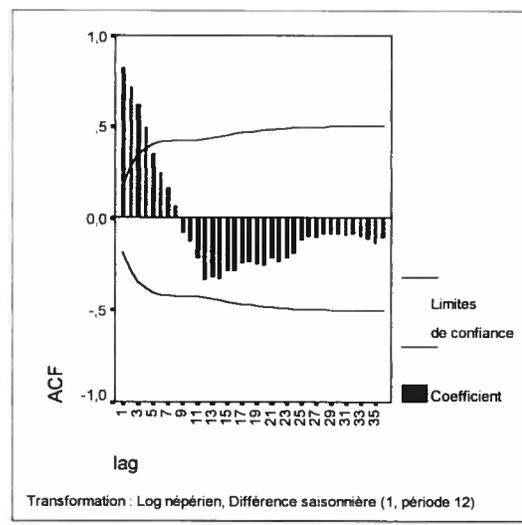
(a)



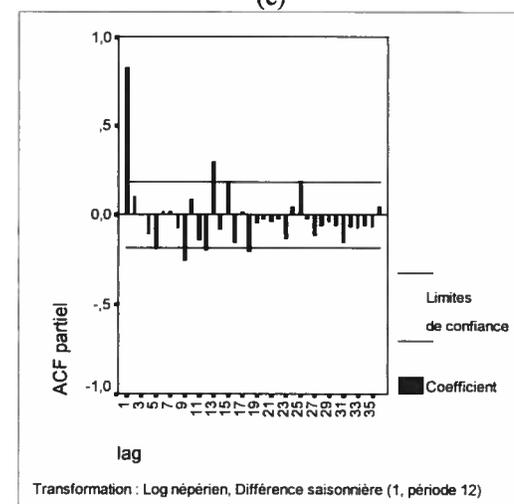
(b)



(c)



(d)



(e)

Incidence mensuelle de la rubéole au Canada (1991-2002) ajustée sur 28 jours et ajustée en ajoutant la constante 1 à toutes les données. (a) série chronologique, (b) série transformée en log, (c) ACF de la série (log), (d) ACF de la série (log) différenciée saisonnièrement (e) PACF de la série (log) différenciée saisonnièrement.

Rubéole 1991-2002 (suite)

Estimation des paramètres

Number of residuals	114
Standard error	,70123908
Log likelihood	-132,41348
AIC	270,82696
SBC	279,03555

Analysis of Variance:			
	DF	Adj. Sum of Squares	Residual Variance
Residuals	111	66,988067	,49173625

Variables in the Model:				
	B	SEB	T-RATIO	APPROX. PROB.
AR1	,81728817	,05075601	16,102293	,00000000
SMA1	,94462971	,34086627	2,771262	,00654899
CONSTANT	-,34501749	,10122144	-3,408542	,00091100

Diagnostic (ACF des résidus)

Lag	Auto-Corr.	Stand. Err.	-1	-.75	-.5	-.25	0	.25	.5	.75	1	Box-Ljung	Prob.
1	-,083	,094					**					,812	,368
2	,016	,094					*					,843	,656
3	,052	,094					*					1,168	,761
4	,126	,095					***					3,073	,546
5	-,055	,096					*					3,433	,633
6	,061	,096					*					3,884	,692
7	-,005	,097					*					3,888	,793
8	,121	,097					**					5,705	,680
9	-,135	,098					***					8,015	,533
10	,039	,100					*					8,205	,609
11	-,025	,100					*					8,287	,687
12	,015	,100					*					8,316	,760
13	,031	,100					*					8,441	,814
14	-,122	,100					**					10,402	,732
15	,036	,101					*					10,577	,782
16	-,113	,101					**					12,302	,723
17	,106	,102					**					13,845	,678
18	,093	,103					**					15,034	,660
19	-,089	,104					**					16,143	,648
20	-,118	,105					**					18,106	,580
21	,093	,106					**					19,345	,563
22	-,187	,107					****					24,380	,328
23	-,090	,109					**					25,550	,323
24	-,079	,110					**					26,461	,330
25	,137	,111					***					29,267	,253
26	-,055	,112					*					29,723	,279
27	-,117	,112					**					31,797	,240
28	-,005	,113					*					31,800	,283
29	,015	,113					*					31,836	,327
30	,085	,113					**					32,972	,324
31	-,024	,114					*					33,066	,367
32	,015	,114					*					33,100	,413
33	-,062	,114					*					33,727	,432
34	-,068	,114					*					34,502	,444
35	-,082	,115					**					35,618	,439
36	,089	,115					**					36,953	,425

Plot Symbols: Autocorrelations \* Two Standard Error Limits. Standard errors are based on the Bartlett (MA) approximation.

Total cases: 126 Computable first lags: 113

