

**Direction des bibliothèques**

**AVIS**

Ce document a été numérisé par la Division de la gestion des documents et des archives de l'Université de Montréal.

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**

This document was digitized by the Records Management & Archives Division of Université de Montréal.

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

**Facteurs agro-environnementaux associés à la résistance  
antimicrobienne d'*Escherichia coli* dans l'eau potable des  
puits du sud de l'Ontario**

Par

Rith Soulyvane Nguon

Département de pathologie et microbiologie

Faculté de médecine vétérinaire

Mémoire présenté à la Faculté des études supérieures  
en vue de l'obtention du grade Maître ès sciences (M.Sc.)

en sciences vétérinaires

option épidémiologie

Août, 2007

© Rith Soulyvane Nguon, 2007



Université de Montréal  
Faculté des études supérieures

Ce mémoire intitulé :  
Facteurs agro-environnementaux associés à la résistance antimicrobienne d'*Escherichia coli* dans l'eau potable des puits du sud de l'Ontario

présenté par :  
Rith Soulyvane Nguon

a été évalué par un jury composé des personnes suivantes :

Dr Michel Bigras-Poulin, président-rapporteur

Dr Pascal Michel, directeur de recherche

Dr André Ravel, membre du jury

## Résumé

Le but de cette étude est d'évaluer des associations entre des facteurs agro-environnementaux et la résistance antimicrobienne d'*Escherichia coli* dans l'eau potable. Des puits privés ont été échantillonnés du 1<sup>er</sup> mai 2004 au 30 avril 2005 en Ontario. Cette étude se base sur des données de 354 puits contaminés par *E. coli* et distribués sur 120 subdivisions de recensement unifiées (SRU). La sensibilité à 16 antibiotiques a été investiguée. Divers patrons de résistance ont été définis. Les variables dépendantes (#puits résistants/#puits total) et d'exposition ont été mesurées à l'échelle des SRU. Les deux antibiotiques démontrant les fréquences d'échantillons résistants les plus élevées furent la tétracycline (151/354) et le sulfaméthoxazole (97/354). Des modèles de régression logistique sur des données binaires groupées ont été construits. Trois jugés adéquats ont été retenus. Tous portaient sur le patron défini par la résistance à au moins un antibiotique. Chacun démontrait une association avec une variable agricole liée aux porcs et une de précipitation totale annuelle. Les variables agricoles ayant démontré une association significative dans chacun des trois modèles étaient respectivement la densité porcine (rapport de cotes : 2.6170), l'épandage de fumier liquide en surface (RC : 4.4027) et le rapport population porcine/population humaine (RC : 3.9780). La direction et la force d'association avec la variable de précipitation variaient selon son degré. Les conclusions doivent être nuancées étant donné la faible puissance de l'étude. Cependant, elles justifient toute de même une inquiétude face au rôle possible de l'agriculture dans la présence de l'antibiorésistance dans l'environnement.

**Mots-clés** : agriculture, antibiotique, environnement, épidémiologie, géospatiale, maladie hydrique

## Abstract

The objective of this study is to explore the associations between agroenvironmental factors and antimicrobial resistant *Escherichia coli* in drinking water. Private wells were sampled from May 1<sup>st</sup> 2004 to April 30<sup>th</sup> 2005 in Ontario. This study is based on samples taken from 354 wells contaminated by *E. coli* and distributed on 120 census consolidated subdivisions (CCS). Sensitivity to 16 antimicrobials has been tested. Various antimicrobial patterns were defined. The exposure-independent variables and the outcomes (#resistant wells/#total wells) were measured on the CCS scale. The two antibiotics showing the highest frequency of resistant samples were tetracycline (151/354) and sulfamethoxazole (97/354). The multivariable associative analyses were evaluated with logistic regression models on grouped binary data. Three models judged adequate were obtained. They all concerned the pattern defined by the resistance to at least one antibiotic. These models show an association with an agricultural variable linked to swine and the variable of total annual precipitation. The agricultural variables that demonstrated a significant association in each of the three models were respectively swine density (odds ratio: 2.6170), liquid manure spreading on the surface (OR: 4.4027) and ratio swine population/human population (OR: 3.9780). The direction and the force of the association with the variable of precipitation varied depending on the level of the total annual precipitation in the region. Although these conclusions are to be considered moderately because of the low power of the study, they can still justify a concern on the possible contribution of agriculture on antimicrobial resistance found in the environment.

**Keywords:** agriculture, antibiotic, environment, epidemiology, geospatial, waterborne disease

# Table des matières

Résumé.....	iii
Abstract.....	iv
Table des matières.....	v
Liste des tableaux.....	vii
Liste des figures.....	viii
Liste des sigles et des abréviations.....	ix
Remerciements.....	xii
Introduction.....	1
Partie 1 Recension de littérature.....	4
1.1. Impacts de la résistance antimicrobienne.....	4
1.1.1. Impacts économiques.....	4
1.1.2. Impacts sur la santé publique.....	5
1.2. Transfert de la résistance antimicrobienne des animaux aux humains.....	6
1.2.1. Voies de transmission.....	6
1.2.2. Contamination de l'eau potable.....	9
1.2.3. Mécanismes de transfert de la résistance antimicrobienne des animaux aux humains.....	14
1.3. Escherichia coli, une bactérie indicatrice.....	18
1.4. Eau potable en Ontario.....	20
1.4.1. Importance des puits en Ontario.....	20
1.4.2. Source de l'eau souterraine.....	21
1.4.3. Types de puits.....	21
1.4.4. Conditions des puits.....	22
Partie 2 Méthodologie.....	24
2.1. Données.....	24
2.1.1. Sources des échantillons.....	24
2.1.2. Analyse de laboratoire.....	25
2.1.3. Banques de données.....	31
2.1.4. Géocodage.....	32
2.1.5. Unités géographiques.....	33

2.1.6. Variables dépendantes.....	35
2.1.7. Variables indépendantes.....	36
2.2. Analyse descriptive.....	43
2.2.1. Transformation des variables.....	44
2.3. Modèle de régression.....	45
2.3.1. Construction du modèle de régression.....	46
2.3.2. Évaluation des modèles.....	48
2.4. Seuil alpha de décision.....	48
2.5. Logiciel et application.....	49
<b>Partie 3 Résultats.....</b>	<b>50</b>
3.1. Analyse descriptive.....	50
3.1.1. Variables dépendantes.....	50
3.1.2. Variables indépendantes.....	59
3.1.3. Association univariée.....	62
3.2. Modèle de régression.....	68
3.2.1. Corrélation.....	69
3.2.2. Modèles de régression.....	73
3.2.3. Évaluation des modèles.....	76
3.2.4. Association.....	79
<b>Partie 4 Discussion générale des résultats.....</b>	<b>80</b>
<b>Conclusions.....</b>	<b>99</b>
<b>Bibliographie.....</b>	<b>101</b>
<b>Annexes.....</b>	<b>1</b>

## Liste des tableaux

Tableau I : Fréquence de soumissions d'échantillons au projet ARO par laboratoire régional de santé publique	26
Tableau II : Seuil de concentration minimale inhibitrice selon l'antibiotique pour la méthode de diffusion en milieu gélosé	27
Tableau III : Seuil de concentration inhibitrice minimale pour diagnostiquer de la résistance antimicrobienne selon l'antibiotique (NARMS)	29
Tableau IV : Définitions des patrons de résistance antimicrobienne de <i>E. coli</i>	36
Tableau V : Définitions des variables indépendantes d'agriculture	38
Tableau VI : Définitions des variables indépendantes socio-économiques	39
Tableau VII : Définitions des variables indépendantes météorologiques	40
Tableau VIII: Fréquence de puits contaminés par <i>E. coli</i> résistant selon l'antibiotique	52
Tableau IX : Fréquence de puits positifs selon le patron de résistance au niveau de l'ensemble de la région d'étude	55
Tableau X: Résumé de l'analyse descriptive régionale des cinq variables dépendantes	55
Tableau XI : Fréquence de SRU selon la catégorie de qualité de drainage du sol	59
Tableau XII : Résumé de l'analyse descriptive des variables explicatrices	60
Tableau XIII : Définitions des classes de chaque variable indépendante	61
Tableau XIV : Résultats de l'analyse univariée des variables retenues	65
Tableau XV : Coefficient de Pearson par paire de variables indépendantes	70
Tableau XVI : Modèles initiaux à tester selon le patron de résistance	71
Tableau XVII : Statistique de tolérance des variables des combinaisons D, E et F avec et sans l'inclusion de la variable BoHu	72
Tableau XVIII: Variables significatives suite à la sélection rétrograde de variables appliquée sur chaque modèle initial déterminé au départ	74
Tableau XIX : Statistiques d'ajustement selon le modèle	76

## Liste des figures

Figure 1 : Modèle de transmission de la résistance antimicrobienne (inspiré de Witte [15])	8
Figure 2 : Résumé de la méthodologie employée avant la manipulation des données pour l'étude présente (année d'étude 2004)	30
Figure 3 : Régions agricoles de recensement et les SRU comprises dans la région d'étude	34
Figure 4 : Qualité de drainage du sol selon les SRU	42
Figure 5 : Diagramme causal	46
Figure 6 : Région d'étude et localisation des puits échantillonnés	51
Figure 7 : Fréquence de SRU selon le nombre de puits échantillonnés	53
Figure 8 : Distribution géo-spatiale de la fréquence absolue de puits échantillonnés par SRU	54
Figure 9: Distribution géo-spatiale de la fréquence de puits contaminés par <i>E. coli</i> résistant au patron 1 (un antibiotique ou plus)	56
Figure 10 : Distribution géo-spatiale de la fréquence de puits contaminés par <i>E. coli</i> résistant au patron 2 (un ou deux antibiotiques)	57
Figure 11: Distribution géo-spatiale de la fréquence de puits contaminés par <i>E. coli</i> résistant au patron 3 (multirésistance)	58
Figure 12: Fréquence relative de puits résistants au patron 1 par SRU et catégorisation des SRU quant à l'épandage de fumier liquide en surface	66
Figure 13 : Fréquence relative de puits résistants au patron 1 par SRU et catégorisation des SRU quant au rapport nb de porcins / nb d'humains	67

## Liste des sigles et des abréviations

%	Pourcent
$\alpha$	Alpha
/	Division
$\mu\text{g}$	Microgramme
$\mu\text{L}$	Microlitre
CDC	<i>Centers for Disease Control and Prevention</i>
CMI	Concentration minimale inhibitrice
DC	<i>Differential coliform</i>
<i>E. coli</i>	<i>Escherichia coli</i>
ha	Hectare
I.C.	Intervalle de confiance
INSPQ	Institut national de santé publique du Québec
km	Kilomètre
LZA	Laboratoire de lutte contre les zoonoses d'origine alimentaire
Max	Maximum
Méd	Médiane
Min	Minimum
mL	Millilitre
Moy	Moyenne
MOHLTC	<i>Ministry of Health and Long Term Care of Ontario</i>
NARMS	<i>National Antimicrobial Resistance Monitoring System</i>
Nb	Nombre
no	Numéro
P	Probabilité
PICRA	Programme intégré canadien de surveillance de la résistance aux antimicrobiens
PLPH(M)	<i>Provincial Laboratory for Public Health (Microbiology)</i>

Projet ARO	<i>Prospective Multi-Province Surveillance for Antimicrobial-Resistant Escherichia coli in Drinking and Recreational Source Waters: Impact on Humans and the Environment</i>
RC	Rapport de cotes
RAR	Région agricole de recensement
SDR	Subdivision de recensement
SIG	Système d'information géographique
SRU	Subdivision de recensement unifiée
STEC	Shiga-toxine <i>E. coli</i>
SWU	<i>Food Safety and Safe Water Unit</i>
TSA	<i>Tryptic Soy Agar</i>
TSB	<i>Tryptic Soy Broth</i>

## Remerciements

Cette étude fait partie d'un projet canadien intitulé *Prospective Multi-Province Surveillance for Antimicrobial-Resistant Escherichia coli in Drinking and Recreational Source Waters: Impact on Humans and the Environment* (projet ARO). Cette étude a été rendue possible grâce au travail de tous les collaborateurs et participants du groupe de recherche ARO formé par I. Barette, M. Boily, P. Boleszczuk, S. Braithwaite, M. Buzzelli, P. Cantin, A. Chagla, P. Chevalier, B. Ciebin, K. Coleman, C. Coulombe, B. Crago, J. Csubak, D. Daignault, C. Dalpe, R. Eddington, L. Gagnon, V. Germain, B. Goodbrand, K. Grimsrud, K. Guay, C. Guénette, M. Guilbeault, E. Harris, C. Hilner, O. Hsueh, R. Irwin, F. Jamieson, M. Jerrett, A. Joyce, C. Kalapos, N. Latimer, M. Lavoie, D. Ledrew, C. Letoumeau, P. Levallois, M. Louie, A. Maki, S. McEwen, A. McGeer, P. Michel, M. Mulvey, M. Murphy, N. Neumann, P. O'Brian, Z. Rajda, J. Ramos, R. Retallick, M. Ronald, N. Rowell, F. Ruf, M. Salvadori, P. Sankar, B. Skora, D. Strong, P. Verbrugghe. Le financement accordé par les Instituts de recherche en santé du Canada a été obtenu par l'investigatrice principale, le Dre Marie Louie que je tiens remercier.

Je voudrais remercier particulièrement le Dr Pascal Michel pour m'avoir dirigé dans ma formation en épidémiologie à l'aide de ce projet de recherche.

Je tiens aussi à remercier Caroline Guénette, géographe médicale à l'Université de Montréal, pour sa disponibilité et son aide très précieuse particulièrement concernant la composante géospatiale de l'étude, ainsi que Guy Beauchamp, biostatisticien à l'Université de Montréal, pour son aide avec les analyses statistiques.

Je remercie aussi les membres de mon comité conseil, le Dre Denise Bélanger, le Dre Marie Archambault, le Dr Alex Thompson, également professeur à l'Université de Montréal au moment de l'étude.

Je remercie le Dre Danielle Daignault, microbiologiste vétérinaire et Julie Roy, technicienne, toutes deux au Laboratoire de lutte contre les zoonoses d'origine alimentaire à l'Agence de santé publique du Canada, pour leur aide à la rédaction de la composante microbiologique de la méthodologie.

En dernier lieu, je tiens à remercier ma famille et mes proches pour leur appui.

## Introduction

Parmi les maladies hydriques, les maladies entériques représentent une proportion importante au Canada. La plupart de ces éclosions implique de l'eau de surface, surtout dans les régions rurales [1]. Une certaine proportion de ces maladies entériques est causée par un agent bactérien zoonotique. Une analyse portant sur 288 éclosions reliées à l'eau potable au Canada de 1974 à 2001 a rapporté que les agents infectieux les plus souvent en cause sont, en ordre croissant : *Giardia*, *Campylobacter*, *Cryptosporidium*, virus Norwalk-like, *Salmonella* et virus de l'hépatite A. Quarante-huit pourcent des éclosions sont reliées à un système d'eau parapublique, 34% à un système public et 18% à un système privé [2]. Il est rapporté que pour chaque cas de gastro-entérite infectieuse aiguë rapporté aux autorités provinciales de santé publique et confirmé au laboratoire, il y avait probablement plus de 300 autres cas non déclarés [2]. Au Canada, l'éclosion de maladies hydrique la plus célèbre et la plus médiatisée est probablement celle de Walkerton (Ontario) en 2000. Cette éclosion a impliqué entre autre *Escherichia coli* O157:H7 et *Campylobacter* spp. Un événement de l'envergure de celui-ci n'avait jamais été répertorié au Canada depuis le début du traitement de l'eau potable au chlore (début du 20<sup>e</sup> siècle) [3]. Cet évènement a fait 1 346 malades, dont 65 hospitalisations. Vingt-sept personnes ont développé le syndrome urémique hémolytique et six sont décédés [4].

Une gastro-entérite est généralement auto-limitante et ne requière qu'un traitement symptomatique sans antibiotique. Cependant, de plus en plus de gastro-entérites résistantes aux antibiotiques font surface. Ces infections causées par une bactérie résistante nécessitent souvent un traitement antibiotique prolongé ou requièrent d'autres antibiotiques plus dispendieux [5, 6]. De plus, un nombre croissant d'agents infectieux autrefois simples à éliminer sont maintenant devenus multirésistants. En effet, la résistance multiple peut provenir de plusieurs années de pressions de sélection dues à un usage abusif d'antibiotique [6]. Habituellement, la résistance bactérienne est reconnue problématique en médecine

lorsqu'un pathogène persiste malgré un traitement antibiotique prescrit. Par contre, lorsque qu'une résistance antimicrobienne émerge chez un agent pathogène humain, autrement que par une sélection rapide d'un chromosome mutant chez une bactérie, il est possible d'assumer que plusieurs autres bactéries ont aussi acquis le même mécanisme de résistance. Ces bactéries peuvent être des organismes commensaux provenant de l'environnement et des animaux. Souvent, il y a un laps de temps potentiellement long entre l'apparition du mécanisme de résistance et son observation médicale [5].

À ce jour, il existe plus de 15 classes d'antimicrobiens [7]. Plus de 150 antibiotiques sont disponibles sur le marché. Cependant, les découvertes et les entrées sur le marché de nouvelles molécules efficaces sont moins fréquentes qu'autrefois [8]. La difficulté grandissante de traiter certaines infections est de plus en plus perçue comme étant une menace à la santé publique. Cette difficulté est expliquée par plusieurs aspects de la résistance antimicrobienne, incluant son développement, sa dissémination à partir des hôpitaux vers les populations en santé, la découverte de nouveaux mécanismes de résistance, le développement de la résistance multiple par certains agents pathogènes et la facilité avec laquelle le matériel génétique codant pour la résistance peut dans certains cas se propager horizontalement entre différentes espèces bactériennes [5]. Certains auteurs proposent que les *E. coli* résistants aux antibiotiques d'origine animale et contaminant notre environnement constituent une source potentielle importante de plasmides pour les *E. coli* indigènes présents dans le tractus intestinal des humains et subséquentement pour les agents pathogènes aux humains [9, 10].

La résistance antimicrobienne et le développement de résistance croisée sont des phénomènes très complexes à comprendre. Les contributions de la médecine humaine, de la médecine vétérinaire et de l'agriculture à l'exacerbation de ce phénomène sont peu comprises. Alors que nous dépendons beaucoup de l'eau, il n'existe aucun programme de surveillance et d'étude systématique et détaillée pour déterminer l'impact de la résistance antimicrobienne présente dans l'eau potable et de récréation au Canada (Pascal Michel et Marie Louie, communication personnelle), d'où le besoin de créer le projet intitulé

*Prospective Multi-Province Surveillance for Antimicrobial-Resistant Escherichia coli in Drinking and Recreational Source Water : Impact on Humans and the Environment* (projet ARO). Cette initiative canadienne a pour but d'évaluer les multiples facettes de la problématique de la résistance antimicrobienne retrouvée dans l'eau et ses impacts sur la santé publique. Trois provinces participent à ce projet : l'Alberta, l'Ontario et le Québec.

La présente étude est une composante de cette initiative canadienne. Elle porte plus spécifiquement la contamination des puits privés par *E. coli* résistant aux antibiotiques dans le Sud de l'Ontario. L'Ontario est une province du Canada, situé entre le Manitoba et le Québec. Elle a une population de plus de 12 millions d'habitants, ce qui représente près du tiers de la population canadienne [11]. Plus de trois millions d'Ontariens dépendent de la nappe phréatique comme source principale d'eau [12].

L'objectif général de la présente étude est d'explorer la distribution spatiale de la présence de la résistance antimicrobienne dans l'eau destinée à la consommation humaine. Plus précisément, cette étude a pour but de déterminer les associations possibles au niveau régional entre divers facteurs agro-environnementaux, tels les densités animales, l'utilisation des terres, facteurs météorologiques, etc., et la résistance de la bactérie indicatrice à divers patrons d'antibiotiques.

# **Partie 1 Recension de littérature**

## **1.1. Impacts de la résistance antimicrobienne**

La résistance antimicrobienne est un phénomène dont la société d'aujourd'hui s'inquiète de plus en plus.

### **1.1.1. Impacts économiques**

Tout d'abord, il y a les impacts économiques. À ce jour, il existe plus de 15 classes d'antimicrobiens [7]. Plus de 150 antibiotiques sont disponibles sur le marché [8]. Cependant, les découvertes et entrées sur le marché de nouvelles molécules efficaces se font moins nombreuses qu'autrefois [8]. Ainsi, les infections résistantes aux antibiotiques sont malheureusement de plus en plus fréquentes. Ce type d'infection demande souvent un traitement de longue durée. Un traitement antibiotique de longue durée, plus difficile, nécessitant parfois un changement de molécules ou des combinaisons, occasionne une augmentation des coûts monétaires pour la société. Ceci est surtout le cas des infections à résistance multiple. À titre d'exemple, les coûts directs de la résistance antimicrobienne dans les hôpitaux américains sont estimés entre 100 millions et 10 milliards \$US par année [8]. Le *US Office of Technology Assessment* a estimé le coût minimal à l'hôpital causé par la résistance antimicrobienne de cinq types d'infection nosocomiale à 4.5 milliards \$US par année [8]. Selon un rapport du Comité canadien sur la résistance aux antibiotiques [13], il est estimé que les infections résistantes aux antibiotiques ajoutent annuellement au système de santé au moins 14.2 à 25.5 millions \$CAN en coûts directs dus aux hospitalisations. Le dépistage des porteurs de microorganismes résistants ajoute environ 10.3 millions \$CAN et les mesures pour prévenir la propagation de ces microorganismes qui colonisent les porteurs sans infection active ajoutent environ 15.9 millions \$CAN. Selon ce rapport, si la

prévalence canadienne de la résistance antimicrobienne augmentait et atteignait les niveaux élevés semblables à ceux retrouvés aux États-Unis, les coûts annuels directs s'élèveraient à entre 103.9 et 187.1 millions \$CAN, les coûts de dépistage des porteurs demeureraient les mêmes et les coûts destinés à la prévention augmenteraient à 157.2 millions \$CAN.

### **1.1.2. Impacts sur la santé publique**

Aux impacts économiques, s'ajoutent malheureusement ceux sur la santé publique. Le principal impact est la difficulté et parfois l'impossibilité de traiter des infections bactériennes. Une gastro-entérite est généralement auto-limitante et ne requière qu'un traitement symptomatique sans antibiotique. Cependant, si la maladie est causée par une bactérie résistante et doit être traitée à l'aide d'un antibiotique, le traitement peut requérir une prolongation [5, 6]. De plus, certains agents pathogènes résistants peuvent avoir une virulence ou une pathogénicité plus importante chez les humains que ceux qui sont sensibles causant ainsi une maladie plus grave, plus longue et plus difficile à traiter [6]. Le deuxième impact est la possibilité d'une propagation accrue de l'infection ou d'une durée prolongée de l'excrétion fécale des agents pathogènes résistants au sein des populations d'animaux recevant l'antibiotique, ce qui pourrait augmenter la disponibilité de ces organismes pour l'infection des humains [6]. Finalement, il y a le développement du phénomène complexe de la résistance multiple. Cette dernière peut provenir de plusieurs années de pressions de sélection de bactéries résistantes provenant de différents fermes, espèces animales (incluant les humains) et pays [6]. Une pression de sélection fait référence aux conditions environnementales permettant aux organismes avec de nouvelles mutations ou ayant nouvellement acquis des traits de résistance de survivre et de proliférer [14]. La meilleure façon de prévenir le développement de la multirésistance est de diminuer les sources de pression de sélection en diminuant l'utilisation d'antibiotiques dans tous les domaines possibles [6]. Bien que toutes les bactéries aient la capacité de développer une forme de résistance, certaines espèces ou souches semblent plus enclines à développer une résistance multiple. Ceci semble entre autre être le cas de *Salmonella* Typhimurium et *Campylobacter jejuni*, deux agents pathogènes zoonotiques [6].

## **1.2. Transfert de la résistance antimicrobienne des animaux aux humains**

Pathogène ou non, les bactéries résistantes aux antimicrobiens peuvent se transmettre des animaux aux humains (et vice versa).

### **1.2.1. Voies de transmission**

Les bactéries peuvent se transmettre entre animaux et humains par différentes voies. Le modèle (voir Figure 1, p. 8) inspiré de Witte illustre bien le modèle de transmission de bactéries résistantes aux antibiotiques [15]. Selon ce modèle, il y a trois réservoirs principaux de bactéries résistantes aux antibiotiques : les animaux de consommation sains ou malades, les patients hospitalisés et les humains sains ou malades dans la communauté. Il est à noter que la résistance aux antibiotiques apparaît surtout lorsque les antibiotiques sont utilisés abondamment et que les bactéries peuvent se transmettre rapidement entre les individus. Ainsi, les hôpitaux et les exploitations intensives de bétail sont des environnements très propices au développement de la résistance antimicrobienne [6]. Les bactéries du réservoir des patients hospitalisés subissent une pression de sélection occasionnée par l'utilisation d'antibiotiques en prophylaxie et en thérapie. Celles du réservoir composé par les animaux de consommation subissent une pression de sélection occasionnée par l'utilisation d'antibiotiques comme promoteurs de croissance, en prophylaxie et en thérapie. Suite à l'épandage de fumier et à l'évacuation des eaux usées, les bactéries provenant de ces réservoirs peuvent se retrouver dans l'environnement, dont l'eau de surface et les nappes phréatiques. Les animaux se contaminent suite à la consommation de végétaux et d'eau contaminés et les humains, suite à l'ingestion de végétaux, d'eau ou de viandes contaminés. De plus, les animaux et les humains peuvent se contaminer directement entre eux ou d'une espèce à l'autre [15].

En conséquence, les agents entéro-pathogènes zoonotiques et les agents commensaux antibiorésistants sont à surveiller car ces bactéries peuvent coloniser le tractus digestif des humains en passant par la chaîne alimentaire [10]. Ce passage via la chaîne alimentaire peut se faire directement par la consommation de viande. Il est à noter que de la présence de résistance antimicrobienne chez des animaux de consommation et sur la viande est l'objet de plusieurs programmes de surveillance de la résistance antimicrobienne, dont le Programme intégré canadien de surveillance de la résistance aux antimicrobiens (PICRA) [16], le *National Antimicrobial Resistance Monitoring System* (NARMS) [17] et le *Danish Integrated Antimicrobial resistance Monitoring and Research Programme* (Danmap) [18]. Un autre passage via la chaîne alimentaire peut se faire par l'ingestion directe ou indirecte de l'eau contaminée par des déchets fécaux ou animaux [1]. Il y a plusieurs voies de transmission hydrique d'agents pathogènes possibles. Ces voies sont entre autre, la consommation de l'eau potable contaminée (ingestion directe) ou d'aliments ayant été contaminés par de l'eau contaminée et l'exposition à de l'eau contaminée lors d'activités récréatives (ingestion indirecte). Les eaux usées par les humains, les fuites des systèmes septiques et le ruissellement des fumiers à partir de terres agricoles et des déchets fécaux des animaux sauvages peuvent contaminer l'eau de surface qui sera par la suite utilisée comme eau potable [1]. Les nappes phréatiques peuvent être contaminées par la contamination en surface des puits, le flot souterrain, les champs septiques incorrectement situés ou les fuites à partir de ceux-ci [1].

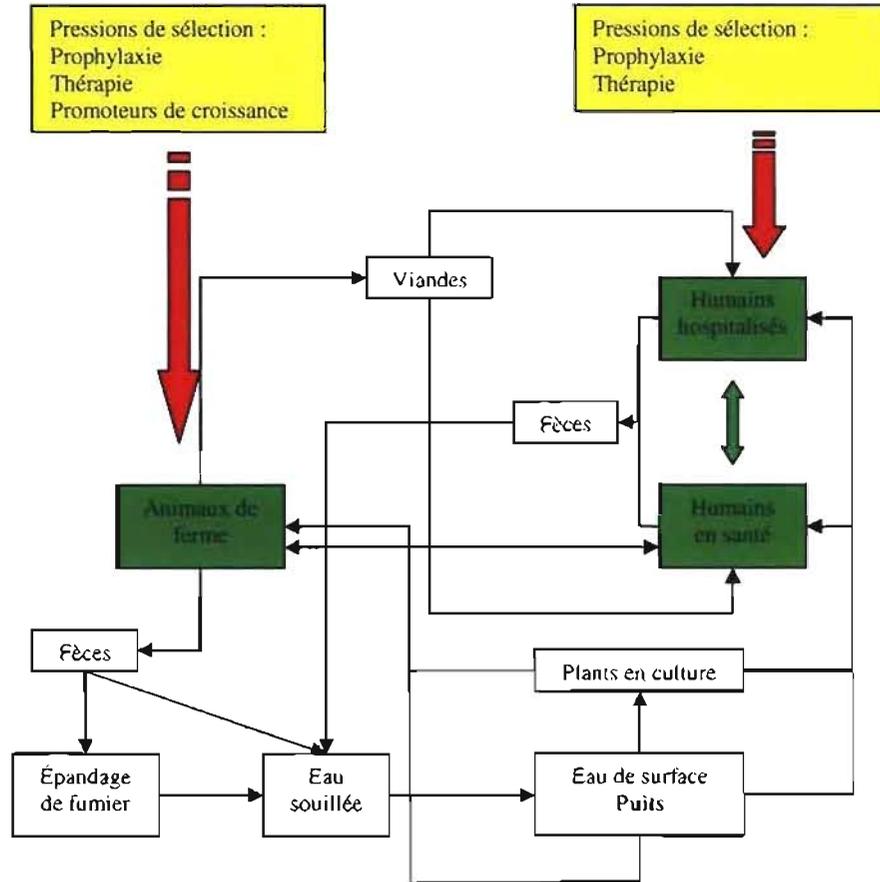


Figure 1 : Modèle de transmission de la résistance antimicrobienne (inspiré de Witte [15])

Aux trois réservoirs principaux énumérés par Witte, on pourrait ajouter celui de l'environnement. Sous des conditions expérimentales, plusieurs résidus d'antibiotiques non métabolisés après leur usage initial et se biodégradant difficilement semblent posséder la capacité de se retrouver et de persister dans un environnement aquatique [19]. Par exemple, les quinolones adsorbent fortement aux eaux usées, aux sols et aux sédiments. La sarafloxacin, un antibiotique utilisé pour la prévention de maladies dans la production aviaire, est un bon exemple de fluoroquinolone qui se biodégrade peu dans l'environnement [19]. Après 80 jours, seulement moins d'un pourcent est éliminé de l'environnement. Ceci pourrait être expliqué par sa grande capacité à se lier au sol. La virginiamycine, un promoteur de croissance, se biodégrade dans différents types de sol avec une très longue demi-vie [19]. Plusieurs antibiotiques utilisés chez les animaux de ferme ne sont que très

peu absorbés par leur tractus digestif. En fait, il est estimé que 25 à 75% des antimicrobiens administrés aux animaux dans les parcs d'engraissement pourrait demeurer non altéré à leur excrétion dans l'environnement et persister dans le sol après leur épandage sur les terres agricoles [20]. Il est estimé que près de 180 millions de tonnes de matières sèches de déchets de bétails et de volailles sont produits chaque année en sol américain. Associé à l'usage d'antibiotiques, ces déchets constituent une source potentielle considérable d'antibiotiques et de bactéries résistantes relâchés dans l'environnement [20]. Les antibiotiques retrouvés dans l'environnement exercent donc également une pression de sélection sur les populations bactériennes dans le sol et l'eau.

## **1.2.2. Contamination de l'eau potable**

Il a déjà été démontré que certains facteurs agricoles favorisent la présence de bactéries dans l'environnement avoisinant les puits tandis que les conditions météorologiques extrêmes favorisent le ruissellement de ces bactéries vers les puits. Sans égard à la présence de résistance antimicrobienne, la contamination de l'eau et les gastro-entérites ont souvent été associées à ces deux catégories de facteurs (agricoles et météorologiques).

### **1.2.2.1. Rôle des animaux de production dans la transmission hydrique de bactéries**

#### *Transmission hydrique de bactéries en général*

Des études ont démontré que le niveau régional d'infections humaines par *E. coli* pathogène est possiblement associé à certaines activités agricoles intenses dans la région. Se basant sur des données de 1990 à 1995 en Ontario, une étude de Michel et al. [21] évaluait l'association spatiale entre l'incidence des cas d'infection à shiga toxine *E. coli*

(STEC) dans les populations humaines et des facteurs d'exposition agro-environnementaux en Ontario. Vivre dans une région rurale où des bovins sont élevés avait été identifié comme facteur de risque potentiel important pour l'acquisition d'une infection à STEC. Les auteurs suggéraient que l'incidence de cas d'infection dans la population humaine était plus élevée dans les régions rurales que dans les régions urbaines. Selon eux, la densité d'animaux de rente, particulièrement celle des bovins, était une variable d'exposition significative dans plusieurs régions de l'Ontario. Les résultats supportaient l'influence possible du contact direct ou indirect avec un réservoir d'animaux comme mode important de la transmission. Ce contact dans les régions à haute densité de bétail pouvait être dû à des facteurs comme la contamination de l'eau de surface et de puits peu profonds par du fumier de bovins utilisé comme fertilisant, le fait de travailler avec ou être en contact avec les bovins, ainsi que la consommation de produits alimentaires produits et transformés localement. Plus tard, une autre étude également effectuée en Ontario étudia l'association entre des indicateurs d'intensité d'animaux de ferme et l'incidence d'infection humaine à STEC [22]. Selon les auteurs, cette étude était la première à indiquer l'application de fumier sur les terres comme facteur de risque potentiel pour une infection endémique chez les humains. Les auteurs ont expliqué cette association par la possibilité de ruissellement à partir de terres agricoles traitées avec du fumier et la contamination, par la suite, des eaux de surface et des puits servant à la consommation.

L'écllosion de gastro-entérites hydriques impliquant *E. coli* O157:H7 et *Campylobacter* spp. associées à la contamination d'un réseau d'aqueduc municipal à Walkerton, en Ontario, est un bon exemple illustrant le rôle du secteur agricole dans la contamination de l'eau potable [4]. Une analyse moléculaire a révélé que les isolats provenant d'une ferme adjacente au puits soupçonné être la cause de l'épidémie étaient identiques à ceux retrouvés chez la plupart des humains malades. De plus, les données recueillies suggéraient que les agents pathogènes ayant contaminé le puits en question provenaient du fumier de bovins de cette ferme. Un modèle de simulation illustrant l'écoulement de l'eau après une pluie abondante aux alentours de ce puits indiquait que l'eau tombant sur la grange et les champs avoisinants aurait ruisselé vers le puits. La

présence d'*E. coli* O157:H7 et de *Campylobacter* spp. dans le fumier bovin environnant fut identifiée comme facteur prédisposant à la contamination de l'eau potable.

#### *Transmission hydrique de bactéries résistantes*

D'après une étude de Kaspar et al. publiée en 1990 [23], il y avait une différence significative entre le pourcentage d'*E. coli* résistants aux antibiotiques dans les eaux des régions urbaines et celui des régions rurales. Le pourcentage des souches résistantes aux antibiotiques était plus élevé en milieu urbain qu'en milieu rural. Par contre, l'origine des souches résistantes aux antibiotiques n'avait pas été étudiée. Plus tard, les trouvailles provenant de l'étude de McKeon, Calabrese et Bissonnette [24], ont plutôt suggéré que la présence de bactéries résistantes aux antibiotiques n'était pas un évènement rare dans un environnement rural. De hautes proportions d'isolats de coliformes et non coliformes étaient résistantes à la novobiocine et/ou céphalotine et/ou ampicilline. Plus de 90% des isolats étaient résistant à au moins un antibiotique et 78% étaient résistant à deux antibiotiques ou plus. Quarante-six pourcent des isolats d'*E. coli* étaient résistants à au moins un antibiotique, tandis que 14% l'étaient à deux antibiotiques ou plus et 11% l'étaient à cinq antibiotiques ou plus. De plus, cette étude suggère aussi que sous des conditions expérimentales favorables, les bactéries possédant des facteurs de résistance pourraient transférer les caractéristiques de résistance à d'autres bactéries réceptrices.

Chee-Sanford et al. [20], quant à eux, ont mené une étude portant sur la présence de gènes de résistance dans les nappes phréatiques et le lien possible avec l'entreposage de purin dans des lagunes. Les auteurs ont utilisé la résistance aux tétracyclines pour surveiller les gènes de résistance pertinents parce que l'usage de cet antibiotique dans l'industrie porcine était fréquent. La détection des gènes dans les lagunes indique qu'il y a eu une sélection pour la résistance à cette drogue et qu'un certain nombre de gènes de résistance peuvent être maintenu dans les populations microbiennes existantes. De plus, un grand nombre de gènes ont été détectés dans les nappes phréatiques en aval des lagunes. La

fréquence relative et la diversité des gènes de résistance à la tétracycline suivent un gradient spatial dont les valeurs maximales sont situées au niveau des lagunes et diminuent graduellement selon la direction du flot de l'eau dans le sol. Des gènes ont tout de même été détectés à 250 m en aval des lagunes. Ceci suggère une mobilité probablement importante des gènes de résistance dans l'environnement. De plus, la présence de gènes de résistance a été retrouvée de façon plus importante dans les puits profonds que dans les puits superficiels. Ainsi, les contaminants peuvent être verticalement mobile dépendamment de l'hydrologie de la localisation donnée. Suite aux résultats obtenus, les auteurs ont soulevé la possibilité d'acquisition des gènes de résistance par les humains et les animaux attribuée à la consommation d'eau potable. De plus, dans une autre étude associée, certains des gènes prédominants dans les lagunes et les puits sont aussi prédominants dans la flore intestinale des porcs et bouvillons, suggérant ainsi que les gènes dans l'environnement peuvent provenir des animaux environnants (ou l'inverse) [20].

#### **1.2.2.2. Rôle des événements météorologiques dans la transmission hydrique de bactéries**

Selon une étude de Schuster et al., la majorité des éclosions ont lieu au printemps et en été [2]. Plusieurs hypothèses peuvent expliquer ces patrons saisonniers. Par exemple, la fonte des neiges au printemps peut fournir un ruissellement concentré de pathogènes aux ruisseaux et aux rivières [2]. L'augmentation des flots terrestres résultant de la fonte des neiges et des précipitations de pluie au printemps, ainsi que les précipitations sporadiques de pluie intenses séparées par des conditions de sécheresse en été peuvent augmenter la fréquence des maladies durant ces saisons [2]. Par contre, selon une étude de Rodgers et al., la contamination des cours d'eau étudiés est plus importante pendant les mois d'été. De plus, les concentrations maximales de contamination fécale ont été mesurées lors d'événements importants de précipitation de pluie [25]. L'incidence des cas d'infection à *E. coli* pathogène suit également un patron saisonnier avec un pic en été. Par exemple, le pic d'incidence de STEC en Ontario dans l'étude de Michel et al. a eu lieu au mois de juillet [21].

Plusieurs recherches révèlent que les précipitations excessives de pluie contribueraient à l'éclosion de maladies transmissibles par l'eau. Auld, MacIver et Klaassen, ont répertorié plusieurs de ces éclosions [26]. Il y a entre autres l'éclosion de *Cryptosporidium* en 1993 à Milwaukee, au Wisconsin, causant la mort de 54 personnes et ayant rendu plus de 403 000 personnes malades [26]. Il y a aussi la plus importante éclosion d'*E. coli* O157:H7 dans l'état de New York en septembre 1999. Il est à noter que la précipitation de pluie très forte avait été précédée d'une période de sécheresse [26]. Une étude américaine majeure a rapporté une association entre une précipitation de pluie excessive et les éclosions de maladies transmissibles par l'eau [26]. Cette étude tenait compte de 548 éclosions rapportées aux États-Unis entre 1948 et 1994. Cinquante-et-un pourcent de celles-ci ont eu lieu suite à une période de précipitations de pluie extrêmes définie par un événement de précipitations d'un niveau supérieur au 90<sup>e</sup> percentile pour la région donnée. Soixante-huit pourcent étaient précédé d'un événement de précipitations d'un niveau supérieur au 80<sup>e</sup> percentile pour la région donnée. Les éclosions associées à la consommation d'eau de surface démontrent une forte association avec des précipitations excessives de pluie durant le mois de l'éclosion, alors que les contaminations des eaux profondes étaient précédées d'une accumulation de pluie ayant eu lieu deux mois auparavant. Il est à noter que les auteurs soulignent que la fonte des neiges ou la fonte des neiges additionnée des précipitations de pluie, des facteurs ayant le potentiel d'augmenter l'association significative entre les précipitations et les éclosions de maladies transmissibles par l'eau, n'ont pas été tenues compte dans l'étude.

Au Canada, la célèbre éclosion d'*E. coli* O157:H7 et de *Campylobacter* à Walkerton a été en partie attribuée à des précipitations excessives de pluie [26]. Ces précipitations ont eu lieu du 9 au 12 mai 2000, alors que le début de la majorité des cas est apparu après le 12 mai, avec un pic entre le 17 et 19 mai. L'hypothèse est que les fortes pluies de mi-mai ont été responsables d'une contamination significative du système de distribution d'eau, causant alors la majorité des cas de maladies gastro-intestinales. Une précipitation totale record pour un mois de mai depuis 1972, i.e. de 206 mm, a été enregistrée à la station météorologique la plus proche de Walkerton. La moitié de la précipitation totale du mois s'est accumulée entre le 8 et le 12 mai. Dans l'enquête, le rôle de la température n'a pas été

déterminé. Par contre, il a été répertorié que la température s'était réchauffée les mois précédant la contamination. En résumé, l'accumulation de pluie significative causée par les fortes précipitations étendues sur plusieurs jours et la saturation du sol par la suite ont augmenté les risques de contamination des nappes phréatiques.

En somme, les conditions météorologiques tendent à exacerber les vulnérabilités créées par une protection inadéquate de l'eau [2]. La contamination de l'eau se fait souvent au début du printemps après la fonte des neiges, et en été suite à une longue période de sécheresse ou à de fortes pluies. Les facteurs agro-environnementaux favorisent la présence de bactéries dans l'environnement avoisinant les puits tandis que les conditions météorologiques extrêmes favorisent le ruissellement de ces bactéries vers les puits.

### **1.2.3. Mécanismes de transfert de la résistance antimicrobienne des animaux aux humains**

Les réservoirs animaux et humains du modèle vu précédemment peuvent contenir des bactéries qui sont commensales et d'autres qui sont pathogènes. Deux mécanismes de transfert de la résistance antimicrobienne à partir des animaux aux humains fréquemment décrits impliquent l'un ou l'autre type de bactéries [5, 27].

#### **1.2.3.1. Transfert par des bactéries pathogènes zoonotiques**

Certains agents pathogènes provenant des animaux sont qualifiés zoonotiques. Un transfert direct aux humains d'agents pathogènes zoonotiques résistants aux antibiotiques par l'exposition aux animaux contaminés ou produits de consommation d'origine animale est possible. Quelques publications ont suggéré l'association entre la consommation de divers produits animaux contaminés par des agents résistants aux antibiotiques et les infections humaines à ces agents résistants. Ces agents zoonotiques sont, entre autres :

Salmonella DT 10 multirésistant [28], *E. coli* O157:H7 [29] et *Campylobacter* [30, 31]. Concernant ce dernier agent zoonotique, une étude menée en Angleterre à un moment où l'enrofloxacin n'était pas encore homologué comme promoteur de croissance chez le poulet, contrairement à d'autres pays d'Europe, a rapporté que 4.1% des infections chez les humains était tout de même résistant au ciprofloxacine alors qu'aucun des patients n'avait consommé cet antibiotique. Il y a eu une augmentation du nombre de gastro-entérites où une souche de *Campylobacter* résistant au ciprofloxacine a été isolée. Les auteurs de cette étude ont suggéré que les facteurs de risque sont un voyage récent à l'extérieur du pays et possiblement la consommation de poulets importés. Selon eux, l'utilisation vétérinaire de quinolones, dont l'enrofloxacin, exerce une pression de sélection pour l'émergence d'isolats résistants au ciprofloxacine chez les humains [30].

### 1.2.3.2. Transfert de bactéries commensales

Mis à part l'acquisition d'agents pathogènes zoonotiques, l'humain peut être affecté par la résistance antimicrobienne via des bactéries commensales qui colonisent sa flore intestinale. Cet autre mécanisme implique un transfert de bactéries commensales et un transfert de leur gène codant pour la résistance. Une bactérie commensale a typiquement de longues périodes de persistance, des transmissions entre humains fréquentes et de grandes charges bactériennes sans causer de maladie. Le danger de ce mécanisme de transfert indirect réside surtout dans la possibilité de transférer du matériel génétique mobile codant pour des mécanismes de résistance antimicrobienne à des agents pathogènes humains déjà multirésistants. Certains évènements ou publications supportent la possibilité de transfert de gènes de résistance à partir de bactéries commensales vers des agents pathogènes. Citons quelques exemples reconnus dans la littérature. Le premier exemple est celui de la molécule de nouséothricine utilisée en Allemagne [15]. Cette molécule a été utilisée comme promoteur de croissance en Allemagne de l'Est de 1983 à 1990. La résistance à cet antibiotique parmi les *Enterobacteriaceae* était négligeable en 1983 chez les humains et les animaux. Deux ans après l'introduction de ce promoteur de croissance, des *E. coli* résistants à cet antibiotique ont été retrouvés dans la flore intestinale de porcs et sur la viande. Cette

résistance est liée à un gène particulier. En 1990, la résistance s'est propagée aux *E. coli* de la flore intestinale des producteurs de porcs, de leurs familles, d'autres membres des communautés et de patients souffrant d'infection urinaire. En 1987, le même déterminant de résistance a été détecté chez d'autres agents pathogènes entériques dont *Shigella*, un microorganisme retrouvé seulement chez l'humain. Le deuxième exemple est l'étude de Johnson et al. [32]. Dans cette étude, des isolats d'*E. coli* résistants à la gentamicine et à l'apramycine ont été retrouvés chez des humains. Plusieurs de ceux-ci étaient résistants à l'hygromycine B. Tous les isolats résistants à la gentamicine/apramycine possédaient le gène codant pour l'enzyme AAC(3)IV qui transmet la résistance à la gentamicine/apramycine aux bactéries isolées à partir des animaux. La résistance à la gentamicine/apramycine et à l'hygromycine B était co-transférable dans la plupart des isolats. L'apramycine (un aminoglycoside) et l'hygromycine B (un antihelminthique) n'avaient jamais été utilisés chez les humains. L'explication la plus probable selon les auteurs de l'étude est que le déterminant génétique de la résistance a été acquis directement ou indirectement à partir de bactéries résistantes à la gentamicine et à l'apramycine d'origine vétérinaire. Même si l'infection ou la colonisation de la flore intestinale de l'humain avec les organismes d'origine animale était transitoire, les gènes codant pour la résistance à l'apramycine ont pu être transférés par conjugaison à d'autres souches d'*Enterobacteriaceae* présents dans la flore intestinale normale des humains. Cependant, les auteurs font aussi part de la possibilité que le trait de résistance ait pu évoluer à cause de la pression de sélection occasionnée par l'usage de la gentamicine et de la tobramycine en médecine humaine, indépendamment des bactéries d'origine animale. Par contre, si cela s'avérait vrai, les auteurs s'attendraient à ce que d'autres espèces d'*Enterobacteriaceae* autres qu'*E. coli* et *Salmonella* spp démontrent aussi de la résistance à la gentamicine et l'apramycine causée par la production d'AAC(3)IV. Selon ce qui était connu, ceci ne semblait pas vraiment être le cas. Par contre, la résistance à l'apramycine des *E. coli* et *Salmonella* spp, deux espèces reconnues pour se propager facilement des animaux aux humains et pour être isolées chez ces derniers, est bien documentée dans la littérature [32].

### 1.2.3.3. Autre opinion

Présentement, il n'y a pas de consensus dans la communauté scientifique concernant l'importance à accorder à chacun de ces deux mécanismes dans la problématique globale du développement de la résistance antimicrobienne dans la population humaine. Comme il a été mentionné précédemment, certains accordent de l'importance à la possibilité de transfert de résistance antimicrobienne des animaux aux humains. Selon d'autres, comme Phillips et al. [33, 34], la communauté scientifique ne possède pas suffisamment d'information scientifique pour conclure à un lien entre l'utilisation des antibiotiques en agriculture et le développement de résistance antimicrobienne dans la population humaine. La résistance antimicrobienne retrouvée chez les humains serait en majeure partie due à l'utilisation d'antibiotiques en médecine humaine [33]. Bien que des bactéries résistantes aux antibiotiques contaminant la viande puissent atteindre la flore intestinale humaine, une cuisson adéquate devrait normalement les détruire [33]. De plus, la fréquence de colonisation et le transfert de gènes de résistance sont inconnus [33]. Selon ces auteurs, en l'absence d'études épidémiologiques conventionnelles et moléculaires adéquates, il faut être conscient qu'il n'y a aucune évidence de l'ampleur à laquelle les entérocoques et les *E. coli* des animaux sont capable de coloniser le tractus intestinal humain [33]. Selon eux, il n'est pas suffisant de simplement affirmer qu'il est «probable» que les espèces commensales résistantes soient d'importants réservoirs de gènes de résistance pouvant être transmis à la flore intestinale des humains [34]. Considérant toutes ces raisons, même si certains agents réussissaient à atteindre l'humain, les conséquences cliniques d'une résistance seraient probablement minimales. Ainsi, se concentrer sur la contribution hypothétique de l'usage des antibiotiques chez les animaux pourrait distraire les gens des vrais enjeux qui devraient préoccuper la santé publique [33].

### 1.3. *Escherichia coli*, une bactérie indicatrice

*E. coli* est une bactérie fortement étudiée et populaire dans le domaine de la recherche depuis le début du 20<sup>e</sup> siècle. Dernièrement, la popularité d'*E. coli* a été renforcée par la découverte de la conjugaison, un moyen d'échange génétique [35]. Cet organisme est un indicateur employé dans plusieurs études concernant la résistance antimicrobienne. Les bactéries commensales sont souvent considérées comme des sentinelles, i.e. des bactéries indicatrices de la situation de la résistance antimicrobienne dans l'environnement qui pourrait affecter la santé des humains [6].

*E. coli* appartient à la famille des *Enterobacteriaceae* qui regroupe les bactéries bâtonnets gram-négatives pouvant croître rapidement. Cette famille est très large et diversifiée. Elle compte plusieurs espèces vivant librement dans la nature. D'autres sont en association très proche avec les humains et autres vertébrés. Elles peuvent faire partie de leur flore intestinale normale ou être des agents pathogènes causant des maladies variées de peu sévères à très menaçantes [35].

*E. coli* a une structure qui balance très bien les éléments nécessaires à sa protection et à sa survie et ceux assurant une prolifération rapide. Il a une capacité remarquable pour coloniser les intestins des mammifères et des oiseaux, mais il doit survivre et croître à l'extérieur de l'animal hôte afin d'effectuer une propagation inter-hôtes efficace. Bien qu'il survive mal dans le sol et dans l'eau, il a tout de même une meilleure habileté à se développer sous une variété de conditions chimiques et physiques (pH, température, pression osmotique) que la plupart des cellules eukaryotiques. De plus, bien que sa capacité de survie aux conditions extrêmes ne soit pas comparable à celle des bactéries sporulantes, *E. coli*, comme la plupart des eubactéries, peut se différencier d'une forme désignée pour la croissance rapide et efficace à une forme de non croissance adaptée pour une survie à long terme [35].

Au Canada, les indicateurs de contamination fécale les plus communément utilisés pour déterminer la qualité microbiologique et la sécurité de l'eau potable sont les coliformes totaux et *E. coli* [36]. Les coliformes totaux se retrouvent naturellement dans le sol et le tube digestif des humains et des animaux. Ainsi, la présence de ces bactéries peut suggérer une contamination fécale [36]. Par contre, *E. coli* n'est retrouvé naturellement que dans la flore intestinale des humains et des animaux, ainsi sa présence dans un puits indiquerait une contamination fécale récente et la présence possible d'autres agents infectieux potentiellement pathogènes. Pour cette raison, *E. coli* est un meilleur indicateur de pollution fécale de l'eau que les coliformes fécaux étant donné que ces derniers peuvent se répliquer dans l'environnement [36, 37]. La concentration minimale d'*E. coli* acceptable établie au Canada est de 0 par 100 mL d'eau [36]. Une étude longitudinale révélait une association entre la présence d'*E. coli* dans l'eau provenant de puits et les maladies gastro-intestinales en Ontario. Les auteurs suggéraient qu'*E. coli* pouvait être un indicateur utile pour détecter les puits posant une menace potentielle sur la santé publique dans les régions rurales [38].

En plus d'être une bactérie indicatrice de la contamination fécale de choix, *E. coli* est aussi très utile dans l'étude du transfert de résistance antimicrobienne puisqu'il est adapté au tractus intestinal à la fois des humains et des animaux. Ainsi, il est fréquemment exposé à une variété de traitements antibiotiques. *E. coli* est un porteur très important de déterminants de la résistance antimicrobienne parmi les bactéries commensales et est capable de transférer des gènes de résistance à des agents pathogènes présents dans la flore intestinale comme *E. coli* O157 et *Salmonella* spp. [39]. Une étude menée sur une ferme a isolé des coliformes multirésistants provenant de fèces de bovins exposés à des antibiotiques, et d'humains associés à ces animaux. Des isolats de ces deux groupes portaient un plasmide R (pTMS1) qui, selon les auteurs, a pu être transféré entre eux à cause d'une pression de sélection dans l'environnement de la ferme [40]. Sous des conditions naturelles simulées, il a été démontré que le transfert horizontal de plasmides R entre différentes souches bactériennes provenant de niches écologiques différentes, soit des humains, animaux ou poissons, est possible [9]. Certains auteurs proposent que les *E. coli* résistants aux antibiotiques d'origine animale constituent une source potentielle importante

de plasmides pour les *E. coli* indigènes présents dans le tractus intestinal des humains et subséquemment pour les agents qui sont pathogènes pour les humains [9, 10]. Par conséquent, la surveillance de la résistance chez les bactéries commensales telles que *E. coli* est importante parce que ces bactéries peuvent constituer des réservoirs de déterminants de résistance [10]. Pour récapituler, la flore intestinale des animaux traités avec des antibiotiques peut servir de réservoir de déterminants de résistance. L'intérêt porte donc particulièrement sur les entérocoques et *E. coli* qui peuvent jouer un rôle dans la transmission de gènes mobiles de résistance [10]. Il est reconnu que la prévalence de la résistance antimicrobienne chez les bactéries commensales humaines et animales est considérée comme un bon indicateur de la pression de sélection de l'usage d'antibiotiques, et reflète donc le potentiel de résistance des infections futures [41].

## **1.4. Eau potable en Ontario**

L'eau est une matière essentielle à la vie. Comme nous l'avons vu, elle peut être contaminée par des bactéries résistantes aux antibiotiques provenant entre autres des animaux de ferme.

### **1.4.1. Importance des puits en Ontario**

En Ontario, les puits privés ne sont pas assujettis aux règlements sur l'eau potable [11]. Présentement, il y a environ 750 000 puits enregistrés en Ontario. Par contre, il y en aurait approximativement 1.5 millions supplémentaires qui ne sont pas enregistrés. Ils sont principalement situés dans le sud de l'Ontario (Mary Jane Conboy, Well Wise Center, communication personnelle). Environ 10 000 à 20 000 nouveaux puits sont construits annuellement dans cette province. De ce nombre, on ne peut connaître lesquels sont à ajouter au dénombrement total de puits et lesquels ne servent qu'à remplacer d'autres puits désuets ou abandonnés [12]. Les puits municipaux et privés sont la source principale d'eau

potable pour les Ontariens. Environ 3 millions de personnes en dépendent [12]. Les nappes phréatiques répondent à 30% des besoins en Ontario. Cependant, 90% des puits se retrouvent en zones rurales et 95% des gens habitant ces régions dépendent de puits privés.

### **1.4.2. Source de l'eau souterraine**

En Ontario, il y a une disponibilité assez généralisée d'aquifères locaux à haut rendement entre 20 à 50 m de profondeur. Dans ces régions, les puits privés sont souvent superficiels et certains sont de large diamètre [12].

La nappe phréatique est une ressource renouvelable en Ontario, étant donné le climat semi-humide unique qu'on y retrouve. Il y a près d'un mètre de précipitations annuellement. Cela fournit une quantité importante d'eau disponible pour recharger les aquifères. La moitié de la recharge annuelle se fait durant deux mois au printemps. Ce pic de recharge est expliqué par l'accumulation de neige fondue et l'absence de végétation croissante qui normalement consommerait beaucoup d'eau [12].

### **1.4.3. Types de puits**

Il y a trois types de construction prédominants en Ontario : puits creusé, puits foré à la tarière et puits foré à la sondeuse [12, 42].

Les deux premiers types sont des puits de surface ou peu profond et ont un diamètre variant entre 1 et 1.5 m. Ce diamètre peut être plus grand chez les plus vieux puits. La grande majorité d'entre eux ont une profondeur de moins de 15 m. Environ 27% des puits en Ontario sont de ces types. L'étanchéité du revêtement peut ne pas demeurer après de longues périodes de temps [12]. L'eau tirée provient d'une nappe peu profonde qui peut

subir plus directement l'influence des pluies et des activités menées à la surface [42, 43]. Dans le contexte de la contamination, l'eau des puits de moins de 40 m de profondeur est généralement considérée comme une eau de surface [43].

Les puits forés à la sondeuse sont de plus petit diamètre, soit généralement entre 0.1 et 0.15 m. Le revêtement est souvent en métal ou en PVC [12]. Ce type de puits protège généralement mieux l'eau contre les sources de pollution superficielle car ils sont souvent forés plus profond et sont souvent plus étanche que les deux premiers types de puits [42]. Leur approvisionnement en eau est donc plus fiable en qualité et en quantité [42]. Cependant, un puits foré à la sondeuse peut avoir une finition qui se termine sous le sol ou qui se termine au-dessus du sol. Le premier type de finition est souvent moins sécuritaire que le second car le fossé peut se remplir d'eau, de débris ou autre. Ceci contribue à une corrosion prématurée du revêtement ou du couvert, laissant ainsi un accès à l'eau de surface possiblement contaminée. Les puits forés à la sondeuse sont les plus communs en Ontario (45 à 71%). Selon des estimés, plus de la moitié de ces puits serait finis sous le sol [12].

#### **1.4.4. Conditions des puits**

Contrairement aux puits municipaux, la condition, l'âge et l'espérance de vie de la plupart des puits privés sont inconnus des autorités. Il n'existe aucun système provincial de collectes de données. Par contre, un système d'étiquetage des puits privés établi en 2004 devrait corriger la situation pour le futur. Présentement, les propriétaires de puits privés ne connaissent généralement pas l'historique de leur puits. Certains puits sont d'ailleurs situés dans des emplacements hydrologiques déficients où le potentiel de contamination est élevée [12]. Les études se sont souvent concentrées sur la qualité de l'eau, mais il y a très peu d'études sur la qualité de l'infrastructure des puits privés. Ainsi, il existe très peu d'information sur la vulnérabilité des puits face à la contamination en surface et des risques de contamination des aquifères à partir des puits abandonnés à travers la province [12]. Il est à noter que 89% des puits en Ontario auraient besoin de réparations [44]. Certains

estiment qu'il y aurait au moins 500 000 puits abandonnés. Leur localisation serait inconnue. Les puits abandonnés présentent un risque important de contamination des aquifères car ils procurent une entrée directe à la contamination de la surface vers l'aquifère. La majorité des puits sont également construits avec des installations pauvrement protégées. L'entretien est généralement rare ou absent. Les interventions faites sur les puits se font souvent plus en réponse à un problème, plutôt que préventivement [12]. Les Canadiens ne testent leurs puits que de façon intermittente. En fait, 53% ne testent pas leur puits sur une base régulière [43]. En Ontario et aux États-Unis, au moins 30% des puits dans les régions rurales sont contaminés par une source d'origine fécale et excèdent les standards gouvernementaux de qualité de l'eau potable pour une consommation sécuritaire [45]. Plus de 20% des éclosions de maladies hydriques au pays, entre 1974-1996, ont été causées par un approvisionnement privé en eau. Environ 45% des éclosions rapportées, principalement dans les régions rurales et les régions éloignées, proviennent de systèmes non municipaux [43].

En résumé, beaucoup de citoyens dépendent de puits privés en Ontario. Par contre, comme démontré par les statistiques, un grand nombre de ces puits ne sont pas en bonne condition et sont vulnérables à la contamination. Cette contamination peut se faire par des bactéries résistantes aux antibiotiques qui pourraient éventuellement s'établir dans la flore intestinale des humains.

## **Partie 2 Méthodologie**

### **2.1. Données**

#### **2.1.1. Sources des échantillons**

Les données de cette étude proviennent d'un projet pancanadien sur la surveillance de la résistance antimicrobienne d'*Escherichia coli* intitulé *Prospective Multi-Province Surveillance for Antimicrobial-Resistant Escherichia coli in Drinking and Recreational Source Waters: Impact on Humans and the Environment* (projet ARO). Par contre, les échantillons d'eau du projet ARO, quant à eux, proviennent de programmes provinciaux de surveillance de la qualité de l'eau de puits privés déjà existants permettant la soumission volontaire d'échantillons d'eau de puits pour des analyses bactériologiques.

La période d'échantillonnage du projet ARO comprend la période du projet pilote s'étalant du 1<sup>er</sup> avril 2003 au 31 octobre 2003 et la période du projet d'étude s'étalant du 1<sup>er</sup> mai 2004 au 30 septembre 2006 (année d'étude 2004 et année d'étude 2005). Les échantillons obtenus pour le projet pilote proviennent des laboratoires régionaux de London et Hamilton. Les échantillons obtenus pour le projet d'étude, quant à eux, proviennent des laboratoires régionaux de Hamilton, Kingston, London, Orillia, Ottawa, Peterborough, Toronto et Windsor. Les analyses spécifiques à ce projet de maîtrise portent sur l'année d'étude 2004 exclusivement. L'année d'étude 2004 est définie par la période s'étalant du 1<sup>er</sup> mai 2004 au 30 avril 2005.

En Ontario, la soumission de l'eau des puits n'est pas mandatée par les autorités de santé publique. Tous les échantillons de n'importe quel type de puits sont soumis sur une base volontaire à des programmes de surveillance de la qualité de l'eau (Étape 1 : Figure 2,

p.30). Dans cette première étape, les échantillons d'eau ont été recueillis par les propriétaires des puits à l'aide d'une trousse fournie par les laboratoires régionaux de santé publique. Ces échantillons ont été acheminés par les propriétaires aux divers laboratoires régionaux de santé publique participant aux programmes de surveillance de la qualité de l'eau. Les échantillons ont généralement été apportés, pour analyse microbiologique de l'eau (comptage de coliformes et de *E. coli*), au laboratoire régional le plus proche. Il est à noter que les échantillons recueillis par ces programmes de surveillance et la banque de données sur la contamination des échantillons étaient gérés par le *Food Safety and Safe Water Unit (SWU)* du *Ministry of Health and Long Term Care of Ontario (MOHLTC)*.

### **2.1.2. Analyse de laboratoire**

Les soumissions d'échantillons aux laboratoires régionaux de santé publique devaient se faire dans les 24 heures suivant la récolte afin d'être analysés pour le dépistage d'*E. coli*. Pour chaque soumission, une filtration sur membrane a été effectuée sur un échantillon de 100 mL d'eau dans les laboratoires régionaux. Ces membranes ont par la suite été placées sur une gélose DC (*differential coliform*) permettant de différencier les *E. coli* et coliformes (Étape 2 : Figure 2, p. 30). Les cultures ont été incubées pendant une nuit à 35 ( $\pm 1$ ) °C. Jusqu'à cinq colonies typiques d'*E. coli* (bleues ou vertes bleues) bien isolées et de différentes morphologies provenant de la gélose DC ont été sélectionnées et inoculées sur un milieu gélosé TSA (*Tryptic Soy Agar*) en pente. Chaque laboratoire régional a ainsi soumis hebdomadairement un nombre prédéterminé d'échantillons d'isolats de présumés *E. coli* au projet ARO (Étape 2 : Figure 2, p. 30). Il est à noter que la sélection des échantillons à envoyer au projet ARO s'est fait par un échantillonnage aléatoire pondéré. Ainsi, le nombre total d'échantillons soumis par laboratoire régional (Tableau I, p. 26) est proportionnel au volume des soumissions de la région desservie et à la proportion d'échantillons contaminés à *E. coli* durant l'année 2002. La taille prédéterminée de l'échantillonnage hebdomadaire varie aussi selon la saison afin de refléter les soumissions normalement plus nombreuses durant les périodes chaudes (1<sup>er</sup> juin au 30 septembre) et moins nombreuses durant les périodes froides (1<sup>er</sup> octobre au 31 mai). Ainsi, cet échantillonnage a été pondéré de façon à refléter, dans le temps et l'espace, les volumes

d'échantillons contaminés par *E. coli* soumis dans chaque laboratoire. Les milieux de culture ainsi sélectionnés ont été incubés 18 à 24 heures à 35°C puis gardés réfrigérés jusqu'à deux semaines à 4 °C, jusqu'à ce qu'ils soient envoyés au *Provincial Laboratory for Public Health (Microbiology) (PLPH(M))*, Calgary, Alberta). L'envoi de ces milieux s'est fait le plus souvent par FedEx (courrier en une nuit) dans un contenant de transport en polystyrène avec des poches de glace. Au total, 1 564 échantillons présumés positifs à *E. coli* provenant de puits distincts ont été acheminés au *PLPH(M)*.

Tableau I : Fréquence de soumissions d'échantillons au projet ARO par laboratoire régional de santé publique

Laboratoire régional	Fréquence absolue	Fréquence relative (%)
Hamilton	236	15.1
Kingston	294	18.8
London	501	32.0
Orillia	148	9.5
Ottawa	142	9.1
Peterborough	147	9.4
Toronto	90	5.8
Windsor	6	0.4
<b>Total</b>	<b>1564</b>	<b>100.0</b>

Au *PLPH(M)*, un dépistage préliminaire de bactéries potentiellement résistantes a été effectué par la méthode de l'étalement sur des géloses supplémentées de différents antibiotiques (Étape 3 : Figure 2, p. 30). Les échantillons ont subi cette épreuve préliminaire dans les deux semaines suivant leur réception. Afin de promouvoir la croissance bactérienne, la méthode de détection préliminaire a nécessité que les isolats d'*E. coli* présents sur le milieu de transport soient inoculés dans un bouillon *Tryptic Soy Broth* (TSB) et incubés à 35 °C pendant quatre à six heures. Puis, 0.01 mL de ce bouillon TSB a été ensemencé sur une gélose MacConkey de contrôle, sans antibiotique, et sur sept autres contenant chacune un antibiotique différent (Tableau II, p. 27). Ces géloses ont par la suite été incubées 18 à 24 heures à 35 °C. Les géloses ensemencées ont été observées afin de noter la présence ou l'absence de croissance bactérienne. La morphologie des colonies présentes a également été documentée. Les colonies typiques d'*E. coli* ont été isolées puis inoculées sur un milieu de transport, une gélose TSA en pente, incubées 18 à 24 heures à 35

°C, puis réfrigérées à 4 °C jusqu'à leur envoi au Laboratoire de lutte contre les zoonoses d'origine alimentaire (LZA) situé à Saint-Hyacinthe, Québec. Il est à noter qu'un contrôle de la qualité de ces géloses de dépistage a été effectué sur une base hebdomadaire à l'aide de sept souches différentes d'*E. coli*, d'une souche de *Pseudomonas aeruginosa* et d'une souche de *Klebsiella pneumoniae* dont les profils de résistance étaient connus. La méthode de détection préliminaire ainsi appliquée aux 1 564 échantillons a montré que 1 111 isolats (71.0%) étaient sensibles à tous les antibiotiques testés et que 453 (29.0%) étaient présumés résistants à un ou plusieurs des antibiotiques ajoutés aux géloses. Ces 453 échantillons ont été envoyés au LZA de Saint-Hyacinthe pour une analyse finale des profils de résistance aux antibiotiques. Les envois, sur glace, se sont fait environ toutes les deux semaines par transport aérien.

Tableau II : Seuil de concentration minimale inhibitrice selon l'antibiotique pour la méthode de diffusion en milieu gélosé

Antibiotique	Seuil de CMI (µg/ml)	Période
Amikacine	> = 16	Discontinué le 02/03/2005
Ampicilline	> = 8	Depuis le 23/02/2005
Céphalothine	> = 8	Depuis le début au 12/09/2004
	> = 16	Depuis le 13/09/2004
	> = 32	Depuis le 26/01/2005
Gentamicine	> = 4	Depuis le début au 03/04/2005
	> = 8	Depuis le 04/04/2005
Acide nalidixique	> = 4	Depuis le début
Streptomycine	> = 32	Depuis le début
Sufaméthoxazole	> = 128	Depuis le début
Tétracycline	> = 4	Depuis le début

Il est à noter que le seuil de CMI utilisé lors de la méthode de l'étalement sur géloses supplémentées de certains antibiotiques a été modifié à certains moments de l'étude. Les seuils trop faibles ne discriminaient pas suffisamment les isolats résistants des isolats sensibles. Trop d'isolats sensibles se retrouvaient alors envoyés à Saint-Hyacinthe pour subir les analyses par la méthode de microdilution en bouillon, ce qui augmentait trop les coûts totaux. Dans une perspective de surveillance, il a été décidé d'augmenter les seuils de CMI afin de diminuer le nombre d'analyses inutiles.

À leur réception au LZA, les souches ont été transférées sur une gélose MacConkey. Lorsque la culture était pure après 18 à 24 heures d'incubation à 35°C, une colonie isolée était alors transférée sur une gélose TSA additionnées de 5% de sang de mouton. Lorsqu'elle n'était pas pure, une colonie était repiquée sur une seconde gélose MacConkey jusqu'à l'obtention de la pureté. L'identification des cultures ainsi obtenues a été confirmée à l'aide de la galerie API 20<sup>E</sup>® (bioMérieux, Marcy l'Étoile, France) (Étape 4 : Figure 2, p. 30). Lors de cette étape, une solution 0.5 McFarland de la culture à tester a servi à ensemercer la galerie et à inoculer une gélose TSA additionnée de 5% de sang de mouton afin d'en vérifier la pureté. Les cultures mixtes ont été retravaillées jusqu'à l'obtention de pureté. Seuls les résultats d'identification d'*E. coli* qualifié «identification acceptable», «bonne identification», «très bonne identification» ou «excellente identification» par l'épreuve API 20<sup>E</sup>® ont été retenus pour la suite des procédures. Ainsi, des 453 échantillons présumés résistants à au moins un antibiotique par la méthode préliminaire, 72 (15.9%) isolats ont été exclus de l'étude car ils n'étaient finalement pas identifié comme étant *E. coli*. 381 (84.1%) ont ainsi été identifiés comme étant *E. coli* avec une identification qualifiée d'au moins acceptable. Ces 381 colonies ont été conservées dans 1 mL de bouillon Brucella additionné de 15% de glycérol et congelées à -70° jusqu'à la réalisation du test final de résistance aux antibiotiques.

L'établissement du profil final de résistance aux antibiotiques a été effectué par la méthode de microdilution en bouillon à l'aide des plaques Sensititre® CMV7CNCD et CMV1AGNF NARMS de la compagnie Trek Diagnostics System (Cleveland, USA) (Étape 5 : Figure 2, p. 30). Les seize antibiotiques testés sont l'amikacine, l'amoxicilline-acide clavulanique, l'ampicilline, la céfoxitine, le ceftiofur, le ceftriaxone, la céphalothine, le chloramphénicol, la ciprofloxacine, la gentamicine, la kanamycine, l'acide nalidixique, la streptomycine, le sulphaméthoxazole, la tétracycline et le triméthoprime-sulphaméthoxazole. La souche à tester a d'abord été décongelée puis soumise à deux passages sur une gélose Mueller-Hinton incubée à 35°C pendant 18-24 heures. Une solution 0.5 MacFarland a ensuite été préparée à l'aide de colonies et d'eau déminéralisée. 10 µL de ce bouillon a ensuite été transféré dans 10 mL de bouillon Mueller-Hinton. 50 µL de ce dernier bouillon a été distribué dans les puits des plaques Sensititre et incubé à 35°C

pendant 18 heures. Puis, la lecture de la croissance minimale inhibitrice (CMI) des bactéries dans les différents puits a été effectuée. Les résultats de CMI ont été interprétés selon des seuils pertinents à la santé publique humaine selon les directives du *Clinical and Laboratory Standards Institute (CLSI)* (M7100, version 2004). Il est à noter que la banque de données des résultats du profil final de résistance antimicrobienne obtenu suite à la méthode de microdilution en bouillon était gérée par le projet ARO.

Tableau III : Seuil de concentration inhibitrice minimale pour diagnostiquer de la résistance antimicrobienne selon l'antibiotique (NARMS)

Antibiotique	Seuil de CMI (µg/ml)	Nombre de puits par plaque	Étendu (µg/ml)
Acide nalidixique	≥ 32	7	0.5-32
Amikacine	≥ 64	4	0.4-4
Amoxicilline / Acide clavulanique	≥ 32/16	6	1-32
Ampicilline	≥ 32	6	1-32
Céfoxitine	≥ 32	7	0.5-32
Ceftiofur	≥ 8	7	0.12-8
Ceftriaxone	≥ 64	9	0.25-64
Céphalothine	≥ 32	5	2-32
Chloramphénicol	≥ 32	5	2-32
Ciprofloxacine	≥ 4	9	0.015-4
Gentamicine	≥ 16	7	0.25-16
Kanamycine	≥ 64	4	8-64
Streptomycine	≥ 64	2	32-64
Sulphaméthoxazole	≥ 512	6	16-512
Tétracycline	≥ 16	3	8-32
Triméthoprime / Sulphaméthoxazole	≥ 4/76	6	0.12-4
Contrôle positif		3	
Nombre total de puits		96	

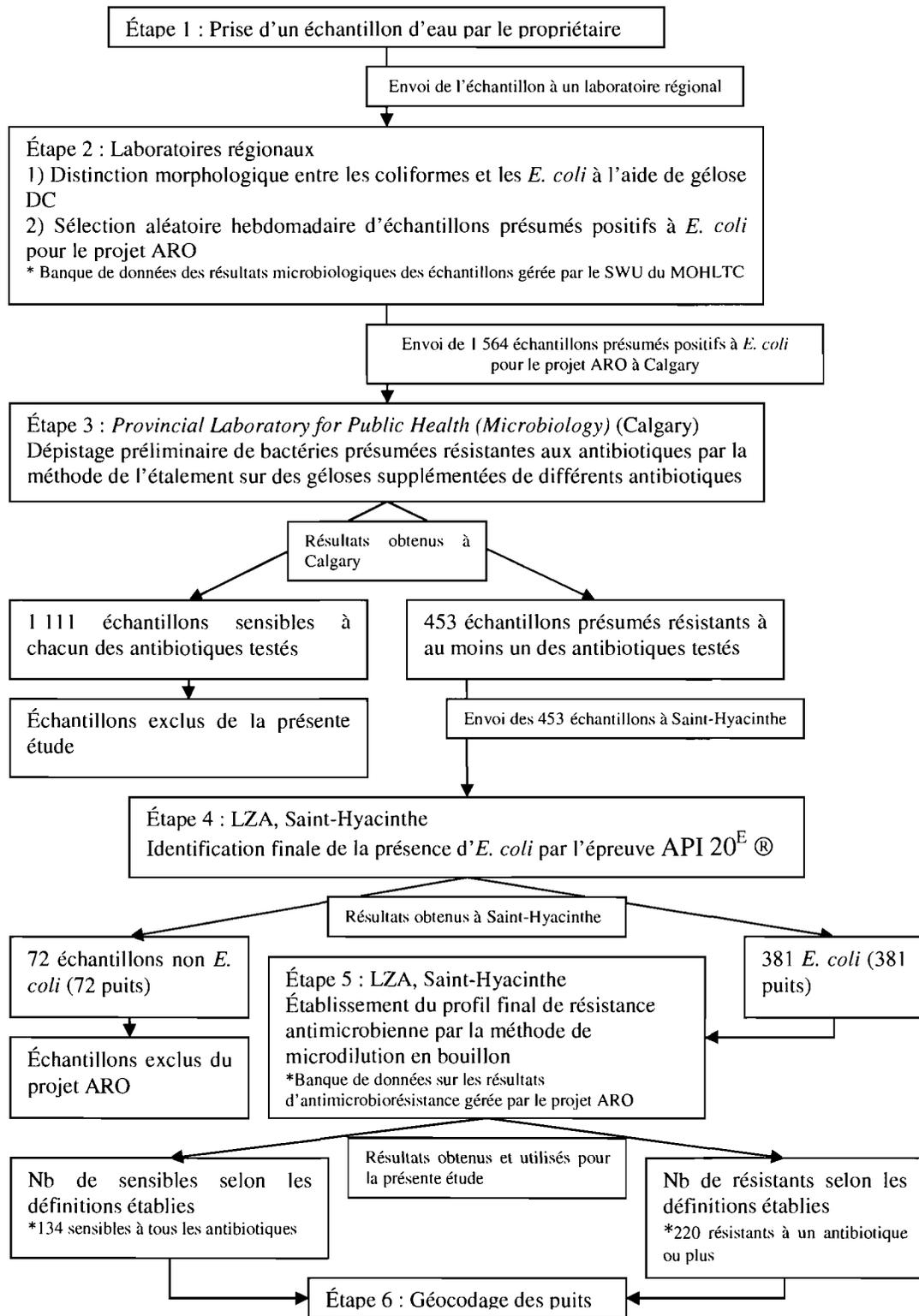


Figure 2 : Résumé de la méthodologie employée avant la manipulation des données pour l'étude présente (année d'étude 2004)

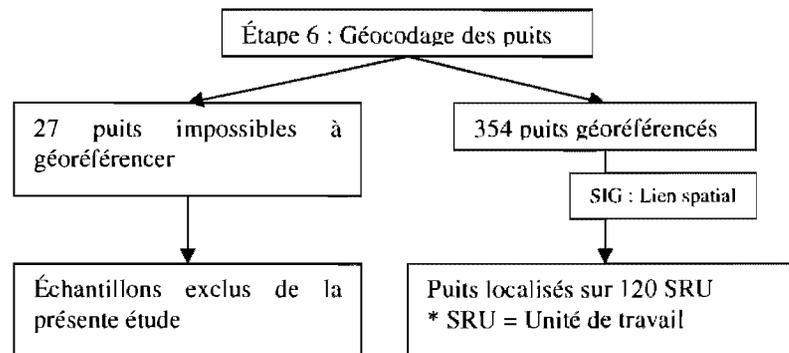


Figure 2 (suite) : Résumé de la méthodologie employée avant la manipulation des données pour l'étude présente (année d'étude 2004)

### 2.1.3. Banques de données

La présente étude s'est basée sur trois banques de données. L'une d'elles était gérée par le SWU du MOHLTC et portait sur les résultats microbiologiques de tous les échantillons d'eau soumis par les propriétaires des puits durant l'année d'étude 2004 (280 139 échantillons) (Étape 2 : Figure 2, p. 30). Les champs d'information inclus avec chaque résultat décrivaient diverses dates (dates de collecte de l'échantillon, de réception de l'échantillon au laboratoire et d'analyse), la provenance de l'échantillon (le laboratoire ayant reçu l'échantillon à tester, l'adresse civique, le numéro de lot et concession, la ville/municipalité, le numéro de localisation d'urgence, le comté, le code postal et le numéro de l'unité de santé publique correspondants au puits), les résultats microbiologiques de l'échantillon (le code numérique du statut de contamination, le résultat textuel du statut de contamination, le comptage de coliforme total et le comptage de *E. coli*) et l'identifiant unique de l'échantillon d'eau. La description de ces champs est donnée plus en détail à la page 4 de l'**Annexe 2**. Un puits peut avoir été échantillonné plusieurs fois. Il n'y avait pas de moyen d'identifier quels puits l'avaient déjà été car il n'y avait pas d'identifiant unique

pour chaque puits. De plus, les données d'adresse civique n'étaient pas toujours textuellement les mêmes pour chaque échantillon provenant d'un même puits.

Les données concernant les résultats finaux du profil de résistance des échantillons envoyés au LZA étaient enregistrées dans une deuxième banque de données. Cette banque de données sur les résultats d'antibiorésistance était gérée par le projet ARO (Étape 5 : Figure 2, p. 30). Les champs d'information inclus dans cette banque décrivaient, entre autre, les résultats de l'identification de *E. coli* par l'épreuve API 20<sup>E</sup>®, la concentration inhibitrice minimale des échantillons identifiés *E. coli* à chaque antibiotique testé et le numéro d'identification de l'échantillon. Cet identifiant n'était pas le même que celui utilisé dans la première banque. Contrairement à la première banque, il n'y avait pas de données sur la provenance de l'échantillon. De plus, il n'existait qu'un lien indirect entre la base de données du *SWU* et celle des résultats d'antibiorésistance (projet ARO). Ce lien a été fourni par les registres des laboratoires qui avaient effectué les tests de résistance antimicrobienne. En effet, ces registres possédaient les numéros d'identification utilisés dans la banque de contamination de l'eau du *SWU* et les numéros d'identification utilisés dans la banque du projet ARO. C'est ce lien qui a permis de géoréférencer les données de la banque provenant du projet ARO.

La troisième banque de données utilisée dans ce projet de maîtrise est celle des variables indépendantes. Elle a été créée dans le cadre de la présente étude. Les données calculées et incluses dans cette banque découlent de données provenant de diverses sources. Cette banque sera expliquée en détail plus tard.

#### **2.1.4. Géocodage**

La base de données du *SWU* a été géoréférencée à l'aide des critères suivant : adresses civiques, codes postaux, municipalité ou numéros de lot et concession d'où provenait le puits échantillonné. La priorité a été donnée à la référence la plus précise.

L'**Annexe 1** décrit le détail technique utilisé pour le géocodage permettant la localisation géospatiale des données de contamination de l'eau. L'application de cette méthode et les résultats obtenus sont décrits dans l'**Annexe 2**. Le démarche permettant de faire le lien entre les résultats des tests de résistance antimicrobienne de *E. coli* (banque de données du projet ARO) et la banque de données des résultats du SWU une fois géoréférencées est présenté dans l'**Annexe 3**. Les résultats obtenus suite à l'application de cette technique sont fournis dans l'**Annexe 4**. En sommes, des 381 échantillons, 354 (92.9%) ont pu être géocodés selon les critères mentionnés ci-dessus (Étape 6 : Figure 2, p. 30). Cette présente étude s'est donc basée sur ces 354 échantillons. Les coordonnées géographiques référencées ont par la suite permis de situer les puits dans les subdivisions de recensement unifiées (SRU) à l'aide d'un système d'information géographique (SIG). Chacun de ces points a ensuite été attribué à l'intérieur d'une SRU grâce à la fonction de lien spatial du SIG.

### **2.1.5. Unités géographiques**

Cette présente étude est de type écologique. Le niveau d'analyse est une unité géographique appelée la subdivision de recensement unifiée (SRU), une unité créée par Statistique Canada [46]. L'étude a été limitée aux régions agricoles de recensement (RAR) [46] Sud, Ouest, Centre et Est (Figure 3, p. 34). La RAR Nord a été exclue de cette étude. Au total, en 2001, 254 SRU composaient les quatre RAR de l'étude. Par contre, dans le Recensement de l'agriculture de Statistique Canada, certaines SRU ont été fusionnées ensemble afin de conserver la confidentialité des exploitations agricoles (254 SRU en 224 SRU) [47]. Il a donc fallu faire de même dans cette présente étude, car les données de plusieurs variables indépendantes provenaient du recensement agricole. Pour les besoins de la cause, des SRU ont été fusionnées par des manipulations à l'aide d'un SIG. Les échantillons dont le géocodage était possible et valide proviennent de 120 SRU une fois fusionnées. L'unité de travail de cette étude est la SRU une fois fusionnée (n=120).

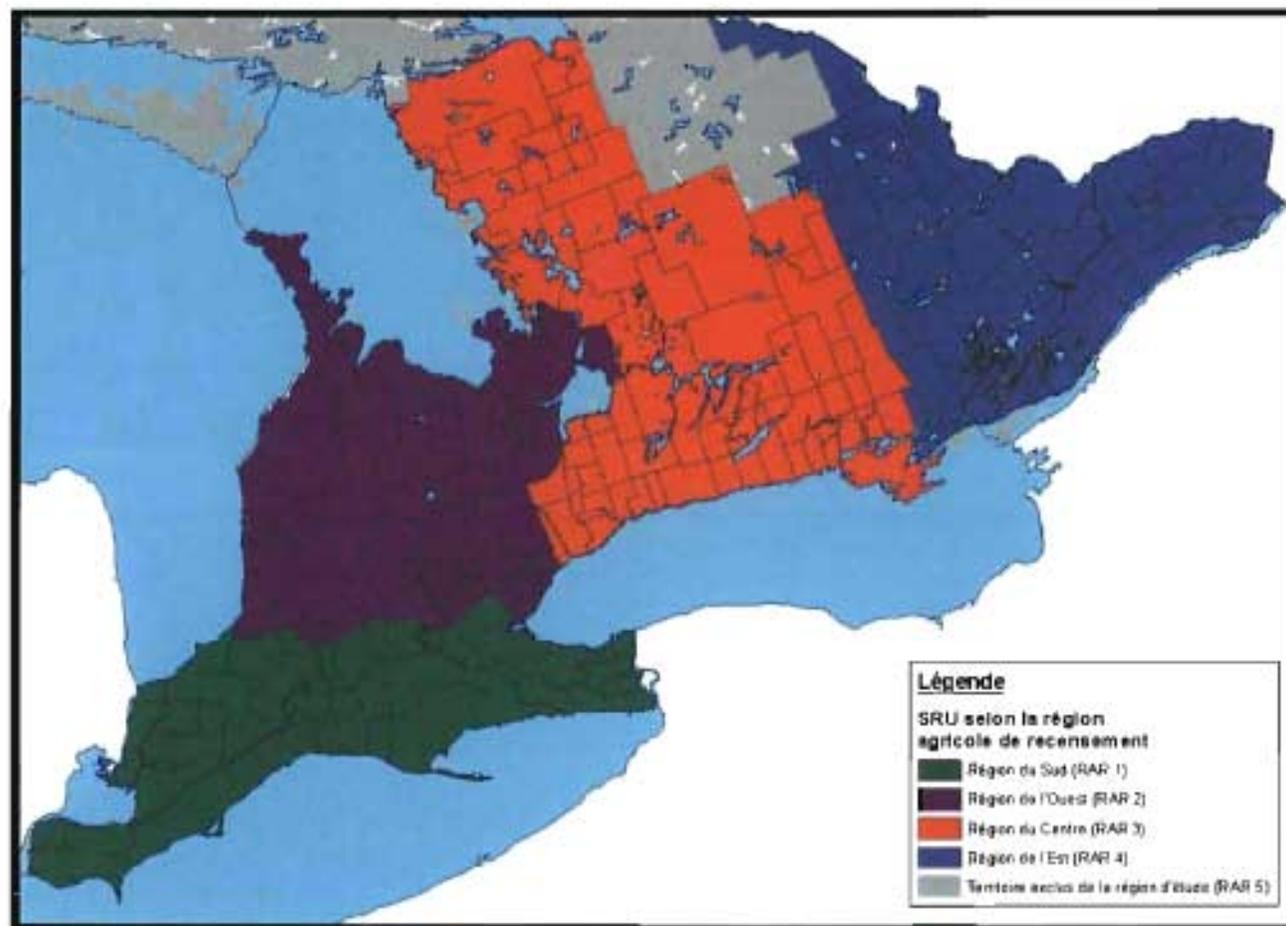


Figure 3 : Régions agricoles de recensement et les SRU comprises dans la région d'étude

### 2.1.6. Variables dépendantes

Au total, cinq définitions de patron de résistance antimicrobienne régionale ont été retenues. Bien qu'elles se chevauchent par leurs définitions, ces cinq définitions sont des composantes de cinq variables dépendantes différentes et non d'une seule variable dépendante à cinq catégories. Chacune des cinq variables dépendantes a été construite comme étant la proportion du nombre de puits répondant respectivement aux cinq définitions sur le nombre total de puits échantillonné par SRU ( $\# \text{ puits positifs} / \# \text{ puits total}$  dans une SRU). Les données par SRU ont été obtenues à l'aide d'un SIG, par la fonction d'un joint spatial entre données provenant de la banque de données d'antibiorésistance (projet ARO) géoréférencées et la couche des polygones de SRU.

Le Tableau IV (p.36) présente la définition des cinq patrons. Trois patrons ont été définis par le nombre d'antibiotiques auxquels *E. coli* est résistant. Deux patrons ont été définis par l'usage important fait en médecine humaine ou en médecine vétérinaire. Le patron humain a été composé par des fluoroquinolones, une catégorie d'antibiotiques homologués uniquement pour un usage chez les humains, les chiens et les chats domestiques [6]. Le patron vétérinaire a été composé par des antibiotiques homologués pour un usage chez les animaux de ferme (ou de même classe que d'autres antibiotiques homologués) et presque exclusivement utilisés chez ceux-ci [6, 16]. En effet, chacun des antibiotiques de ce patron représentait seulement moins d'un pourcent des doses thérapeutiques quotidiennes d'antibiotiques oraux destinés aux humains et délivrés par les pharmacies de détail au Canada (sauf pour les combinaison de sulfamidés et de triméthoprime incluant leurs dérivés avec 5.29%) [16].

Tableau IV : Définitions des patrons de résistance antimicrobienne de *E. coli*

Type de patron	Appellation	No	Définition
Défini par un nombre d'antibiotiques	Résistance générale	1	<i>E. coli</i> résistant à un antibiotique ou plus (patron no 2 + patron no 3)
	Résistance	2	<i>E. coli</i> résistant à un ou deux antibiotiques
	Multirésistance	3	<i>E. coli</i> résistant à trois antibiotiques ou plus
Défini par l'usage fait	Patron humain	4	<i>E. coli</i> résistant à ciprofloxacine et/ou acide nalidixique
	Patron vétérinaire	5	<i>E. coli</i> résistant à sulfaméthoxazole et/ou streptomycine et/ou gentamicine et/ou kanamycine et/ou triméthoprime-sulphaméthoxazole

### 2.1.7. Variables indépendantes

Les variables indépendantes regroupaient certains aspects agro-environnementaux, socio-économiques, météorologiques et pédologiques pertinents à la problématique de la contamination de l'eau des puits et à la résistance antimicrobienne des bactéries. Par la fonction d'un joint de tableaux à l'aide d'un SIG, les données obtenues par SRU de la banque de données des variables indépendantes montée à l'aide des différentes sources mentionnées ci-dessous ont été jointes à la banque de données des variables dépendantes par SRU obtenue précédemment.

#### 2.1.7.1. Sources des données des variables indépendantes

Les données des variables indépendantes ont été obtenues de plusieurs sources. Les données agricoles ont été tirées du Recensement de l'agriculture 2001 (Statistique Canada), le recensement le plus récent disponible au moment de l'étude et le plus rapproché dans le temps de l'année d'étude 2004. Ces données étaient disponibles à l'échelle des SRU. Par contre, comme il a été expliqué plus tôt, certaines SRU ont été fusionnées afin de protéger la confidentialité de l'identité de certaines fermes peu nombreuses dans certaines régions.

Les données socio-économiques ont été tirées du Recensement du Canada 2001 (Statistique Canada), recensement de la population le plus récent disponible au moment de l'étude. Ces données étaient disponibles à l'échelle des subdivisions de recensement (SDR), une échelle plus petite que celle des SRU. Ces données ont donc été agrégées au niveau des SRU à l'aide d'un SIG, tout en tenant compte de la fusion des certaines SRU pour garder la confidentialité de certaines fermes.

Les données météorologiques ont été tirées des sommaires climatologiques d'Environnement Canada disponible sur Internet [48]. Les données provenaient de multiples stations réparties un peu partout en Ontario. Les stations fournissant les données variaient d'un mois à l'autre.

Les données pédologiques ont été fournies par Agriculture et Agroalimentaire Canada (Le groupe de travail du Pédo-paysage du Canada, 2005. Pédo-paysage du Canada version 3.0. Agriculture et Agroalimentaire du Canada. (carte numérique et base de donnée à l'échelle de 1 : 1 million)). Les données ont été manipulées à l'aide d'un SIG afin d'obtenir une valeur unique à l'échelle des SRU.

#### **2.1.7.2. Données d'agriculture**

Le Tableau V (p. 38) présente les 18 variables agro-environnementales identifiées pour l'étude. Ces variables ont entre autre été inspirées des variables souvent considérées dans les publications sur le lien entre la contamination de l'eau par *E. coli* et l'agriculture [21, 22, 25, 49]. Trois catégories de variables ont été considérées parmi les variables d'agriculture : densité animale, utilisation des terres et rapport population animale/population humaine. Les données de densité ont été calculées à l'aide du nombre d'animaux de l'espèce en question sur la superficie totale des terres agricoles de la SRU. Les données de l'utilisation des terres ont été calculées à l'aide de la superficie de terres utilisées pour l'usage en question sur la superficie totale des terres agricoles.

Tableau V : Définitions des variables indépendantes d'agriculture

Variable	Définition	Calcul
<b>DBo</b>	Densité bovine (Animaux/ha)	Nombre (Nb) de bovins et de veaux/Aire totale des terres agricoles
<b>DPo</b>	Densité porcine (Animaux/ha)	Nb de porcs/Aire totale des terres agricoles
<b>DVo</b>	Densité de volailles (Animaux/ha)	Nb de poules et de poulets/Aire totale des terres agricoles
<b>DEq</b>	Densité équine (Animaux/ha)	Nb de chevaux et poneys/Aire totale des terres agricoles
<b>DRu</b>	Densité de ruminants autres que les bovins (Animaux/ha)	Nb de chèvres, de bisons, de chevreuils, d'élans, de lamas, d'alpagas, de moutons et d'agneaux/Aire totale des terres agricoles
<b>UCult</b>	Utilisation des terres pour la culture	Aire de terres utilisées pour la culture/Aire totale des terres agricoles
<b>UJach</b>	Utilisation des terres en jachère	Aire de terres en jachère/Aire totale des terres agricoles
<b>UPatCE</b>	Utilisation des terres pour le pâturage cultivé ou ensemencé	Aire de terres utilisées pour le pâturage cultivé ou ensemencé/Aire totale des terres agricoles
<b>UPatNat</b>	Utilisation des terres pour le pâturage naturel	Aire de terres utilisées pour le pâturage naturel/Aire totale des terres agricoles
<b>UEpFS</b>	Utilisation des terres pour l'épandage de fumier solide	Aire de terres utilisées pour l'épandage de fumier solide/Aire totale des terres agricoles
<b>UEpFSI</b>	Utilisation des terres pour l'épandage des terres à l'aide d'un système d'irrigation	Aire de terres utilisées pour l'épandage des terres à l'aide d'un système d'irrigation/Aire totale des terres agricoles
<b>UEpFLS</b>	Utilisation des terres pour l'épandage de fumier liquide en surface	Aire de terres utilisées pour l'épandage de fumier liquide en surface/Aire totale des terres agricoles
<b>UEpFLI</b>	Utilisation des terres pour l'épandage de fumier liquide injecté	Aire de terres utilisées pour l'épandage de fumier liquide injecté/Aire totale des terres agricoles
<b>BoHu</b>	Rapport bovins/humains	Nb de bovins et veaux/nb d'humains
<b>PoHu</b>	Rapport porcins/humains	Nb de porcs/nb d'humains
<b>VoHu</b>	Rapport volailles/humains	Nb de poules et poulets/nb d'humains
<b>EgHu</b>	Rapport équins/humains	Nb de chevaux et poneys/nb d'humains
<b>RuHu</b>	Rapport ruminants autres que les bovins/humains	Nb de chèvres, de bisons, de chevreuils, d'élans, de lamas, d'alpagas, de moutons et d'agneaux/nb d'humains

### 2.1.7.3. Données socio-économiques

Les variables socio-économiques ont été considérées comme facteurs de confusion possibles car le niveau socio-économique d'une région peut potentiellement influencer le niveau régional de détection d'*E. coli* résistant aux antibiotiques tout en influençant les variables d'agriculture. Le Tableau VI (p. 39) présente la description de ces variables d'intérêt. Le Recensement du Canada 2001 fournissait la période médiane de construction des logements dans chaque SDR. Pour obtenir l'âge médian des logements d'une SDR, l'année médiane de la période de construction a été soustraite de 2001. L'âge médian approximatif des logements au niveau des SRU a été obtenu à l'aide de la moyenne des âges médians des SDR constituant la SRU. Étant donné que les données de l'âge des puits d'une région étaient souhaitées mais impossible à obtenir, l'âge médian des logements a été choisi, entre autre, pour représenter une approximation possible de l'âge des puits. Le revenu moyen approximatif des ménages a aussi été obtenu à l'aide de la moyenne des revenus moyens des SDR constituant la SRU. Le nombre de personnes par SRU sans grade, diplôme ou certificat a été obtenu à l'aide de la somme du nombre de personnes respectant ces critères de toutes les SDR constituant la SRU. Le nombre total de personnes de 15 ans et plus par SRU a été obtenu par l'addition du nombre total de personnes de 15 ans et plus de toutes les SDR constituant la SRU. La superficie habitée utilisée pour calculer la densité humaine est celle de l'écoumène de la population telle qu'obtenue par des manipulations à l'aide d'un SIG dans l'évaluation des polygones de ces territoires.

Tableau VI : Définitions des variables indépendantes socio-économiques

Variable	Définition	Calcul
<b>LogeA</b>	Âge médian des logements (années)	* 2001 – l'année médiane de la période médiane de construction des logements (niveau SDR) * Moyenne de l'âge médian des SDR comprises dans la SRU
<b>Rev</b>	Revenu moyen des ménages en 2000 (\$)	Moyenne des revenus médians des ménages des SDR comprises dans la SRU
<b>Scol</b>	Proportion de la population qui est peu scolarisée	Nb d'humains sans grade, certificat ou diplôme / nb d'humains de 15 ans et plus
<b>DHu</b>	Densité humaine (Humains/ha)	Nb d'humains / aire de l'écoumène de la population

#### 2.1.7.4. Données météorologiques

Les variables météorologiques de cette étude ont été choisies en fonction de leur influence possible sur le niveau de ruissellement de l'eau dans l'environnement. Il est à noter qu'une variable impliquant le nombre de jours avec des précipitations extrêmes était souhaitée mais impossible à obtenir dans le cadre de cette étude. C'est pour cette raison que la variable de précipitation annuelle totale a été choisie afin d'étudier l'influence du niveau de précipitations. Dans les cas où plusieurs stations météorologiques fournissant les informations recherchées se retrouvaient dans une SRU donnée, la valeur attribuée à cette SRU était la moyenne des valeurs des stations situées à l'intérieur de la SRU. S'il n'y avait qu'une seule station météorologique, la valeur unique enregistrée était attribuée à celle-ci. Par contre, si aucune station météorologique n'était située dans la SRU, la valeur attribuée à cette dernière était celle de la station la plus proche de la frontière de la SRU en question. Le Tableau VII (p. 40) présentait une description des variables météorologiques d'intérêt.

Tableau VII : Définitions des variables indépendantes météorologiques

Variable	Définition	Calcul
<b>PrTo</b>	Précipitation annuelle totale (mm)	Addition des précipitations mensuelles (liquide ou solide, calculé en équivalent d'eau) enregistrées au cours des 12 mois de la période d'échantillonnage 2004
<b>NMarAv</b>	Fonte de neige en avril (cm)	Différence de hauteur de neige au sol à la fin du mois entre mars 2005 et avril 2005
<b>FonN</b>	Fonte maximale de neige entre 2 mois consécutif (cm)	Valeur maximale différence de hauteur de neige au sol à la fin du mois entre deux mois consécutifs au cours de la période d'échantillonnage 2004, peu importe ces mois (cm)

### **2.1.7.5. Données pédologiques**

La variable choisie pour représenter l'influence pédologique possible est la qualité de drainage du sol. Puisque cette variable peut à la fois influencer les variables dépendantes et les variables d'agriculture, elle a été considérée comme facteur de confusion potentiel. La valeur de cette variable a été attribuée de façon à ce qu'il n'y ait qu'une valeur unique par SRU. Pour ce faire, la valeur qualitative occupant la plus grande superficie sur la SRU a été attribuée comme étant la valeur de la qualité de drainage du sol de la SRU donnée (Figure 4, p. 42). Des manipulations à l'aide d'un SIG ont permis d'obtenir des valeurs de superficie.

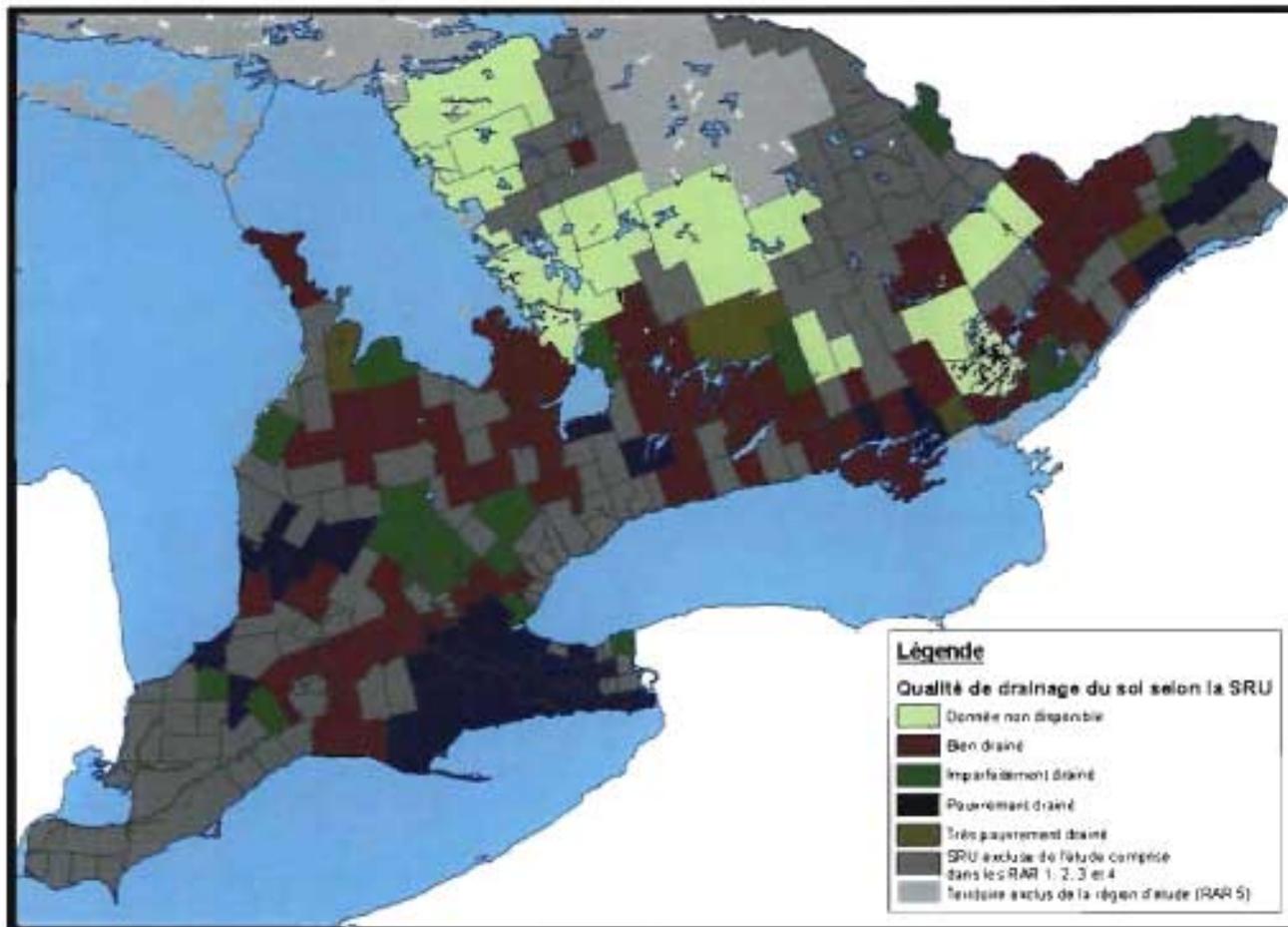


Figure 4 : Qualité de drainage du sol selon les SRU

## 2.2. Analyse descriptive

Pour chaque variable dépendante, le compte de puits positifs dans l'ensemble de la région d'étude, la moyenne, la médiane, l'écart-type, le minimum et le maximum de chaque variable dépendante ont été calculés. Des cartes choroplèthes ont aussi été produites afin de visualiser la distribution géospatiale de ces variables.

Pour chaque variable indépendante, le compte de données manquantes, moyenne, médiane, écart-type, étendue, minimum, maximum et histogramme de la distribution de fréquence ont aussi été calculés. Ces statistiques ont par la suite permis d'orienter la forme (catégorisation et transformation) de ces variables pour l'analyse associative. L'impact des valeurs extrêmes de ces variables sur les variables dépendantes a été évalué par une analyse d'association non conditionnelle. Lorsque cette dernière était significative, des cartes choroplèthes ont été produites afin d'apprécier visuellement le contraste entre les extrêmes. Pour ce faire, les variables dépendantes ont aussi été catégorisées.

Cette étude ne comprend que les SRU ayant eu au moins un échantillon contaminé par *E. coli* dont le profil de résistance a été testé dans le cadre du projet ARO. Il a été jugé pertinent de comparer les SRU prises en compte dans l'étude et celles qui ont été exclues. Ceci a été fait à l'aide d'un test statistique (test de T ou test d'Aspin-Welch) sur l'hypothèse nulle d'égalité de la moyenne pour chacune des variables d'agriculture (exception faite de celles de la catégorie des rapports population animale / population humaine) entre le groupe des SRU incluses dans l'étude et celui des SRU exclues. Par exemple, pour la variable DBo, le test a évalué l'hypothèse nulle qu'il n'y avait pas de différence entre la densité bovine moyenne des SRU incluses dans l'étude et celle des SRU exclues. Avant d'effectuer ces tests d'égalité, il a fallu vérifier si la distribution de la variable testée était normale dans chacun des groupes (SRU incluses et SRU exclues) à l'aide du test d'Omnibus. Si les distributions n'étaient pas normales, mais que le nombre de SRU dans chaque groupe était considéré élevé ( $n > 30$ ), les distributions pouvaient être considérées approximativement

normales par le théorème de la tendance centrale limite. Ensuite, il a fallu vérifier l'homogénéité des variances entre les deux groupes. Ceci s'est fait à l'aide du test de Levene modifié. Lorsque l'homogénéité des variances était démontrée, la comparaison entre les moyennes de la variable testée des deux groupes se faisait par le test de T. Dans le cas contraire, le test alternatif utilisé était le test d'Aspin-Welch. La comparaison pour les variables non-agricoles et les rapports population animale / population humaine n'a pas été faite car les données de ces variables n'étaient pas disponibles pour les SRU exclues au moment de cette analyse. Il est à noter que ces analyses se sont faites sur les données quantitatives avant leur transformation en format catégoriques et que le seuil alpha utilisé était de 0.05.

### **2.2.1. Transformation des variables**

Initialement, toutes les variables indépendantes étaient quantitatives, sauf celle de la qualité du drainage du sol qui était qualitative. Cette dernière a subi une restructuration des catégories afin d'obtenir suffisamment d'observations dans chaque catégorie.

Avant de construire les modèles de régression logistique à l'aide des variables quantitatives, il a tout d'abord fallu évaluer la condition de linéarité de l'association entre les variables indépendantes et le log odds de la variable dépendante. Pour ce faire, chaque variable indépendante a été catégorisée en quatre selon les quartiles. Puis, le log odds de la variable dépendante a été généré en fonction de chaque variable indépendante catégorisée. Un diagramme du log odds obtenu en fonction du point milieu des catégories de la variable indépendante a alors fourni une bonne appréciation visuelle de la linéarité de cette relation.

La distribution des variables continues a été étudiée afin de les catégoriser en se basant sur les cassures naturelles d'histogramme ou des classes liées aux percentiles lorsque la condition de linéarité n'était pas respectée. Lorsque la distribution de fréquence des SRU

en fonction d'une variable donnée était fortement asymétrique, la variable a été transformée en variable à deux catégories car son histogramme nous permettait de distinguer facilement deux classes (faible et élevée). S'il y avait plus d'une cassure importante dans la distribution, la cassure suivant le mode de la distribution et possédant la valeur la plus élevée était choisie pour départager les deux catégories. Les valeurs choisies dans le cadre de cette étude étaient celles du 5<sup>e</sup> percentile le plus proche de la valeur de la cassure d'histogramme déterminée visuellement. Si le percentile correspondant était plus de 90, la valeur du 90<sup>e</sup> percentile était tout même choisie afin d'obtenir un nombre suffisant d'observations dans une des deux classes. La cassure d'histogramme choisie a toujours suivi la valeur de tendance centrale déterminée par le mode, sauf dans le cas de la variable du revenu où la cassure d'histogramme était choisie pour une valeur inférieure au mode afin de mettre l'emphase sur un faible revenu moyen. Les variables dont la distribution de fréquence de SRU était approximativement normale ont été catégorisées en quatre catégories égales selon les quartiles.

### **2.3. Modèle de régression**

Il est à noter que dans le cadre de cette étude, il n'y a eu aucune mesure directe des facteurs indépendants au niveau des puits. Les expositions, présumées homogènes pour tous les puits d'une SRU donnée, étaient mesurées au niveau régional. Étant donné la nature des variables dépendantes (# de puits positifs / # total de puits) présumant un risque homogène des puits dans une région donnée, une approche binomiale a été employée. Ainsi, considérant la structure primaire des données affichant un seul patron de covariables par SRU, les analyses statistiques ont été faites en utilisant un modèle de régression logistique sur des données binaires groupées.

### Modèle de régression logistique :

$$\ln(p / 1 - p) = \beta_0 + \sum \beta_j X_j$$

$p$  = proportion de puits positifs selon les patrons de résistance définis dans une SRU

$j = j^{\text{ème}}$  région

## 2.3.1. Construction du modèle de régression

### 2.3.1.1. Modèles initiaux

La première étape dans la construction du modèle de régression était d'identifier la variable dépendante et toutes les variables indépendantes que l'on désirerait considérer. Au total, 26 variables indépendantes potentielles ont été identifiées [Tableau V (p. 38), Tableau VI (p. 39), Tableau VII (p. 40), Figure (p. 49), Figure 5 (p. 46)]. Les variables socio-économiques et pédologiques ont été considérées comme facteurs de confusion possibles.

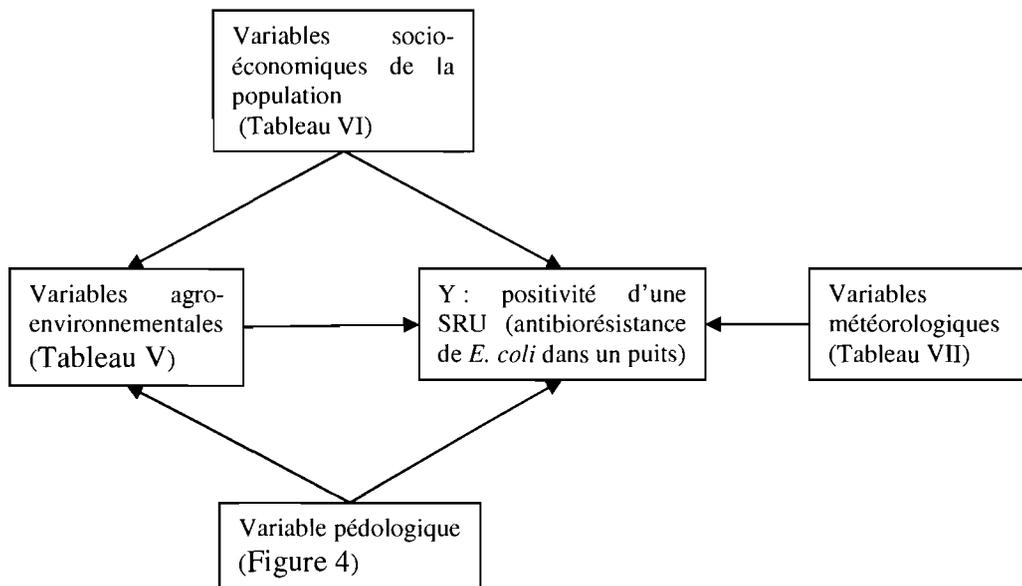


Figure 5 : Diagramme causal

Seules les variables dont l'analyse d'association non conditionnelle (régression logistique univariée sur des données binaires groupées) était significative à un seuil  $\alpha$  de 0.15 ont été retenues. Certaines variables ont été éliminées parce qu'il y avait trop de données manquantes. Il a été déterminé que si une variable manquait plus de 20% des données, soit plus de 24 observations sur 120, elle serait éliminée du processus de sélection des variables. Parmi les variables restantes, des analyses de corrélation ont été effectuées afin d'évaluer la colinéarité possible entre elles. Cette analyse a été faite à l'aide du coefficient de Pearson entre les différentes paires parmi les variables retenues jusqu'à ce moment. Le seuil limite déterminant la présence de colinéarité a été fixé à 0.7. Il a été déterminé que si une paire de variables démontrait de la colinéarité, les deux variables de la paire devraient être séparées dans deux modèles initiaux différents (ex. : A et B, C et D, ...) afin de ne pas les intégrer ensemble dans le même processus de sélection de variables. Nous avons alors obtenu plusieurs modèles initiaux différents à tester. Une fois que les variables indépendantes démontrant de la colinéarité eurent été départagées en plusieurs modèles initiaux, la multicollinéarité des variables d'exposition à inclure dans le processus de sélection des variables dû également être considérée. Elle a été diagnostiquée par de la statistique de tolérance. Le seuil limite pour déterminer la présence de multicollinéarité a été fixé à 0.4.

#### **2.3.1.2. Stratégie de sélection**

La sélection des variables indépendantes a été effectuée à l'aide d'une stratégie d'élimination rétrograde. En premier lieu, toutes les variables retenues selon le modèle ont été incluses dans le processus. Ensuite, la variable dont le coefficient était le moins statistiquement significatif a été éliminée du modèle. Ce processus a été réitéré jusqu'à ce qu'il ne reste que les variables possédant un coefficient statistiquement significatif. Cette stratégie a permis de construire un modèle d'effet principal. Le test statistique utilisé pour exclure une variable du modèle fut le Wald chi-carré dont le seuil  $\alpha$  a été fixé à 0.05. Il est à noter que cette procédure a été répétée sur chacun des modèles initiaux obtenus suite aux partages des variables présentant de la colinéarité (modèle A, B, C, ...).

### **2.3.1.3. Modèle incluant des interactions**

Des interactions à deux voies entre les variables indépendantes qui étaient significatives dans le modèle d'effet principal ont été créées et évaluées.

### **2.3.1.4. Modèle incluant les variables de confusion**

L'effet des variables de confusion potentielles (socio-économiques et pédologiques) a été évalué en incluant une variable à la fois dans le modèle d'effet principal obtenu suite à la sélection rétrograde des variables. Il a été déterminé que si l'inclusion de cette variable modifiait de façon notable, i.e. plus de 25%, l'estimé du rapport de cotes des paramètres des variables d'agriculture retenues, cette variable serait incluse dans le modèle final.

## **2.3.2. Évaluation des modèles**

L'évaluation d'un modèle s'est fait en deux étapes. Premièrement, il a fallu déterminer si le modèle était bien ajusté en général en utilisant des mesures d'ajustement du modèle (tests Pearson  $\chi^2$ , de déviance  $\chi^2$  et de Hosmer-Lemeshow). Deuxièmement, il a fallu déterminer s'il y avait des observations (ou des groupes d'observations) spécifiques qui ne correspondaient pas au modèle ou qui avaient une influence excessive sur le modèle. Ceci a été effectué à l'aide d'une évaluation des résidus de déviance. Puis, une évaluation des delta-betas des observations sur le modèle a été effectuée afin de déterminer si ces données avaient un effet excessif sur le modèle.

## **2.4. Seuil alpha de décision**

Le seuil  $\alpha$  de décision a été fixé à 0.05, sauf lors des analyses d'association non conditionnelle où le seuil avait été fixé à 0.15.

## 2.5. Logiciel et application

Les analyses statistiques descriptives ont été générées à l'aide des logiciels NCSS (Hintze, J. (2001). NCSS and PASS. Number Cruncher Statistical Systems. Kaysville, Utah. [www.NCCS.com](http://www.NCCS.com)) (Descriptive Stats), SAS (SAS®8, SAS Institute Inc., Cary, NC, USA, 1999-2001) (proc means) et Excel (Microsoft® Excel 2002) (graphiques). Les analyses univariées, les coefficients de corrélation de Person et les statistiques de tolérance ont été générés à l'aide du logiciel SAS (respectivement proc logistic, proc corr et proc reg). Les sélections rétrogrades ont été effectuées à l'aide du logiciel SAS (proc logistic). Les analyses de régression logistique sur des données binaires groupées ont été générées à l'aide des logiciels SAS (proc genmod) et Egret (Egret®, Cystel Statistical Software, Cambridge, MA) (Logistic regression model). L'évaluation des modèles a été effectuée à l'aide du logiciel SAS (proc logistic) pour les statistiques d'ajustement et évaluation des résidus de déviance et du logiciel Egret (Logistic regression model) pour l'évaluation des delta-betas. Les manipulations des données géospatiales et les cartes choroplètes ont été effectuées à l'aide du SIG ArcGIS (ArcGIS™ version 9, ESRI®, CA, 2004).

## **Partie 3 Résultats**

### **3.1. Analyse descriptive**

#### **3.1.1. Variables dépendantes**

La Figure 6 (p. 51) présente la localisation des 354 échantillons soumis provenant des 120 SRU. Il est à noter que cette figure ne démarque pas nécessairement les 354 puits de façon distincte pour la raison que certains puits possédaient les mêmes coordonnées géographiques attribuées selon l'échelle de référence choisie.

Les trois antibiotiques pour lesquels les puits contaminés avec *E. coli* résistant étaient les plus fréquents sont la tétracycline (151/354), le sulphaméthoxazole (97/354) et la streptomycine (88/354). Il y a très peu de puits contaminés avec *E. coli* résistant au ciprofloxacine (3/354) et aucun au ceftriaxone (0/354) (Tableau VIII, p. 52).

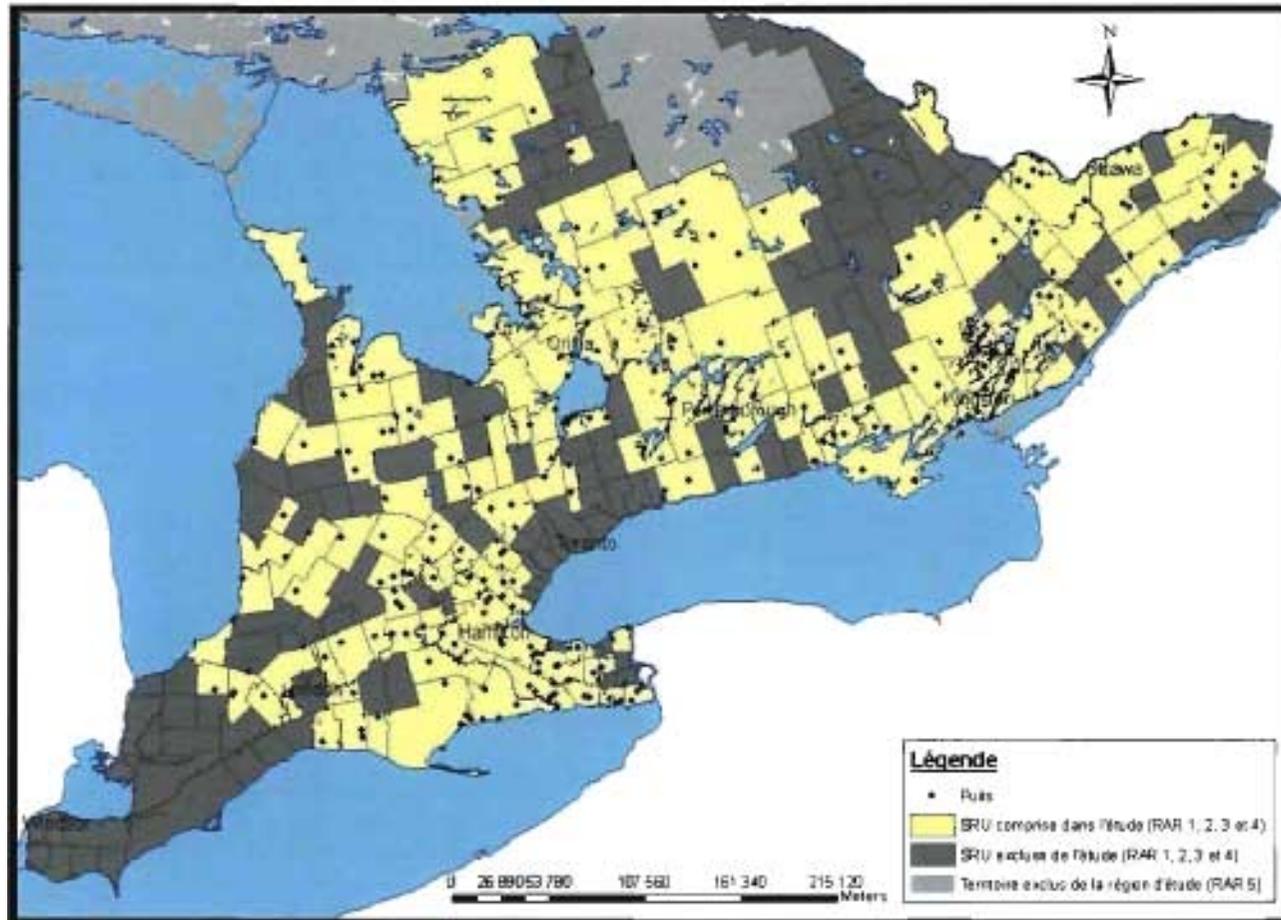


Figure 6 : Région d'étude et localisation des puits échantillonnés

Tableau VIII: Fréquence de puits contaminés par *E. coli* résistant selon l'antibiotique

Antibiotique	Fréquence absolue de puits positifs	Fréquence absolue de puits échantillonnés	Fréquence relative de puits positifs (%)
Tétracycline	151	354	42.7
Sulfaméthoxazole	97	354	27.4
Streptomycine	88	354	24.9
Ampicilline	67	354	18.9
Triméthoprim-Sulphaméthoxazole	51	354	14.4
Céphalothine	50	354	14.1
Chloramphénicol	23	354	6.5
Kanamycine	23	354	6.5
Céfoxitin	16	354	4.5
Amoxicilline-Acide clavulanique	14	354	4.0
Acide nalidixique	13	354	3.7
Ceftiofur	11	354	3.1
Gentamicine	8	354	2.3
Amikacine	8	354	2.3
Ciprofloxacine	3	354	0.8
Ceftriaxone	0	354	0.0

Les puits soumis à la caractérisation de la sensibilité aux divers antibiotiques et dont la localisation géospatiale a été possible étaient répartis sur 120 SRU. Le nombre de puits soumis différait d'une SRU à l'autre (d'un à 19 puits par SRU). Un seul puits était échantillonné dans 46 SRU (38.3%), deux puits dans 25 SRU (20.8%), trois puits dans 19 SRU (15.8%), quatre puits dans 11 SRU (9.2%). 19 SRU (15.8%) possédaient cinq puits soumis ou plus (Figure 7, p. 53). La distribution spatiale de ces fréquences d'échantillons soumis par SRU était légèrement hétérogène, i.e. le nombre de puits soumis semble légèrement aggloméré à certaines régions (Figure 8, p. 54).

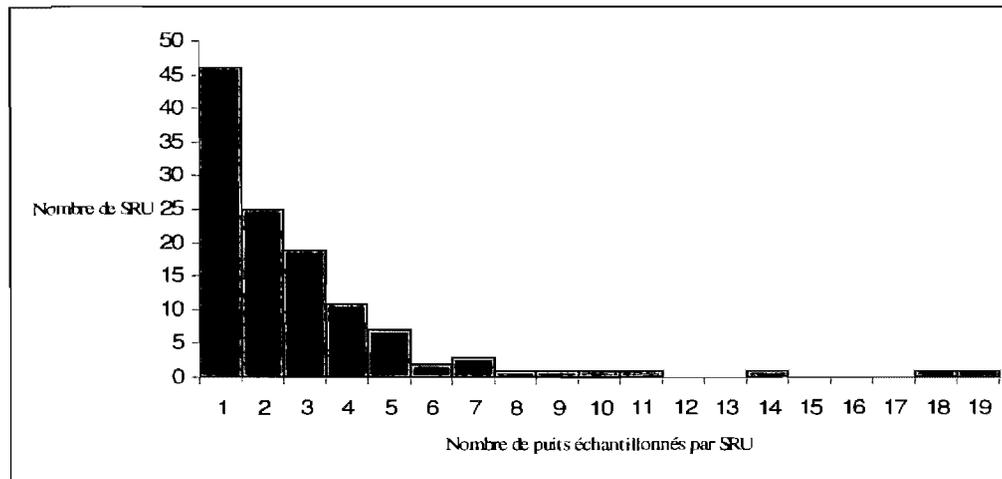


Figure 7 : Fréquence de SRU selon le nombre de puits échantillonnés

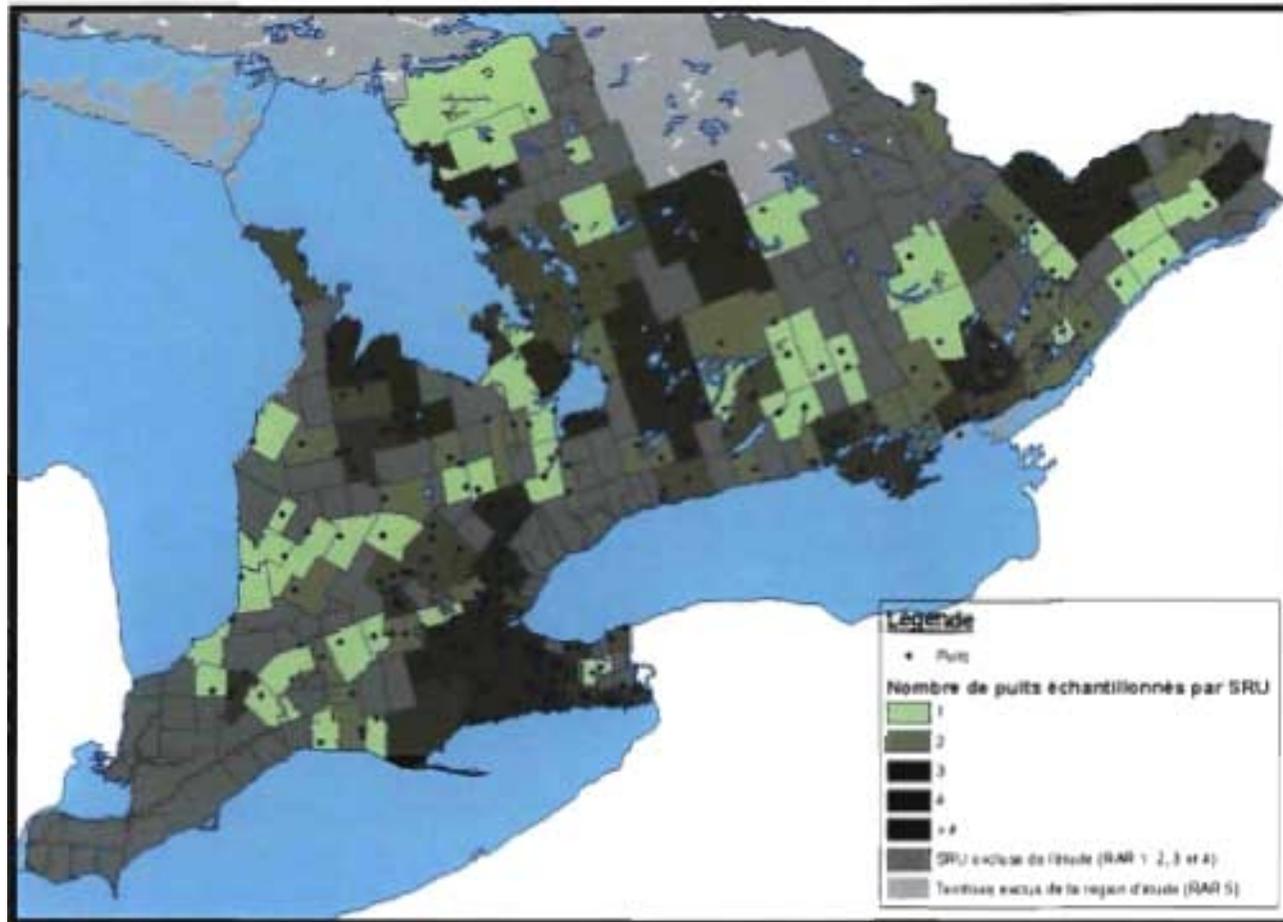


Figure 8 : Distribution géo-spatiale de la fréquence absolue de puits échantillonnés par SRU

Le Tableau IX (p. 55) présente la fréquence de puits contaminés par *E. coli* résistant selon le patron donné dans l'ensemble de la région d'étude. Dans le cas de plusieurs SRU, 100% des puits soumis étaient résistants au patron 1 (un antibiotique ou plus). Ces SRU étaient nombreuses dans le sud, l'ouest et l'est de la région d'étude. Le pourcentage de puits résistants au patron 1 de la plupart des SRU situées au centre de la région d'étude semblait plus faible (Figure 9, p.56). La résistance au patron 2 (un ou deux antibiotiques) ne semblait pas avoir un patron géo-spatiale spécifique (Figure 10, p.57). Par contre, la multirésistance (patron 3) semblait plutôt être concentrée dans l'ouest et l'est de la région d'étude (Figure 11, p. 58). Il y avait très peu de puits positifs au patron 4 (humain) et patron 5 (vétérinaire) (Tableau IX, p. 55), seulement 2.0 % et 3.2 % respectivement l'étaient. Il n'y avait aucune SRU dont la totalité des puits échantillonnés était contaminée par *E. coli* résistant au patron 5. Étant donné la faible fréquence de puits positifs à ces patrons, la construction d'un modèle de régression multiple n'a pas été tentée. L'Annexe 5 présente un tableau de fréquence de puits contaminés par *E. coli* résistant à chaque patron par SRU.

Tableau IX : Fréquence de puits positifs selon le patron de résistance au niveau de l'ensemble de la région d'étude

Patron	Nb de puits négatifs	Nb de puits positifs	Nb de puits échantillonnés	Fréquence relative de puits positifs (%)
1	134	220	354	62.1
2	234	120	354	33.9
3	254	100	354	28.2
4	347	7	354	2.0
5	343	11	354	3.2

Tableau X: Résumé de l'analyse descriptive régionale des cinq variables dépendantes

Patron	Moyenne (%)	Médiane (%)	Écart-type (%)	Minimum (%)	Maximum (%)
1	63.0	68.4	38.0	0	100
2	35.5	33.3	37.1	0	100
3	27.5	71.5	34.9	0	100
4	2.5	0	11.6	0	100
5	2.2	0	8.1	0	50

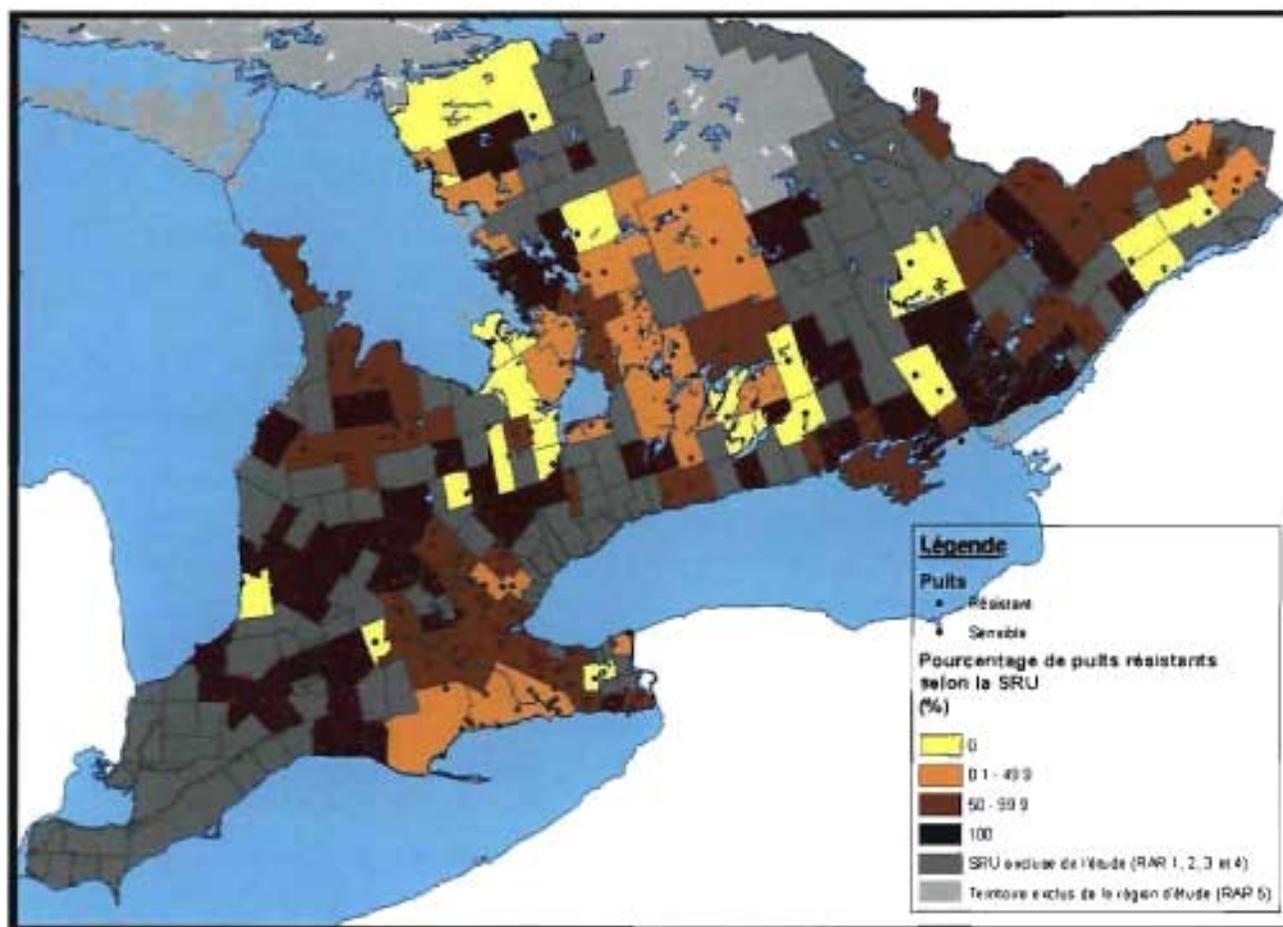


Figure 9: Distribution géo-spatiale de la fréquence de puits contaminés par *E. coli* résistant au patron 1 (un antibiotique ou plus)

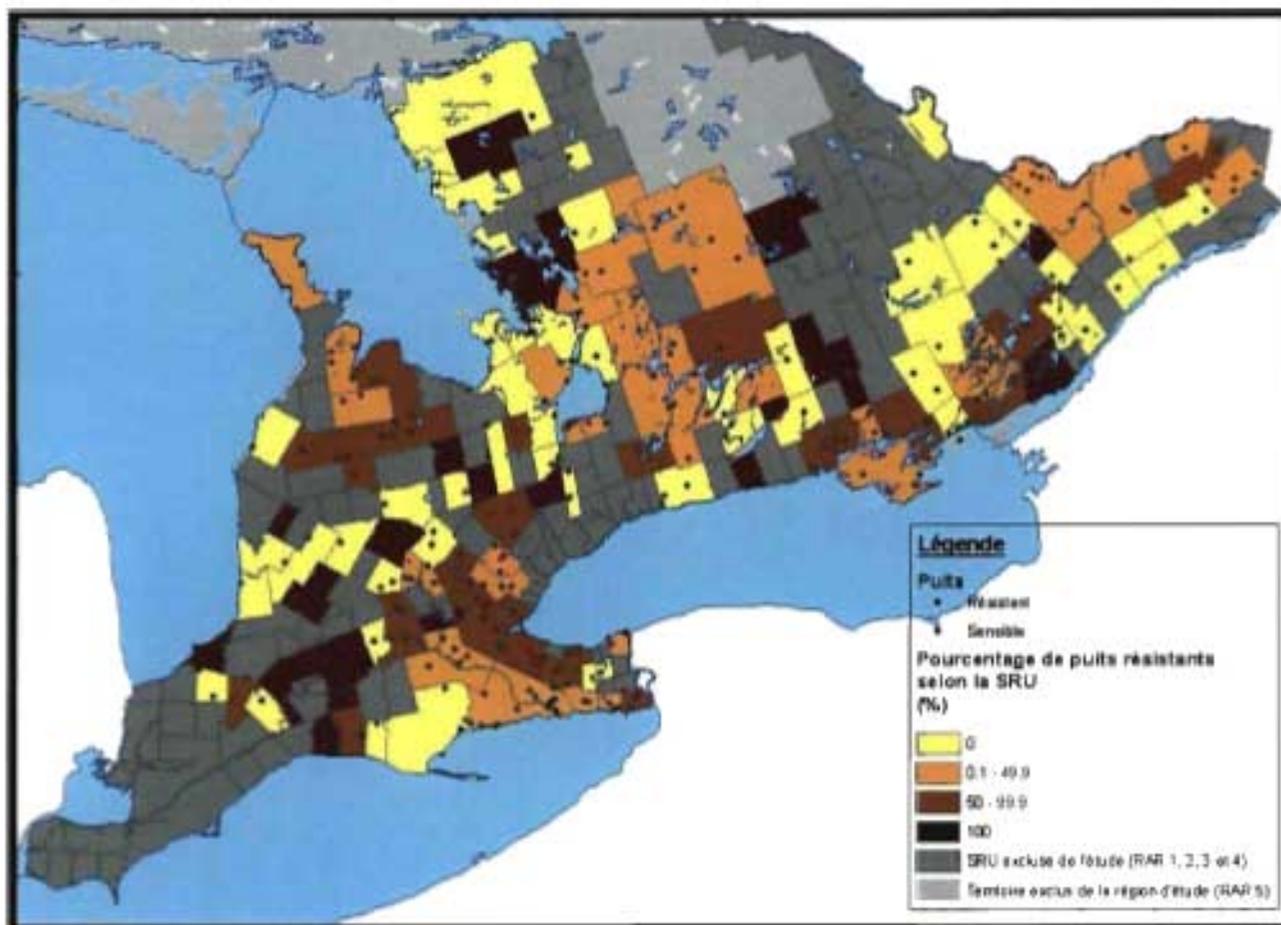


Figure 10 : Distribution géo-spatiale de la fréquence de puits contaminés par *E. coli* résistant au patron 2 (un ou deux antibiotiques)

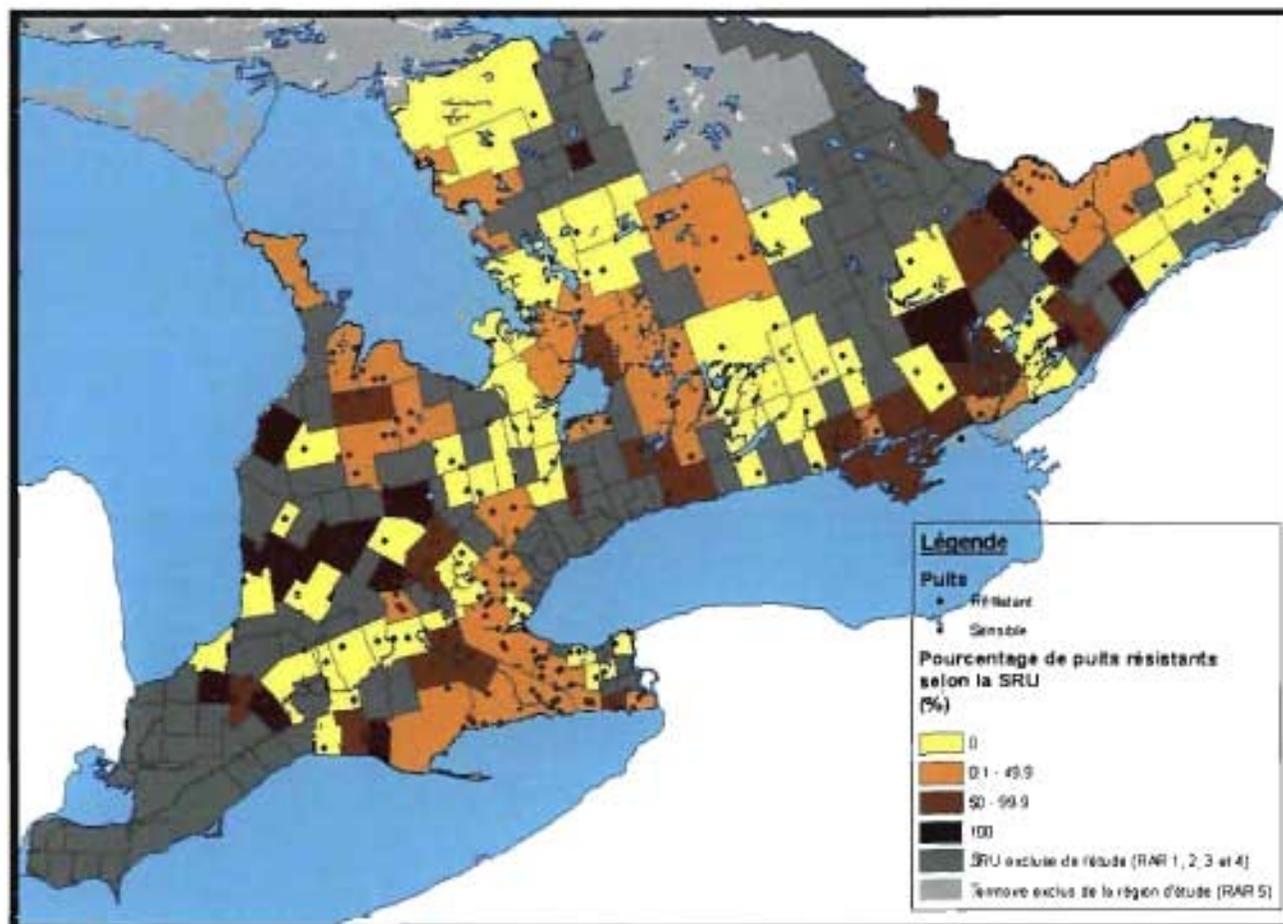


Figure 11: Distribution géo-spatiale de la fréquence de puits contaminés par *E. coli* résistant au patron 3 (multirésistance)

### 3.1.2. Variables indépendantes

Toutes les variables indépendantes (Annexe 5), sauf une, étaient d'abord quantitatives continues. Seule la variable Drain était catégorique ordinale. Les SRU dont la donnée n'était pas disponible ou ne pouvait pas s'appliquer ont été groupées dans la catégorie «donnée non disponible». Pour les besoins statistiques, les catégories de sol drainé «pauvrement» et «très pauvrement» ont été fusionnées.

Tableau XI : Fréquence de SRU selon la catégorie de qualité de drainage du sol

Drain	Fréquence absolue	Fréquence relative (%)
<b>Bien drainé</b>	57	54.8
<b>Imparfaitement drainé</b>	19	18.3
<b>Pauvrement drainé</b>	24	23.1
<b>Très pauvrement drainé</b>	4	3.8
<b>Total</b>	104	100

Le Tableau XII (p. 60) présente un résumé de l'analyse descriptive de chaque variable indépendante antérieurement identifiée pour l'évaluation de leur association potentielle avec le niveau de résistance antimicrobienne régionale. Par contre, plusieurs variables possédaient plus de données manquantes que le seuil défini (DPo, DVo, UJach, UEpFLI, PoHu et VoHu). Aucune variable continue ne respectait la condition de linéarité entre la variable indépendante et le log odds de la variable dépendante. Elles ont toutes été transformées en variables catégoriques. La plupart des variables possédait une distribution des fréquences asymétrique à droite, sauf la variable UCult qui avait une distribution asymétrique à gauche et les variables LogeA et PrTo qui avaient une distribution approximativement normale. Toutes les variables possédant une distribution asymétrique ont été transformées en variables dichotomiques et les variables LogeA et PrTo, en variables catégoriques à quatre catégories de valeurs possibles. Le Tableau XIII (p.61) présente les catégories créées, leurs valeurs quantitatives correspondantes et le percentile associé.

Tableau XII : Résumé de l'analyse descriptive des variables explicatives

Variable	n	Moy	Méd	Écart-type	Min	Max	Étendu
<b>DBo</b>	120	0.3759	0.3577	0.2360	0.0416	1.7178	1.6762
<b>DPo</b>	90	0.6015	0.2290	0.8898	0	4.0665	4.0665
<b>DVo</b>	93	11.0698	3.8754	15.3330	0.0225	67.3757	67.3532
<b>DEq</b>	118	0.0201	0.0150	0.0172	0.0022	0.1019	0.0997
<b>DRu</b>	120	0.0777	0.0657	0.0686	0	0.3294	0.3294
<b>UCult</b>	120	0.6059	0.6601	0.2144	0.0911	0.8859	0.7949
<b>UJach</b>	87	0.0046	0.0027	0.00834	0	0.0720	0.0720
<b>UPatCE</b>	119	0.0601	0.0534	0.0360	0.0096	0.1880	0.1785
<b>UPatNat</b>	120	0.1149	0.0924	0.0974	0.0070	0.4991	0.4922
<b>UEpFS</b>	120	0.0914	0.0827	0.0428	0.0167	0.3036	0.2869
<b>EpFSI</b>	96	0.0044	0	0.0081	0	0.0416	0.0416
<b>UEpFLS</b>	104	0.0339	0.0211	0.0354	0	0.1530	0.1530
<b>UEpFLI</b>	84	0.0023	0	0.0047	0	0.0234	0.0234
<b>LogeA</b>	120	35.0	35.0	5.9	20.0	48.0	28.0
<b>Rev</b>	120	50392.2	48231.5	12645.4	5274.8	84381.0	79106.2
<b>Scol</b>	120	0.3649	0.3626	0.0654	0.2170	0.5962	0.3792
<b>DHu</b>	120	0.7196	0.2519	1.4697	0.0083	12.5364	12.5282
<b>BoHu</b>	120	0.9028	0.5386	1.0607	0.0013	5.5996	5.5983
<b>PoHu</b>	90	2.2112	0.2300	4.9911	0	26.4521	26.4521
<b>VoHu</b>	93	24.478	3.101	41.724	0.013	209.834	209.821
<b>EqHu</b>	118	0.0293	0.0223	0.0302	0.0007	0.2201	0.2194
<b>RuHu</b>	120	0.1707	0.0984	0.2169	0	0.9767	0.9767
<b>PrTo</b>	120	978.6	1003.0	116.1	682.7	1227.6	544.9
<b>NMarAv</b>	120	3.04	0	9.06	0	45.00	45.00
<b>FonN</b>	120	18.54	17.00	8.28	5.00	45.00	40.00
<b>Drain</b>	104		Impar-fait				

Tableau XIII : Définitions des classes de chaque variable indépendante

Groupe	Variable	Transformation	Valeurs	Percentile	
Densité animale	DBo	Faible	$\leq 0.6506633$	Q.90	
		Élevée	$> 0.6506633$		
	DPo	Faible	$\leq 0.7888818$	Q.75	
		Élevée	$> 0.7888818$		
	DVo	Faible	$\leq 18.85143$	Q. 80	
		Élevée	$> 18.85143$		
	DEq	Faible	$\leq 0.04089894$	Q.90	
		Élevée	$> 0.04089894$		
	DRu	Faible	$\leq 0.1196513$	Q.85	
		Élevée	$> 0.1196513$		
	Utilisation des terres	UCult	Faible	$\leq 0.7258843$	Q.60
			Forte	$> 0.7258843$	
UJach		Faible	$\leq 0.007174198$	Q.85	
		Forte	$> 0.007174198$		
UPatCE		Faible	$\leq 0.1167579$	Q.90	
		Forte	$> 0.1167579$		
UPatNat		Faible	$\leq 0.2463039$	Q.90	
		Forte	$> 0.2463039$		
UEpFS		Faible	$\leq 0.1444738$	Q.90	
		Forte	$> 0.1444738$		
UEpSI		Faible	$\leq 0.002934521$	Q.70	
		Forte	$> 0.002934521$		
UEpFLS		Faible	$\leq 0.06363109$	Q.85	
		Forte	$> 0.06363109$		
UEpFLI		Faible	$\leq 0.001229201$	Q.70	
		Forte	$> 0.001229201$		
Socio-économiques		LogeA	Jeune	$\leq 31$	Q.25, Q.50, Q.75
			Moyen	$> 31$ et $\leq 35$	
	Vieux		$> 35$ et $\leq 39$		
	Très vieux		$> 39$		
	Rev	Faible	$< 36686.95$	Q.10	
		Élevé	$\geq 36686.95$		
	Scol	Faible	$\leq 0.4354112$	Q.90	
		Élevée	$> 0.4354112$		
	DHu	Faible	$\leq 0.6717562$	Q.75	
		Élevée	$> 0.6717562$		
Rapport animaux/humains	BoHu	Faible	$\leq 2.336542$	Q.90	
		Élevé	$> 2.336542$		
	PoHu	Faible	$\leq 2.39508$	Q.80	
		Élevé	$> 2.39508$		
	EqHu	Faible	$\leq 0.06497419$	Q.90	
		Élevé	$> 0.06497419$		
	VoHu	Faible	$\leq 9.638886$	Q.60	
		Élevé	$> 9.638886$		
	RuHu	Faible	$\leq 0.2177678$	Q.75	
		Élevé	$> 0.2177678$		

Suite du Tableau XIII : Définitions des classes de chaque variable indépendante

Groupe	Variable	Transformation	Valeurs	Percentile
Météorologique	PrTo	Faible	< = 908.25	Q.25, Q.50, Q.75
		Moyenne	> 908.25 et < = 1003	
		Élevée	> 1003 et < = 1046.425	
		Très élevée	> 1046.425	
	NMarAv05	Faible	< = 2.283305	Q.85
		Importante	> 2.283305	
	FonN	Faible	< = 31	Q.90
Importante		> 31		
Pédologique	Drain	Bien drainé	Bien drainé	
		Drainage imparfait	Drainage imparfait	
		Drainage pauvre	Drainage pauvre Drainage très pauvre	

Les tests de T et d'Aspin-Welch ont déterminé que l'on ne pouvait rejeter l'hypothèse nulle d'égalité des moyennes de chacune des variables testées, à l'exception de deux, entre le groupe des SRU incluses dans l'étude et le groupe des SRU exclues de l'étude. Seules les moyennes des variables DRu et UEpFS étaient différentes entre les deux groupes. La densité des ruminants autres que les bovins semblait être plus élevée dans le groupe des SRU incluses dans l'étude. À l'opposé, l'épandage de fumier solide semblait être plus important dans les SRU exclues de l'étude.

### 3.1.3. Association univariée

Le Tableau XIV (p. 65) présente les variables indépendantes retenues suite à l'analyse des associations univariées. Le patron 1 présentait le plus grand nombre de variables indépendantes ayant une association non conditionnelle significative avec la variable dépendante. Selon l'une des conditions d'inclusion établies, les variables DPo et PoHu devaient être exclues du processus de sélection des variables pour la construction du modèle de régression. Par contre, l'association non conditionnelle de ces variables avec la variable dépendante 1 suggérait que ces variables étaient chacune fortement associées à la variable dépendante. En effet, les rapports de cotes de ces variables étaient assez élevés. Ainsi, une SRU possédant une DPo élevée aurait un risque 2.8575 (I.C. 95% : 1.3209-

6.1814) fois plus élevés que *E. coli* contaminant le ou les puits soit résistant au patron 1 qu'une SRU possédant une DPo faible. Ce risque serait plus élevé lorsque l'on considère la variable PoHu [RC : 5.5673 (I.C. 95% : 1.6404-18.8951)]. Pour cette raison, il a été jugé bon de tenir compte de ces variables représentant une influence à caractère porcin. Ainsi, bien que les variables DPo et PoHu devaient être retirées du processus de sélection des variables pour la construction du modèle de régression selon la condition préétablie d'inclusion des variables, elles ont tout de même été retenues. La même exception a été appliquée à la variable DVo (selon le patron 1,  $p = 0.0868$  et selon le patron 2,  $p = 0.0862$ ). Par contre, les rapports de cotes obtenus pour cette variable étaient plus faibles que ceux des variables représentant une influence porcine. Afin d'explorer son influence, le processus de sélection rétrograde des variables pour la construction des modèles de régression a été appliqué à des modèles initiaux comprenant la variable DVo et d'autres l'excluant.

En fonction de la résistance au patron 1, ce sont des variables ayant un lien avec l'élevage porcin qui avait les rapports de cotes les plus élevés, soit UEpFLS (4.3241) et PoHu (5.5673). En effet, les SRU fortement utilisées pour l'épandage de fumier liquide en surface étaient surtout situées dans le sud et l'ouest de la région d'étude. Un grand nombre de ces SRU possédait une proportion importante de ses puits qui étaient contaminés par *E. coli* résistant au patron 1 (Figure 12, p. 66). Un constat semblable pouvait s'appliquer à la variable PoHu (Figure 13, p. 67). Les SRU où l'épandage de fumier liquide en surface était important et les SRU possédant un nombre élevé de porcs par rapport aux humains se retrouvaient sensiblement dans la même région, soit dans le sud et l'ouest de la région d'étude.

Quant au patron 2, le rapport de cote le plus élevé était retrouvé avec la variable UEpFLS. Par contre, les intervalles de confiance des rapports de cotes de toutes les variables retenues incluaient la valeur 1. L'association entre UEpFLS et la proportion de puits possédant *E. coli* résistant au patron 2 était visuellement moins évidente sur les cartes choroplèthes. Il y avait beaucoup moins de SRU fortement utilisées pour l'épandage de

fumier liquide en surface dont la totalité des puits échantillonnés était positive au patron 2, comparativement au patron 1. D'ailleurs, plusieurs de ces SRU ne possédaient aucun puits échantillonné avec *E. coli* résistant au patron 2.

Aucune variable d'épandage de fumier n'était significative lorsque le patron 3 était considéré. Par contre, d'après la régression logistique univariée, il semblait y avoir une association avec des variables représentant les facteurs bovins (DBo) et porcins (DPo et PoHu). Les indications de production bovine importante semblaient aussi être concentrées dans l'ouest de la région d'étude, mais un peu plus au nord que celles liées aux porcs. D'un point de vue subjectif, les associations visualisées spatialement entre ces variables et la variable dépendante 3 étaient moins évidentes à visualiser sur les cartes choroplèthes comparativement à celles entre la résistance au patron 1 et les variables UEpFLS et PoHu.

Tableau XIV : Résultats de l'analyse univariée des variables retenues

Patron	Variable indépendante	Coefficient (probabilité)	Rapport de cotes (I.C.95%)
1	DBo	1.0531 (0.0382)	Élevée vs faible = 2.8667 (1.0588-7.7616)
	DPo	1.0499 (0.0077)	Élevée vs faible = 2.8575 (1.3209-6.1814)
	DVo	0.5637 (0.0868)	Élevée vs faible = 1.7572 (0.9218-3.3499)
	UJach	-0.4824 (0.1282)	Élevée vs faible = 0.6173 (0.3315-1.1493)
	UEpFS	0.7982 (0.0725)	Élevée vs faible = 2.2216 (0.9298-5.3081)
	UEpFLS	1.4642 (0.0079)	Élevée vs faible = 4.3241 (1.4669-12.7463)
	LogeA	(ensemble : 0.0494)	Moyen vs jeune = 1.3976 (0.7899-2.4727) Vieux vs jeune = 2.2040 (1.2043-4.0336) Très vieux vs jeune = 1.8678 (1.0108-3.4514)
	Scol	1.0088 (0.1208)	Faible vs forte = 2.7423 (0.7668-9.8072)
	BoHu	1.2325 (0.0269)	Élevé vs faible = 3.4296 (1.1510-10.2196)
	PoHu	1.7169 (0.0059)	Élevé vs faible = 5.5673 (1.6404-18.8951)
	EqHu	1.2220 (0.0283)	Élevé vs faible = 3.3939 (1.1387-10.1158)
	PrTo	(ensemble : 0.0015)	Moyenne vs faible = 0.6627 (0.3746-1.1723) Élevée vs faible = 2.1880 (1.1607-4.1247) Très élevée vs faible = 1.5316 (0.8177-2.8688)
2	DVo	0.5171 (0.0862)	Élevée vs faible = 1.6771 (0.9291-3.0272)
	UEpFLS	0.7669 (0.0510)	Élevée vs faible = 2.1530 (0.9968-4.6505)
	EqHu	0.6223 (0.1358)	Élevée vs faible = 1.8632 (0.8224-4.2212)
	PrTo	(ensemble : 0.0333)	Moyenne vs faible = 0.9804 (0.5185-1.8537) Élevée vs faible = 1.8301 (0.9874-3.3918) Très élevée vs faible = 2.0465 (1.0861-3.8562)
3	DBo	0.9406 (0.0202)	Élevée vs faible = 2.5616 (1.1582-5.6654)
	DPo	0.5902 (0.0744)	Élevée vs faible = 1.8044 (0.9434-3.4511)
	UJach	-0.9139 (0.0347)	Élevée vs faible = 0.4010 (0.1717-0.9362)
	Scol	1.2498 (0.0160)	Élevée vs faible = 3.4898 (1.2628-9.6446)
	PoHu	0.7862 (0.0493)	Élevé vs faible = 2.1950 (1.0025-4.8059)
4	UCult	1.4533 (0.0851)	Élevée vs faible = 4.2773 (0.8180-22.3670)
	NMarAv	1.4709 (0.0589)	Importante vs faible = 4.3529 (0.9462-20.0255)
5	DVo	1.2490 (0.0694)	Élevée vs faible = 3.4868 (0.9055-13.4265)
	DEq	1.1841 (0.0908)	Élevée vs faible = 3.2679 (0.8285-12.8889)
	UJach	1.2216 (0.0665)	Élevée vs faible = 3.3926 (0.9201-12.5095)
	DHu	1.0844 (0.0789)	Élevé vs faible = 2.9576 (0.8823-9.9140)

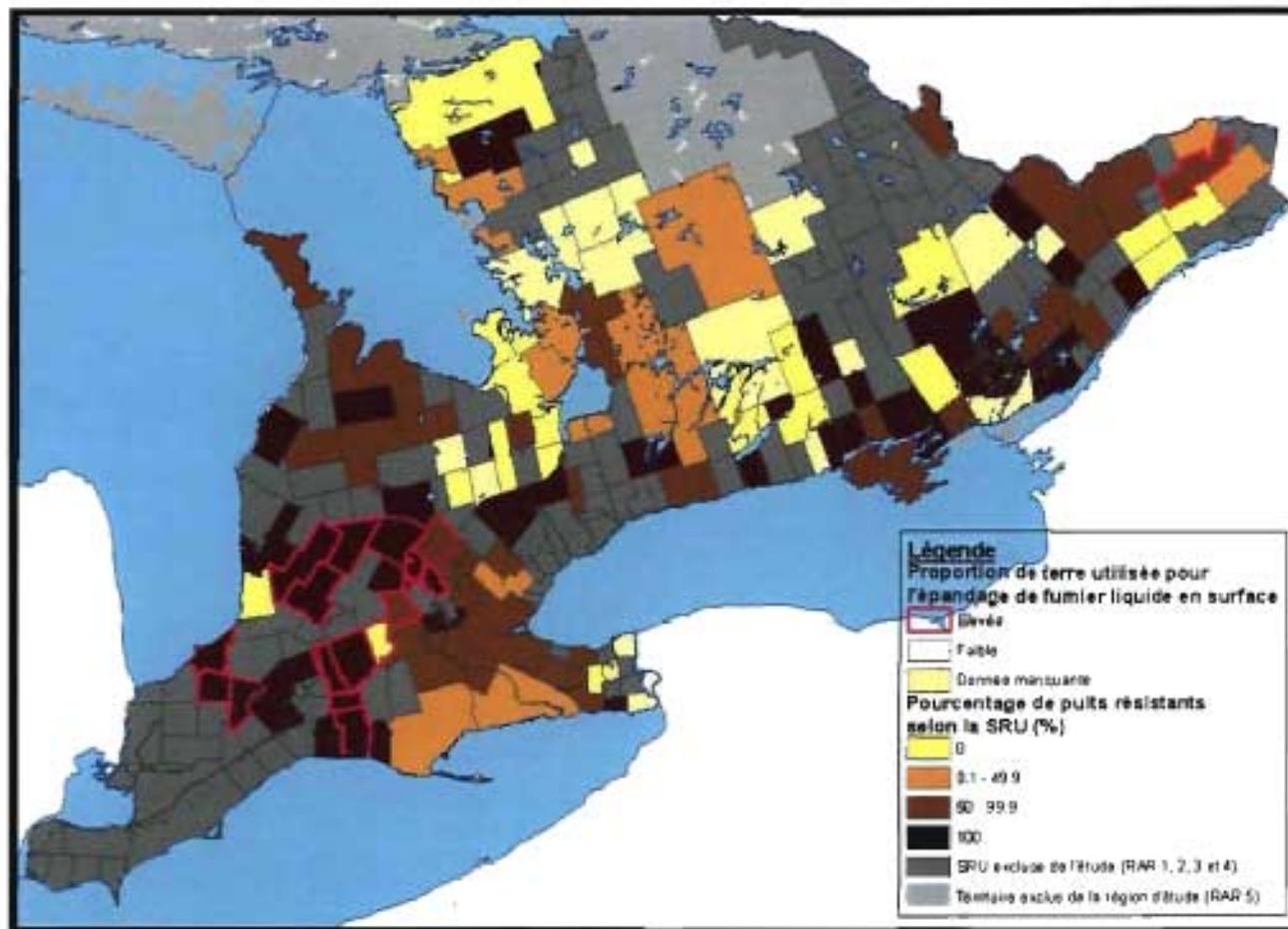


Figure 12: Fréquence relative de puits résistants au patron 1 par SRU et catégorisation des SRU quant à l'épandage de fumier liquide en surface

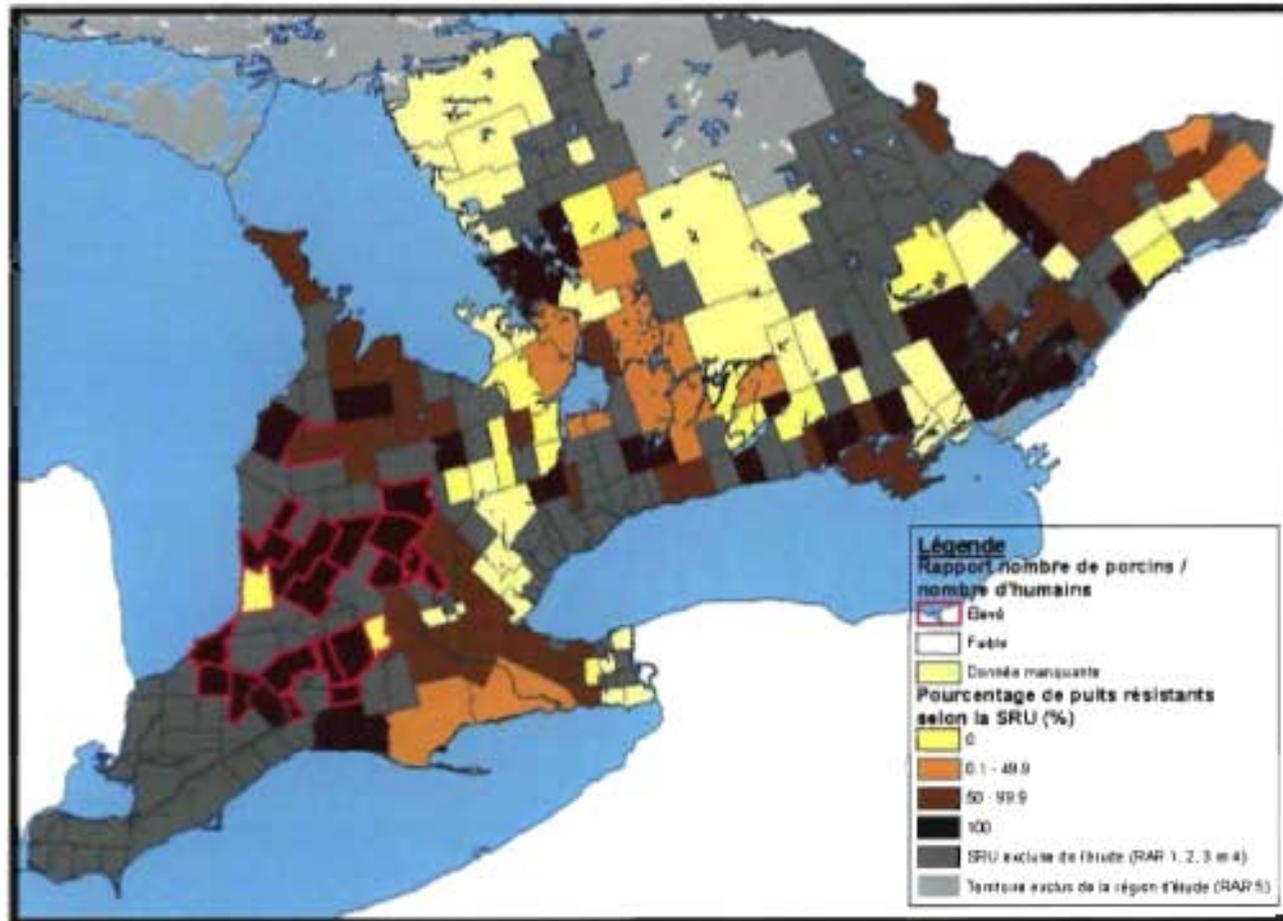


Figure 13 : Fréquence relative de puits résistants au patron I par SRU et catégorisation des SRU quant au rapport nb de porcins / nb d'humains

Aucune variable d'agriculture n'avait d'influence sur la résistance antimicrobienne d'*E. coli* dans les puits au patron humain (patron 4). Les seules variables ayant une association non conditionnelle significative étaient NMarAv ( $p = 0.0589$ ) et UCult ( $p = 0.0851$ ). Aucune association significative de ces variables avec la résistance antimicrobienne d'*E. coli* dans les puits aux autres patrons n'a été retrouvée (Tableau XIV, p. 65).

À l'opposé, certaines variables d'agriculture et une variable de population humaine avaient une association non conditionnelle significative avec la variable dépendante 5 (patron vétérinaire) (Tableau XIV, p. 65). Ces variables étaient DVo ( $p = 0.0694$ ), DEq ( $p = 0.0908$ ) et UJach ( $p = 0.0665$ ). De plus, les rapports de cotes de chacune de ces variables étaient situés entre 3 et 4 avec la valeur 1 incluse dans leurs intervalles de confiance. La variable significative de la catégorie socio-économique était DHu ( $p = 0.0789$ ).

Il est à noter que, concernant les patrons 4 et 5, la valeur de P de ces associations dites significatives à un seuil alpha de 0.15 était toutes supérieures à 0.05.

### **3.2. Modèle de régression**

Pour le patron 1, les variables DBo, DPo, DVo, UEpFS, UEpFLS, BoHu, PoHu, EqHu et PrTo étaient considérées comme variables indépendantes potentielles. Pour le patron 2, les variables DVo, UEpFLS, EqHu et PrTo étaient considérées comme variables indépendantes potentielles. Finalement, pour le patron 3, les variables DBo, DPo et PoHu étaient jugées comme variables indépendantes potentielles.

### 3.2.1. Corrélation

Le Tableau XV (p. 70) présente les coefficients de corrélation de Pearson entre chaque variable indépendante dont l'association non conditionnelle avec la variable dépendante d'un des patrons de résistance était significative. DBo et UEpFS semblaient démontrer de la colinéarité ( $Rho = 0.76977$ ). Il en était de même entre DBo et BoHu ( $Rho = 0.71665$ ), des variables représentant une influence bovine. Il semblait aussi y avoir de la colinéarité entre DPo et UEpFLS ( $Rho = 0.80623$ ), DPo et PoHu ( $Rho = 0.90187$ ) et UEpFLS et PoHu ( $Rho = 0.72766$ ), des variables représentant une influence porcine. Par contre, il ne semblait pas y avoir de colinéarité prononcée entre UEpFS et BoHu ( $Rho = 0.55488$ ). Ainsi, pour chaque patron, les modèles de régression possibles ne devaient contenir qu'une seule des variables par paires de variables démontrant de la colinéarité. Le Tableau XVI (p. 71) présente les combinaisons possibles de variables (appelées modèles initiaux) sur lesquelles se basait la sélection de variables pour construire les modèles de régression multiple concernant les patrons de résistance 1, 2 et 3. Pour le patron 1, 12 combinaisons étaient identifiées, deux pour le patron 2 et deux pour le patron 3. Concernant les patrons 1 et 2, la moitié des combinaisons incluait DVo, alors que l'autre moitié l'excluait.

Tableau XV : Coefficient de Pearson par paire de variables indépendantes

	Coefficient de corrélation de Pearson (Rho) (Probabilité)											
	DBo	DPo	DVo	UEpFS	UEpFLS	LogeA	Scol	BoHu	PoHu	EqHu	PrTo	NMarAv
<b>DBo</b>	1	0.39773 0.0001	0.10869 0.2997	<b>0.76977</b> <b>&lt;0.0001</b>	0.43460 <0.0001	0.18201 0.0466	0.14924 0.1038	<b>0.71665</b> <b>&lt;0.0001</b>	0.22483 0.0331	0.61358 <0.0001	0.01028 0.9112	-0.30270 0.0008
<b>DPo</b>	<b>0.39773</b> <b>0.0001</b>	1	0.34674 0.0031	0.28463 0.0065	<b>0.80623</b> <b>&lt;0.0001</b>	0.41681 <0.0001	0.35814 0.0005	0.53182 <0.0001	<b>0.90187</b> <b>&lt;0.0001</b>	0.34560 0.0008	0.13247 0.2133	-0.15530 0.1438
<b>DVo</b>	<b>0.10869</b> <b>0.2997</b>	<b>0.34674</b> <b>0.0031</b>	1	0.36401 0.0003	0.34844 0.0017	0.19816 0.0569	-0.05447 0.6041	0.11368 0.2779	0.20359 0.0886	0.07609 0.4710	0.07784 0.4583	-0.23162 0.0255
<b>UEpFS</b>	<b>0.76977</b> <b>&lt;0.0001</b>	<b>0.28463</b> <b>0.0065</b>	<b>0.36401</b> <b>0.0003</b>	1	0.28766 0.0031	0.14453 0.1153	0.13776 0.1335	0.55488 <0.0001	0.09919 0.3523	0.58721 <0.0001	0.04280 0.6425	-0.25045 0.0058
<b>UEpFLS</b>	<b>0.43460</b> <b>&lt;0.0001</b>	<b>0.80623</b> <b>&lt;0.0001</b>	<b>0.34844</b> <b>0.0017</b>	<b>0.28766</b> <b>0.0031</b>	1	0.26770 0.0060	0.11193 0.2580	0.53156 <0.0001	<b>0.72766</b> <b>&lt;0.0001</b>	0.27290 0.0053	0.10835 0.2736	-0.21436 0.0289
<b>LogeA</b>	<b>0.18201</b> <b>0.0466</b>	<b>0.41681</b> <b>&lt;0.0001</b>	<b>0.19816</b> <b>0.0569</b>	<b>0.14453</b> <b>0.1153</b>	<b>0.26770</b> <b>0.0060</b>	1	0.36913 <0.0001	0.45256 <0.0001	0.46507 <0.0001	0.18969 0.0396	0.42642 <0.0001	-0.16690 0.0685
<b>Scol</b>	<b>0.14924</b> <b>0.1038</b>	<b>0.35814</b> <b>0.0005</b>	<b>-0.05447</b> <b>0.6041</b>	<b>0.13776</b> <b>0.1335</b>	<b>0.11193</b> <b>0.2580</b>	<b>0.36913</b> <b>&lt;0.0001</b>	1	0.34224 0.0001	0.31534 0.0025	0.24248 0.0082	0.27475 0.0024	0.24625 0.0067
<b>BoHu</b>	<b>0.71665</b> <b>&lt;0.0001</b>	<b>0.53182</b> <b>&lt;0.0001</b>	<b>0.11368</b> <b>0.2779</b>	<b>0.55488</b> <b>&lt;0.0001</b>	<b>0.53156</b> <b>&lt;0.0001</b>	<b>0.45256</b> <b>&lt;0.0001</b>	<b>0.34224</b> <b>0.0001</b>	1	0.54804 <0.0001	0.69008 <0.0001	0.23892 0.0086	-0.19018 0.0375
<b>PoHu</b>	<b>0.22483</b> <b>0.0331</b>	<b>0.90187</b> <b>&lt;0.0001</b>	<b>0.20359</b> <b>0.0886</b>	<b>0.09919</b> <b>0.3523</b>	<b>0.72766</b> <b>&lt;0.0001</b>	<b>0.46507</b> <b>&lt;0.0001</b>	<b>0.31534</b> <b>0.0025</b>	<b>0.54804</b> <b>&lt;0.0001</b>	1	0.26828 0.0106	0.23072 0.0287	-0.10783 0.3117
<b>EqHu</b>	<b>0.61358</b> <b>&lt;0.0001</b>	<b>0.34560</b> <b>0.0008</b>	<b>0.07609</b> <b>0.4710</b>	<b>0.58721</b> <b>&lt;0.0001</b>	<b>0.27290</b> <b>0.0053</b>	<b>0.18969</b> <b>0.0396</b>	<b>0.24248</b> <b>0.0082</b>	<b>0.69008</b> <b>&lt;0.0001</b>	<b>0.26828</b> <b>0.0106</b>	1	0.02384 0.7978	-0.11820 0.2024
<b>PrTo</b>	<b>0.01028</b> <b>0.9112</b>	<b>0.13247</b> <b>0.2133</b>	<b>0.07784</b> <b>0.4583</b>	<b>0.04280</b> <b>0.6425</b>	<b>0.10835</b> <b>0.2736</b>	<b>0.42642</b> <b>&lt;0.0001</b>	<b>0.27475</b> <b>0.0024</b>	<b>0.23892</b> <b>0.0086</b>	<b>0.23072</b> <b>0.0287</b>	<b>0.02384</b> <b>0.7978</b>	1	0.15783 0.0851
<b>NMarAv</b>	<b>-0.30270</b> <b>0.0008</b>	<b>-0.15530</b> <b>0.1438</b>	<b>-0.23162</b> <b>0.0255</b>	<b>-0.25045</b> <b>0.0058</b>	<b>-0.21436</b> <b>0.0289</b>	<b>-0.16690</b> <b>0.0685</b>	<b>0.24625</b> <b>0.0067</b>	<b>-0.19018</b> <b>0.0375</b>	<b>-0.10783</b> <b>0.3117</b>	<b>-0.11820</b> <b>0.2024</b>	<b>0.15783</b> <b>0.0851</b>	1

Tableau XVI : Modèles initiaux à tester selon le patron de résistance

Patron	Modèle initial	Variables potentiellement incluses
1	A	DBo, DPo, EqHu, PrTo
	B	DBo, UEpFLS, EqHu, PrTo
	C	DBo, PoHu, EqHu, PrTo
	D	DPo, UEpFS, BoHu, EqHu, PrTo
	E	UEpFS, UEpFLS, BoHu, EqHu, PrTo
	F	UEpFS, BoHu, PoHu, EqHu, PrTo
	G	DBo, DPo, DVo, EqHu, PrTo
	H	DBo, DVo, UEpFLS, EqHu, PrTo
	I	DBo, DVo, PoHu, EqHu, PrTo
	J	DPo, DVo, UEpFS, BoHu, EqHu, PrTo
	K	DVo, UEpFS, UEpFLS, BoHu, EqHu, PrTo
	L	UEpFS, DVo, BoHu, PoHu, EqHu, PrTo
	2	M
N		DVo, UEpFLS, EqHu, PrTo
3	O	DBo, DPo
	P	DBo, PoHu

Seule la variable BoHu dans les combinaisons D, E, F, J, K et L possédait une valeur de la statistique de tolérance inférieure au seuil fixé de 0.4. Lorsque la variable BoHu était retirée de ces modèles, aucune variable n'avait de valeur de tolérance inférieure au seuil fixé et la valeur de tolérance des autres variables d'agriculture augmentait (Tableau XVII, p. 72). La variable BoHu a donc été éliminée avant le processus de sélection des variables pour la construction des modèles de régression. Quant aux combinaisons non présentées dans le Tableau XVII, (p. 72) aucune variable n'avait de valeur de la statistique inférieure au seuil fixé. Ainsi, toutes les variables de ces combinaisons étaient retenues pour le processus de sélection des variables pour la construction des modèles de régression.

Tableau XVII : Statistique de tolérance des variables des combinaisons D, E et F avec et sans l'inclusion de la variable BoHu

Modèle initial	Variable	Tolérance avec BoHu	Tolérance sans BoHu
<b>D</b>	DPo	0.71595	0.85582
	UEpFS	0.57290	0.62332
	BoHu	<b>0.36479</b>	
	EqHu	0.45145	0.59650
	PrTo	0.89235	0.98094
<b>E</b>	UEpFS	0.56134	0.59815
	UEpFLS	0.69466	0.89438
	BoHu	<b>0.34351</b>	
	EqHu	0.44272	0.60308
	PrTo	0.87495	0.98760
<b>F</b>	UEpFS	0.52789	0.62390
	BoHu	<b>0.31121</b>	
	PoHu	0.61994	0.86865
	EqHu	0.45030	0.58303
	PrTo	0.88816	0.94193
<b>J</b>	DPo	0.56817	0.75143
	DVo	0.72419	0.75409
	UEpFS	0.44274	0.50773
	BoHu	<b>0.30694</b>	
	EqHu	0.40673	0.51307
	PrTo	0.90623	0.97800
<b>K</b>	DVo	0.74923	0.79777
	UEpFs	0.46312	0.52606
	UEpFLS	0.55865	0.81550
	BoHu	<b>0.27731</b>	
	EqHu	0.40442	0.55661
	PrTo	0.86545	0.98025
<b>L</b>	DVo	0.75882	0.79673
	UEpFS	0.36972	0.48610
	BoHu	<b>0.24172</b>	
	PoHu	0.47800	0.80275
	EqHu	0.40527	0.50816
	PrTo	0.90349	0.93560

### **3.2.2. Modèles de régression**

#### **3.2.2.1. Sélection rétrograde pour les modèles d'effet principal et ajout d'interaction**

Le processus de sélection rétrograde a été appliqué sur chaque modèle initial. Le Tableau XVIII (p. 74) présente les sélections significatives pour chaque modèle et patron. Dans les cas où le processus de sélection rétrograde sélectionnait les mêmes variables à partir de combinaisons de variables cependant différentes, seul la sélection obtenue avec la taille d'échantillons le plus élevé (n) était considéré pour la poursuite de la construction du modèle puisque l'ajustement des modèles aux données serait de toute manière évalué par après. Pour chacun de ces modèles d'effet principal, une interaction entre deux variables présentes était ajoutée et testée. Aucune interaction dans ces modèles de régression n'était significative.

Tableau XVIII : Variables significatives suite à la sélection rétrograde de variables appliquée sur chaque modèle initial déterminé au départ

Modèle initial	N	Pseudo-R <sup>2</sup>	Paramètre	Coefficient (p)	Erreur-type	RC (I.C. 95%)
<b>A (D)</b>	90	0.1202	Intercepte	0.3465 (0.1596)	0.2464	
			DPo élevée	0.9620 (0.0182)	0.4075	2.6170 (1.1774-5.8168)
			PrTo moyenne	-0.5988 (0.0758)	0.3373	0.5495 (0.2837-1.0642)
			PrTo élevée	0.8703 (0.0287)	0.3979	2.3877 (1.0947-5.2081)
			PrTo très élevée	0.3889 (0.3015)	0.3765	1.4754 (0.7054-3.0859)
<b>B (E)</b>	103	0.1395	Intercepte	0.2636 (0.2034)	0.2073	
			UEpFLS élevée	1.4822 (0.0094)	0.5703	4.4027 (1.4397-13.4642)
			PrTo moyenne	-0.6673 (0.0336)	0.3140	0.5131 (0.2773-0.9495)
			PrTo élevée	1.0942 (0.0026)	0.3635	2.9868 (1.4647-6.0903)
			PrTo très élevée	0.3465 (0.3010)	0.3351	1.4142 (0.7333-2.7271)
<b>C (F)</b>	90	0.1203	Intercepte	0.3363 (0.1741)	0.2474	
			PoHu élevé	1.3808 (0.0295)	0.6345	3.9780 (1.1470-13.7966)
			PrTo moyenne	-0.4826 (0.1517)	0.3367	0.6172 (0.3191-1.1939)
			PrTo élevée	0.8286 (0.0385)	0.4003	2.2900 (1.0449-5.0189)
			PrTo très élevée	0.4931 (0.1861)	0.3729	1.6374 (0.7883-3.4009)
<b>G (J)</b>	71	0.1207	Intercepte	0.3996 (0.1294)	0.2635	
			DPo élevée	0.8820 (0.0348)	0.4178	2.4158 (1.0652-5.4788)
			PrTo moyenne	-0.6693 (0.0691)	0.3682	0.5121 (0.2488-1.0539)
			PrTo élevée	0.7919 (0.0529)	0.4090	2.2075 (0.9903-4.9208)
			PrTo très élevée	0.4543 (0.2820)	0.4222	1.5750 (0.6885-3.6032)
<b>H (K)</b>	79	0.1349	Intercepte	0.1525 (0.5134)	0.2333	
			UEpFLS élevée	1.3320 (0.0202)	0.5734	3.7888 (1.2315-11.6567)
			PrTo moyenne	-0.3625 (0.2922)	0.3442	0.6959 (0.3545-1.3663)
			PrTo élevée	1.1502 (0.0025)	0.3799	3.1587 (1.5002-6.6507)
			PrTo très élevée	0.7691 (0.0549)	0.4006	2.1578 (0.9840-4.7315)

Suite du Tableau XVIII : Variables significatives suite à la sélection rétrograde de variables appliquée sur chaque modèle initial déterminé au départ

Modèle initial	N	Pseudo-R <sup>2</sup>	Paramètre	Coefficient (p)	Erreur-type	RC (I.C. 95%)
<b>I (L)</b>	71	0.1211	Intercepte	0.3884 (0.1423)	0.2647	
			PoHu élevé	1.2617 (0.0487)	0.6402	3.5313 (1.0070-12.3835)
			PrTo moyenne	-0.5364 (0.1439)	0.3670	0.5848 (0.2849-1.2007)
			PrTo élevée	0.7545 (0.0665)	0.4112	2.1266 (0.9499-4.7611)
			PrTo très élevée	0.5781 (0.1652)	0.4166	1.7827 (0.7879-4.0334)
<b>M</b>	103	0.0442	Intercepte	-1.0578 (<0.0001)	0.2321	
			PrTo moyenne	-0.1462 (0.6806)	0.3551	0.8640 (0.4308-1.7330)
			PrTo élevée	0.7076 (0.0320)	0.3300	2.0291 (1.0628-3.8740)
			PrTo très élevée	0.7011 (0.0384)	0.3385	2.0160 (1.0383-3.9142)
<b>N</b>	79	0.0760	Intercepte	-1.3927 (<0.0001)	0.2896	
			DVo élevée	0.6581 (0.0467)	0.3309	1.9312 (1.0095-3.6941)
			PrTo moyenne	0.1888 (0.6341)	0.3967	1.2078 (0.5551-2.6281)
			PrTo élevée	0.9173 (0.0132)	0.3700	2.5024 (1.2117-5.1680)
			PrTo très élevée	1.0527 (0.0080)	0.3971	2.8655 (1.3158-6.2402)
<b>O (P)</b>	90	0.0377	Intercepte	-1.0211 (<0.0001)	0.1425	
			DBo élevée	1.1881 (0.0062)	0.4337	3.2809 (1.4021-7.6770)

Patron d'antibiotiques évalués selon le modèle de régression (voir Tableau XVI, p. 71)

Modèle A-B-C-D-E-F-G-H-I-J-K-L : patron 1 (résistance à un antibiotique ou plus)

Modèle M-N : patron 2 (résistance à un ou deux antibiotiques)

Modèle O-P : patron 3 (résistance à trois antibiotiques ou plus)

### 3.2.2.2. Modèles incluant les variables confondantes

Pour chacun des modèles d'effet principal, l'ajout de chacune des variables confondantes potentielles ne modifiait l'estimé des rapports de cotes que de moins de 25%. Ainsi, aucune variable préalablement identifiée comme variable de confusion potentielle n'a été considérée.

### 3.2.3. Évaluation des modèles

Les statistiques de déviance  $\chi^2$ , de Pearson  $\chi^2$  et de Hosmer-Lemeshow (Tableau XIX, p. 76) de tous les modèles concernant le patron 1 (A, B et C) concluaient toutes à un non rejet de l'hypothèse nulle d'ajustement adéquat des modèles aux données. Par contre, les valeurs de probabilité de deux tests sur trois [déviance  $\chi^2$  (0.1669) et Pearson  $\chi^2$  (0.2018)] étaient un peu faibles dans le cas du modèle C. Ceci pouvait suggérer un problème potentiel dans l'ajustement de ces modèles. Les valeurs de probabilité obtenues des trois tests du modèle N (résistance à un ou deux antibiotiques) étaient aussi assez faibles. Une évaluation plus approfondie pourrait nous révéler un problème d'ajustement de ce modèle. Les valeurs des statistiques des modèles M et O étaient toutes incohérentes, suggérant ainsi fort probablement un mauvais ajustement de ces modèles.

Tableau XIX : Statistiques d'ajustement selon le modèle

Modèle	Déviance $\chi^2$			Pearson $\chi^2$			Hosmer-Lemeshow		
	$\chi^2$	dl	P	$\chi^2$	dl	P	Valeur	dl	P
A	0.5068	3	0.9174	0.4933	3	0.9204	0.3738	4	0.9846
B	4.2926	3	0.2315	2.9968	3	0.3921	1.2990	4	0.8615
C	5.0673	3	0.1669	4.6207	3	0.2018	0.0863	3	0.9934
M	0	0	.	0	0	.	0.0000	2	1.0000
N	5.2748	3	0.1527	5.3444	3	0.1482	5.0279	4	0.2844
O	0	0	.	0	0	.	0.0000	0	.

Concernant le modèle A, il ne semblait pas y avoir de valeur résiduelle aberrante. Les valeurs étaient toutes situées entre -3 et 3. Il n'y avait aucun patron systématique visualisé, i.e. que la dispersion des points semblait être homogène. Il y avait une seule observation qui semblait avoir une influence plus forte que les autres dans le modèle sur le coefficient de DPo. Dans ce cas, l'étape suivante serait d'enlever cette observation et de recommencer la construction du modèle à partir du processus de sélection rétrograde des variables et d'évaluer s'il était jugé nécessaire d'enlever cette observation pour la construction de ce modèle. Puisque les évaluations préliminaires des tests d'ajustement du modèle et des résidus de déviance étaient satisfaisantes, cette étape n'a pas été poursuivie.

Concernant le modèle B, les valeurs résiduelles étaient aussi toutes situées entre -3 et 3. Deux observations semblaient avoir une valeur résiduelle légèrement aberrante. Il n'y avait aucun patron systématique visualisé. Par contre, la dispersion des points semblait être un peu moins homogène que celle obtenue pour le modèle A. Seulement trois observations semblaient avoir une influence plus forte que les autres dans le modèle sur le coefficient de la variable UEpFLS. Étant donné que les résultats de deux des trois tests d'ajustement du modèle étaient jugés adéquats, le modèle obtenu fut conservé tel quel.

Concernant le modèle C, les valeurs résiduelles étaient aussi toutes situées entre -3 et 3. La dispersion des points semblait homogène, ou sinon l'écart par rapport à zéro des valeurs résiduelles négatives semblait légèrement supérieur à celui des valeurs positives. Trois observations semblaient avoir une influence plus forte que les autres dans le modèle sur le coefficient de PoHu. Étant donné que les résultats d'un des trois tests d'ajustement du modèle laissaient à désirer, l'étape suivante serait d'exclure ces observations et de recommencer la construction du modèle C à partir du processus de sélection rétrograde des variables et d'évaluer s'il était jugé nécessaire d'enlever cette ou ces observations pour la construction de ce modèle. Par manque de temps et parce que l'évaluation des valeurs résiduelles de déviance était tout de même jugée satisfaisante, cette étape n'a pas été poursuivie.

Concernant le modèle M, les valeurs étaient situées entre environ -2 et 2. Il semblait y avoir un léger patron systématique des valeurs résiduelles entre -1.2 et -0.5. Aucune variable d'agriculture n'avait été retenue lors de la construction de ce modèle. Certaines observations semblaient avoir une plus grande influence sur le coefficient de chacune des catégories de la variable PrTo. Les observations influentes étaient généralement les mêmes pour chaque catégorie de la variable. Pour ces raisons et à cause des résultats incohérents des tests d'ajustement du modèle, l'ajustement du modèle M a été jugé inadéquat.

Concernant le modèle N, les valeurs résiduelles étaient aussi toutes situées entre -2 et 2. Il n'y avait aucun patron systématique visualisé. La dispersion des points semblait relativement homogène. Une observation semblait influencer de façon très importante l'estimé du coefficient de la variable DVo comparativement aux autres observations. Cette observation était celle de la SRU de Hamilton. Cette dernière possédait une forte densité de volailles et cinq des neuf puits échantillonnés étaient positifs au patron 2. Pour cette raison, et étant donné que les résultats des trois tests d'ajustement du modèle suggéraient un problème dans l'ajustement du modèle N aux données, l'observation problématique en question a été retirée et la construction du modèle a été recommencée à partir du processus de sélection rétrograde des variables afin d'évaluer si le modèle était mieux ajusté sans cette observation. Il apparut que le modèle obtenu sans cette observation était complètement autre que celui obtenu avec cette observation. Suite à l'exclusion de cette observation, la variable DVo n'était plus sélectionnée. Seulement la variable PrTo demeurait dans le modèle. Le modèle final se retrouvait alors à être le même que le modèle M. Par contre, le modèle M n'était pas ajusté adéquatement aux données et l'interprétation de celui-ci s'en retrouvait donc affectée.

Concernant le modèle O, les valeurs résiduelles étaient situées entre -2 et environ 2. La dispersion des points était légèrement hétérogène ; il semblait y avoir une concentration importante de points avec une valeur résiduelle entre -1.5 et -0.75. Quatre observations semblaient influencer l'estimé du coefficient de DBo de façon plus importante que les autres. Étant donné les résultats des tests d'ajustement du modèle et les évaluations des

résidus et de l'influence des données sur l'estimé du coefficient de la variable DBo, le modèle obtenu ne semblait pas adapté adéquatement aux données observées. Pour cette raison, le modèle obtenu ne pouvait pas être interprété de manière satisfaisante.

En résumé, nous avons tenté de construire des modèles de régression multiples pour les patrons 1, 2 et 3. Ces constructions se sont basées sur différents ensembles de variables. Trois modèles étaient obtenus pour le patron 1 et contenaient tous une variable d'exposition représentant une influence d'origine porcine et une variable météorologique de précipitation totale annuelle. L'ajustement de ces trois modèles était satisfaisant. Par contre, les modèles obtenus pour les patrons 2 et 3 laissaient plutôt à désirer. Leur ajustement aux données observées était jugé insatisfaisant.

### **3.2.4. Association**

Seuls les modèles obtenus avec le patron 1 (A, B et C) pouvaient être interprétés. Dans chaque modèle final concernant ce patron, une variable liée à la production porcine se retrouvait significative. Le risque de retrouver *E. coli* résistant au patron 1 dans un puits contaminé par *E. coli* était plus élevé lorsqu'une SRU présentait une indication d'une influence liée à une production porcine élevée (densité porcine élevée, épandage de fumier liquide en surface important ou rapport entre le nombre de porcs et d'humains élevé) comparativement à une SRU avec une indication faible. En général, le rapport de cotes de la variable de l'influence porcine se situait entre 2 et 5 dépendamment de la variable considérée. La variable ayant le rapport de cotes le plus élevé était UEpFLS, soit de 4.4027 (I.C. 95% 1.4397-13.4642) pour le modèle B (Tableau XVIII, p. 74). De plus, dans chacun des modèles obtenus, la variable météorologique de précipitation totale annuelle semblait aussi jouer un rôle. Par contre, la force et la direction de son association avec la «positivité d'une SRU» variaient selon le degré de la quantité annuelle de précipitation.

## Partie 4 Discussion générale des résultats

L'Ontario est une province canadienne subdivisée physiographiquement entre le nord et le sud avec la délimitation située approximativement entre le bouclier précambrien et les dépôts glaciaux non consolidés, de sorte que la province ne peut généralement pas être étudiée et analysée partout de la même façon [50]. La présente étude porte sur le sud de l'Ontario (région agricole de recensement 1, 2, 3 et 4), où il y a une concentration des activités démographiques et agricoles. Cette étude vise à déterminer les associations existantes entre le niveau de résistance antimicrobienne et des facteurs agro-environnementaux au niveau régional.

En Ontario, il y a approximativement 750 000 puits enregistrés et environ 1.5 million supplémentaires qui ne le sont pas (Mary Jane Conboy, Well Wise Center, communication personnelle). Les puits sont surtout situés dans le sud de l'Ontario, ce qui correspond à notre région d'étude et en milieu rural [12]. Des auteurs estiment qu'il y a environ 500 000 puits abandonnés en Ontario [12]. L'unité de travail de cette étude est la SRU fusionnées en respect du principe de confidentialité. Dans le sud de l'Ontario, il y avait 224 SRU de la sorte. Parmi elles, 120 ont fait partie de notre étude. En restant conservateur, nous pouvons dire que la présente étude est valide pour les 120 régions ayant présenté des échantillons au projet ARO durant l'année d'étude 2004 et sur lesquelles se basent ce projet de maîtrise.

Dans une approche moins conservatrice, nous pouvons tenter d'élargir la population de référence. Les échantillons soumis aux programmes de surveillance selon les données de l'année d'étude 2004 étaient distribués de façon semblable à la distribution de tous les puits enregistrés en Ontario. Les 354 échantillons soumis au projet ARO et sélectionnés pour cette étude-ci semblaient spatialement distribués de la même manière que les échantillons révélés contaminés à *E. coli* par les programmes de surveillance de la qualité de l'eau durant l'année d'étude 2004. D'ailleurs, les régions sans échantillon dans ce projet de maîtrise sont généralement des régions présentant peu de puits contaminés. Ceci pourrait

être expliqué par le fait que l'échantillonnage a été généré aléatoirement à partir des échantillons soumis aux programmes de surveillance. Ainsi, la population de référence pourrait être étendue aux régions rurales connaissant de forts problèmes de contamination de ses puits. Des raisons expliquant pourquoi certaines régions semblent sous représentées seront mentionnées plus tard, lorsque le sujet de biais causé par un échantillonnage par commodité sera abordé. Il est aussi à noter que d'un point de vue «phénomène antibiorésistance», où on ne considère que les puits contaminés par *E. coli*, il n'y a pas lieu de croire que les puits soumis et échantillonnés se comporteraient différemment des autres puits, contrairement à si l'on considérait plutôt la problématique du «phénomène contamination», où les puits contaminés et non contaminés sont considérés. En d'autres mots, le mécanisme microbiologique du transfert de la résistance à *E. coli* devrait avoir lieu avant la contamination de l'eau comme telle. Ainsi, les résultats de cette étude pourraient potentiellement être applicable à l'ensemble du sud de l'Ontario où de la contamination par *E. coli* (résistant ou non) est présente. Nous tenons à mentionner que cette affirmation est faite sous toute réserve, car en comparant les SRU incluses et les SRU exclues de l'étude, nous avons déterminé une différence statistiquement significative dans la moyenne de densité de ruminants autres que les bovins entre le groupe des SRU incluses et le groupes des SRU exclues de l'étude. Ceci en est de même pour la proportion moyenne de superficie utilisée pour l'application de fumier solide. Les SRU des deux groupes diffèrent donc dans ces deux domaines. Par contre, elles paraissent semblables pour les autres variables agricoles, ce qui nous procure un peu plus de confiance sur l'hypothèse que, dans l'ensemble, les régions comprises dans l'étude et exclues de l'étude ne sont pas trop différentes.

Globalement, les résultats démontrent une présence non négligeable de résistance antimicrobienne dans les puits privée servant d'eau potable en Ontario. Dans l'ensemble de la région d'étude, plusieurs puits ont démontré un certain niveau de résistance antimicrobienne. Cependant, il y avait plus de puits contaminés par *E. coli* résistant à seulement un ou deux antibiotiques qu'il n'y en avait à trois antibiotiques et plus. Cette présence de résistance antimicrobienne est à considérer car la flore intestinale des gens et des animaux consommant de l'eau contaminée pourrait être colonisée par les bactéries

commensales résistantes aux antibiotiques présentes. Par contre, nous sommes conscients qu'il faudrait aussi pousser la réflexion plus loin, à savoir si ces *E. coli* commensaux résistants à ces antibiotiques (surtout tétracycline, sulfaméthoxazole et streptomycine) et colonisant la flore intestinale humaine sont réellement menaçants pour la santé humaine. Nous n'avons pas trouvé de réponse à ce sujet jusqu'à présent. Il est à noter que très peu de puits était contaminé par *E. coli* résistant aux patrons spécifiques à la médecine vétérinaire et à la médecine humaine. Pour cette raison, les efforts d'analyses ont plutôt été concentrés sur les patrons de résistance définis par un nombre déterminé d'antibiotiques. Le niveau régional de résistance antimicrobienne en général, peu importe le patron antimicrobien donné, a semblé plus faible dans le centre de la région d'étude. À l'opposé, certaines zones ont semblé démontrer des proportions plus élevées de résistance antimicrobienne. Il serait possible qu'il y ait une association entre le niveau de résistance antimicrobienne et des facteurs d'exposition définis au niveau régional. C'est ce que nous avons tenté d'explorer.

Bien qu'il y ait eu peu de puits positif au patron vétérinaire, un très grand nombre de puits était contaminé par *E. coli* résistant à des antibiotiques utilisés de façon considérable chez les animaux. À l'inverse, il y avait très peu de puits contaminé par *E. coli* résistant aux fluoroquinolones (acide nalidixique et ciprofloxacine), une classe d'antibiotiques fréquemment prescrite aux humains [16] et qui n'est aucunement homologuée chez les animaux de consommation [6]. Les trois antibiotiques auxquels le plus grand nombre de puits était positif étaient la tétracycline (42.7%), le sulfaméthoxazole (27.4%) et la streptomycine (24.9%). Des molécules de la classe des tétracyclines sont à la fois homologuées chez certains animaux de consommation et chez les humains [6]. Les tétracyclines orales sont un peu plus souvent prescrits aux humains par les médecins que les deux autres antibiotiques [16]. Le sulfaméthoxazole et la streptomycine sont homologués chez les humains mais d'autres molécules de la classe des sulfamides et des aminoglycosides sont aussi homologués chez les animaux de consommation [6]. Il est à noter que, bien que ces molécules soient homologuées chez les humains, elles sont faiblement utilisées en médecine humaine [16]. Les tétracyclines et les sulfamides figurent parmi les antibiotiques les plus fréquemment utilisés en médecine et production porcines [10]. Des tétracyclines sont d'ailleurs homologuées comme promoteur de croissance chez

les porcs, mais aussi chez les poulets, les dindons et les moutons [6]. Dans une étude de Schroeder et al. sur la résistance antimicrobienne d'*E. coli* O157 isolé à partir d'animaux de consommation, de viandes et d'humains [51], les prévalences de résistance antimicrobienne les plus élevées provenaient des porcs. Les deux antibiotiques auxquels il y avait le plus d'isolats résistants provenant de porcs étaient le sulfaméthoxazole et la tétracycline. Les auteurs suggèrent que l'usage de ces antibiotiques chez les porcs pourrait être un facteur dans l'émergence de la résistance antimicrobienne d'*E. coli* O157. Il est à noter que de la résistance à ces antibiotiques avait également été retrouvée parmi les *E. coli* O157 provenant d'autres espèces, dont les humains et les bovins, et que ces deux antibiotiques étaient aussi homologués chez les bovins. Dans une autre étude semblable [52], les fréquences de phénotypes résistants aux antimicrobiens les plus élevées concernaient les isolats d'*E. coli* provenant des dindes et des humains. Cinquante-neuf pourcent et 84% des isolats provenant respectivement des humains et des dindes étaient résistants au sulfaméthoxazole, 59% et 82% à la streptomycine, 56% et 49% à l'ampicilline et 56% et 71% à la tétracycline. Les profils de résistance étaient semblables entre les isolats provenant de bovins, poulets et porcs. Concernant les isolats provenant des porcs, 81% était résistants à la tétracycline, 62% à la streptomycine et 31% au sulfaméthoxazole. Tel que suggéré par les études mentionnées ci-haut, la résistance aux tétracyclines, au sulfaméthoxazole et à la streptomycine retrouvée dans les puits pourrait être liée aux animaux de consommation présents dans la région. En effet, ces antibiotiques ou autres de mêmes classes sont fréquemment utilisés chez diverses espèces animales, dont les porcs.

Cette possibilité d'association entre les animaux de ferme, plus particulièrement les porcs, et la résistance antimicrobienne présente dans les puits pourrait être supportée par les analyses de régressions univariées et multiples effectuées. Selon les résultats obtenus à partir des régressions logistiques univariées concernant le patron de résistance à au moins un antibiotique, plusieurs facteurs agricoles démontrent une association non conditionnelle significative avec le niveau régional de résistance. Les facteurs démontrant les mesures d'association les plus élevées sont des facteurs en lien avec la production porcine.

Il est possible de constater que les diverses activités agricoles représentées par les variables indépendantes définies n'étaient pas distribuées spatialement de façon homogène dans la région d'étude. L'ouest de la région d'étude semblait être plus agricole que l'est, avec des indices de production bovine et/ou porcine plus importantes. Les cartes choroplèthes ont permis d'explorer la dimension spatiale des résultats. En effet, les régions présentant une proportion élevée des puits échantillonnés contaminés par *E. coli* résistant à au moins un antibiotique étaient surtout situées dans les régions où la production porcine ou la présence de lisier sur les terres agricoles semblaient être des facteurs importants.

Les modèles de régression multiples évalués dans cette présente étude ont permis d'explorer l'influence de certains facteurs agro-environnementaux sur la présence de résistance antimicrobienne dans les puits contaminés par *E. coli* au niveau régional. Bien que nous étions intéressés à étudier différentes mesures régionales du niveau d'antibiorésistance, seule la mesure générale de celle-ci, sous la forme de «présence de résistance à au moins un antibiotique (patron 1)», a donné des résultats probants. Pour cette raison, les analyses et les interprétations ont été concentrées sur cette définition. Tous les modèles de régression multiple dont l'ajustement était jugé satisfaisant (modèles A, B et C) sont arrivés sensiblement à la même conclusion. Aucun modèle n'a semblé supérieur à l'autre en terme d'ajustement (sauf le modèle C qui semblait un peu inférieur aux deux autres), de prédictibilité et de facilité d'interprétation. Pour chacun de ces modèles, une variable définissant une certaine influence liée à la production porcine et une variable de précipitation totale annuelle a démontré une association significative avec la proportion régionale de puits positifs. Il n'a pas semblé y avoir d'interaction entre les deux variables. Le modèle A a présenté l'estimé le plus petit de la variance du paramètre du facteur agricole et l'intervalle de confiance du rapport de cotes le plus restreint comparativement aux deux autres modèles. Les modèles B et C ont présenté les intervalles de confiance du rapport de cotes les plus larges (environ 1 à 14) limitant ainsi un peu plus les conclusions que l'on pouvait en retirer. De plus, le paramètre du facteur agricole présentant la variance la plus élevée s'est retrouvé dans le modèle C. L'estimé de variance des paramètres météorologiques et l'étendu des intervalles de confiance des rapports de cotes concernant ces paramètres ont été cohérents d'un modèle à l'autre.

Comme mentionné ci-haut, chacun des modèles jugé satisfaisant a présenté une variable agricole et une variable de précipitation totale annuelle significatives. Décortiquons maintenant ces modèles en commençant par l'interprétation possible de la variable d'agriculture. En fait, chacun des trois modèles retenus proposait une variable agricole différente mais toujours liée à la production porcine comme facteur de risque pour la présence d'*E. coli* résistant à au moins un antibiotique dans les puits contaminés. Ainsi, selon les modèles obtenus, il serait possible que la proportion de puits contaminés par *E. coli* résistant ait plus de risque d'augmenter dans une région possédant une caractéristique agricole d'origine porcine importante. Cette caractéristique pourrait être exprimée par un épandage important de fumier liquide en surface ou une population porcine importante dans une région. Bien que le lisier puisse provenir d'autres espèces animales (bovins, volaille, ...), le fumier liquide provient généralement en grande partie des porcs. Ainsi, une explication possible des associations positives retrouvées pourrait être la présence de bactéries résistantes aux antibiotiques qui se retrouvent dans le lisier de porc et qui contaminent l'eau de puits pour une raison donnée [20]. La variable représentant l'épandage de fumier liquide en surface pourrait être considérée comme un indice de l'ampleur de l'épandage de lisier de porc dans une région. Une variable considérant plutôt la population porcine pourrait être considérée comme un indicateur de la pression de la production porcine dans la région. La contamination possible des drains et eaux de surface par les bactéries contenues dans le lisier épandu est une préoccupation reconnue, particulièrement en Ontario [53]. Suite au ruissellement, les bactéries, dont *E. coli* résistant et non résistant aux antibiotiques, pourraient contaminer les puits. La présence de bactéries commensales résistantes aux antibiotiques dans le lisier pourrait être expliquée par la pression de sélection exercée par l'utilisation d'antibiotiques dans la production porcine. Nos résultats corroborent ceux de Chee-Sanford et al. [20]. Par contre, leur étude portait non pas sur l'épandage de fumier sur les terres agricoles, mais plutôt sur la présence de gènes de résistance dans les nappes phréatiques et le lien possible avec l'entreposage de lisier dans des lagunes. Des gènes de résistance aux tétracyclines avaient été détectés dans les lagunes (indiquant une sélection pour la résistance à cette drogue) et dans les puits environnants (suggérant une mobilité spatiale des gènes de résistance dans l'environnement à partir des lagunes). Il est à noter que leur étude ne concernait que la résistance aux

tétracyclines et le lisier de porc. Les gènes de la résistance à la tétracycline est l'objet de plusieurs études sur la résistance antimicrobienne retrouvée dans l'environnement et impliquant la production porcine, entre autre parce que l'usage de cet antibiotique est fréquent dans cette industrie [20, 54].

Il y a peu d'études semblables portant sur la résistance antimicrobienne des bactéries dans l'environnement au niveau régional dans la littérature auxquelles les résultats obtenus pourraient être comparés. Par contre, en parcourant la littérature concernant les études faites au sujet de gastro-entérites transmissibles par l'eau, de la contamination de l'eau et de l'évènement historique de Walkerton, on pourrait s'attendre à retrouver des associations avec les mêmes catégories de variables. Contrairement à quelques unes de ces études [4, 21, 22, 49], les modèles de régression multiples obtenus dans notre étude n'ont démontré aucune association significative avec les variables d'influence bovine (intensité de la population bovine ou épandage de fumier bovin). Par contre, les mesures des associations non conditionnelles significatives impliquant une variable reflétant une influence d'origine bovine étaient tout de même statistiquement significatives, bien que plus faibles que celles impliquant l'influence des porcs. Contrairement à l'étude de Valcour et al. [22], la densité porcine d'une région a été associée positivement dans notre étude. Par contre, comme dans leur étude, la proportion de terres agricoles sur laquelle du fumier est appliqué par un épandeur liquide en surface a aussi été associée positivement dans notre étude. Il est à noter qu'il est difficile de comparer nos résultats aux résultats de ces études étant donné que le sujet d'étude n'est pas exactement le même. Ces études concernent la contamination de l'eau ou l'infection humaine par *E. coli* alors que la notre porte plutôt sur la résistance antimicrobienne présente dans l'eau contaminée. Il serait possible que les facteurs de risque expliquant la présence d'*E. coli* résistants aux antibiotiques dans l'eau potable soient simplement différents de ceux expliquant la contamination de l'eau potable. Par contre, on s'attendrait aussi à ce que les résultats obtenus concordent avec ceux du PICRA où les résistances aux tétracycline, aux sulfamides et aux aminoglycosides sont les plus fréquentes dans les échantillons caecaux de bœufs, de poulet et de porcs avec une proportion de résistance respectivement croissante d'une espèce à la suivante [16], mais ceci n'a pas été le cas. Suite aux résultats obtenus par la présente étude et d'après ce que

nous connaissons par la littérature, la conclusion générale que l'on peut tirer est qu'il semble y avoir une association entre des facteurs représentant des activités agricoles et la résistance antimicrobienne retrouvée dans l'eau potable. Certaines associations seraient potentiellement en lien avec la production porcine dans la région exprimée par l'importance de la population porcine ou par l'importance de l'épandage de lisier. Par contre, comme nous le verrons plus tard, ceci est à considérer avec nuance.

En plus de la variable agricole porcine, nous avons aussi constaté une association significative avec la variable de précipitation totale annuelle dans les modèles de régressions multiples. L'interprétation de l'effet météorologique variait selon le niveau de précipitation totale dans une année dans la région. Généralement, d'un modèle à l'autre, les associations selon le niveau de précipitation abondaient dans la même direction et étaient de force comparable. Les mesures d'association des diverses catégories de niveau de précipitation n'étaient pas toujours statistiquement significatives, pas toujours très élevées et semblaient paradoxales d'un niveau à l'autre. Cette instabilité étant fréquente, l'interprétation de l'effet de ces résultats est difficile à faire. Cependant, étant donné que cette variable apparaissait toujours statistiquement significative dans son ensemble et que les modèles donnaient toujours des mesures d'association cohérents les uns entre les autres, il est tout de même justifié, ou du moins intéressant, de tenter d'interpréter la possibilité du niveau annuel total régional de précipitation comme facteur de risque ou facteur protecteur, tout en gardant en tête que ce ne sont là que des hypothèses. Ainsi, les résultats ont démontré que, comparativement à un niveau faible, un niveau annuel moyen de précipitation dans une région semblerait avoir un effet protecteur sur le niveau régional de résistance antimicrobienne. Par contre, il est à noter que l'estimé de la mesure d'association n'était pas toujours statistiquement significatif. À l'opposé, une précipitation totale annuelle considérée élevée dans une région était associée positivement au niveau de résistance régionale. De plus, dans chacun des modèles, cette association était toujours statistiquement significative. Une région connaissant une précipitation annuelle considérée très élevée était aussi associée positivement au niveau régional de résistance antimicrobienne. La force d'association était par contre moins élevée qu'avec la catégorie précédente et était non statistiquement significative. L'interprétation biologique de cette variable est moins

intuitive que celle développée avec la variable agricole. En effet, on s'attendrait plutôt à ce que plus la précipitation régionale annuelle est importante, plus son association positive avec le niveau régional de résistance antimicrobienne soit forte. Une interprétation hypothétique telle que suggérée plutôt par les résultats obtenus pourrait être qu'une région connaissant une précipitation totale annuelle moyenne ait un risque plus faible qu'une région connaissant une précipitation totale annuelle faible car le sol de cette région posséderait peut-être plus de fissures étant donné une certaine sécheresse pendant l'année. Ces fissures rendraient le sol plus perméable en permettant plus facilement une percolation de l'eau vers les nappes phréatiques, ce qui faciliterait la contamination de l'eau potable par le lisier de porcs contenant des *E. coli* résistants aux antibiotiques. Ainsi, une région ayant connu une précipitation totale annuelle moyenne posséderait un sol moins fissuré, et serait donc un peu plus imperméable à la contamination rapide de l'eau, ce qui pourrait expliquer l'aspect protecteur détecté dans les régions à précipitation totale annuelle moyenne. Ensuite, lorsque la précipitation totale dans une année est considérée élevée, il y aurait potentiellement un excès d'eau par rapport à la capacité de filtration, augmentant ainsi le risque de contamination des nappes phréatiques. En effet, les précipitations de pluie pourraient mobiliser les bactéries retenues dans le sol et permettre leur déplacement vers les nappes phréatiques [55]. De plus, *E. coli* survit plus longtemps dans un sol humide [55]. Le sol généralement moins humide dans une région ayant connu une précipitation moins importante n'offre pas les conditions optimales pour la survie de la bactérie. Cependant, une précipitation totale annuelle plus importante considérée très élevée sur l'ensemble de l'année, bien que contribuant aussi à la contamination des puits, résulterait potentiellement en un effet de dilution des contaminants, expliquant ainsi pourquoi le niveau de résistance de la région serait plus faible ou peut-être même nul. Ce type de raisonnement a d'ailleurs déjà été soulevé dans une étude pour expliquer la diminution de la concentration de coliforme et de streptocoques dans des ruisseaux durant le mois d'avril où le ruissellement était important et entraînait ainsi une dilution des bactéries [56]. Concernant notre étude, il se pourrait que le ruissellement de microorganismes à partir des eaux de surface contaminées et des terres sur lequel du fumier contaminé était épandu se retrouve dilué par l'accumulation extrême des précipitations au courant de l'année ou par des précipitations extrêmes spécifiquement à certaines périodes de l'année. De plus, il a déjà été décrit dans la

littérature que les coliformes fécaux pénètrent plus profondément dans le sol et atteignent les nappes phréatiques en quantité plus importante tout de suite après le début d'une inondation plutôt qu'après une inondation ayant perduré pendant une certaine période de temps [55]. Une précipitation définie extrême sur une période d'un an pourrait être due à une accumulation d'eau à long terme plutôt qu'à une inondation soudaine expliquant ainsi un risque présent mais faible de contamination des puits par *E. coli* résistant aux antibiotiques. En résumé, le niveau annuel de précipitation dans une région semblerait jouer un rôle; une région connaissant une précipitation totale annuelle élevée sans être extrême aurait un risque plus élevé de présenter des puits contaminés par *E. coli* résistant aux antibiotiques que les autres régions. Par contre, l'effet de l'accumulation de précipitation atteindrait un certain maximum, après quoi, l'augmentation de l'accumulation d'eau aurait moins d'effet. Pour les raisons mentionnées plus haut, nous sommes conscients que l'interprétation ainsi faite peut être discutable.

L'interprétation que l'on peut faire des résultats a certaines limites étant donné le devis de l'étude. Tout d'abord, une étude écologique est très statique. Il n'y avait pas assez d'observation répartie dans le temps dans chaque région afin que l'on puisse analyser la dimension temporelle. Ainsi, cette étude essaie de trouver une relation entre une mesure ponctuelle d'un phénomène qui pourrait être temporellement dynamique avec des variables loco-régionales mesurées sur une échelle de temps annuel. Il est généralement entendu que le phénomène de contamination des eaux est un phénomène temporellement dynamique. Cela pourrait être de même pour le phénomène de la résistance antimicrobienne retrouvée dans l'eau. Par contre, nous ne savons pas jusqu'à quel point. Le phénomène de la présence de la résistance antimicrobienne dans l'eau potable n'agit peut-être pas temporellement comme le phénomène de la contamination de l'eau. En d'autres mots, la présente étude tient compte d'*E. coli* qui est déjà présent dans l'eau, du fait qu'il est résistant ou non aux antibiotiques et non de la présence/absence de *E. coli* dans l'eau. Dans un tel contexte, nous ne pouvons dire à quel point l'aspect temporel est important à considérer.

Le devis de cette étude est dit écologique. Une étude écologique concerne une investigation impliquant des agrégats d'individus comme unité d'analyse alors que l'unité visée est l'individu [57, 58]. La caractéristique primaire d'une telle étude est que l'on ne connaît pas la distribution commune des variables indépendantes et des variables dépendantes dans chaque groupe [57, 59]. Ce type d'étude comporte des avantages. Dans le cas spécifique de la présente étude, il a permis d'utiliser des sources de données déjà existantes (Statistique Canada) et est moins dispendieux que ceux utilisant l'individu comme unité d'analyse [58]. Par contre, l'inconvénient majeur de cette étude est que les résultats sont plus sujets à des biais considérables parce que l'on doit assumer que ce qui est vrai au niveau du groupe est aussi vrai au niveau de l'individu puisque les données à cette échelle sont inconnues. Cette supposition peut malheureusement être fausse [58]. Ici, les données des variables indépendantes sont connues à l'échelle régionale, mais nous ne connaissons aucunement les données au niveau du puits même. Nous reviendrons plus en détail sur les biais écologiques auxquels notre étude a été exposée.

Un des points faibles de l'étude est sa puissance limitée étant donné la taille instable et parfois très faible d'échantillons dans chaque région. De plus, il y avait très peu de puits pris en compte dans cette étude comparativement au nombre total de puits dans le sud de l'Ontario. Afin de soutirer toute la puissance possible des données à notre disposition, tous les échantillons disponibles dans une région ont été considérés, justifiant ainsi le choix d'une approche binomiale avec le modèle de régression logistique sur des données binaires groupées. Une puissance d'étude réduite limiterait la détection des associations potentielles et augmenterait l'incertitude des mesures d'association. Ainsi, étant donné la faible puissance de cette étude et la stratégie de sélection des variables qui est de type automatique, les résultats mathématiques que l'on a obtenus pourraient ne pas représenter la réalité biologique. Pour cette raison, il serait possible que le facteur porcin ressorte statistiquement toujours fortement significatif, excluant alors les autres facteurs. Il est aussi à noter que, selon le test d'Aspin-Welch, la moyenne des proportions de terres utilisées pour l'épandage de fumier solide est plus faible dans les SRU incluses dans l'étude que dans les SRU exclues de l'étude. Ceci pourrait aussi expliquer pourquoi le facteur bovin ne ressort pas assez fort dans nos analyses et que l'association statistique que nous détectons

toujours est celle avec un facteur porcin. Ainsi, il faut interpréter les résultats obtenus avec nuance. Certes, il y a probablement un facteur agricole qui influence le niveau régional de résistance antimicrobienne et cette étude semble indiquer une association statistiquement positive avec le facteur porcin. Par contre, l'association biologique positive avec le facteur porcin est plutôt incertain. Ainsi, le fait de trouver une association positive avec le facteur porcin pourrait tout simplement dire qu'il existe une association avec le milieu agricole.

Il est à noter qu'avec les régressions univariées, aucune variable n'a démontré d'association significative avec les patrons humain et vétérinaire si l'on fixe le seuil alpha à 0.05. Ceci pourrait être dû à un mauvais choix de modèle de régression ou à une mauvaise définition des patrons. Il est à noter qu'il a été difficile de définir des patrons d'antibiotiques à usage exclusif aux humains et exclusif aux animaux tout en considérant la possibilité des résistances croisées. Avec l'aide de plusieurs personnes ressources, dont des microbiologistes, nous avons tenté de définir les meilleurs, bien qu'imparfait, sous-ensembles possibles d'antibiotiques qui représenteraient les patrons humain et vétérinaire à partir des antibiotiques testés dans le cadre du projet ARO. Dans le cas d'évènements rares comme c'est le cas des puits positifs aux patrons spécifiques aux humains et spécifiques aux animaux, un autre modèle de régression aurait aussi peut-être été plus justifié et adapté que le modèle de régression logistique. Avec les données disponibles et la puissance de l'étude, il est probablement impossible de détecter certaines associations si elles existent vraiment. Ce raisonnement pourrait aussi s'appliquer au patron défini par un ou deux antibiotiques et trois antibiotiques et plus.

Comme toute étude épidémiologique, cette étude a été soumise à certains biais, dont le biais d'information. Un biais d'information peut provenir d'une mauvaise classification des observations. Les biais de mauvaise classification présents dans cette étude concernent surtout la classification des régions selon l'exposition. Ce type de biais est à considérer entre autre pour les variables météorologiques. En effet, certaines régions ne possédaient aucune station météorologique. La valeur des variables météorologiques de ces régions était celle de la station la plus proche d'une des frontières de la région concernée. Cette valeur

pouvait donc être non représentative de la vraie valeur pour la région en question. Ceci est probablement le cas dans les régions moins peuplées qui se retrouvent aussi probablement à être les régions où il y a eu moins d'échantillon de puits. Les biais de mauvaise classification des variables météorologiques pourraient expliquer pourquoi les résultats obtenus concernant la variable de précipitation annuelle totale sont difficiles à interpréter. Une autre raison pouvant expliquer le paradoxe dans les résultats météorologiques obtenus pourrait être le mauvais choix de variable (précipitation totale annuelle) pour représenter le facteur météorologique de la problématique. En fait, cette variable comprenait non seulement les précipitations de pluie, mais aussi celles de la neige en hiver. Il serait possible que les régions ayant connu une précipitation totale annuelle très élevée aient plutôt connu une précipitation totale annuelle de neige très élevée et une précipitation annuelle de pluie faible. Une variable de choix aurait été le nombre de jours avec des précipitations de pluie extrêmes. Celle-ci ne nous était malheureusement pas accessible. La seule variable semblable en notre possession était le nombre de jours avec 1.0 mm ou plus de précipitation, une quantité insuffisante pour être qualifiée d'extrême. Dans la même optique, les observations concernant la variable de drainage pourraient aussi avoir été victimes de mauvaise classification. En effet, un puits pourrait être situé dans un sol tout autre que la donnée attribuée à la région selon notre règle proposée a priori. Cette règle définissait que la donnée de drainage attribuée à une SRU était la catégorie de drainage prédominante dans cette SRU. Par contre, plusieurs régions n'étaient composées que d'un seul type de drainage ou avaient un type de drainage largement prédominant sur l'autre. L'impact d'une mauvaise classification de la variable pédologique est probablement minime.

Un autre biais d'information potentiel à considérer provient de la source des données des variables indépendantes. Les variables d'agriculture et socio-économiques provenaient des recensements de l'année 2001, soit 3 ans avant le début de la période concernée par cette étude. On pourrait s'attendre à ce que le portrait agricole et celui de la population aient changé en trois ans. Par contre, la catégorisation des variables employées avait pour but de mettre l'emphasis sur les extrêmes. Ainsi, une région jugée

extrême pour une variable indépendante donnée en 2001 le serait aussi probablement en 2004, ou du moins devrait probablement être proche de l'être.

Comme mentionné plus haut, les biais écologiques sont à considérer. Il est possible, par exemple, qu'un puits que l'on présumait dans une région à densité porcine élevée était, en réalité, situé dans une région où la production porcine était concentrée dans un élevage intensif considérablement éloigné du puits en question. Ce puits contribuerait ainsi à une possibilité d'inférence erronée des résultats. Pour faire attention à ce phénomène, il est important, premièrement, d'évaluer si les résultats obtenus sont logiques [59]. Concernant notre étude, les associations retrouvées avec les variables d'agriculture l'étaient. Les associations avec la variable de précipitation étaient logiques, mais aussi questionnables pour les raisons mentionnées plus tôt. Deuxièmement, la possibilité d'inférence erronée a lieu surtout lorsque la composition du groupe n'est pas homogène en regard d'un facteur [57, 59]. Par contre, il est généralement considéré que plus une association au niveau du groupe est forte, plus il est probable que l'association au niveau de l'individu soit vraie [58]. Troisièmement, un biais écologique est généralement moindre lorsque la variance de l'exposition à l'intérieure des groupes est petite [57]. Ceci devrait être le cas de notre étude. En effet, l'unité géographique retenue dans le cadre de cette étude est la SRU, l'unité la plus petite que l'on pouvait utiliser avec les données des recensements disponibles. Finalement, le biais écologique est aussi considéré moindre lorsque la variance entre les groupes du facteur d'exposition est grande [57]. Cette étude devrait respecter cette condition puisque la distribution de fréquence de la majorité des variables indépendantes était fortement asymétrique et la catégorisation de ces variables était développée selon cette caractéristique afin de mettre l'emphase sur les extrêmes. En sommes, la faible taille d'échantillonnage de certains groupes et la possibilité de mauvaise classification de l'exposition des puits pourraient facilement mener à une fausse inférence, d'autant plus que les rapports de cotes obtenus ne sont pas extraordinairement élevés. L'interprétation des résultats de cette étude doit donc se faire avec vigilance et demeurer conservatrice. Les résultats semblent suggérer que les régions connaissant de fortes activités agricoles à caractère porcin ont plus de probabilité qu'une proportion élevée de leurs puits contaminé par *E. coli* le soit par *E. coli* résistant à au moins un antibiotique.

Les biais de mauvaise classification et les biais écologiques sont les deux biais majeurs à considérer dans cette étude. À ceux-ci s'ajoutent les biais de sélection. Cette étude est basée sur un échantillonnage par commodité. Les données provenaient d'un programme déjà établi de surveillance passive de la contamination de l'eau de puits. Parce que les puits n'avaient pas tous la même chance d'être soumis au programme pour diverses raisons, l'échantillonnage n'était pas nécessairement représentatif de la population de puits du sud de l'Ontario. Il serait possible que les associations retrouvées dans la population d'étude diffèrent de celle de la population cible. Par contre, il faut rappeler que cette étude évalue l'association possible de facteurs agro-environnementaux avec la résistance antimicrobienne et non avec la contamination de l'eau. Un échantillonnage par commodité de la sorte serait plus questionnable dans une étude concernant la contamination de l'eau ou la prévalence de résistance antimicrobienne. Ainsi, bien que la population d'étude n'ait pas été nécessairement représentative de la population cible, les conclusions retirées de l'exploration des associations entre les facteurs d'exposition et le niveau régional de résistance antimicrobienne d'*E. coli* dans les puits privés demeurent des indicateurs à considérer. Autrement dit, pour une étude dont le but était d'explorer les associations possible entre le niveau de résistance antimicrobienne régionale et des facteurs agro-environnementaux, un échantillonnage aléatoire provenant d'un programme de surveillance volontaire de la qualité de l'eau représente de façon satisfaisante la population cible de l'étude. Malheureusement, une autre composante de la sélection des puits pourrait aussi contribuer au biais de sélection causé par l'étape de présélection des puits possédant des bactéries potentiellement résistantes à un ou des antibiotiques. Les puits contaminés par une bactérie où aucune résistance n'était retrouvée lors de cette étape se sont retrouvés exclus du projet. Les puits provenant des régions possédant moins de puits contaminés par *E. coli* résistant aux antibiotiques avaient alors moins de chance d'être échantillonnés pour notre étude puisque les échantillons négatifs étaient systématiquement exclus du bassin d'échantillons disponibles pour l'étude. Cependant, si on assume que cette présélection d'échantillons contaminés par des bactéries sensibles dans les régions exposées et les régions non exposées sont proportionnellement les mêmes que lors du test final, i.e. la méthode de microdilution en bouillon, cette étape de présélection a probablement une répercussion minimale. Le but de l'étude était de trouver des associations. Cette étape aurait

été plus problématique si l'on avait effectué une étude sur la prévalence de puits contaminés par *E. coli* résistant par exemple.

Le changement de seuils de CMI pour détecter des échantillons présumés résistants lors de l'étape de présélection des échantillons occasionnerait probablement une erreur sur la mesure. En augmentant la CMI de certains antibiotiques au fil du temps, nous avons diminué le nombre d'échantillons possiblement sensibles à tous les antibiotiques de la présente étude puisqu'ils ont systématiquement été rejetés du projet ARO. Ainsi, la proportion d'échantillons sensibles n'était pas nécessairement stable à travers l'année d'étude. Après consultation auprès de quelques personnes ressources, nous reconnaissons qu'il pourrait y avoir un certain impact, mais nous ne savons pas à quel point. Par contre, on peut supposer que l'impact est minime. En effet, la période d'étude est plutôt restreinte, i.e. du 1<sup>er</sup> mai 2004 au 30 avril 2005. L'amikacine a été retiré du test de présélection seulement vers la fin de l'année d'étude. Le changement du seuil de CMI de la gentamicine a aussi été réalisé vers la fin de l'année d'étude. Ces faits concernant ces deux antibiotiques diminuent probablement l'impact que l'on pourrait envisager. Le seul antibiotique présentant des changements bien répartis sur toute la période d'étude est la céphalothine. Nous ne connaissons malheureusement pas l'impact précis des changements de CMI de cet antibiotique sur les résultats de notre étude.

Étant donné la structure des données, il est possible que les observations ne soient pas nécessairement indépendantes et que des observations partagent des caractéristiques communes non explicitement prises en compte par les variables indépendantes identifiées au départ. C'est ce qu'on appelle des données corrélées (Egret<sup>®</sup>, Cystel Statistical Software, Cambridge, MA). On pourrait s'attendre à une dépendance entre les réponses à l'intérieure d'une même région parce qu'elles partageaient des caractéristiques qui mènent à des réponses plus similaires que d'autres. Ce phénomène d'agglomération pourrait être expliqué par la proximité géographique de certains puits. La conséquence majeure de ce phénomène est que l'erreur-type de l'estimé d'un paramètre soit biaisé, surtout si la taille d'échantillonnage est petite. En fait, la variance est souvent sous-estimée dans ces cas. Une

variance véritablement plus élevée dans l'estimé du paramètre signifierait que l'estimé du paramètre pourrait être plus loin de la vraie valeur, mais ceci n'apparaîtrait pas dans un modèle de régression conventionnel et résulterait en des estimés biaisés [57]. L'exploration de la possibilité de corrélation entre les données a été faite à l'aide des tests statistiques d'ajustement des modèles aux données. La conclusion obtenue de ces tests est que les modèles concernant le patron 1 étaient ajustés de façon satisfaisante aux données. Il semblait, par contre, y avoir un problème avec les modèles des autres patrons. Cependant, cette conclusion d'insatisfaction d'ajustement des modèles obtenue avec ces tests statistiques ne signifie pas nécessairement une corrélation entre les observations, mais plutôt une possibilité de corrélation. Dans la même optique, des analyses d'autocorrélation spatiale auraient été intéressantes à explorer.

Malgré les limites et les biais, cette étude comporte aussi ses forces. L'une d'elles est la robustesse de la conclusion, i.e. qu'elle est constante d'un modèle à l'autre, peu importe l'ensemble de variables considérées au départ (modèles initiaux). De plus, cette étude apporte une nouveauté car il y a très peu d'étude sur le sujet. La population s'inquiète de plus en plus du développement de la résistance antimicrobienne car il reste beaucoup d'inconnus dans cette problématique affectant la santé publique. Les études écologiques peuvent jouer un rôle dans l'investigation des associations, mais elles sont souvent plus utiles comme génératrices d'hypothèses nécessitant une évaluation plus poussée, i.e. une évaluation mesurant l'exposition et la réponse au niveau de l'unité concerné, et ne devraient pas être considérées comme des études finales sur le sujet [58]. Il faut considérer cette étude dans cette ligne de pensée. Le devis de l'étude a été développé de façon à minimiser les sources du biais écologique dans la mesure du possible. Ceci a été fait à l'aide du choix de l'unité géographique pour minimiser la variation intra-groupe et de la construction des variables indépendante pour tenir compte de la variation inter-groupes.

À la lumière de cette étude, certaines suggestions pourraient être faites. Étant donné que les puits contaminés par *E. coli* résistants à la tétracycline sont les plus fréquents dans cette étude-ci, il serait possible que l'association retrouvée entre le niveau régional de la

résistance à un antibiotique ou plus et le facteur porcin soit en fait secondaire à la contamination importante de l'eau par le lisier porcin qui contient beaucoup de bactéries commensales résistantes à la tétracycline. D'ailleurs, dans le rapport 2004 du PICRA [16] concernant la surveillance en abattoir, un pourcentage élevé d'isolats d'*E. coli* générique résistant à la tétracycline provenait des porcs. Par contre, la proportion chez les poulets à griller était aussi considérable. Il serait intéressant d'examiner l'association entre la contamination de l'eau et les mêmes variables étudiées avec des données de la même période, ainsi que de répéter le même exercice avec un patron de résistance défini par la résistance à la tétracycline si les données le permettaient. Il serait également pertinent d'utiliser les données de l'année d'étude 2005 du projet ARO pour valider les modèles que nous avons obtenus. D'autre part, l'ensemble des antibiotiques testés pourrait être étendu à des promoteurs de croissance. En effet, la galerie d'antibiotiques proposés (NARMS) était principalement composée d'antibiotiques intéressants pour un usage en médecine humaine. Un ensemble plus large d'antibiotiques testés aurait permis entre autres de définir un patron plus spécifique à la médecine vétérinaire et à l'agriculture. Afin d'inclure les promoteurs de croissance dans une telle étude, il faudrait ajouter une bactérie indicatrice gram-positif, tel qu'*Enterococcus* spp par exemple. Il aurait aussi été intéressant d'investiguer des patrons d'antibiotiques spécifiques à certaines espèces animales, incluant l'espèce porcine.

La problématique de la résistance antimicrobienne est complexe. Les humains et les animaux sont inévitablement interreliés. Il est donc difficile d'identifier une cause claire de la résistance antimicrobienne retrouvée dans l'environnement et d'identifier la part de responsabilité de l'usage des antibiotiques chez les animaux de consommation à cette problématique. Étant donné la complexité des possibilités de transmission de résistance antimicrobienne et de développement de résistance croisée, il nous était impossible d'identifier la ou les sources d'*E. coli* résistant aux antibiotiques retrouvé dans l'eau potable. Par contre, cette étude a permis d'identifier les facteurs agricoles comme un facteur de risque d'émergence de résistance antimicrobienne d'*E. coli* contaminant l'eau potable. De plus, les résultats des autres volets du projet ARO permettront probablement de répondre à diverses interrogations de la problématique complexe du développement de la résistance antimicrobienne dans l'environnement et de son impact sur la santé publique.

Malgré les limites soulevées, cette étude justifierait une surveillance de la résistance antimicrobienne présente dans l'environnement compte tenu de la détection constante d'une association entre la présence de résistance antimicrobienne dans les puits d'une région et des variables agricoles. La résistance antimicrobienne présente dans l'environnement représente une problématique dont l'impact sur la santé publique est encore peu compris, mais elle mériterait un effort aussi considérable que celui accordé jusqu'à présent dans la surveillance des résidus antimicrobiens dans la viande. À la lueur des résultats obtenus, une surveillance des antibiotiques utilisés en médecine vétérinaire en Ontario, incluant la médecine et production porcine, serait justifiée. Pour cela, la prescription vétérinaire devrait être obligatoire et contrôlée. Même si aucun antibiotique exclusivement promoteur de croissance n'a été évalué dans cette étude, à la lumière des différentes découvertes, des faits répertoriés dans la littérature et des résultats obtenus dans cette étude, par principe de précaution, un programme canadien d'abolition graduelle des promoteurs de croissance pourrait aussi être recommandé, tel que cela s'est déjà fait dans plusieurs pays d'Europe.

## Conclusions

*E. coli* est une bactérie commensale indicatrice fréquemment utilisée pour documenter la contamination fécale de l'eau. Pour plusieurs raisons, elle est aussi souvent utilisée pour étudier la résistance antimicrobienne chez les humains, chez les animaux et dans l'environnement.

L'objectif était de déterminer l'association possible entre divers facteurs agro-environnementaux et le niveau de résistance de la bactérie indicatrice *E. coli* à divers patrons d'antibiotiques dans les puits privés à un niveau régional. Cet objectif a été rencontré, à un certain degré, pour le patron 1 (résistance à un antibiotique ou plus). Des associations positives statistiquement significatives avec certains facteurs agro-environnementaux ont été détectées. Des analyses avec des modèles de régression mieux adaptés aux patrons 2 et 3 permettraient probablement de retirer des conclusions pertinentes.

Les principaux résultats de la présente étude suggèrent que le niveau régional de la résistance antimicrobienne (à un antibiotique ou plus) d'*E. coli* dans l'eau des puits privés des régions sur laquelle se base cette étude est positivement associé à des activités agricoles. Les régressions multiples suggèrent plus spécifiquement des facteurs représentant la production porcine. Les variables indicatrices de l'importante influence de la production porcine dans une région étaient la densité porcine, le rapport entre la population porcine et humaine et l'importance de l'épandage du fumier liquide en surface sur les terres agricoles dans la région. Par contre, la faible puissance de l'étude peut mettre en doute l'association avec les facteurs porcins faisant en sorte que l'association retrouvée devrait possiblement être élargie au facteur agricole en général. Comme dans la problématique de contamination des eaux, la précipitation totale annuelle semble aussi être un facteur à considérer. Son impact change selon la quantité annuelle de précipitation. À ce sujet également, nous avons émis une interprétation hypothétique qui peut être discutable pour certains.

En conclusion, cette étude présente des pièces complémentaires à la compréhension de la problématique complexe qu'est la résistance antimicrobienne retrouvée dans l'environnement et l'implication de l'agriculture dans le développement de celle-ci. Par contre, étant donné les limites posées par le devis d'étude, elle ne permet pas de faire la lumière sur l'impact réel de l'agriculture sur la santé publique. Des résultats futurs du Projet ARO permettront probablement d'éclairer d'avantage le sujet de l'impact de la présence d'*E. coli* résistant à des antibiotiques sur la santé de la population humaine dans les régions à risque, comme les régions identifiées dans cette étude.

## Bibliographie

1. Charron, D., et al., *Vulnerability of waterborne diseases to climate change in Canada: a review*. J Toxicol Environ Health A, 2004. **67**(20-22): p. 1667-77.
2. Schuster, C.J., et al., *Infectious disease outbreaks related to drinking water in Canada, 1974-2001*. Can J Public Health, 2005. **96**(4): p. 254-8.
3. McQuigge, M., *The Walkerton disaster and family physicians*. Can Fam Physician, 2002. **48**: p. 1596-7, 1605-7.
4. *Waterborne outbreak of gastroenteritis associated with a contaminated municipal water supply, Walkerton, Ontario, May-June 2000*. Can Commun Dis Rep, 2000. **26**(20): p. 170-3.
5. Acar, J. and B. Rostel, *Antimicrobial resistance: an overview*. Rev Sci Tech, 2001. **20**(3): p. 797-810.
6. Comité consultatif sur l'utilisation d'antimicrobiens chez les animaux et les conséquences pour la résistance et la santé humaine, *L'utilisation au Canada d'antimicrobiens chez les animaux destinés à l'alimentation : les conséquences pour la résistance et la santé humaine*. 2002.
7. Organisation mondiale de la santé. *L'utilisation des antimicrobiens en dehors de la médecine humaine et les résistances qui en résultent chez l'homme*. 2002 [page Internet consultée en janvier 2007]; URL: <http://www.who.int/mediacentre/factsheets/fs268/fr/index.html>.
8. Khachatourians, G.G., *Agricultural use of antibiotics and the evolution and transfer of antibiotic-resistant bacteria*. Cmaj, 1998. **159**(9): p. 1129-36.
9. Kruse, H. and H. Sorum, *Transfer of multiple drug resistance plasmids between bacteria of diverse origins in natural microenvironments*. Appl Environ Microbiol, 1994. **60**(11): p. 4015-21.
10. McEwen, S.A. and P.J. Fedorka-Cray, *Antimicrobial use and resistance in animals*. Clin Infect Dis, 2002. **34 Suppl 3**: p. S93-S106.
11. Gouvernement de l'Ontario. 2006 [page Internet consultée en novembre 2006]; URL: [www.ontario.ca](http://www.ontario.ca).
12. Novokowski, K., et al., *Water well sustainability in Ontario*. 2006.
13. Canadian Committee on Antimicrobial Resistance. *Antimicrobial Resistance: A Deadly Burden No Country Can Afford to Ignore*. 2002 [page Internet consultée en Février 2007]; URL: <http://www.ccar-ccra.com/french/word/FinalReport2.doc>.
14. Tenover, F.C. and J.E. McGowan, Jr., *Reasons for the emergence of antibiotic resistance*. Am J Med Sci, 1996. **311**(1): p. 9-16.
15. Witte, W., *Medical consequences of antibiotic use in agriculture*. Science, 1998. **279**(5353): p. 996-7.
16. Gouvernement du Canada. Programme intégré canadien de surveillance de la résistance aux antimicrobiens, *Programme intégré canadien de surveillance de la résistance aux antimicrobiens (PICRA) 2004*. Guelph (Ontario): Agence de la santé publique du Canada, 2006

17. National Antimicrobial Resistance Monitoring System-Enteric Bacteria (NARMS), *2003 Executive Report*. Rockville, MD:U.S. Department of Health and Human Services, U.S. FDS, 2006.
18. Danmap 2006, *Use of antimicrobial agents and occurrence of antimicrobial resistance in bacteria from food animals, foods and humans in Denmark*. ISSN 1600-2032.
19. Kummerer, K., *Significance of antibiotics in the environment*. J Antimicrob Chemother, 2003. **52**(1): p. 5-7.
20. Chee-Sanford, J.C., et al., *Occurrence and diversity of tetracycline resistance genes in lagoons and groundwater underlying two swine production facilities*. Appl Environ Microbiol, 2001. **67**(4): p. 1494-502.
21. Michel, P., et al., *Temporal and geographical distributions of reported cases of Escherichia coli O157:H7 infection in Ontario*. Epidemiol Infect, 1999. **122**(2): p. 193-200.
22. Valcour, J.E., et al., *Associations between indicators of livestock farming intensity and incidence of human Shiga toxin-producing Escherichia coli infection*. Emerg Infect Dis, 2002. **8**(3): p. 252-7.
23. Kaspar, C.W., et al., *Antibiotic resistance indexing of Escherichia coli to identify sources of fecal contamination in water*. Can J Microbiol, 1990. **36**(12): p. 891-4.
24. McKeon, D.M., J.P. Calabrese, and G.K. Bissonnette, *Antibiotic resistant gram-negative bacteria in rural groundwater supplies*. Water Res, 1995. **29**(8): p. 1902-8.
25. Rodgers, P., et al., *Spatial and temporal bacterial quality of a lowland agricultural stream in northeast Scotland*. Sci Total Environ, 2003. **314-316**: p. 289-302.
26. Auld, H., D. MacIver, and J. Klaassen, *Heavy rainfall and waterborne disease outbreaks: the Walkerton example*. J Toxicol Environ Health A, 2004. **67**(20-22): p. 1879-87.
27. Smith, D.L., et al., *Animal antibiotic use has an early but important impact on the emergence of antibiotic resistance in human commensal bacteria*. Proc Natl Acad Sci U S A, 2002. **99**(9): p. 6434-9.
28. Threlfall, E.J., et al., *Increasing spectrum of resistance in multiresistant Salmonella typhimurium*. Lancet, 1996. **347**(9007): p. 1053-4.
29. Kim, H.H., et al., *Characteristics of antibiotic-resistant Escherichia coli O157:H7 in Washington State, 1984-1991*. J Infect Dis, 1994. **170**(6): p. 1606-9.
30. Gaunt, P.N. and L.J. Piddock, *Ciprofloxacin resistant Campylobacter spp. in humans: an epidemiological and laboratory study*. J Antimicrob Chemother, 1996. **37**(4): p. 747-57.
31. Saenz, Y., et al., *Antibiotic resistance in Campylobacter strains isolated from animals, foods, and humans in Spain in 1997-1998*. Antimicrob Agents Chemother, 2000. **44**(2): p. 267-71.
32. Johnson, A.P., et al., *Gentamicin resistance in clinical isolates of Escherichia coli encoded by genes of veterinary origin*. J Med Microbiol, 1994. **40**(3): p. 221-6.
33. Phillips, I., et al., *Does the use of antibiotics in food animals pose a risk to human health? A critical review of published data*. J Antimicrob Chemother, 2004. **53**(1): p. 28-52.
34. Phillips, I., et al., *Antibiotic use in animals*. J Antimicrob Chemother, 2004. **53**(5): p. 885; author reply 886.

35. Neidhardt, F.C., *Escherichia coli and Salmonella : cellular and molecular biology*. 2nd ed. 1996, Washington, D.C.: ASM Press. 2 v.
36. Santé Canada. *Qu'est-ce qu'il y a dans votre puits? Un guide de traitement et d'entretien de l'eau de puits* 2005-08-09 [page Internet consultée en novembre 2006]; URL: [http://www.hc-sc.gc.ca/ewh-semt/water-eau/drink-potab/well\\_water-eau\\_de\\_puits\\_f.html](http://www.hc-sc.gc.ca/ewh-semt/water-eau/drink-potab/well_water-eau_de_puits_f.html).
37. McLellan, S.L., A.D. Daniels, and A.K. Salmore, *Clonal populations of thermotolerant Enterobacteriaceae in recreational water and their potential interference with fecal Escherichia coli counts*. Appl Environ Microbiol, 2001. **67**(10): p. 4934-8.
38. Raina, P.S., et al., *The relationship between E. coli indicator bacteria in well-water and gastrointestinal illnesses in rural families*. Can J Public Health, 1999. **90**(3): p. 172-5.
39. Edge, T.A. and S. Hill, *Occurrence of antibiotic resistance in Escherichia coli from surface waters and fecal pollution sources near Hamilton, Ontario*. Can J Microbiol, 2005. **51**(6): p. 501-5.
40. Oppegaard, H., T.M. Steinum, and Y. Wasteson, *Horizontal transfer of a multi-drug resistance plasmid between coliform bacteria of human and bovine origin in a farm environment*. Appl Environ Microbiol, 2001. **67**(8): p. 3732-4.
41. Anderson, A.D., et al., *Public health consequences of use of antimicrobial agents in food animals in the United States*. Microb Drug Resist, 2003. **9**(4): p. 373-9.
42. Ministère de l'agriculture de l'alimentation et des affaires rurales. *Fiche d'information no 2 - Les puits*. 2006 [page Internet consultée novembre 2007]; URL: [http://www.omafra.gov.on.ca/french/environment/efp/infosheet\\_2.htm#210](http://www.omafra.gov.on.ca/french/environment/efp/infosheet_2.htm#210).
43. Corkal, D., W.C. Schutzman, and C. Hilliard, *Rural water safety from the source to the on-farm tap*. J Toxicol Environ Health A, 2004. **67**(20-22): p. 1619-42.
44. Well Wise Center. [page Internet consultée en novembre 2007]; URL: [www.wellwise.ca](http://www.wellwise.ca).
45. Strauss, B., et al., *A prospective study of rural drinking water quality and acute gastrointestinal illness*. BMC Public Health, 2001. **1**: p. 8.
46. Statistique Canada. *Glossaire illustré*. 2003-01-20 [page Internet consultée en juin 2007]; URL: [http://geodepot.statcan.ca/Diss/Reference/COGG/Index\\_f.cfm](http://geodepot.statcan.ca/Diss/Reference/COGG/Index_f.cfm).
47. Statistique Canada. 2007 [page Internet consultée en novembre 2007]; URL: [http://www.statcan.ca/francais/freepub/95-629-XIF/2007000/general\\_f.htm#confid](http://www.statcan.ca/francais/freepub/95-629-XIF/2007000/general_f.htm#confid).
48. Environnement Canada. *Sommaire climatologique*. 2002 2005-04-11 [page Internet consultée en Janvier 2006]; URL: [http://www.climat.meteo.ec.gc.ca/prods\\_servs/cdn\\_climate\\_summary\\_f.html](http://www.climat.meteo.ec.gc.ca/prods_servs/cdn_climate_summary_f.html).
49. Johnson, J.Y., et al., *Prevalence of Escherichia coli O157:H7 and Salmonella spp. in surface waters of southern Alberta and its relation to manure sources*. Can J Microbiol, 2003. **49**(5): p. 326-35.
50. Goss, M.J., D.A.J. Barry, and D.L. Rudolph, *Contamination in Ontario farmstead domestic wells and its association with agriculture : 1. Results from drinking water wells*. J Contam Hydrol, 1998. **32**(3): p. 267-293.
51. Schroeder, C.M., et al., *Antimicrobial resistance of Escherichia coli O157 isolated from humans, cattle, swine, and food*. Appl Environ Microbiol, 2002. **68**(2): p. 576-81.

52. Schroeder, C.M., et al., *Antimicrobial resistance of Escherichia coli O26, O103, O111, O128, and O145 from animals and humans*. *Emerg Infect Dis*, 2002. **8**(12): p. 1409-14.
53. Agriculture et Agroalimentaire Canada (Direction générale de la recherche). *Stratégie de recherche sur la gestion du lisier de porc au Canada*. 1998 [page Internet consultée en avril 2007]; URL: <http://dsp-psd.tpsgc.gc.ca/Collection/A42-77-1998F.pdf>.
54. Jindal, A., et al., *Antimicrobial use and resistance in swine waste treatment systems*. *Appl Environ Microbiol*, 2006. **72**(12): p. 7813-20.
55. Bitton, G. and C.P. Gerba, *Groundwater pollution microbiology*. 1984, New York: Wiley. xii, 377 p.
56. Chokmani, K., *Caractérisation de la qualité de l'eau de surface de deux bassins versants en production animale*. 1996, Ottawa: National Library of Canada = Bibliothèque nationale du Canada. 2 microfiches.
57. Dohoo, I., W. Martin, and H. Stryhn, *Veterinary epidemiologic research*. 2003, Charlottetown, P.E.I.: AVC Inc. xx, 706.
58. Martin, S.W., A.H. Meek, and P. Willeberg, *Veterinary epidemiology : principles and methods*. 1st ed. 1987, Ames: Iowa State University Press. xi, 343.
59. Kleinbaum, D.G., L.L. Kupper, and H. Morgenstern, *Epidemiologic research : principles and quantitative methods*. 1982, Belmont, Calif. ; Toronto New York: Lifetime Learning Publications ; Van Nostrand Reinhold. xix, 529.

# **Annexes**

## **Annexe 1**

Détails techniques du géocodage de la banque de données concernant la contamination de l'eau

Document :

Developing a geocoding algorithm and geocoding a database related to water testing, version 2.2 (Caroline Guénette, mars 2005)

*Version 2.2*

Developing a geocoding algorithm and geocoding a database related to water testing  
Contract Reference Number :4500090637

By Caroline Gu enette, M.Sc.

March 2005

Report presented to Dr. Pascal Michel, DMV, Ph.D.  
Laboratory for Foodborne Zoonoses  
Public Health Agency of Canada

<b>1. OVERVIEW .....</b>	<b>3</b>
1.1. DATABASE DESCRIPTION .....	3
1.2. DATABASE ACCURACY CHECK .....	4
1.3. TOOLS .....	4
<b>2. PRE-PROCESSING .....</b>	<b>5</b>
2.1. CLEANING THE DATABASE ENTRIES.....	6
<i>Postal Boxes</i> .....	6
<i>Standardisation of the civic address field</i> .....	10
<i>Standardisation of the lot-concession field</i> .....	13
<i>Standardisation of street and municipality appellations</i> .....	18
<i>Inconsistencies</i> .....	18
2.2. NARROWING THE DATABASE UPON WATER TEST RESULTS.....	20
2.3. PRE-PROCESSING RESULTS .....	20
<b>3. GEOCODING METHODS USED .....</b>	<b>22</b>
<b>4. GEOCODING USING GEOPINPOINT .....</b>	<b>22</b>
4.1. USING CIVIC ADDRESS.....	23
4.2. USING POSTAL CODE .....	24
4.3. USING MUNICIPALITIES .....	25
4.4. GEOCODING RESULTS USING GEOPINPOINT.....	25
<b>5. GEOCODING USING LOT AND CONCESSION NUMBERS .....</b>	<b>26</b>
5.1. GEOCODING RESULTS FOR LOT/CONCESSION .....	28
<b>6. SELECTING A GEOCODING RESULT .....</b>	<b>28</b>
6.1. ASSIGNING AN URBAN OR RURAL LOCATION CODE.....	28
6.2. THE PRECISION SEQUENCES .....	32
6.3. SELECTING WHICH GEOCODING RESULT TO USE .....	32
<b>7. SELECTED GEOCODING RESULTS.....</b>	<b>39</b>
<b>8. VISUALISING THE POINTS ON A MAP .....</b>	<b>39</b>
<b>9. APPENDIXES .....</b>	<b>40</b>
9.1. APPENDIX A : PUBLIC HEALTH UNITS LIST .....	40
9.2. APPENDIX B : STATUS CODES LIST .....	41
9.3. APPENDIX C : STATISTICS ON REJECTION OF UNTESTED WATER SAMPLES .....	42
9.4. APPENDIX D : RESULTING MAP .....	50
9.5. APPENDIX E : (ELECTRONIC SUPPORT) GEOCODED DATABASE (ACCESS FORMAT). 51	

## 1. OVERVIEW

In order to achieve the objectives of a research project, a database related to water testing needed to be geocoded. Water samples and their analysis results needed to be geographically referenced and displayed on a map, using the location of the water source itself, in order to allow interpretation suitable to the research project.

This document aims at describing the process that was used and the obtained results. It will consequently highlight steps to follow and tips on geocoding a database related to water testing. The resulting map is attached as an appendix.

### 1.1. Database description

The provided database came from the Food Safety and Safe Water Unit (SWU) of the Ministry of Health and Long Term Care of Ontario (MOHLTC). It contained 129 019 entries, all in Access format. It gathered drinking water testing performed by the Central and Regional Public Health Laboratories of Ontario between May and September 2003.

Each line (entry) of the database refers to a drinking water sample that was sent to the MOHLTC laboratories for testing.

The database contained several fields, listed below with their characteristics. Fields had to be renamed for shorter/simpler ones upon reception of the database to allow the use of SAS statistics software

<b>Field name</b>	<b>Data type</b>	<b>original name</b>	<b>Description</b>
Drecei	DD-MM-YY	Date received	Date the sample arrived at the laboratory for testing
labo	text	Lab	Name of the laboratory which received the sample for testing
adress	text	SrcStreet	Street address of the water source location
lot	text	SrcLotConc	Lot, concession of the water source location
city	text	SrcCityTownMun	Township/Municipality of the water source location
emerg	numeric	Emerg#Num	Emergency locator number of the water source location
cnty	text	SrcCnty	County of the water source location
postal	TNT-TNT	SrcPostalcode	Postal code ( six digits) of the water source location
Dcoll	DD-month-YY	Date Collected	Collection date of the sample at the water source
phu	numeric	srcPHU	Number of the Public Health Unit where the water source is located (see list in appendix A for description)
stati	text	Status Int#	Result of the water testing in text description
statn	numeric	Status	Result of the water testing in numeric code (see table in appendix B for code description)
coliform	numeric	Coliform	Coliform count (total number per 100ml) as determined by water testing
ecoli	numeric	E.coli	Escherichia coli count (total number per 100ml) as determined by water testing
dread	DD-MM-YY	Date Read	Date the sample was processed at the analysing laboratory
access	numeric	Accession number	Numeric string of 6 digits being a unique identification number for each laboratory by season.

Other informations were collected on the requisition form filled by the water source owner. They were excluded from the provided database in order to ensure confidentiality. Although, they can explain some database anomalies caused by field confusion or inversion. Below is a partial list of database fields that were removed to ensure confidentiality before the database was delivered to us:

Name of the person requesting the water testing (water source owner);  
Return civic address;  
Return address municipality;  
Return address province;  
Return address postal code;  
Water source owner phone number.

### **1.2. Database accuracy check**

Fields of the database were typed-in by an employee of MOHLTC as they appeared on the analysis requisition form filled by the water source owner and provided with the water sample. No cross validation was performed by the MOHLTC to ensure data accuracy. The database was delivered as is.

### **1.3. Tools**

The database Access (Access 2000, Microsoft Corporation, 1992-1999) format allowed manipulating the data and performing pre-processing in its original file format. To avoid syntax problem, the computer language must be set to English before using Access and the 'maxlock per file' parameter contained in the Jet4.0 parameter was set to 2,000,000 to allow Access to process requested tasks without problem.

SAS statistical software release 8.1(SAS Institute Inc, Cary, NC, USA, 1999-2000) was used for statistical analysis of the entries distribution.

The geocoding process using the address and/or postal codes was realized using GeoPinPoint software (GeoPinpoint™ – Windows Standalone Desktop Version, DMTI Spatial Inc. 2004). The assignation of the geographical location for lot/concession numbers was done by database merging using Access.

The lot/concession file of the Ontario Base Map product provided by the Ministry of Natural Resources of Ontario was used as the reference database for geocoding the lot/concession information.

Once a geographical location was attributed to each entry, the GIS software entitled ArcGIS (ArcGIS™ 8 , ESRI, USA, 2002) was used to map the different geographical locations as points.

## 2. PRE-PROCESSING

In order to ensure the possibility to trace back all modifications/corrections done to the database entries, the fields containing spatial information were copied under another name and then modified ('gcd' prefix stands for geocoding). The geocoding process was performed on those modified fields reflecting the changes made in the cleaning process.

Use the original table *WaterTest* : Create <WT>

Query \ make table query \ design view

Field : *WaterTest*\* \ table : *WaterTest*

Field : *gcdaddress*:*Adress* \ table : *WaterTest*

Field : *gcdlot*:*Lot* \ table : *WaterTest*

Field : *gcdcity*:*City* \ table : *WaterTest*

Field : *gcdpostal*:*Postal* \ table : *WaterTest*

Name the new table : *WT* \ Run

Save query as <*qr\_mkt\_addgcd*>

Close the <*WaterTest*> table \ open the freshly created <*WT*> table \ verify field names syntax.

Field:	<i>WaterTest</i> .*	<i>gcdaddress</i> : <i>Adress</i>	<i>gcdlot</i> : <i>Lot</i>	<i>gcdcity</i> : <i>City</i>	<i>gcdpostal</i> : <i>Postal</i>
Table:	<i>WaterTest</i>	<i>WaterTest</i>	<i>WaterTest</i>	<i>WaterTest</i>	<i>WaterTest</i>
Sort:					
Show:	<input checked="" type="checkbox"/>				
Criteria:					
or:					

Total record number: 129 019

All double entries have been put aside from the database. In the Access software, a sub-table have been created with no double. It is that new table without double that was used for the geocoding process.

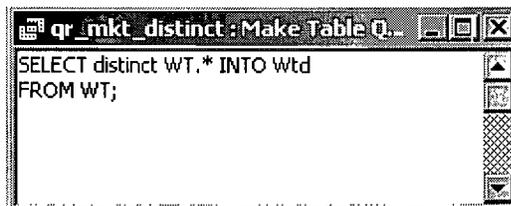
Use the <WT> table : create <WTd>

Query \ make table query \ design view

select \* for all fields \ SQL view \ use the formulation <select distinct WT>

Name the new table : *WTd* \ Run

Save as <*qr\_mkt\_distinct*>



Number of double entries: 1

Total distinct record number (no double) : 129 018

A unique number (ranging from 1 to 129018) was assigned to each record in order to facilitate merges in the database manipulations.

*Use the <WTd> table*

*Open the table design view \ drag fields in the order you wish them to appear \ Insert a field row \ Name the filed 'ID' \ in the data type choose 'AutoNumber' \ save.*

A column was inserted to attribute an exclusion criteria code to some entries. It is under the Field name 'gcdcode'

The resulting database named 'WTd' therefore contained the following fields:

<b>Field name</b>	<b>Data type</b>	<b>Description</b>
ID	numeric	unique ID number added to facilitate merges
adress	text	Street address of the water source location
lot	text	Lot, concession of the water source location
city	text	Township/Municipality of the water source location
emerg	numeric	Emergency locator number of the water source location
cnty	text	County of the water source location
postal	TNT-TNT	Postal code ( six digits) of the water source location
gcdadress	text	Cleaned field-Street address of the water source location
gcdlot	text	Cleaned field-Lot, concession of the water source location
gcdcity	text	Cleaned field-Township/Municipality of the water source location
gcdpostal	TNT-TNT	Cleaned field-Postal code ( six digits) of the water source location
gcdcode	text	Exclusion criteria code assigned to some entries
Drecei	DD-MM-YY	Date the sample arrived at the laboratory for testing
labo	text	Name of the laboratory which received the sample for testing
Dcoll	DD-month-YY	Collection date of the sample at the water source
phu	numeric	Number of the Public Health Unit where the water source is located (see list in appendix A for description)
stati	text	Result of the water testing in text description
statn	numeric	Result of the water testing in numeric code (see table in appendix B for code description)
coliform	numeric	Coliform count (total number per 100ml) as determined by water testing
ecoli	numeric	Escherichia coli count (total number per 100ml) as determined by water testing
dread	DD-MM-YY	Date the sample was processed at the analysing laboratory
access	numeric	Numeric string of 6 digits being a unique identification number for each laboratory by season.

## 2.1. Cleaning the database entries

### *Postal Boxes*

The geographical location of interest for this project is the one of the water source. The provided database contains only the address field associated to the water source location.

The presence of a postal box address reflects either extra information that was accidentally typed in from the 'return address' field of the requisition form, or a switch in the 'return address' and 'location of the water source address' fields of the requisition form. The entries containing a civic address and a postal box in the address field were cleaned from their postal box component and kept. All entries having for only address information a postal box had been attributed a <gcdcode> of 'POBOXONLY'. They were considered as improper for this project as the location of a water source is unlikely to be a postal box.

Identify P.O. Box-alike entries in the <WTd> table:

```
Query \ selection query \ design view
Field: gcdaddress \ Criteria: <like '*po box*'>, <like '*P.O. Box*'>, <like '*box*'>
```

Use the following command to isolate the PO Box information from the rest of the text string in the <gcdaddress>field:

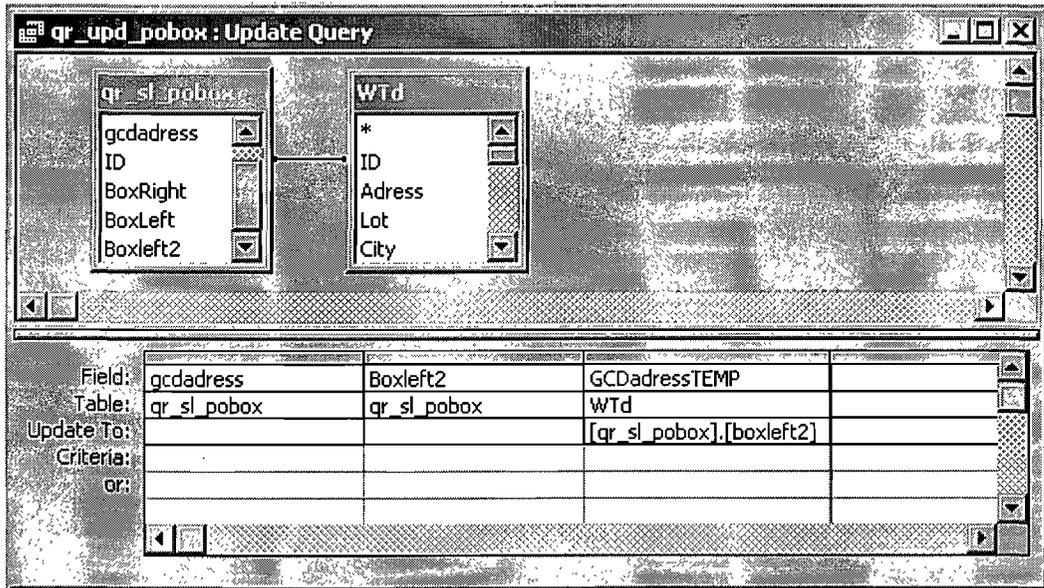
```
Field: ID
Field: BoxRight: Right([gcdaddress]*Len([gcdaddress])-
  InStr(1,[gcdaddress],"Box")+1)
Field: BoxLeft: Mid([gcdaddress],1,InStr(1,[gcdaddress],"Box")-1)
Field: Boxleft2: IIf([BoxLeft]="", "POBOXONLY", IIf([boxleft]="PO
  ", "POBOXONLY", IIf([boxleft]="P. O. ",
  "POBOXONLY", IIf([boxleft]="P. O.
  ", "POBOXONLY", IIf([boxleft]="P O
  ", "POBOXONLY", IIf([boxleft]="PO
  ", "POBOXONLY", IIf([boxleft]="P.O " , "POBOXONLY", [Boxleft])))
Save as <qr_sl_pobox> \ Run
```

```
Create a new colum in the <WTd> table \ rename colum <GCDadressTEMP>
save \ close <WTd>
```

Copy the modifications performed to P.O. Box alike in the query to the 'temp' column of the table:

```
Query \ update query \ design view
link the query containing all POBox alike entries <qr_sl_pobox> with the <WTd>
table using the <ID> field \ use the following command to copy modified address
information to the freshly created <GCDadressTEMP> field:
```

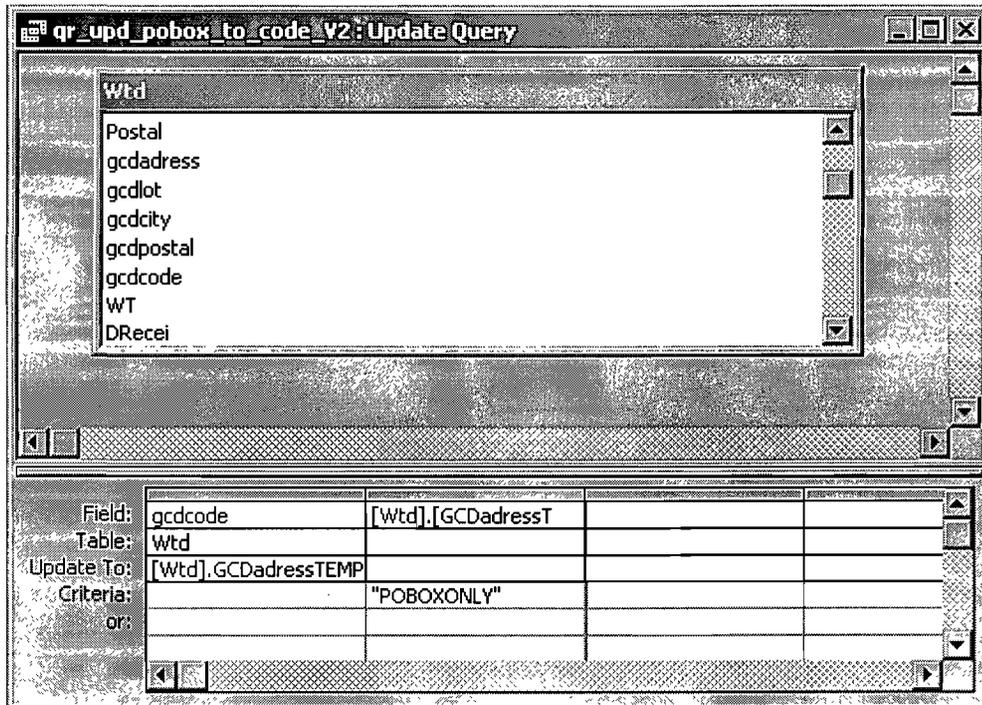
```
Field: gcdaddress \ table: qr_sl_pobox
Field: boxleft2 \ table: qr_sl_pobox
Field: gcdadressTEMP \ table: WTd \ updateto: qr_sl_pobox.boxleft2
Save query as <qr_update_pobox> \ Run \ Save \ close all queries
```



Open <WTd> table, manually compare entries containing data in the field <GCDadressTEMP> to corresponding data in the field <GCDadress> to confirm that the good part of the text string got preserved. Edit the <GCDadressTEMP> field if necessary.

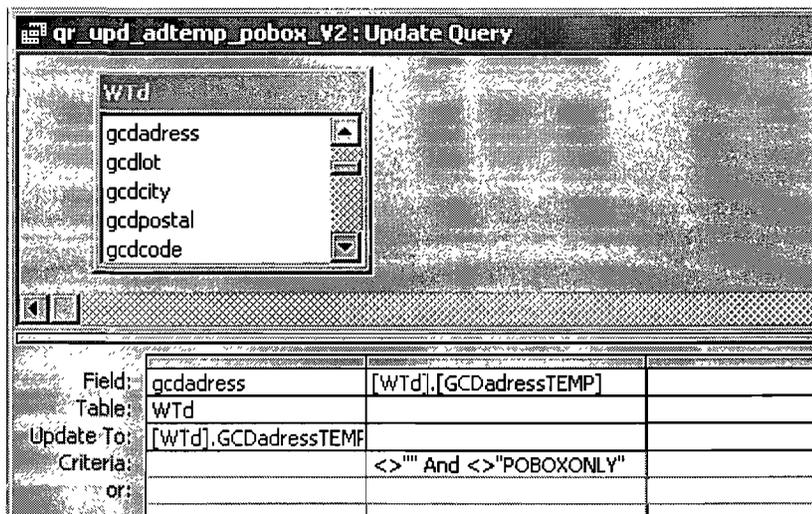
Perform an update query to assign the code 'POBOXONLY' in the <gcdcode> field at each line that was identified by 'POBOXONLY' in the previous step. They are the ones containing only P.O Box-like information in the address field.

Query \ update query \ design view  
 Field : gcdcode \ Table: WTd \ Update to: [WTd].GCDadressTEMP  
 Field : [WTd].[GCDadress] \ Criteria: 'POBOXONLY'  
 Save query as <Or\_upd\_pobox to code>



Copy the content of the <GCDadresTEMP> column to the <GCDadres> field with all the performed modifications:

Query \ update query \ design view  
 Field : gcdaddress \ update to : Wtd.gcdaddressTEMP  
 Field : gcdaddressTEMP \ criteria : < > « » And < > "POBOXONLY"  
 Save as <qr\_upd\_adtemp\_pobox> \ run  
 Clear the <GCDadresstemp> field.



The way this procedure was performed implies that all entries containing a PO Box component first and a civic address component after in the <GCDaddress> field were treated as containing only a PO Box component. We realize that this represents a loss of information but the goal is to develop an algorithm that would optimize the information in the geocoding procedure. As the number of entries containing a PO Box component first and a civic address component following often represent the civic address of the post office or mail depot, this procedure was considered the optimal solution.

Number of record containing PO Box alike entries in the address field :2308

Number of records modified to remove the extra PO Box alike information : 1783

Number of records classified as 'POBOXONLY' because containing only PO Box-like information in the address field : 525

### ***Standardisation of the civic address field***

In order to increase the geocoding efficiency, the civic address field was standardized by removing all text in brackets or between quotation marks. The correspondence with the reference database was then increased.

The appellations 'cottage', 'house', 'kitchen', 'well', 'tap' or 'barn' could not be used to sort entries containing extra information in the address field as too many street/roads names include that specific text string. Same problem with using the heifen, too many street/road names include that character, which makes it impossible to perform the corrections automatically. It would be possible to increase precision by making those corrections but they would have to be done manually, which was not considered optimal in this project. A good proportion of those anomalies were picked up anyways by the selection on brackets and quotation marks.

*Perform a select query to modify the entries containing brackets :*

*Use <Wtd>*

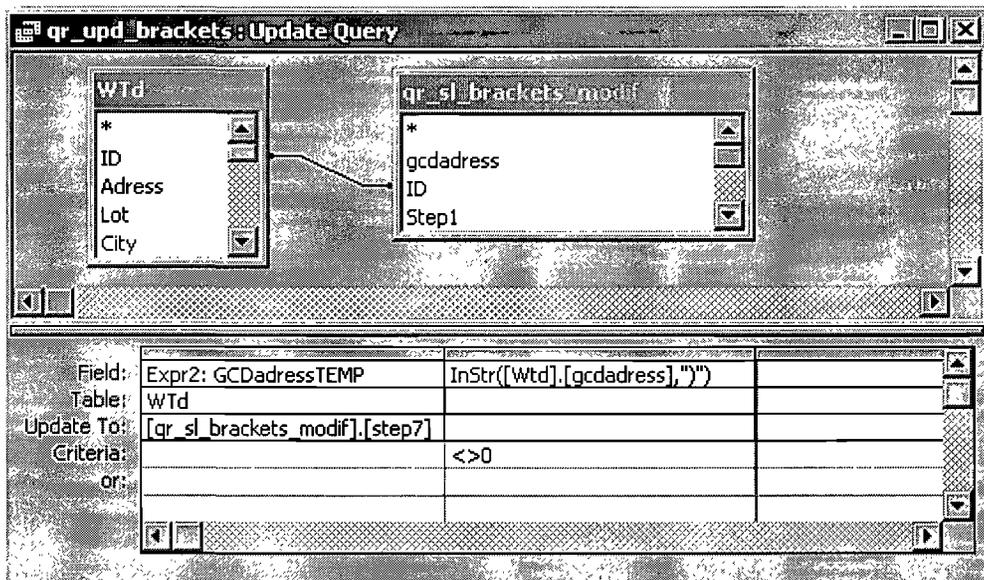
```
Query \ select query \ design view
Field : ID \ table : Wtd
Field : gcdaddress \ table : Wtd
Field : Step1 : InStr(1,[gcdaddress],"(")
Field : Step2 : InStr(1,[gcdaddress],")")
Field : Step3 : (InStr(1,[gcdaddress],")"))-(InStr(1,[gcdaddress],"("))
Field : Step4 : Left([gcdaddress],InStr(1,[gcdaddress],"(")-1)
Field : step5 : mid([gcdaddress],InStr(1,[gcdaddress],
"("),(InStr(1,[gcdaddress],")"))-(InStr(1,[gcdaddress],"(")-1))
Field : step6 : Mid([gcdaddress],InStr(1,[gcdaddress],")")+1,Len([gcdaddress]))
Field : step7 : Trim(Left([gcdaddress],InStr(1,[gcdaddress],"(")-1) & " " &
Mid([gcdaddress],InStr(1,[gcdaddress],")")+1,Len([gcdaddress])))
Save as <qr sl brackets modif> \ run
```

Correction for brackets was performed on 456 lines.

Use an update query to copy the content of the modification performed in the previous select query to the <GCDadressTEMP> field. Use a second update query to copy non-null entries of the <GCDadressTEMP > field to the <GCDadress> one.

```

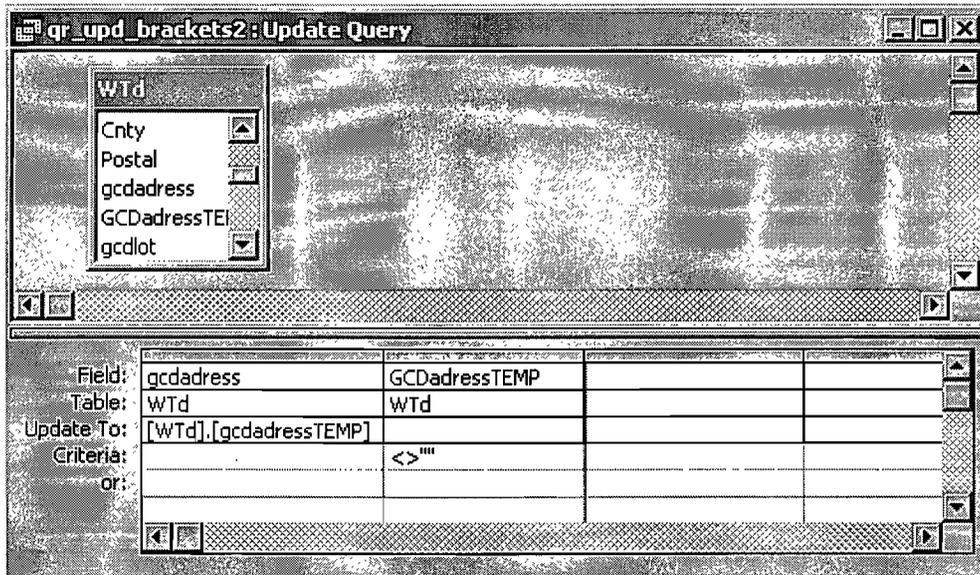
Query \ update query \ design view
Link table <WTd> with query <qr_sl_brackets_modif> using the field <ID> they
have in common
Field: Expr2: Wtd.GCDadressTEMP
      update to : [qr_sl_brackets_modif].[step7]
Field: InStr([Wtd].[gcdadress],")") \ criteria : <>0
Save as <qr_upd_brackets> \ run.
    
```



```

Query \ update query \ design view
Field: gcdadress \ Table: WTd \ update to : [WTd].gcdadressTEMP
Field: gcdadressTEMP \ table: WTd \ criteria : <> « «
Save as <qr_upd_brackets2> \ run.
    
```

Clear the <GCDadressTEMP> field. \ Save.



Repeat the process for the quotation marks instead of the brackets, using a slightly different query, as quotation marks are the same character used twice

Perform a select query to modify the entries containing quotation marks :

Use <WTd> table

Query \ select query \ design view

Field : ID \ table : WTd

Field : Expr1 : gcdaddress \ table : WTd

Field : WTd.gcdaddressTEMP

Field : Step1 : InStr(1,[gcdaddress],Chr(34))

Field : Step2 : InStr(InStr(1,[gcdaddress],Chr(34))+1,[gcdaddress],Chr(34))

Field : Step3 : InStr(InStr(1,[gcdaddress],Chr(34))+1,[gcdaddress],Chr(34))-  
InStr(1,[gcdaddress],Chr(34))

Field : Step4 : Left([gcdaddress],InStr(1,[gcdaddress],Chr(34))-1)

Field : step5 : Mid([gcdaddress],InStr(1,[gcdaddress],Chr(34)),InStr(InStr(1,[gcdadr  
ess],Chr(34))+1,[gcdaddress],Chr(34))-InStr(1,[gcdaddress],Chr(34))+1)

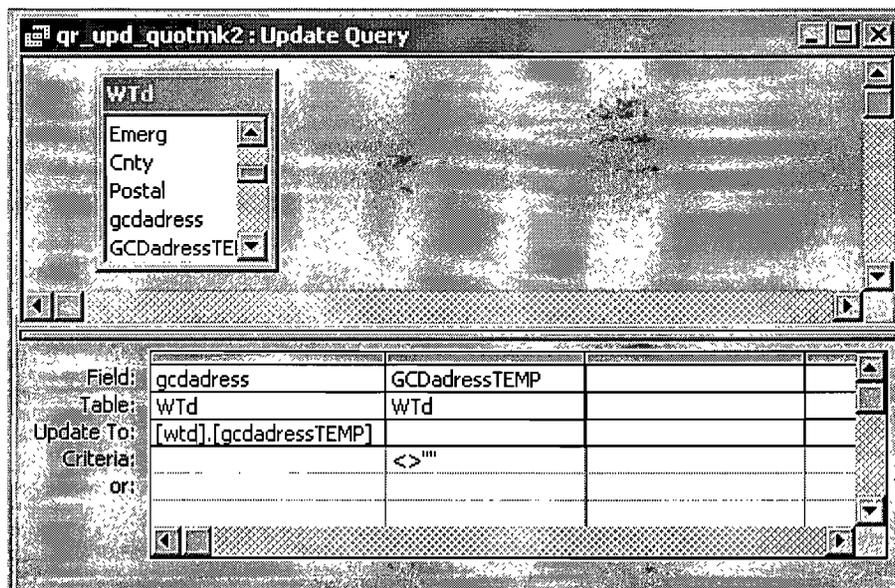
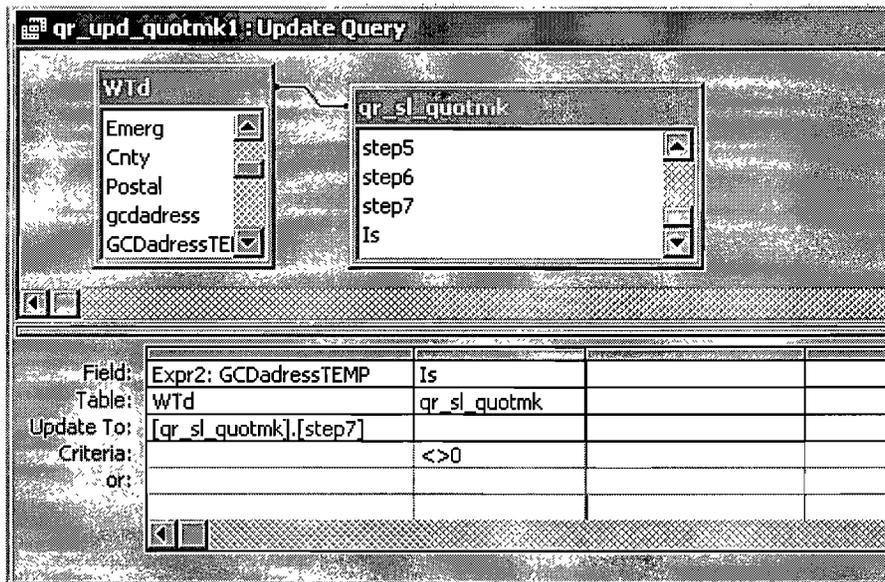
Field : step6 : Mid([gcdaddress],InStr(InStr(1,[gcdaddress],Chr(34))+1,[gcdaddress],  
Chr(34))+1,Len([gcdaddress]))

Field : step7 : Trim(Left([gcdaddress],InStr(1,[gcdaddress],Chr(34))-1) & " " &  
Mid([gcdaddress],InStr(InStr(1,[gcdaddress],Chr(34))+1,[gcdaddress],  
Chr(34))+1,Len([gcdaddress])))

Field : Is : InStr([gcdstreet],Chr(34))

Save as <qr\_sl quotmk> \ run

Use update queries as pictured below to update the WTd table with the modified quotation marks entries.



*Clear the <GCDadrestTEMP> field \ Save.*

Number of corrections performed for quotation marks : 32 entries

***Standardisation of the lot-concession field***

The lot and concession numbers were standardized to match the most often used structure in the reference database used for geocoding: 'LOT xx, CON xx'. Although some other structures or appellation are present in the reference database used for geocoding, they represent a small proportion of the polygons and their spatial distribution was uneven. The choice was made to concentrate only on the above structure for standardisation of the field.

Use a select query to identify the entries having the appellation 'Lt' instead of 'lot'

```
Field: gcdlot \ Table: Wtd \ Criteria: Like 'Lt*Con*'  
Save as <qr_sl Lt*Con*>  
Run \ Save table.
```

In the query window, use a find and replace function to transform the 'Lt' entries in 'LOT'

```
Edit menu \ replace  
Find What: Lt  
Replace With: LOT  
Look In: gcdlot  
Match: Any part of the field  
Replace All
```

The <gcdlot> field in the <Wtd> table is automatically updated with the changes.

Repeat the process to change the different typographies of concession to 'Con'

```
Field: gcdlot \ Table: Wtd \ Criteria: Like '*Conc*' or Like '*concession*'  
Save as <qr_sl *Concession*>  
Run \ Save table.
```

Find and replace function in the resulting query table

```
Edit menu \ replace  
Find What: Concession  
Replace With: CON  
Look In: gcdlot  
Match: Any part of the field  
Replace All
```

```
Find What: Conc  
Replace With: CON  
Look In: gcdlot  
Match: Any part of the field  
Replace All
```

Use a separate query to select entries typed in the form of 'L ?? C??'

```
Field: gcdlot \ Table: Wtd  
Criteria: Like 'L?? C??' or Like 'L??, C??' or Like 'L? C?' or Like 'L?? C?' or  
Like 'L? C??' or Like 'L??, C?' or Like 'L ?? C ??' or Like 'C?? L??' or  
Like 'C??, L??' or Like 'C? L?' or Like 'C?? L?' or Like 'C? L??' or  
Like 'C??, L?' or Like 'C?? L??' or Like 'L??' or Like 'L?' or
```

Like 'C??' or Like 'C?'  
Save as <qr\_sl\_1\_C> \Run

Find and replace function in the resulting query table

Edit menu \ replace  
Find What: L  
Replace With: 'LOT ' (LOT and a space)  
Look In: gcdlot  
Match: Any part of the field  
Replace All

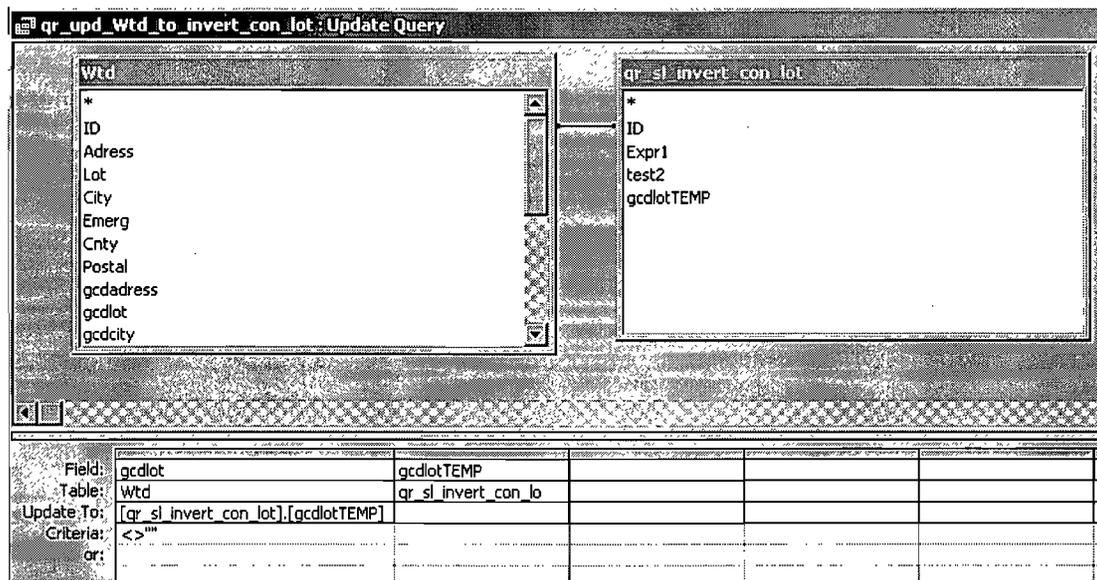
Find What: C  
Replace With: 'CON ' (CON and a space)  
Replace All  
Look In: gcdlot  
Match: Any part of the field  
Replace All

Perform a query to switch the lot and concession info when they appear in the con – lot format.

Query \ Select query \ Design view  
Field: ID \ Table: Wid  
Field: Expr1: gcdlot \ Table: Wid  
Field: test2: InStr([Wid].[gcdlot], "LOT") \ criteria: <> 0  
Field: gcdlotTEMP: Mid([gcdlot], InStr([Wid].[gcdlot], "LOT"), 25) & ", " & Mid([gcdlot], 1, InStr([Wid].[gcdlot], "LOT") - 2)  
Field: gcdlot \ Table: Wid \ criteria: Like 'Con\*'  
Save as: <qr\_sl\_invert\_con\_lot> \Run

Perform an update query (see picture below) to copy the non-null results of the previous inversion query to the <gcdlot> field.

Query \ update query \ Design view  
Field: gcdlot \ Table: Wid  
Field: gcdlot \ Table: Wid \ Update to: [qr\_sl\_invert\_con\_lot].[gcdlotTEMP]  
Criteria: <> ""  
Save as <qr\_upd\_Wid\_to\_invert\_con\_lot>  
Run \ Save table

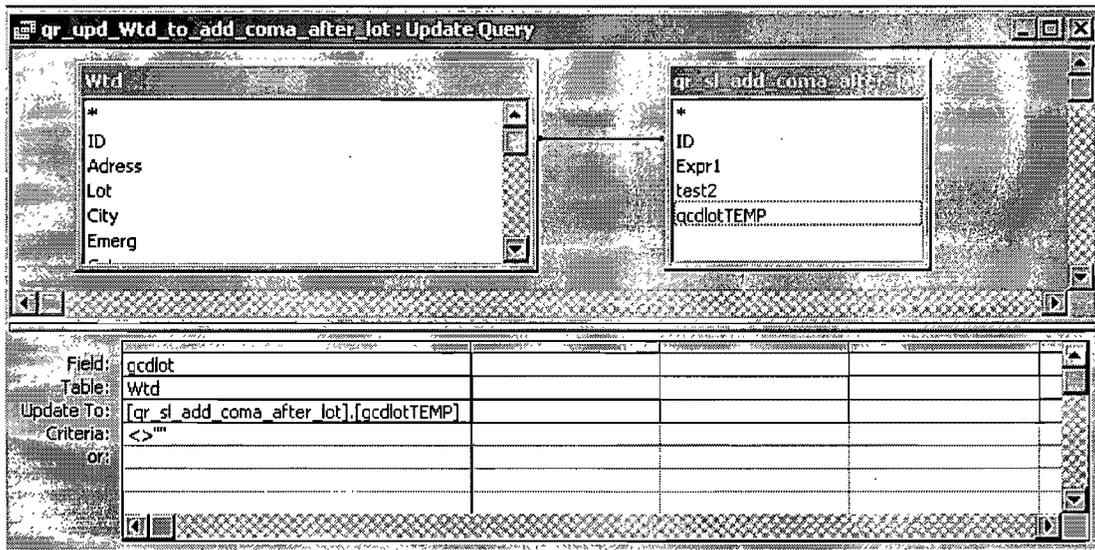


Use a query to add a coma after the lot number/letter

```
Field: ID \ Table: Wtd
Field: Expr1: gcdlot \ Table: Wtd
Field: test2: InStr([Wtd].[gcdlot], "CON") \ criteria: <>0
Field: gcdlotTEMP: Mid([gcdlot], 1, InStr([Wtd].[gcdlot], "CON")-2) & ", " &
Mid([gcdlot], InStr([Wtd].[gcdlot], "CON"), 25)
Field: gcdlot \ Table: Wtd \ Criteria: Like "Lot*"
Save as <qr_sl_add_coma_after_lot>
Run \ Save table.
```

Use an update query (as detailed in the previous step) to update the <gcdlot> field to the <gcdlotTEMP> column in the <qr\_sl\_add\_coma\_after\_lot> query just created. (See picture below)

```
Save as <qr_upd_Wtd_to_add_coma_after_lot>
Run \ Save table.
```



This update query added comas after the lot number/letters. In the cases where a coma was already present, it is now doubled. Use a find and replace function to solve this problem.

*Use the <Wtd> table*

*Select the 'gcdlot' field \ Edit \ Replace  
Find What: ,, (double coma)  
Replace with: , (single coma)  
Look in: gcdlot  
Match: Any part of the filed  
Replace All*

*Select the 'gcdlot' field \ Edit \ Replace  
Find What: ,, (coma space coma)  
Replace with: ,  
Look in: gcdlot  
Match: Any part of the filed  
Replace All*

Repeat the process to remove the double spaces

*Find What: (double space)  
Replace with: (single space)  
Look in: gcdlot  
Match: Any part of the filed  
Replace All  
Repeat until no double spaces are found. It will allow to get rid of triple and multiples spaces.*

### **Standardisation of street and municipality appellations**

There was no need to standardize the street appellation, suffix or prefix, either for municipality as the software GeopinPoint allows to relax the matching criteria to fit to street alias and populated place names.

### **Inconsistencies**

Records containing no spatial information were sorted from the database. At least one level of geographical information is needed for each entry so it becomes possible to associate a geographic reference to them. The entries containing no information in the <gcdaddress>, <gcdlot>, <gcdpostal>, and <gcdcity> fields were assign the 'NOGEO' label in the <gcdcode> field.

*Use the <Wtd> table*

```
Query \ update query \ design view
Field : gcdcode \ Table : Wtd \ Update to: 'NOGEO'
Field : gcdaddress \ Table : Wtd \ criteria: ""
Field : gcdlot \ Table : Wtd \ criteria: ""
Field : gcdpostal \ Table : Wtd \ criteria: ""
Field : gcdcity \ Table : Wtd \ criteria: ""
Save query as <qr upd nogeo> \ run
```

Number of record to which code 'NOGEO' is attributed because they did not contain any usable geographic information : 839

In order to be used as spatial information, the lot/concession number needs to be attributed to a specific municipality. The Records containing no civic address and no postal code, but a lot/concession number without a municipality were also assigned the 'NOGEO' code.

*Use the <Wtd> table*

```
Query \ update query \ design view \ include all fields
Field : gcdcode \ table : Wtd \ Update to: « NOGEO »
Field : gcdaddress \ table : Wtd \ Criteria : = » »
Field : gcdlot \ Table : Wtd \ Criteria : <> » »
Field : gcdcity \ Table : Wtd \ Criteria : = » »
Field : gcdpostal \ Table : Wtd \ Criteria : = » »
Save query as <qr upd lotnocity> \ run
```

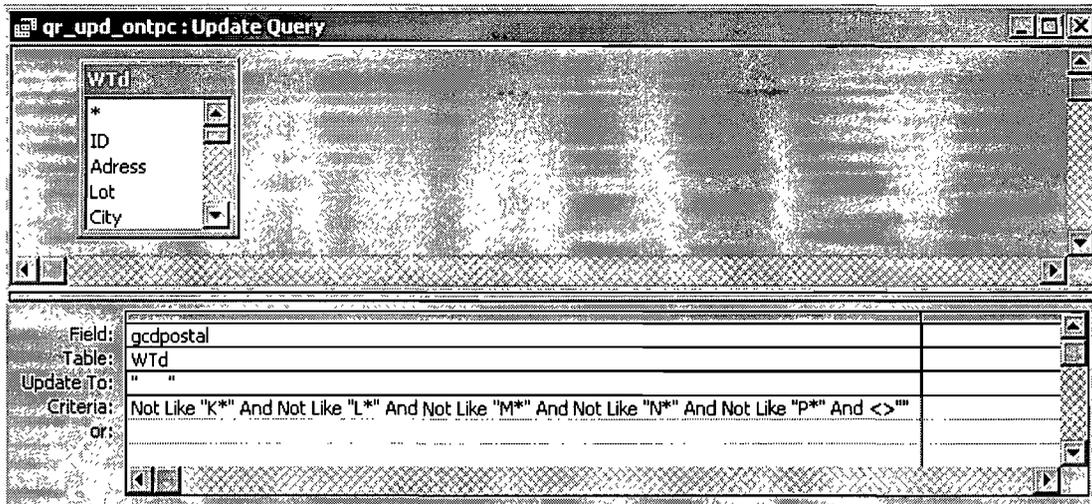
Number of records classified as 'NOGEO' because they contained lot/concession number as the only spatial information but no municipality name : 118

All entries were verified for the sample's province of origin, using the first digit of their postal code. Within this study, only samples from Ontario are of interest. Consequently, entries containing invalid Ontario postal code got modified in the cleaning process to ensure that invalid postal code would not be used in the geocoding process. Blank postal code fields were kept, as they could not lead to improper geocoding.

Canada Post states that the postal codes of Ontario are all starting by one of the following letter: K, L, M, N or P. Consequently, letter\number strings contained in the postal code field starting with another letter than the five selected ones got replaced by a blank field. It might have reflected situations where the postal code of the water source location was unknown, postal code of the return address outside province was used, or an invalid text\number string was accidentally entered.

The following query allows identifying the problematic entries and replacing them by a blank field:

```
Query \ update query \ design view  
Field: gcdpostal \ Table : WTd \ update to : « »  
criteria : Not Like "K*" And Not Like "L*" And Not Like "M*" And Not  
Like "N*" And Not Like "P*" And <> ""
```



```
Save as <qr upd ontpe> \ run.
```

Number of entries that contained invalid Ontario postal codes and got replaced by a blank field : 40

## 2.2. Narrowing the database upon water test results

Samples on which water testing was not performed were as well identified. Different reasons can explain the rejection of a water sample at the testing process, as showed in the status code table in appendix B. All samples having a numeric status code between 5 and 21 (included) were attributed a 'NWT' in the <WT> column of the database.

Appendix C is a statistical description of the non-tested water samples.

*Add a column called <WT> to specify the entries having no water test results.*

```
Query \ update query \ design view
Field: statN \ criteria: 5 or 6 or 7 or 8 or 10 or 11 or 12 or 13 or 14 or 15 or 16
or 17 or 18 or 19 or 20 or 21
Field: WT \ update to: 'NWT'
Save query as <qr_upd nottested to code> \ run
```

Number of entries classified as 'NWT' (absence of water testing results) : 10 332

Total number of entries having a status code of 1,2,3,4 or 22 : 118 118

## 2.3. Pre-processing results

Of the initial 129 019 entries in the database :

- 1 was removed because it was a double;
- 525 were labelled 'POBOXONLY' in the gcdcode column as they contained PO Box alike information only in the address field;
- 839 were labelled 'NOGEO' as they contained no usable spatial information;
- 118 were labeled 'NOGEO' as they contained lot/concession numbers as only spatial identifier but no municipality information;
- 10 332 were labelled 'NWT' as they did not contained any water test results;
- 118 118 records (91.6%) are potentially usable for geocoding and do have a water testing result.

Of the 129 018 distinct entries in the database, 40 380 (~31%) were modified to better match the reference databases at geocoding :

- 1783 were modified to remove the extra PO Box alike information;

- 455 were modified because they contained brackets in the address field;
- 32 were modified because they contained quotation marks in the address field;
- 40 were modified as they contained invalid Ontario postal codes;
- 38 070 were modified to standardize the lot/concession information.

### **3. GEOCODING METHODS USED**

Two major geocoding processes were used to attribute geographical reference to the data. The first process implies the GeoPinpoint software as an interface to link data with the reference database called CanStreetFile. This process is used to perform geocoding based on civic address, postal codes, Forward Sortation Areas (FSA) and municipalities.

The second process is used to attribute geographical locations based on lot/concession numbers. The link with the reference database, in this case the Ontario Base Map – lot/concession thematic layer provided by the Ministry of Natural Resources of Ontario, is done using Access software.

Six geocoding possibilities were applied to each entry: based on civic address; six digits postal code; five digit postal code; FSA; lot/concession numbers and based on municipality. It is after performing all possibilities on every entries of the database that a decision tree is applied to determine in each case which geocoding result is the most appropriate to use.

### **4. GEOCODING USING GEOPINPOINT**

GeoPinpoint uses its own geographic coordinate reference database made by DMTI and updated quarterly. The said database, CanStreetFiles, is included in the software, so using the later version becomes an issue. We used the release 5.2 from May 2004 that was the latest one at the time of the data collection.

The Geopoint software is an interface that allows comparing a specific address type entry with a reference database. The reference database is structured to allow interpolations, which makes possible to attribute to an address an interpolated localization within a street segment. Consequently, it is possible to relax the matching criteria to the database to increase the number of matching entries.

Please refer to the Geopoint user guide for more details about the possibilities of the software.

Add columns to the Access database to write the X,Y, Result code and Precision code generated by Geopoint or answer yes when the software will prompt your authorization to add them to the database.

All geocoding sequences were done on the same database, avoiding using the 'overwrite existing coordinates' option in the 'output specification' tab. A separate report is automatically generated with each geocoding sequences (Appendix E, electronic format).

## 4.1. Using civic address

### Strict address

The first sequence of geocoding was based on the strict civic address (strict being opposed to relaxed). It ensures that whenever the given address is found exactly as is in the reference database, it is used without relaxing or approximation. The 'street alias' option was used. It allows use of different appellation for the same street segment (ex. Highway 17 and Main Street), therefore it increases the number of matches.

**Data Source Tab**  
 Geocoding Path: Address Geocoder (will all nods included)

**Input Specifications Tab**  
 Un-parsed Address: gcdaddress  
 Municipality: gcdcity  
 Province: none  
 Postal Code: gcdpostal  
 Use Un-Parsed address field: checked  
 Use constant for Province: checked – ON (for Ontario)  
 Geocode to Street Alias: checked  
 Refine Address by Postal Code: checked

**Output Specifications Tab**  
 Longitude: Adr\_X  
 Latitude: Adr\_Y  
 Result Code: Adr\_Rcode  
 Precision Code: Adr\_Prescode

### Relaxed address

The relaxed criteria allow geocoding based on address when the street prefix, street type or street direction does not meet perfectly the reference database.

**Data Source Tab**  
 Geocoding Path: Address Geocoder (will all nods included)

**Input Specifications Tab**  
 Un-parsed Address: gcdaddress  
 Municipality: gcdcity  
 Province: none  
 Postal Code: gcdpostal  
 Use Un-Parsed address field: checked  
 Use constant for Province: checked – ON (for Ontario)  
 Geocode to Street Alias: checked  
 Relax matching on S\_Type, S\_Direction, S\_Prefix: checked  
 Refine Address by Postal Code: checked  
 Closest tolerance: 5  
 Opposite side of Street: checked

**Output Specifications Tab**  
 Longitude: Adr\_X

Latitude : Adr\_Y  
Result Code : Adr\_Rcode  
Precision Code : Adr\_Prescode

## 4.2. Using postal code

### Using the full postal code ( six digits)

#### Data Source Tab

Geocoding Path : Postal Code Geocoder (will all nodes included)

#### Input Specifications Tab

Un-parsed Address : gcdaddress

Municipality : gcdcity

Province : none

Postal Code : gcdpostal

Use Un-Parsed address field : checked

Use constant for Province : checked - ON (for Ontario)

Precision Code : 600 - checked

#### Output Specifications Tab

Longitude : 6PC\_X

Latitude : 6PC\_Y

Result Code : 6PC\_Rcode

Precision Code : 6PC\_Prescode

### Using the first five digits

#### Data Source Tab

Geocoding Path : Postal Code Geocoder (will all nodes included)

#### Input Specifications Tab

Un-parsed Address : gcdaddress

Municipality : gcdcity

Province : none

Postal Code : gcdpostal

Use Un-Parsed address field : checked

Use constant for Province : checked - ON (for Ontario)

Precision Code : 500 - checked

#### Output Specifications Tab

Longitude : 5PC\_X

Latitude : 5PC\_Y

Result Code : 5PC\_Rcode

Precision Code : 5PC\_Prescode

### Using the first three digits (RTA)

#### Data Source Tab

Geocoding Path : Boundary Geocoder To FSA Centroid

#### Input Specifications Tab

Un-parsed Address : gcdaddress

Municipality : gcdcity

Province : none  
Postal Code : gcdpostal  
Use Un-Parsed address field : checked  
Use constant for Province : checked – ON (for Ontario)  
Output Specifications Tab  
Longitude : RTA   
Latitude : RTA   
Result Code : RTA Rcode  
Precision Code : RTA Prescode

### 4.3. Using municipalities

Data Source Tab  
Geocoding Path: Boundary Geocoder \ To PPN Points + To Municipal Centroid  
Input Specifications Tab  
Un-parsed Address : gcdaddress  
Municipality : gcdcity  
Province : none  
Postal Code : gcdpostal  
Use Un-Parsed address field : checked  
Use constant for Province : checked – ON (for Ontario)  
Output Specifications Tab  
Longitude : Mun   
Latitude : Mun   
Result Code : Mun Rcode  
Precision Code : Mun Prescode

### 4.4. Geocoding results using GeoPinpoint

- 25,861 (~20%) Have a successful geocoding result using the strict civic address
- 4,468 (~3%) Have a successful geocoding result using the relaxed civic address
- 16,823 (~13%) Have a successful geocoding result using six digits postal code
- 33,553 (~26%) Have a successful geocoding result using five digits postal code
- 103,615 (~80%) Have a successful geocoding result using tree digits postal code
- 52,169 (~75%) Have a successful geocoding result using municipality

As these numbers show, many entries had successful geocoding results using different information sources. In a further step, the most appropriate result is selected for use.

## 5. GEOCODING USING LOT AND CONCESSION NUMBERS

The 'Ontario Base Map' product, specifically the lot and concession numbers layer provided by the Ministry of Natural resources of Ontario was used as the reference database. The said database contains two distinct field with municipality information: the <official\_name> and <city\_name> fields. One refers to the usual name of the locality while the other represents the name of the administrative jurisdiction in place.

Using ArcGIS, add the unprojected longitude and latitude information to each polygon by using the 'add XY' feature. Name columns <LOT\_X\_COORD> and <LOT\_Y\_COORD>. Export the attribute table in dbf format.

Import the dbf attribute table in Access. Name <TwCLCt\_modif>

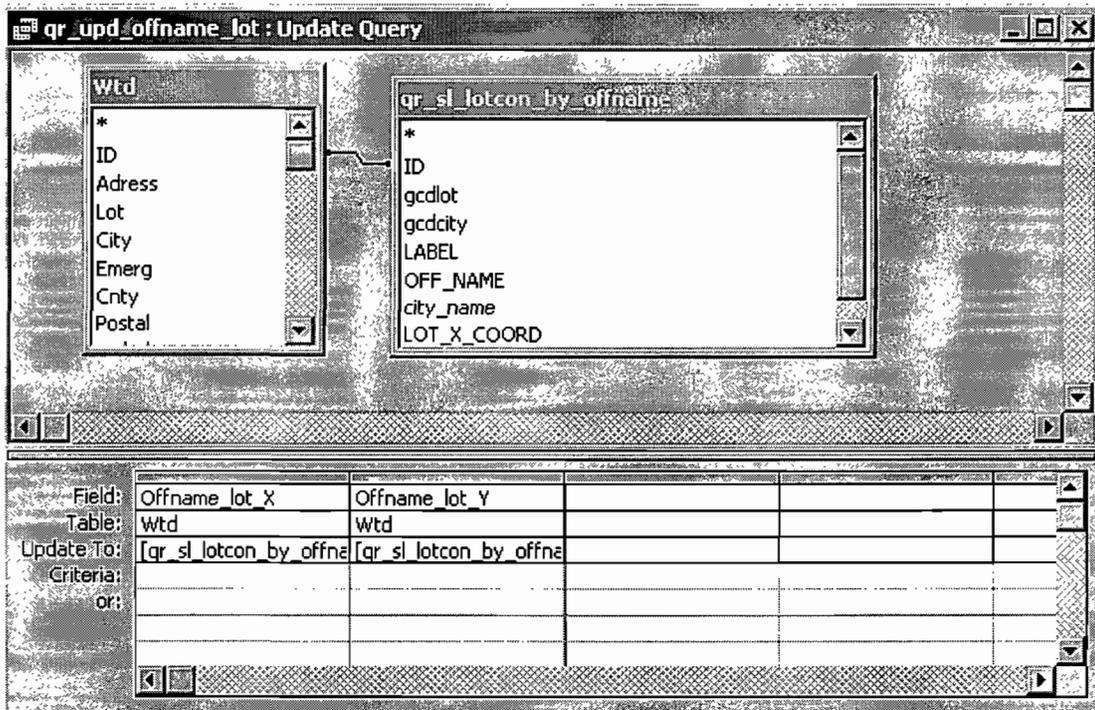
Create a query to link the <TwCLCt\_modif> table to the <Wtd> table using the <label> and <official\_names> columns.

```
Query \ Select query \ Design view
Show tables: Wtd and TwCLCt_modif
Link tables using the fields: <gcdlot> with <label> and <gcdcity> with
                             <off_name>
Field: ID \ Table: Wtd
Field: LOT_X_COORD \ Table: TwCLCt_modif
Field: LOT_Y_COORD \ Table: TwCLCt_modif
Save as <qr_sl_lotcon_by_offname> \ Run
```

Add four columns to the <Wtd> table: Offname\_lot\_X ; Offname\_lot\_Y; City\_lot\_X and City\_lot\_Y.

Use an update query to append the LOT\_X\_COORD and LOT\_Y\_COORD to the <Wtd> entries when there is a match using the <off\_name> column.

```
Field: Offname_lot_X \ Table: Wtd
Update to: [qr_sl_lotcon_by_offname].[LOT_X_COORD]
Field: Offname_lot_Y \ Table: Wtd
Update to: [qr_sl_lotcon_by_offname].[LOT_Y_COORD]
```



Repeat the process using the 'city\_name' field for linkage as it represents different appellations of the same location. It will allow different matches.

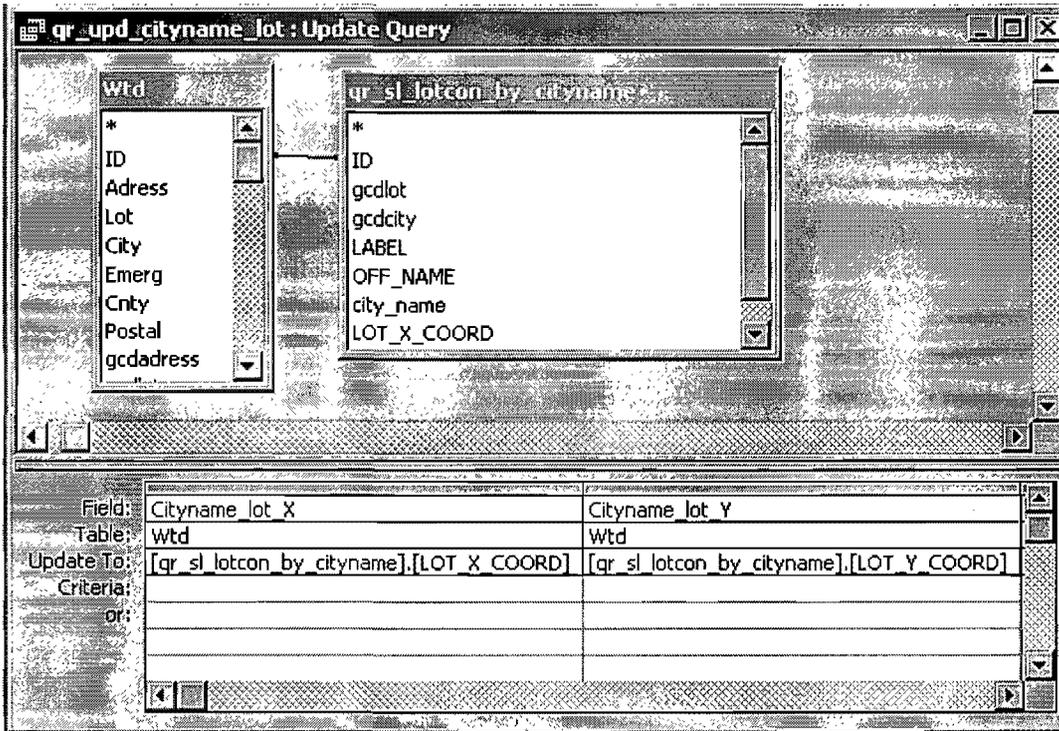
```

Query \ Select query \ Design view
Show tables: Wtd and TwCLCt_modif
Link tables using the fields: <gcdlot> with <label> and <gcdcity> with
<city_name>
Field: ID \ Table: Wtd
Field: LOT_X_COORD \ Table: TwCLCt_modif
Field: LOT_Y_COORD \ Table: TwCLCt_modif
Save as <qr_sl_lotcon by cityname> \ Run.
    
```

Use an update query to append the LOT\_X\_COORD and LOT\_Y\_COORD to the Wtd entries when there is a match using the 'city\_name' column.

```

Field: Cityname_lot_X \ Table: Wtd
Update to: [qr_sl_lotcon by cityname].[LOT X COORD]
Field: Cityname_lot_Y \ Table: Wtd
Update to: [qr_sl_lotcon by cityname].[LOT Y COORD]
    
```



Perform a 'make table query' to create a separate table containing all the longitude/latitude data.

*Query \ make table query \ design view  
Use every field of the <Wtd> table  
Save new table as <Gcd\_main> \ Run*

### 5.1. Geocoding results for Lot/Concession

- 9,980 (~8%) Have a successful geocoding result using the lot/concession numbers in the official municipality name
- 8,206 (~6%) Have a successful geocoding result using the lot/concession numbers in the city name appellation

## 6. SELECTING A GEOCODING RESULT

### 6.1. Assigning an urban or rural location code

In order to retain the longitude/latitude representing the more precise location for each entry, distinction must be made between urban and rural areas.

Crete a column in the <Gcd\_main> table. Name it <Urban\_code>

Perform an update query to assign the letter 'R' to any entry having a zero as second digit of their postal code.

```
Query \ Update query \ Design view
Field: Urban_code \ Table: Gcd_main
Update to: 'R'
Field: gcdpostal \ Table: Gcd_main \ Criteria: Like '0*'
Save as <qr_upd_r_urban_code> \ Run
```

Perform an update query to assign the letter 'U' to any entry having a number between 1 and 9 as second digit of their postal code.

```
Query \ Update query \ Design view
Field: Urban_code \ Table: Gcd_main
Update to: 'U'
Field: gcdpostal \ Table: Gcd_main \ Criteria: Like '?1*' or Like '?2*' or Like '?3*' or Like '?4*' or Like '?5*' or Like '?6*' or Like '?7*' or Like '?8*' or Like '?9*'
Save as <qr_upd_u_urban_code> \ Run
```

Create a new table containing only the entries having no data in the <gcdpostal> field.

```
Query \ Make table query \ Design view
Field: ID \ Table: Gcd_main
Field: gcdpostal \ Table: Gcd_main \ Criteria: ""
Field: gcdaddress \ Table: Gcd_main
Field: gcdlot \ Table: Gcd_main
Field: gcdcity \ Table: Gcd_main
Field: Adr_X \ Table: Gcd_main
Field: Adr_Y \ Table: Gcd_main
Field: Mun_X \ Table: Gcd_main
Field: Mun_Y \ Table: Gcd_main
Field: Offname_lot_X \ Table: Gcd_main
Field: Offname_lot_Y \ Table: Gcd_main
Field: Cityname_lot_X \ Table: Gcd_main
Field: Cityname_lot_Y \ Table: Gcd_main
Save as <qr_mkt_TEMPgcd> \ Run
```

In this database, 25 233 entries are selected for the <TEMPgcd> table. Those 25 233 entries have no postal code information allowing to assign them an urban/rural code.

Create a new table with the entries having no postal code information but a successful geocoding result using the civic address:

```
Query \ Make table query \ Design view
```

```
Query \ Make table query \ Design view
Name table as: TEMPgcd_address
Field: ID \ Table: TEMPgcd
Field: Adr_X \ Table: TEMPgcd \ Criteria: <>0
Field: Adr_Y \ Table: TEMPgcd
Save as <qr_mkt_TEMPgcd_address> \ Run
```

The freshly created <TEMPgcd\_address> table contains 3704 entries

Repeat the process to create a table with entries having no postal code information but successful geocoding by lot/concession:

```
Query \ Make table query \ Design view
Name table as: TEMPgcd_lot_offname
Field: ID \ Table: TEMPgcd
Field: Offname_lot_X \ Table: TEMPgcd \ Criteria: Is not null
Field: Offname_lot_Y \ Table: TEMPgcd
Field: Adr_X \ Table: TEMPgcd \ Criteria: =0
Field: Adr_Y \ Table: TEMPgcd
Save as <qr_mkt_TEMPgcd_lot_offname> \ Run
```

The freshly created <TEMPgcd\_lot\_offname> table contains 1969 entries

```
Query \ Make table query \ Design view
Name table as: TEMPgcd_lot_cityname
Field: ID \ Table: TEMPgcd
Field: Adr_X \ Table: TEMPgcd \ Criteria: =0
Field: Adr_Y \ Table: TEMPgcd
Field: Offname_lot_X \ Table: TEMPgcd \ Criteria: Is null
Field: Offname_lot_Y \ Table: TEMPgcd
Field: Cityname_lot_X \ Table: TEMPgcd \ Criteria: Is not null
Field: Cityname_lot_Y \ Table: TEMPgcd
Save as <qr_mkt_TEMPgcd_lot_cityname> \ Run
```

The freshly created <TEMPgcd\_lot\_cityname> table contains 328 entries

Repeat the process to create a table with entries having no postal code information but successful geocoding by municipality:

```
Query \ Make table query \ Design view
Name table as: TEMPgcd_muni
Field: ID \ Table: TEMPgcd
Field: Offname_lot_X \ Table: TEMPgcd \ Criteria: Is null
Field: Offname_lot_Y \ Table: TEMPgcd
Field: Cityname_lot_X \ Table: TEMPgcd \ Criteria: Is null
Field: Cityname_lot_Y \ Table: TEMPgcd
```

```
Field: Adr_X \ Table: TEMPgcd \ Criteria: =0  
Field: Adr_Y \ Table: TEMPgcd  
Field: Mun_X \ Table: TEMPgcd \ Criteria: <>0  
Field: Mun_Y \ Table: TEMPgcd  
Save as <qr_mkt_TEMPgcd_muni> \ Run
```

The freshly created <TEMPgcd\_muni> table contains 12 305 entries

Export the <TEMPgcd\_muni> ; <TEMPgcd\_lot\_cityname> ;  
<TEMPgcd\_lot\_offname>; <qr\_mkt\_TEMPgcd\_address> table in .dbf format.

Import the .dbf tables in ArcGIS. Display the points according to the geocoding result in the proper column and perform a spatial join with the FSA layer to assign each entry in a FSA.

Import in Access the resulting dbf tables (4) and rename them as <address>; <muni>;  
<cityname>; <offname>.

Create a new column in the <Gcd\_main> table called <spatialFSA>.

Perform update queries to bring the FSA assigned by the spatial join to the main  
<Gcd\_main> table and be able to attribute an urban/rural code to those entries.

```
Query \ Update query \ Design view  
Tables involved: <Gcd_main> and <address>  
Link by the filed: ID  
Field: spatialFSA \ Table: Gcd_main \ Update to: [address].[FSA]  
Field: FSA \ Table: address \ Criteria: Is not null  
Save query as: <qr_upd_adr_to_spatialFSA> \ Run
```

Repeat the process with the other 3 tables:

```
Query \ Update query \ Design view  
Tables involved: <Gcd_main> and <muni>  
Link by the filed: ID  
Field: spatialFSA \ Table: Gcd_main \ Update to: [muni].[FSA]  
Field: FSA \ Table: address \ Criteria: Is not null  
Save query as: <qr_upd_muni_to_spatialFSA> \ Run
```

```
Query \ Update query \ Design view  
Tables involved: <Gcd_main> and <cityname>  
Link by the filed: ID  
Field: spatialFSA \ Table: Gcd_main \ Update to: [cityname].[FSA]  
Field: FSA \ Table: address \ Criteria: Is not null  
Save query as: <qr_upd_cityname_to_spatialFSA> \ Run
```

```
Query \ Update query \ Design view
Tables involved: <Gcd main> and <offname>
Link by the filed: ID
Field: spatialFSA \ Table: Gcd main \ Update to: [offname].[FSA]
Field: FSA \ Table: address \ Criteria: Is not null
Save query as: <qr upd offname to spatialFSA> \ Run
```

Perform an update query to assign the letter 'R' to any entry having a zero as second digit of their postal code as determined by the spatial joint.

```
Query \ Update query \ Design view
Field: Urban code \ Table: Gcd main
Update to: 'R'
Field: spatialFSA \ Table: Gcd main \ Criteria: Like '0*'
Save as <qr upd r urban code from spatialFSA> \ Run
```

Perform an update query to assign the letter 'U' to any entry having a number between 1 and 9 as second digit of their postal code as determined by the spatial joint.

```
Query \ Update query \ Design view
Field: Urban code \ Table: Gcd main
Update to: 'U'
Field: spatialFSA \ Table: Gcd main \ Criteria: Like '1*' or Like '2*' or Like '3*' or Like '4*' or Like '5*' or Like '6*' or Like '7*' or Like '8*' or Like '9*'
Save as <qr upd u urban code from spatialFSA> \ Run
```

## 6.2. The precision sequences

Distinct precision sequences were created, depending if the entry is considered to be rural or urban. In both cases, the geocoding based on civic address being considered as the most precise localization and primary used if present.

Rural location:

Civic address ⇒ Lot/Concession ⇒ 6 digits postal code ⇒ 5 digits postal code ⇒ municipality ⇒ FSA

Urban location:

Civic address ⇒ 6 digits postal code ⇒ 5 digits postal code ⇒ FSA ⇒ Lot/Concession ⇒ Municipality

## 6.3. Selecting which geocoding result to use

Create <select\_X> and <select\_Y> columns to copy the geocoding results most appropriate for each entry.

Create <select\_code> column to mention which geocoding result was selected as the most useful one.

Use an update query to select and copy the entries having a successful geocoding result based on the civic address. The same query will also put the 'CIVICADR' mention in the <select\_code> column.

```
Query \ Update query \ Design view
Field: select_code \ Table: Gcd_main \ Update to: 'CIVICADR'
Field: Adr_X \ Table: Gcd_main \ Criteria: <>0
Field: Adr_Y \ Table: Gcd_main \ Criteria: <>0
Field: select_X \ Table: Gcd_main \ Update to: [Gcd_main].[Adr_X]
Field: select_Y \ Table: Gcd_main \ Update to: [Gcd_main].[Adr_Y]
Save as <qr upd select adr> \ Run
```

Repeat the process for the entries having no geocoding result by civic address but a successful geocoding result based on the six digits postal code and classified as 'urban'

```
Query \ Update query \ Design view
Field: select_code \ Table: Gcd_main \ Update to: 'PC6'
Field: urban_code \ Table: Gcd_main \ Criteria: ='U'
Field: Adr_X \ Table: Gcd_main \ Criteria: =0
Field: Adr_Y \ Table: Gcd_main \ Criteria: =0
Field: 6PC_x \ Table: Gcd_main \ Criteria: <>0
Field: 6PC_y \ Table: Gcd_main \ Criteria: <>0
Field: select_X \ Table: Gcd_main \ Update to: [Gcd_main].[6PC_x]
Field: select_Y \ Table: Gcd_main \ Update to: [Gcd_main].[6PC_y]
Save as <qr upd select PC6urban> \ Run
```

Use the following query for the 'urban' entries having no geocoding result by civic address or by six digits postal code:

```
Query \ Update query \ Design view
Field: select_code \ Table: Gcd_main \ Update to: 'PC5'
Field: urban_code \ Table: Gcd_main \ Criteria: ='U'
Field: Adr_X \ Table: Gcd_main \ Criteria: =0
Field: Adr_Y \ Table: Gcd_main \ Criteria: =0
Field: 6PC_x \ Table: Gcd_main \ Criteria: =0
Field: 6PC_y \ Table: Gcd_main \ Criteria: =0
Field: 5PC_x \ Table: Gcd_main \ Criteria: <>0
Field: 5PC_y \ Table: Gcd_main \ Criteria: <>0
Field: select_X \ Table: Gcd_main \ Update to: [Gcd_main].[5PC_x]
Field: select_Y \ Table: Gcd_main \ Update to: [Gcd_main].[5PC_y]
Save as <qr upd select PC5urban> \ Run
```

Use the following query for the 'urban' entries having no geocoding result by civic address, by six or five digits postal code:

```
Query \ Update query \ Design view
Field: select_code \ Table: Gcd_main \ Update to: FSA
Field: urban_code \ Table: Gcd_main \ Criteria: = 'U'
Field: Adr_X \ Table: Gcd_main \ Criteria: = 0
Field: Adr_Y \ Table: Gcd_main \ Criteria: = 0
Field: 6PC_x \ Table: Gcd_main \ Criteria: = 0
Field: 6PC_y \ Table: Gcd_main \ Criteria: = 0
Field: 5PC_x \ Table: Gcd_main \ Criteria: = 0
Field: 5PC_y \ Table: Gcd_main \ Criteria: = 0
Field: RTA_x \ Table: Gcd_main \ Criteria: <> 0
Field: RTA_y \ Table: Gcd_main \ Criteria: <> 0
Field: select_X \ Table: Gcd_main \ Update to: [Gcd_main].[RTA_x]
Field: select_Y \ Table: Gcd_main \ Update to: [Gcd_main].[RTA_y]
Save as <qr_upd_select_FSAurban> \ Run
```

Use the following queries to use the lot and concession results for the 'urban' entries having no successful geocoding by civic address or postal code:

```
Query \ Update query \ Design view
Field: select_code \ Table: Gcd_main \ Update to: LOTCON
Field: urban_code \ Table: Gcd_main \ Criteria: = 'U'
Field: Adr_X \ Table: Gcd_main \ Criteria: = 0
Field: Adr_Y \ Table: Gcd_main \ Criteria: = 0
Field: 6PC_x \ Table: Gcd_main \ Criteria: = 0
Field: 6PC_y \ Table: Gcd_main \ Criteria: = 0
Field: 5PC_x \ Table: Gcd_main \ Criteria: = 0
Field: 5PC_y \ Table: Gcd_main \ Criteria: = 0
Field: RTA_x \ Table: Gcd_main \ Criteria: = 0
Field: RTA_y \ Table: Gcd_main \ Criteria: = 0
Field: Offname_lot_x \ Table: Gcd_main \ Criteria: Is Not Null
Field: Offname_lot_y \ Table: Gcd_main \ Criteria: Is Not Null
Field: select_X \ Table: Gcd_main \ Update to: [Gcd_main].[Offname_lot_x]
Field: select_Y \ Table: Gcd_main \ Update to: [Gcd_main].[Offname_lot_y]
Save as <qr_upd_select_LOTCONoffurban> \ Run
```

```
Query \ Update query \ Design view
Field: select_code \ Table: Gcd_main \ Update to: LOTCON
Field: urban_code \ Table: Gcd_main \ Criteria: = 'U'
Field: Adr_X \ Table: Gcd_main \ Criteria: = 0
Field: Adr_Y \ Table: Gcd_main \ Criteria: = 0
Field: 6PC_x \ Table: Gcd_main \ Criteria: = 0
Field: 6PC_y \ Table: Gcd_main \ Criteria: = 0
```

```

Field: 5PC_x \ Table: Gcd_main \ Criteria: =0
Field: 5PC_y \ Table: Gcd_main \ Criteria: =0
Field: RTA_x \ Table: Gcd_main \ Criteria: =0
Field: RTA_y \ Table: Gcd_main \ Criteria: =0
Field: Offname_lot_x \ Table: Gcd_main \ Criteria: Is Null
Field: Offname_lot_y \ Table: Gcd_main \ Criteria: Is Null
Field: Cityname_lot_x \ Table: Gcd_main \ Criteria: Is Not Null
Field: Cityname_lot_y \ Table: Gcd_main \ Criteria: Is Not Null
Field: select_X \ Table: Gcd_main \ Update to: [Gcd_main].[Cityname_lot_x]
Field: select_Y \ Table: Gcd_main \ Update to: [Gcd_main].[Cityname_lot_y]
Save as <qr upd select LOTCONCityurban> \ Run

```

Use the following query to select the municipality geocoding result when an 'urban' entry does not have any successful geocoding result by civic address, postal code or lot/concession numbers.

```

Query \ Update query \ Design view
Field: select_code \ Table: Gcd_main \ Update to: 'MUN'
Field: urban_code \ Table: Gcd_main \ Criteria: ='U'
Field: Adr_X \ Table: Gcd_main \ Criteria: =0
Field: Adr_Y \ Table: Gcd_main \ Criteria: =0
Field: 6PC_x \ Table: Gcd_main \ Criteria: =0
Field: 6PC_y \ Table: Gcd_main \ Criteria: =0
Field: 5PC_x \ Table: Gcd_main \ Criteria: =0
Field: 5PC_y \ Table: Gcd_main \ Criteria: =0
Field: RTA_x \ Table: Gcd_main \ Criteria: =0
Field: RTA_y \ Table: Gcd_main \ Criteria: =0
Field: Offname_lot_x \ Table: Gcd_main \ Criteria: Is Null
Field: Offname_lot_y \ Table: Gcd_main \ Criteria: Is Null
Field: Cityname_lot_x \ Table: Gcd_main \ Criteria: Is Null
Field: Cityname_lot_y \ Table: Gcd_main \ Criteria: Is Null
Field: Mun_x \ Table: Gcd_main \ Criteria: <>0
Field: Mun_y \ Table: Gcd_main \ Criteria: <>0
Field: select_X \ Table: Gcd_main \ Update to: [Gcd_main].[Mun_x]
Field: select_Y \ Table: Gcd_main \ Update to: [Gcd_main].[Mun_y]
Save as <qr upd select MUNurban> \ Run

```

Repeat the process for the entries having a 'Rural' code, but this time in a different order as the precision resulting from geocoding processes differs in rural areas.

Use the following queries to use the lot and concession results for the 'rural' entries having no successful geocoding by civic address :

```

Query \ Update query \ Design view
Field: select_code \ Table: Gcd_main \ Update to: 'LOTCON'
Field: urban_code \ Table: Gcd_main \ Criteria: ='R'

```

```

Field: Adr_X \ Table: Gcd_main \ Criteria: =0
Field: Adr_Y \ Table: Gcd_main \ Criteria: =0
Field: Offname_lot_x \ Table: Gcd_main \ Criteria: Is Not Null
Field: Offname_lot_y \ Table: Gcd_main \ Criteria: Is Not Null
Field: select_X \ Table: Gcd_main \ Update to: [Gcd_main].[Offname_lot_x]
Field: select_Y \ Table: Gcd_main \ Update to: [Gcd_main].[Offname_lot_y]
Save as <qr_upd_select_LOTCONoffrural> \ Run

```

```

Query \ Update query \ Design view
Field: select_code \ Table: Gcd_main \ Update to: 'LOTCON'
Field: urban_code \ Table: Gcd_main \ Criteria: = 'R'
Field: Adr_X \ Table: Gcd_main \ Criteria: =0
Field: Adr_Y \ Table: Gcd_main \ Criteria: =0
Field: Offname_lot_x \ Table: Gcd_main \ Criteria: Is Null
Field: Offname_lot_y \ Table: Gcd_main \ Criteria: Is Null
Field: Cityname_lot_x \ Table: Gcd_main \ Criteria: Is Not Null
Field: Cityname_lot_y \ Table: Gcd_main \ Criteria: Is Not Null
Field: select_X \ Table: Gcd_main \ Update to: [Gcd_main].[Cityname_lot_x]
Field: select_Y \ Table: Gcd_main \ Update to: [Gcd_main].[Cityname_lot_y]
Save as <qr_upd_select_LOTCONCityrural> \ Run

```

Repeat the process for the 'rural' entries having no geocoding result by civic address or lot/concession numbers but a successful geocoding result based on the six digits postal code

```

Query \ Update query \ Design view
Field: select_code \ Table: Gcd_main \ Update to: 'PC6'
Field: urban_code \ Table: Gcd_main \ Criteria: = 'R'
Field: Adr_X \ Table: Gcd_main \ Criteria: =0
Field: Adr_Y \ Table: Gcd_main \ Criteria: =0
Field: Offname_lot_x \ Table: Gcd_main \ Criteria: Is Null
Field: Offname_lot_y \ Table: Gcd_main \ Criteria: Is Null
Field: Cityname_lot_x \ Table: Gcd_main \ Criteria: Is Null
Field: Cityname_lot_y \ Table: Gcd_main \ Criteria: Is Null
Field: 6PC_x \ Table: Gcd_main \ Criteria: <>0
Field: 6PC_y \ Table: Gcd_main \ Criteria: <>0
Field: select_X \ Table: Gcd_main \ Update to: [Gcd_main].[6PC_x]
Field: select_Y \ Table: Gcd_main \ Update to: [Gcd_main].[6PC_y]
Save as <qr_upd_select_PC6rural> \ Run

```

Use the following query for the ones having a successful geocoding based on five digits postal code

```

Query \ Update query \ Design view
Field: select_code \ Table: Gcd_main \ Update to: 'PC5'
Field: urban_code \ Table: Gcd_main \ Criteria: = 'R'

```

```

Field: Adr_X \ Table: Gcd_main \ Criteria: =0
Field: Adr_Y \ Table: Gcd_main \ Criteria: =0
Field: Offname_lot_x \ Table: Gcd_main \ Criteria: Is Null
Field: Offname_lot_y \ Table: Gcd_main \ Criteria: Is Null
Field: Cityname_lot_x \ Table: Gcd_main \ Criteria: Is Null
Field: Cityname_lot_y \ Table: Gcd_main \ Criteria: Is Null
Field: 6PC_x \ Table: Gcd_main \ Criteria: =0
Field: 6PC_y \ Table: Gcd_main \ Criteria: =0
Field: 5PC_x \ Table: Gcd_main \ Criteria: <>0
Field: 5PC_y \ Table: Gcd_main \ Criteria: <>0
Field: select_X \ Table: Gcd_main \ Update to: [Gcd_main].[5PC_x]
Field: select_Y \ Table: Gcd_main \ Update to: [Gcd_main].[5PC_y]
Save as <qr_upd_select_PC5rural> \ Run

```

Use the following query when a 'rural' entry does not have any successful geocoding result by civic address, lot/concession numbers or postal code.

```

Query \ Update query \ Design view
Field: select_code \ Table: Gcd_main \ Update to: 'MUN'
Field: urban_code \ Table: Gcd_main \ Criteria: ='R'
Field: Adr_X \ Table: Gcd_main \ Criteria: =0
Field: Adr_Y \ Table: Gcd_main \ Criteria: =0
Field: 6PC_x \ Table: Gcd_main \ Criteria: =0
Field: 6PC_y \ Table: Gcd_main \ Criteria: =0
Field: 5PC_x \ Table: Gcd_main \ Criteria: =0
Field: 5PC_y \ Table: Gcd_main \ Criteria: =0
Field: Offname_lot_x \ Table: Gcd_main \ Criteria: Is Null
Field: Offname_lot_y \ Table: Gcd_main \ Criteria: Is Null
Field: Cityname_lot_x \ Table: Gcd_main \ Criteria: Is Null
Field: Cityname_lot_y \ Table: Gcd_main \ Criteria: Is Null
Field: Mun_x \ Table: Gcd_main \ Criteria: <>0
Field: Mun_y \ Table: Gcd_main \ Criteria: <>0
Field: select_X \ Table: Gcd_main \ Update to: [Gcd_main].[Mun_x]
Field: select_Y \ Table: Gcd_main \ Update to: [Gcd_main].[Mun_y]
Save as <qr_upd_select_MUNrural> \ Run

```

Use the following query for the 'rural' entries having no geocoding result by civic address, lot/concession numbers either by six or five digits postal code or by municipality:

```

Query \ Update query \ Design view
Field: select_code \ Table: Gcd_main \ Update to: 'FSA'
Field: urban_code \ Table: Gcd_main \ Criteria: ='R'
Field: Adr_X \ Table: Gcd_main \ Criteria: =0
Field: Adr_Y \ Table: Gcd_main \ Criteria: =0

```

```

Field: 6PC_x \ Table: Gcd_main \ Criteria: =0
Field: 6PC_y \ Table: Gcd_main \ Criteria: =0
Field: 5PC_x \ Table: Gcd_main \ Criteria: =0
Field: 5PC_y \ Table: Gcd_main \ Criteria: =0
Field: Offname_lot_x \ Table: Gcd_main \ Criteria: Is Null
Field: Offname_lot_y \ Table: Gcd_main \ Criteria: Is Null
Field: Cityname_lot_x \ Table: Gcd_main \ Criteria: Is Null
Field: Cityname_lot_y \ Table: Gcd_main \ Criteria: Is Null
Field: Mun_x \ Table: Gcd_main \ Criteria: =0
Field: Mun_y \ Table: Gcd_main \ Criteria: =0
Field: RTA_x \ Table: Gcd_main \ Criteria: <>0
Field: RTA_y \ Table: Gcd_main \ Criteria: <>0
Field: select_X \ Table: Gcd_main \ Update to: [Gcd_main].[RTA_x]Field:
select_Y \ Table: Gcd_main \ Update to: [Gcd_main].[RTA_y]
Save as <qr_upd_select_FSArural> \ Run

```

Use a final query to assign the string 'NOGCD' to the entries where the geocoding process was unsuccessful.

```

Query \ Update query \ Design view
Field: select_code \ Table: Gcd_main \ Update to: 'NOGCD'
Field: Adr_X \ Table: Gcd_main \ Criteria: =0
Field: Adr_Y \ Table: Gcd_main \ Criteria: =0
Field: 6PC_x \ Table: Gcd_main \ Criteria: =0
Field: 6PC_y \ Table: Gcd_main \ Criteria: =0
Field: 5PC_x \ Table: Gcd_main \ Criteria: =0
Field: 5PC_y \ Table: Gcd_main \ Criteria: =0
Field: Offname_lot_x \ Table: Gcd_main \ Criteria: Is Null
Field: Offname_lot_y \ Table: Gcd_main \ Criteria: Is Null
Field: Cityname_lot_x \ Table: Gcd_main \ Criteria: Is Null
Field: Cityname_lot_y \ Table: Gcd_main \ Criteria: Is Null
Field: Mun_x \ Table: Gcd_main \ Criteria: =0
Field: Mun_y \ Table: Gcd_main \ Criteria: =0
Field: RTA_x \ Table: Gcd_main \ Criteria: =0
Field: RTA_y \ Table: Gcd_main \ Criteria: =0
Save as <qr_upd_select_NOGCD> \ Run

```

## **7. SELECTED GEOCODING RESULTS**

The columns <select\_x> and <select\_Y> of the <GCD\_main> table contain the geocoding results the most appropriate to use.

Of the 129 018 entries in the initial database,

For 30,329 (~24%) entries, geocoding using the civic address was considered as the best option;

For 12,375 (~10%) entries, geocoding using the six digits postal code was considered as the best option;

For 25,042 (~19%) entries, geocoding using the five digits postal code was considered as the best option;

For 21,031 (~16%) entries, geocoding using the tree digits postal code was considered as the best option;

For 8,115 (~ 6%) entries, geocoding according to lot/concession numbers was considered as the best option;

For 25,166 (~ 20%) entries, geocoding to the centroid of their municipality was considered as the best option;

In total : 122,058 entries (~95%) were geocoded while 6,959 (~5%) were not.

## **8. VISUALISING THE POINTS ON A MAP**

Using Access, records were sorted out to:

- exclude the ones having either 'POBOXONLY' or 'NOGEO' in the <gcdcode> column
- exclude the ones having the 'NWT' label in the 'WT' column

Selected records were brought in the ArcGIS geographical information system software using the following procedure:

Export Access database to .dbf file;

Import the .dbf table in ArcGIS;

Use the 'Add even theme' function of ArcGIS to display the X and Y columns of the .dbf table; Save as a 'shapefile' and create map.

## 9. APPENDIXES

### 9.1. Appendix A : Public Health Units list

<b>Health Unit Number</b>	<b>Health Unit Name</b>
2226	Algoma
2227	Brant County
2240	Chatham-Kent Public Health Division
2230	Durham Region Health Department
2258	Eastern Ontario
2231	Elgin-St. Thomas
2233	Grey Bruce
2238	Hastings & Prince Edward Counties
2234	Haldimand-Norfolk
2235	Haliburton, Kawartha, Pine Ridge District
2236	Halton Region Health Department
2237	City of Hamilton-Public Health & Community Services Dept.
2239	Huron County
2241	Kingston, Frontenac And Lennox & Addington
2242	County of Lambton Community Health Services Department
2243	Leeds, Grenville and Lanark District
2244	Middlesex-London
2245	Muskoka-Parry Sound
2246	Regional Niagara Public Health Department
2247	North Bay & District
2249	Northwestern
2251	City of Ottawa-Public Health & Long Term Care Branch
2252	County of Oxford
2253	Regional Municipality of Peel Health Department
2254	Perth District
2255	Peterborough County-City
2256	Porcupine
2257	Renfrew County & District
2260	Simcoe County District
2261	Sudbury & District
2262	Thunder Bay District
2263	Timiskaming
3895	Toronto Public Health
2265	Region of Waterloo, Public Health
2266	Wellington-Dufferin-Guelph
2268	Windsor-Essex County
2270	York Region Health Services Department

## 9.2. Appendix B : Status Codes list

<b>Status Code</b>	<b>Status Integer</b>
1	No significant evidence of bacterial contamination
2	Significant evidence of bacterial contamination
3	UNSAFE TO DRINK, evidence of sewage contamination
4	Overgrown - May be unsafe to drink
5	The sample was taken from an unprotected source, such as a lake or river
6	The client requested chemical testing
7	The sample was too old
8	The bottle was received broken or damaged
10	The appearance or odour makes the sample unacceptable as drinking water
11	The requisition was received separated from the sample bottle
12	The sample was not collected in the proper bottle
13	The sample was received frozen
14	The sample was collected from a hot water tap
15	Insufficient sample was submitted
17	The sample leaked in transit
18	The sample was collected from a source located outside of Ontario
19	The sample was collected from a municipal water supply
20	Insufficient information was supplied on the sample
21	Not tested for some other reason
22	Overgrown with evidence of sewage contamination

### **9.3. Appendix C : Statistics on rejection of untested water samples**

All statistics were performed on the database containing 127,536 records, being the one cleaned for doubles, PO Box as only spatial info, no usable spatial info, unusable lot/concession information.

To obtain a PHUcalculated (Public Health Unit):

Went back to the initial database after the cleaning process but before narrowing upon water test results.

Performed geocoding, using the same process as described earlier. Using ArcGIS, performed a spatial join to assign a PHU calculated to each point depending on its location.

In Access, created a reference table to convert RSS names in PHU codes. Updated a new 'PHUcalcd' column with the PHU codes when available, if no PHU code was calculated (point did not fall within the boundaries of a PHU or point was not geocoded), the declared PHU was used.

## Distribution of Status Codes by PHU (Public Health Unit)

PHU	Status Code																				Total
	1	2	3	4	5	6	7	8	10	11	12	13	14	15	17	18	19	20	21	22	
0000	198	47	17	11	0	0	7	3	1	0	0	0	1	0	1	16	0	101	16	0	419
2226	1350	383	92	82	0	0	16	7	1	0	0	0	0	0	0	0	2	30	24	0	1987
2227	476	113	32	26	0	0	16	7	1	0	0	0	0	0	0	0	0	38	5	2	716
2230	1999	691	171	89	0	0	63	41	3	0	0	0	0	0	2	0	0	160	38	1	3258
2231	654	229	40	82	0	0	11	12	1	0	0	1	0	0	0	0	0	52	7	8	1097
2233	7162	1704	481	230	2	0	192	114	8	2	1	0	0	0	0	0	0	597	52	31	10576
2234	1143	312	135	55	0	0	46	11	9	0	0	3	0	0	1	0	0	104	23	5	1847
2235	8485	2264	617	244	2	0	96	94	5	1	1	10	0	1	2	0	0	270	92	3	12187
2236	1371	544	161	22	0	0	62	1	7	0	2	0	0	0	1	0	2	154	53	2	2382
2237	1261	385	103	27	0	0	32	2	9	0	1	1	1	0	1	0	0	85	19	0	1927
2238	3771	1412	468	109	0	0	66	29	7	0	0	0	0	0	0	0	0	314	35	30	6241
2239	1359	270	48	37	0	0	42	10	1	0	0	1	0	0	0	0	0	65	12	5	1850
2240	735	75	15	21	1	0	1	1	12	0	0	0	0	0	0	0	0	19	1	1	882
2241	3320	1061	392	73	0	0	39	22	5	0	0	1	0	1	1	0	1	77	12	17	5022
2242	202	63	13	20	0	0	13	11	1	0	0	0	0	0	0	0	0	14	4	2	343
2243	5361	1445	410	196	1	0	70	46	68	0	3	4	0	0	0	0	0	172	20	9	7805
2244	1248	377	73	101	0	0	28	18	4	0	4	1	0	0	0	0	0	93	9	8	1964
2245	4912	1547	415	230	2	0	138	28	34	18	0	2	2	0	2	0	0	313	101	21	7765
2246	869	267	146	22	0	0	32	4	19	0	0	0	0	0	0	0	0	62	33	0	1454
2247	1494	453	122	34	0	0	42	16	3	0	0	0	0	1	0	1	0	134	51	6	2357
2249	563	182	80	37	0	0	7	1	11	0	0	0	0	0	2	0	0	22	14	23	942
2251	3483	772	176	235	0	0	46	36	68	0	1	6	0	0	0	0	0	296	38	0	5157
2252	933	237	42	34	0	0	24	18	2	0	0	3	0	0	2	0	0	86	3	5	1389
2253	760	235	43	24	0	0	13	5	3	0	0	0	0	0	0	0	1	28	15	7	1134
2254	843	206	36	33	0	0	25	18	1	0	0	0	0	0	0	0	0	48	12	0	1222
2255	4485	936	274	96	0	0	62	54	3	1	0	3	0	0	2	0	0	217	29	2	6164
2256	353	59	15	19	0	0	6	3	0	0	0	0	0	0	1	0	0	15	2	0	473
2257	2944	826	218	170	0	1	68	20	73	0	0	0	0	0	1	0	0	239	17	2	4579
2258	3616	1377	422	377	1	2	64	40	113	0	2	0	0	3	1	0	0	197	38	11	6264
2260	8050	2757	391	360	6	0	227	76	35	36	1	0	1	1	1	0	1	407	133	18	12501
2261	1750	410	200	37	6	0	17	37	3	0	0	0	2	1	0	0	0	48	28	10	2549
2262	1458	632	101	69	0	0	11	7	8	0	0	0	0	0	9	0	3	57	16	33	2404
2263	605	103	80	22	2	0	14	0	0	0	0	0	0	0	2	0	0	27	5	4	864
2265	854	311	57	51	0	0	26	13	0	0	2	0	0	0	1	0	0	84	12	6	1417
2266	3074	871	142	205	0	0	132	50	5	0	1	0	0	0	1	0	0	215	38	7	4741
2268	90	21	7	0	0	0	4	0	0	0	0	0	0	0	0	0	0	1	0	0	123
2270	2415	622	129	83	0	0	21	5	2	0	1	0	0	0	1	0	0	126	43	16	3464
3895	36	8	3	2	0	0	1	0	1	0	0	0	0	0	0	0	10	6	1	2	70
Total	83682	24207	6367	3565	23	3	1780	860	527	58	20	36	7	8	35	17	20	4973	1051	297	4327536

Distribution of Status Codes by PHU calculated (Public Health Unit)

PHU calculated	Status Code																				Total
	1	2	3	4	5	6	7	8	10	11	12	13	14	15	17	18	19	20	21	22	
0000	31	12	6	3	0	0	3	0	0	0	0	0	0	0	1	15	0	39	7	0	117
2226	1311	377	91	75	0	0	17	8	1	0	0	0	0	0	0	0	2	29	24	0	1935
2227	424	106	29	22	0	0	12	6	0	0	0	0	0	0	0	0	0	29	5	2	635
2230	1901	669	160	88	0	0	59	38	3	0	1	0	0	0	2	0	0	147	38	1	3107
2231	789	258	50	92	0	0	15	15	1	0	0	1	0	0	0	0	0	65	9	10	1305
2233	7199	1725	490	244	1	0	195	116	8	0	1	0	0	0	0	0	0	600	51	30	10660
2234	1260	331	142	54	0	0	52	14	10	0	0	2	0	0	1	0	0	120	25	6	2017
2235	8140	2078	580	226	2	0	94	86	4	1	0	7	0	1	2	0	0	261	91	4	11577
2236	1423	534	156	24	0	0	66	2	7	0	2	0	1	0	1	0	2	146	56	2	2422
2237	1001	320	79	26	0	0	23	3	6	0	1	1	1	0	1	0	0	74	13	0	1549
2238	4307	1590	513	135	0	0	78	41	17	0	0	3	0	0	1	0	0	345	38	30	7098
2239	1303	287	48	47	0	0	48	9	0	0	0	1	0	0	0	0	0	65	10	4	1822
2240	735	82	17	25	1	0	1	2	12	0	0	0	0	0	0	0	0	21	1	1	898
2241	3554	1152	425	84	0	0	41	21	12	0	0	1	0	1	1	0	1	79	14	17	5403
2242	179	53	10	13	0	0	11	10	1	0	0	0	0	0	0	0	0	14	4	3	298
2243	4910	1321	377	188	1	0	64	43	64	0	3	4	0	0	0	0	0	167	21	8	7171
2244	1097	334	63	85	0	0	22	18	4	0	4	0	0	0	1	0	0	81	13	6	1728
2245	5006	1577	418	236	3	0	138	31	35	18	0	1	2	0	2	0	0	319	102	20	7908
2246	1260	377	176	39	0	0	43	4	24	0	0	0	0	0	0	0	0	81	34	0	2058
2247	1513	461	121	35	0	0	40	18	3	0	0	0	0	1	0	1	0	132	52	6	2383
2249	576	178	77	35	0	0	7	1	10	0	0	0	0	0	2	0	0	23	12	23	944
2251	3034	673	146	186	1	0	41	35	62	0	1	4	0	1	0	0	0	252	36	0	4472
2252	859	221	35	36	0	0	24	14	2	0	0	4	0	0	2	0	0	75	2	2	1276
2253	413	166	24	21	0	0	8	2	1	0	0	0	0	0	0	0	2	24	7	2	670
2254	839	198	35	26	0	0	20	12	1	0	0	0	0	0	0	0	0	40	9	1	1181
2255	4598	1005	297	107	0	0	67	55	3	1	0	5	0	0	1	0	0	225	30	2	6396
2256	321	56	19	15	1	0	5	3	0	0	0	0	0	0	1	0	0	15	3	0	439
2257	2693	762	189	159	0	1	59	17	61	0	0	0	0	0	1	0	0	217	17	4	4180
2258	4356	1536	458	427	1	2	72	45	119	0	2	2	0	3	1	1	0	271	47	12	7355
2260	8195	2786	405	344	5	0	233	74	32	38	1	0	1	0	1	0	1	429	138	21	12704
2261	1702	408	192	35	5	0	21	37	4	0	0	0	1	1	0	0	0	47	25	10	2488
2262	1453	625	105	69	0	0	14	8	8	0	0	0	0	0	9	0	3	59	16	33	2402
2263	534	92	69	20	2	0	9	0	0	0	0	0	0	0	1	0	0	27	5	4	763
2265	881	321	58	50	0	0	21	15	1	0	2	0	0	0	1	0	0	88	13	6	1457
2266	3433	930	172	214	0	0	136	49	7	0	1	0	0	0	1	0	0	229	39	14	5225
2268	91	23	8	0	0	0	4	0	0	0	0	0	0	0	0	0	0	1	0	0	127
2270	2275	567	125	78	0	0	16	7	3	0	1	0	1	0	1	0	2	131	41	12	3260
3895	66	16	2	2	0	0	1	1	1	0	0	0	0	0	0	0	7	6	3	1	106
Total	83682	24207	6367	3565	23	3	1780	860	527	58	20	36	7	8	35	17	20	4973	1051	297	127536

Distribution of Status Codes by laboratory

Laboratory	Satus Code																				Total
	1	2	3	4	5	6	7	8	10	11	12	13	14	15	17	18	19	20	21	22	
Central	4568	1185	287	182	1	0	66	6	5	1	1	0	1	0	3	0	11	257	82	39	6695
Hamilton	4228	1441	519	77	0	0	158	5	49	0	3	8	0	0	2	0	2	355	123	0	6970
Kingston	8807	3196	1005	224	0	0	127	69	17	0	1	2	0	3	1	0	1	413	68	62	13996
London	17419	4484	1012	894	1	0	551	289	22	0	8	7	2	0	5	0	0	1371	151	79	26295
Orillia	12637	4285	749	540	8	0	346	105	71	57	1	0	3	1	3	0	1	709	247	32	19795
Ottawa	12716	3500	1018	913	2	3	219	118	318	0	5	9	0	1	2	12	0	881	93	7	19817
Peterborough	14919	3819	1069	416	2	0	199	194	7	0	1	10	0	1	5	0	0	628	143	1	21414
Sault Ste. Marie	1346	378	93	81	0	0	17	7	1	0	0	0	0	0	0	0	2	29	24	0	1978
Sudbury	3148	836	313	63	6	0	54	55	6	0	0	0	1	2	0	4	0	179	82	16	4765
Thunder Bay	2034	820	182	108	0	0	18	9	19	0	0	0	0	0	11	0	3	81	31	56	3372
Timmins	1030	162	96	45	2	0	19	3	0	0	0	0	0	0	3	1	0	51	6	4	1422
Windsor	830	101	24	22	1	0	6	0	12	0	0	0	0	0	0	0	0	19	1	1	1017
<b>Total</b>	<b>83682</b>	<b>24207</b>	<b>6367</b>	<b>3565</b>	<b>23</b>	<b>3</b>	<b>1780</b>	<b>860</b>	<b>527</b>	<b>58</b>	<b>20</b>	<b>36</b>	<b>7</b>	<b>8</b>	<b>35</b>	<b>17</b>	<b>20</b>	<b>4973</b>	<b>1051</b>	<b>297</b>	<b>127536</b>

Chi-square test by Laboratory:

Statistic	DF	Value	Prob
Chi-Square	11	544.1507	<.0001
Likelihood Ratio Chi-Square	11	563.3377	<.0001
Mantel-Haenszel Chi-Square	1	53.3909	<.0001
Phi Coefficient		0.0653	
Contingency Coefficient		0.0652	
Cramer's V		0.0653	

Sample Size = 127536

Chi-square test by PHU:

Statistic	DF	Value	Prob
Chi-Square	37	1315.7496	<.0001
Likelihood Ratio Chi-Square	37	1142.0206	<.0001
Mantel-Haenszel Chi-Square	1	1.1572	0.2820
Phi Coefficient		0.1016	
Contingency Coefficient		0.1011	
Cramer's V		0.1016	

Sample Size = 127536

Chi-square test by PHU calculated:

Statistic	DF	Value	Prob
Chi-Square	37	1159.2161	<.0001
Likelihood Ratio Chi-Square	37	983.0219	<.0001
Mantel-Haenszel Chi-Square	1	0.0002	0.9882
Phi Coefficient		0.0953	
Contingency Coefficient		0.0949	
Cramer's V		0.0953	

Sample Size = 127536

Confidence interval (95%) by Laboratory: (rejected samples / total samples)

<b>LAB</b>	<b>RISK</b>	<b>CI-Low</b>	<b>CI-High</b>
Windsor	0.0383	0.0274	0.0521
Timmins	0.0598	0.0480	0.0734
ThunderBay	0.0510	0.0438	0.0590
Peterborough	0.0556	0.0525	0.0587
Sudbury	0.0816	0.0740	0.0898
SStMarie	0.0404	0.0322	0.0501
Ottawa	0.0839	0.0801	0.0879
Orillia	0.0784	0.0747	0.0822
London	0.0915	0.0881	0.0951
Kingston	0.0502	0.0466	0.0539
Hamilton	0.1011	0.0942	0.1085
Central	0.0648	0.0590	0.0710

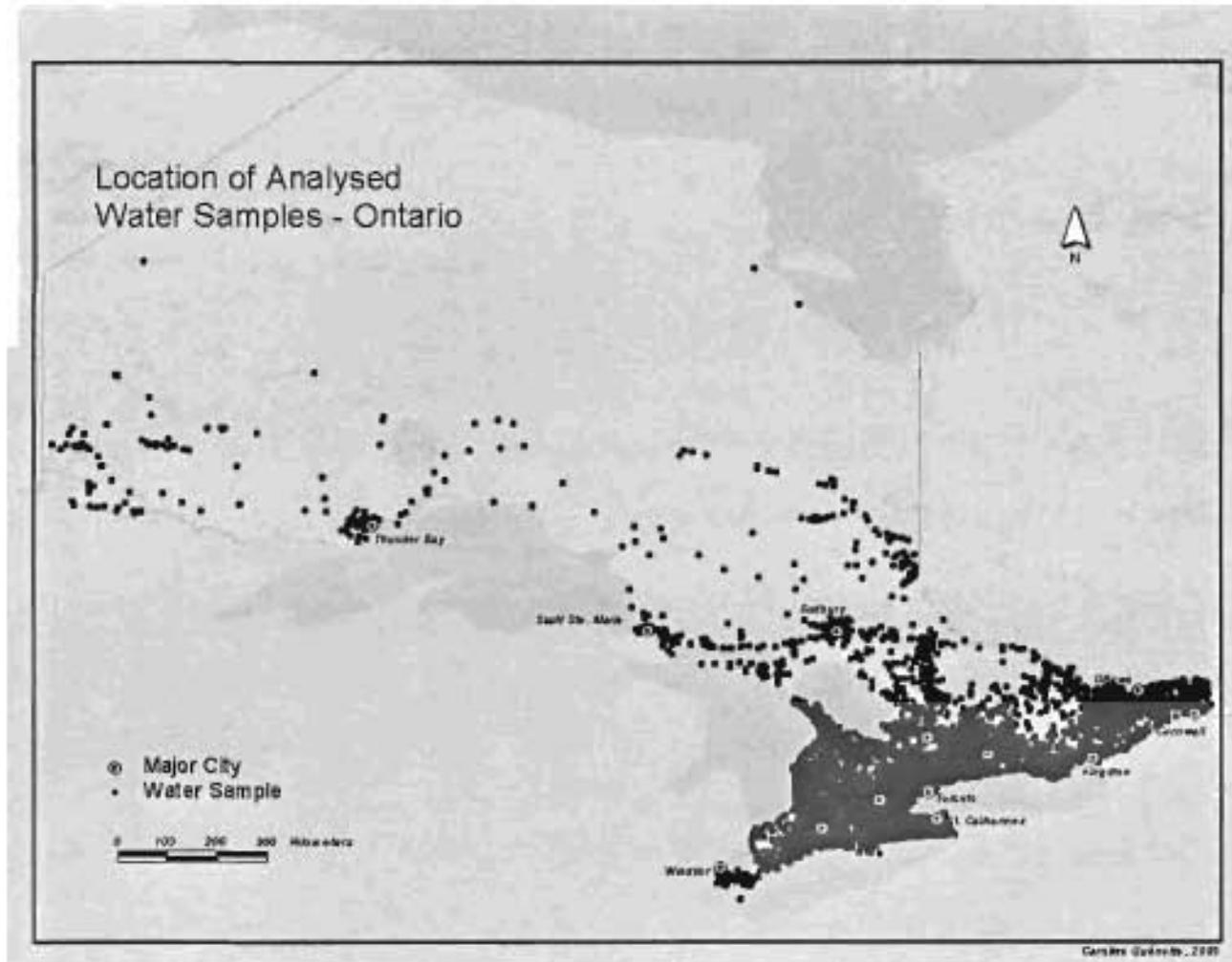
Confidence interval (95%) by PHU: (rejected samples / total samples)

<b>PHU</b>	<b>RISK</b>	<b>CI-Low</b>	<b>CI-High</b>
0000	0.3484	0.3028	0.3962
2226	0.0403	0.0321	0.0499
2227	0.0936	0.0936	0.1173
2230	0.0942	0.0844	0.1048
2231	0.0766	0.0615	0.0939
2233	0.0915	0.0861	0.0972
2234	0.1067	0.0929	0.1216
2235	0.0471	0.0434	0.0510
2236	0.1184	0.1057	0.1320
2237	0.0784	0.0668	0.0913
2238	0.0723	0.0660	0.0790
2239	0.0708	0.0595	0.0835
2240	0.0397	0.0278	0.0548
2241	0.0317	0.0270	0.0369
2242	0.1254	0.0922	0.1651
2243	0.0492	0.0445	0.0542
2244	0.0799	0.0683	0.0928
2245	0.0824	0.0764	0.0888
2246	0.1032	0.0880	0.1200
2247	0.1052	0.0931	0.1183
2249	0.0605	0.0461	0.0777
2251	0.0952	0.0873	0.1035
2252	0.0994	0.0841	0.1163
2253	0.0573	0.0445	0.0725
2254	0.0851	0.0701	0.1022
2255	0.0602	0.0544	0.0664
2256	0.0571	0.0380	0.0820
2257	0.0915	0.0833	0.1002
2258	0.0736	0.0672	0.0803
2260	0.0740	0.0695	0.0787
2261	0.0557	0.0471	0.0653
2262	0.0462	0.0381	0.0553
2263	0.0579	0.0433	0.0756
2265	0.0974	0.0824	0.1140
2266	0.0932	0.0851	0.1019
2268	0.0407	0.0133	0.0923
2270	0.0574	0.0499	0.0657
3895	0.2714	0.1720	0.3910

Confidence interval (95%) by PHU calculated: (rejected samples / total samples)

<b>PHU</b>	<b>RISK</b>	<b>CI-Low</b>	<b>CI-High</b>
0000	0.5556	0.4608	0.6474
2226	0.0419	0.0334	0.0518
2227	0.0819	0.0618	0.1060
2230	0.0927	0.0827	0.1034
2231	0.0812	0.0670	0.0974
2233	0.0912	0.0858	0.0968
2234	0.1111	0.0977	0.1256
2235	0.0474	0.0436	0.0515
2236	0.1168	0.1043	0.1303
2237	0.0794	0.0664	0.0940
2238	0.0737	0.0677	0.0800
2239	0.0730	0.0615	0.0859
2240	0.0423	0.0301	0.0576
2241	0.0316	0.0271	0.0367
2242	0.1342	0.0977	0.1783
2243	0.0512	0.0462	0.0565
2244	0.0828	0.0702	0.0968
2245	0.0823	0.0764	0.0886
2246	0.0904	0.0783	0.1036
2247	0.1037	0.0917	0.1166
2249	0.0583	0.0442	0.0752
2251	0.0968	0.0883	0.1059
2252	0.0964	0.0808	0.1139
2253	0.0657	0.0481	0.0872
2254	0.0694	0.0556	0.0855
2255	0.0605	0.0548	0.0666
2256	0.0638	0.0428	0.0909
2257	0.0892	0.0808	0.0983
2258	0.0770	0.0710	0.0833
2260	0.0750	0.0705	0.0797
2261	0.0567	0.0479	0.0665
2262	0.0487	0.0404	0.0581
2263	0.0577	0.0422	0.0766
2265	0.0968	0.0821	0.1131
2266	0.0884	0.0809	0.0964
2268	0.0394	0.0129	0.0895
2270	0.0623	0.0542	0.0711
3895	0.1792	0.1115	0.2657

### 9.4. Appendix D : Resulting Map



**9.5. Appendix E : (electronic support) Geocoded Database (Access format)**

## **Annexe 2**

Résultats du géocodage de la base de données concernant la qualité de l'eau

Document :

**PERFORMANCE REPORT / Geocoding a database related to water testing / 2004 Dataset**

– ARO Study (Caroline Guénette, septembre 2005)

**PERFORMANCE REPORT**

**Geocoding a database related to water testing**

**2004 Dataset – ARO Study**

**Caroline Guénette, September 2005**

<b>1. OVERVIEW</b> .....	<b>3</b>
1.1. DATABASE DESCRIPTION.....	3
1.2. DATABASE ACCURACY CHECK .....	5
1.3. TOOLS .....	5
<b>2. PRE-PROCESSING RESULTS</b> .....	<b>6</b>
<b>3. GEOCODING RESULTS USING GEOPINPOINT</b> .....	<b>7</b>
<b>4. GEOCODING RESULTS FOR LOT/CONCESSION</b> .....	<b>7</b>
<b>5. SELECTED GEOCODING RESULTS</b> .....	<b>7</b>
<b>6. APPENDIXES</b> .....	<b>9</b>
6.1. APPENDIX A : PUBLIC HEALTH UNITS LIST .....	9
6.2. APPENDIX B : STATUS CODES LIST .....	10
6.3. APPENDIX C : STATISTICS ON REJECTION OF UNTESTED WATER SAMPLES.....	11
6.4. APPENDIX E : (ELECTRONIC SUPPORT) GEOCODED DATABASE (ACCESS FORMAT)	20

## **1. OVERVIEW**

This document describes the initial database that needed to be geocoded and the performances of the geocoding process, as detailed in the previous report entitled *Developing a geocoding algorithm and geocoding a database related to water testing – Caroline Guénette, March 2005*.

### **1.1. Database description**

The provided database came from the Food Safety and Safe Water Unit (SWU) of the Ministry of Health and Long Term Care of Ontario (MOHLTC). It contained 280 139 entries, all in Access format. It gathered drinking water testing performed by the Central and Regional Public Health Laboratories of Ontario between May 1<sup>st</sup> 2004 and April 30<sup>th</sup> 2005.

Each line (entry) of the database refers to a drinking water sample that was send to the MOHLTC laboratories for testing.

The database contained several fields, listed below with their characteristics. Fields had to be renamed for shorter/simpler ones upon reception of the database to allow the use of SAS statistics software

<b>Field name</b>	<b>Data type</b>	<b>original name</b>	<b>Description</b>
Drecei	DD-MM-YY	Date received	Date the sample arrived at the laboratory for testing
labo	text	Lab	Name of the laboratory which received the sample for testing
address	text	SrcStreet	Street address of the water source location
lot	text	SrcLotConc	Lot, concession of the water source location
city	text	SrcCityTownMun	Township/Municipality of the water source location
emerg	numeric	Emerg#Num	Emergency locator number of the water source location
cnty	text	SrcCnty	County of the water source location
postal	TNT-TNT	SrcPostalcode	Postal code (six digits) of the water source location
Dcoll	DD-month-YY	Date Collected	Collection date of the sample at the water source
phu	numeric	srcPHU	Number of the Public Health Unit where the water source is located (see list in appendix A for description)
stati	text	Status Int#	Result of the water testing in text description
statn	numeric	Status	Result of the water testing in numeric code (see table in appendix B for code description)
coliform	numeric	Coliform	Coliform count (total number per 100ml) as determined by water testing
ecoli	numeric	E.coli	Escherichia coli count (total number per 100ml) as determined by water testing
dread	DD-MM-YY	Date Read	Date the sample was processed at the analysing laboratory
access	numeric	Accession number	Numeric string of 6 digits being a unique identification number for each laboratory by season.

Other informations were collected on the requisition form filled by the water source owner. They were excluded from the provided database in order to ensure confidentiality. Although, they can explain some database anomalies caused by field confusion or inversion. Below is a partial list of database fields that were removed to ensure confidentiality before the database was delivered to us:

Name of the person requesting the water testing (water source owner);  
Return civic address;  
Return address municipality;  
Return address province;  
Return address postal code;  
Water source owner phone number.

### **1.2. Database accuracy check**

Fields of the database were typed-in by an employee of MOHLTC as they appeared on the analysis requisition form filled by the water source owner and provided with the water sample. No cross validation was performed by the MOHLTC to ensure data accuracy. The database was delivered as is.

### **1.3. Tools**

The database Access (Access 2000, Microsoft Corporation, 1992-1999) format allowed manipulating the data and performing pre-processing in its original file format.

SAS statistical software release 8.1(SAS Institute Inc, Cary, NC, USA, 1999-2000) was used for statistical analysis of the entries distribution.

The geocoding process using the address and/or postal codes was realized using GeoPinPoint software (GeoPinpoint™ – Windows Standalone Desktop Version, DMTI Spatial Inc. 2004). The assignation of the geographical location for lot/concession numbers was done by database merging using Access.

The lot/concession file of the Ontario Base Map product provided by the Ministry of Natural Resources of Ontario was used as the reference database for geocoding the lot/concession information.

Once a geographical location was attributed to each entry, the GIS software entitled ArcGIS (ArcGIS™ 9.1, ESRI, USA, 2005) was used to map the different geographical locations as points.

## 2. PRE-PROCESSING RESULTS

Of the initial 280 139 entries in the database :

-0 was removed because it was a double;

-855 were labelled 'POBOXONLY' in the gcocode column as they contained PO Box alike information only in the address field;

-1429 were labelled 'NOGEO' as they contained no usable spatial information;

-141 were labelled 'NOGEO' as they contained lot/concession numbers as only spatial identifier but no municipality information;

-20 336 were labelled 'NWT' as they did not contained any water test results;

259 803 records (92.7%) are potentially usable for geocoding and do have a water testing result.

Of the 280 139 distinct entries in the database, 80 172 (~29%) were modified to better match the reference databases at geocoding :

-3899 were modified to remove the extra PO Box alike information;

-813 were modified because they contained brackets in the address field;

-46 were modified because they contained quotation marks in the address field;

-91 were modified as they contained invalid Ontario postal codes;

-75 323 were modified to standardize the lot/concession information.

### **3. GEOCODING RESULTS USING GEOPINPOINT**

- 59,918 (~21%) Have a successful geocoding result using the strict civic address
- 10,155 (~4%) Have a successful geocoding result using the relaxed civic address
- 38,190 (~14%) Have a successful geocoding result using six digits postal code
- 74,012 (~26%) Have a successful geocoding result using five digits postal code
- 236,409 (~84%) Have a successful geocoding result using tree digits postal code
- 199,926 (~71%) Have a successful geocoding result using municipality

### **4. GEOCODING RESULTS FOR LOT/CONCESSION**

- 17,906 (~6%) Have a successful geocoding result using the lot/concession numbers in the official municipality name
- 15,070 (~5%) Have a successful geocoding result using the lot/concession numbers in the city name appellation

### **5. SELECTED GEOCODING RESULTS**

The columns <select\_x> and <select\_Y> of the <GCD\_main> table contain the geocoding results the most appropriate to use.

Of the 280,139 entries in the initial database,

For 70,073 (~25%) entries, geocoding using the civic address was considered as the best option;

For 27,937 (~10%) entries, geocoding using the six digits postal code was considered as the best option;

For 55,927 (~20%) entries, geocoding using the five digits postal code was considered as the best option;

For 51,091 (~18%) entries, geocoding using the tree digits postal code was considered as the best option;

For 14,182 (~ 5%) entries, geocoding according to lot/concession numbers was considered as the best option;

*Version 2.2*

For 47,964 (~ 17%) entries, geocoding to the centroid of their municipality was considered as the best option;

In total : 267,174 entries (~95%) were geocoded while 12,965 (~5%) were not.

## 6. APPENDIXES

### 6.1. Appendix A : Public Health Units list

<b>Health Unit Number</b>	<b>Health Unit Name</b>
2226	Algoma
2227	Brant County
2240	Chatham-Kent Public Health Division
2230	Durham Region Health Department
2258	Eastern Ontario
2231	Elgin-St. Thomas
2233	Grey Bruce
2238	Hastings & Prince Edward Counties
2234	Haldimand-Norfolk
2235	Haliburton, Kawartha, Pine Ridge District
2236	Halton Region Health Department
2237	City of Hamilton-Public Health & Community Services Dept.
2239	Huron County
2241	Kingston, Frontenac And Lennox & Addington
2242	County of Lambton Community Health Services Department
2243	Leeds, Grenville and Lanark District
2244	Middlesex-London
2245	Muskoka-Parry Sound
2246	Regional Niagara Public Health Department
2247	North Bay & District
2249	Northwestern
2251	City of Ottawa-Public Health & Long Term Care Branch
2252	County of Oxford
2253	Regional Municipality of Peel Health Department
2254	Perth District
2255	Peterborough County-City
2256	Porcupine
2257	Renfrew County & District
2260	Simcoe County District
2261	Sudbury & District
2262	Thunder Bay District
2263	Timiskaming
3895	Toronto Public Health
2265	Region of Waterloo, Public Health
2266	Wellington-Dufferin-Guelph
2268	Windsor-Essex County
2270	York Region Health Services Department

## 6.2. Appendix B : Status Codes list

<b>Status Code</b>	<b>Status Integer</b>
1	No significant evidence of bacterial contamination
2	Significant evidence of bacterial contamination
3	UNSAFE TO DRINK, evidence of sewage contamination
4	Overgrown - May be unsafe to drink
5	The sample was taken from an unprotected source, such as a lake or river
6	The client requested chemical testing
7	The sample was too old
8	The bottle was received broken or damaged
10	The appearance or odour makes the sample unacceptable as drinking water
11	The requisition was received separated from the sample bottle
12	The sample was not collected in the proper bottle
13	The sample was received frozen
14	The sample was collected from a hot water tap
15	Insufficient sample was submitted
17	The sample leaked in transit
18	The sample was collected from a source located outside of Ontario
19	The sample was collected from a municipal water supply
20	Insufficient information was supplied on the sample
21	Not tested for some other reason
22	Overgrown with evidence of sewage contamination

### **6.3. Appendix C: Statistics on rejection of untested water samples**

To obtain a PHUcalculated (Public Health Unit):

Geocoding was performed, using the same process as described earlier. Using ArcGIS, a spatial join was used to assign a PHU calculated to each point depending on its location.

For each entry of the database: a new 'PHUcald' column was added with the PHU codes when available, if no PHU code was calculated (point did not fall within the boundaries of a PHU or point was not geocoded), the declared PHU was used.

Version 2.2

Distribution of Status Codes by PHU (Public Health Unit)

PHU	Total of samples	Status Code																				Total Rejecte	Total Tested
		1	2	3	4	5	6	7	8	10	11	12	13	14	15	17	18	19	20	21	22		
0000	1006	183	37	6	12	1	0	11	3	1	0	8	0	0	0	1	28	1	581	132	1	767	239
2226	3693	2642	736	157	74	1	0	25	10	6	0	0	10	0	0	0	0	1	18	13	0	84	3609
2227	2458	1403	538	172	73	0	0	45	15	16	0	1	1	0	0	2	0	0	159	27	6	266	2192
2230	9157	5309	2016	287	519	2	0	198	42	17	2	7	6	2	0	55	0	1	545	113	36	990	8167
2231	2401	1467	501	112	86	0	0	26	4	4	0	1	1	0	0	0	0	1	167	19	12	223	2178
2233	20036	13905	2716	1054	341	1	0	446	94	23	0	2	27	1	1	5	0	2	1248	107	63	1957	18079
2234	4742	3046	751	295	143	1	0	108	19	17	0	1	7	1	1	0	0	0	313	29	10	497	4245
2235	27972	18426	5460	1185	1420	1	0	273	122	23	0	9	50	4	0	88	0	0	577	243	91	1390	26582
2236	5772	3594	1182	226	110	0	0	134	15	11	0	11	2	0	1	0	0	0	430	55	1	659	5113
2237	5116	3267	1021	226	78	0	0	125	20	25	0	9	0	0	0	0	0	3	316	24	2	522	4594
2238	15490	9787	3187	1215	268	0	0	186	52	15	0	2	20	1	2	5	0	0	622	76	52	981	14509
2239	3365	2540	438	86	41	0	0	64	22	4	0	1	3	0	0	1	0	0	139	21	5	255	3110
2240	1133	853	188	33	25	0	0	7	1	1	0	0	2	0	0	0	0	0	17	6	0	34	1099
2241	10792	7663	1805	966	98	0	0	42	31	5	0	1	3	0	0	2	0	4	98	38	36	224	10568
2242	784	470	146	41	31	0	0	20	11	1	0	0	1	0	0	0	0	0	51	7	5	91	693
2243	15934	12143	2046	788	231	1	0	128	71	69	0	7	4	0	0	1	0	1	274	101	69	657	15277
2244	4558	2977	790	189	141	0	0	73	31	6	0	2	4	0	0	0	0	0	296	35	14	447	4111
2245	13644	9241	2354	545	574	2	0	161	28	68	2	1	10	1	0	1	0	0	512	105	39	891	12753
2246	3585	2191	610	307	85	3	0	88	9	18	0	3	0	1	1	0	0	0	216	53	0	392	3193
2247	5823	3925	996	188	70	2	0	103	36	17	1	6	5	0	0	5	5	1	385	77	1	643	5180
2249	2050	1232	426	90	73	0	0	30	2	20	0	1	70	0	0	2	0	3	62	32	7	222	1828
2251	12489	9668	1071	272	306	0	0	148	73	82	0	2	5	1	1	0	2	1	697	112	48	1124	11365
2252	3699	2580	584	102	79	0	0	71	15	7	1	1	0	0	0	1	0	1	223	22	12	342	3357
2253	2902	1834	661	109	67	1	0	59	13	4	1	0	2	0	0	0	0	3	104	40	4	227	2675
2254	2724	2001	332	83	40	0	0	63	18	1	0	0	0	1	0	0	0	0	153	26	6	262	2462
2255	14127	9537	2638	641	649	1	0	109	98	18	0	3	4	5	0	28	0	2	262	86	46	616	13511
2256	851	684	84	25	20	0	0	8	3	2	0	0	0	1	0	1	0	0	17	6	0	38	813
2257	7915	5783	1019	215	229	1	0	109	14	78	0	2	1	0	1	1	1	0	348	73	40	629	7286
2258	12209	8498	1902	543	370	1	0	117	65	92	0	1	68	4	0	1	3	1	302	138	103	793	11416
2260	28692	18598	6276	889	1159	5	0	332	114	147	5	6	11	4	0	11	0	3	889	209	34	1736	26956
2261	5171	3794	813	263	87	3	0	26	49	20	0	1	6	1	1	0	0	1	72	31	3	211	4960
2262	4988	3298	1217	166	111	0	0	25	5	36	0	1	4	0	0	2	0	1	86	20	16	180	4808
2263	1886	1367	186	108	32	0	0	24	7	1	0	0	42	0	0	0	0	2	99	14	4	189	1697
2265	3598	2375	699	113	52	0	0	56	18	2	0	2	1	1	0	0	0	0	234	27	17	342	3256
2266	10908	7724	1714	275	210	0	0	272	54	8	2	1	0	0	0	5	0	0	550	79	14	971	9937
2268	281	199	38	22	5	0	0	3	0	1	0	0	0	0	0	0	0	0	8	5	0	17	264
2270	8148	5577	1684	250	169	0	0	96	22	13	0	1	0	2	1	1	0	3	249	59	21	447	7701
3895	40	12	5	3	0	0	0	0	0	0	0	0	0	0	0	0	0	18	2	0	0	20	20
<b>Total</b>	<b>280139</b>	<b>189793</b>	<b>48867</b>	<b>12247</b>	<b>8078</b>	<b>27</b>	<b>0</b>	<b>3811</b>	<b>1206</b>	<b>879</b>	<b>14</b>	<b>94</b>	<b>370</b>	<b>31</b>	<b>10</b>	<b>219</b>	<b>39</b>	<b>55</b>	<b>11321</b>	<b>2260</b>	<b>818</b>	<b>20336</b>	<b>259803</b>

Distribution of Status Codes by PHU calculated (Public Health Unit)

PHU calculated	Total of samples	Status Code																				Total Rejected	Total Tested
		1	2	3	4	5	6	7	8	10	11	12	13	14	15	17	18	19	20	21	22		
0000	648	21	4	2	2	0	0	2	1	0	0	8	0	0	0	0	26	0	480	102	0	619	29
2222	25995	17202	5095	1100	1280	1	0	242	126	20	0	8	42	5	0	80	0	0	483	223	88	1230	24765
2226	3560	2537	719	153	73	1	0	21	10	6	0	0	10	0	0	0	0	1	15	13	1	77	3483
2227	2004	1203	392	107	58	0	0	39	15	13	0	1	0	0	0	1	0	0	142	27	6	238	1766
2230	9031	5213	1994	280	524	2	0	202	43	18	2	7	6	2	0	56	0	1	538	106	37	983	8048
2231	2767	1694	583	137	94	0	0	33	5	4	0	1	2	0	0	0	2	1	178	21	12	247	2520
2233	19309	13382	2617	1028	335	1	0	431	89	21	0	2	26	1	1	4	0	2	1211	97	61	1886	17423
2234	5142	3225	872	353	162	1	0	111	20	20	0	1	8	1	1	1	0	0	326	30	10	520	4622
2235	761	439	124	34	35	0	0	16	2	3	0	1	0	0	0	0	0	0	78	28	1	128	633
2236	5967	3719	1220	238	106	0	0	135	17	11	0	11	2	0	1	0	0	0	448	58	1	683	5284
2237	4781	3098	917	201	83	0	0	111	23	18	0	6	0	0	0	0	0	3	294	25	2	480	4301
2238	16821	10683	3431	1276	338	1	0	199	52	21	0	3	22	1	2	8	0	0	654	76	54	1039	15782
2239	3530	2662	471	92	45	0	0	68	21	3	0	1	1	0	0	1	0	0	138	23	4	256	3274
2240	1166	874	190	36	24	0	0	6	1	2	0	0	2	0	0	0	0	0	26	5	0	42	1124
2241	11706	8336	1958	1019	110	1	0	40	33	10	0	1	5	0	0	2	0	4	113	37	37	246	11460
2242	780	465	149	42	31	0	0	21	10	1	0	0	1	0	0	0	0	0	48	7	5	88	692
2243	14947	11362	1927	732	220	1	0	128	69	67	0	7	2	1	0	1	0	1	266	100	63	643	14304
2244	4721	3081	807	202	139	0	0	82	34	8	0	2	3	0	0	0	0	0	311	38	14	478	4243
2245	14811	10092	2559	568	589	3	0	173	30	74	2	1	10	1	0	2	0	0	547	121	39	964	13847
2246	4105	2523	711	328	88	3	0	105	10	24	0	3	0	1	1	0	0	0	254	54	0	455	3650
2247	5618	3762	975	188	79	1	0	101	36	16	1	6	5	0	0	4	5	1	367	68	3	611	5007
2249	2019	1208	418	90	73	0	0	27	2	21	0	1	68	0	0	2	0	3	65	33	8	222	1797
2251	12524	9669	1080	280	314	0	0	144	74	79	0	2	7	1	1	0	2	1	705	120	45	1136	11388
2252	3777	2619	610	97	80	0	0	78	15	9	1	1	0	0	0	1	0	1	231	23	11	360	3417
2253	2699	1695	622	100	66	1	0	61	14	4	0	1	2	0	0	0	0	4	93	32	4	212	2487
2254	2576	1905	302	78	35	0	0	60	18	2	0	0	0	1	0	0	0	0	143	25	7	249	2327
2255	14813	9966	2762	668	693	1	0	123	93	19	0	3	12	4	0	33	0	2	297	87	50	674	14139
2256	825	668	79	21	17	0	0	9	3	2	0	0	0	1	0	1	0	0	17	7	0	40	785
2257	7525	5510	946	204	221	0	0	101	13	72	0	2	1	0	1	1	1	0	336	76	40	604	6921
2258	12433	8651	1915	545	378	1	0	123	65	93	0	1	69	3	0	1	3	1	329	146	109	835	11598
2260	27820	18026	6094	850	1130	5	0	323	111	141	5	5	11	4	0	11	0	3	873	197	31	1689	26131
2261	5088	3741	791	261	85	3	0	25	47	21	0	1	6	1	1	0	0	1	71	31	2	208	4880
2262	4882	3231	1186	164	109	0	0	23	5	34	0	1	3	0	0	2	0	1	88	19	16	176	4706
2263	1829	1319	184	103	32	0	0	23	7	1	0	0	42	0	0	0	0	2	99	13	4	187	1642
2265	3748	2438	754	124	57	0	0	58	18	2	0	3	1	1	0	0	0	1	247	26	18	357	3391
2266	10829	7731	1658	268	203	0	0	265	51	6	3	1	1	0	0	6	0	0	538	83	15	954	9875
2268	320	224	50	24	4	0	0	3	0	1	0	0	0	0	0	0	0	0	9	5	0	18	302
2270	8116	5538	1674	249	165	0	0	97	22	11	0	2	0	2	1	1	0	2	258	74	20	470	7646
2295	145	81	27	5	1	0	0	2	1	1	0	0	0	0	0	0	0	19	4	4	0	31	114
3895	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0
<b>Total</b>	<b>280139</b>	<b>189793</b>	<b>48867</b>	<b>12247</b>	<b>8078</b>	<b>27</b>	<b>0</b>	<b>3811</b>	<b>1206</b>	<b>879</b>	<b>14</b>	<b>94</b>	<b>370</b>	<b>31</b>	<b>10</b>	<b>219</b>	<b>39</b>	<b>55</b>	<b>11321</b>	<b>2260</b>	<b>818</b>	<b>20336</b>	<b>259803</b>

Version 2.2

Distribution of Status Codes by laboratory

Laboratory	Total of samples	Status Code																				Total Rejected	Total Tested
		1	2	3	4	5	6	7	8	10	11	12	13	14	15	17	18	19	20	21	22		
Central	14492	9708	2971	535	281	2	0	220	47	13	0	1	4	7	1	2	0	25	507	131	37	960	13532
Hamilton	17908	10834	3568	1083	382	3	0	449	44	83	0	33	10	1	3	1	0	3	1249	160	2	2039	15869
Kingston	32439	22165	5933	2599	325	1	0	250	119	20	0	5	6	1	2	2	5	5	765	143	93	1324	31115
London	56674	39150	8465	2213	1144	2	0	1182	307	63	4	11	27	4	1	13	0	5	3522	397	164	5538	51136
Orillia	42611	27860	8776	1431	1763	6	0	480	140	238	9	7	32	5	0	12	0	4	1440	340	68	2713	39898
Ottawa	39720	29355	4571	1280	1064	3	0	451	177	321	0	8	74	5	2	2	25	2	1709	422	249	3201	36519
Peterborough	50930	33109	9986	2067	2627	4	0	545	262	43	0	19	78	6	0	178	0	2	1364	466	174	2967	47963
Sault Ste. Marie	3642	2605	729	156	74	1	0	24	9	6	0	0	10	0	0	0	0	1	14	13	0	78	3564
Sudbury	10436	7291	1723	444	150	5	0	114	86	32	1	8	10	1	1	4	5	1	455	101	4	824	9612
Thunder Bay	7034	4527	1642	256	184	0	0	55	7	56	0	2	74	0	0	4	0	4	148	52	23	402	6632
Timmins	2854	2147	271	127	60	0	0	31	8	3	0	0	43	1	0	1	4	3	127	24	4	245	2609
Windsor	1399	1042	232	56	24	0	0	10	0	1	0	0	2	0	0	0	0	0	21	11	0	45	1354
<b>Total</b>	<b>280139</b>	<b>189793</b>	<b>48867</b>	<b>12247</b>	<b>8078</b>	<b>27</b>	<b>0</b>	<b>3811</b>	<b>1206</b>	<b>879</b>	<b>14</b>	<b>94</b>	<b>370</b>	<b>31</b>	<b>10</b>	<b>219</b>	<b>39</b>	<b>55</b>	<b>11321</b>	<b>2260</b>	<b>818</b>	<b>20336</b>	<b>259803</b>

Chi-square test by Laboratory:

Statistic	DF	Value	Prob
Chi-Square	11	1937.5089	<.0001
Likelihood Ratio Chi-Square	11	1984.2733	<.0001
Mantel-Haenszel Chi-Square	1	122.5453	<.0001
Phi Coefficient		0.0832	
Contingency Coefficient		0.0829	
Cramer's V		0.0832	

Sample Size = 280139

Chi-square test by PHU:

Statistic	DF	Value	Prob
Chi-Square	37	9979.6129	<.0001
Likelihood Ratio Chi-Square	37	5987.4246	<.0001
Mantel-Haenszel Chi-Square	1	277.1630	<.0001
Phi Coefficient		0.1887	
Contingency Coefficient		0.1855	
Cramer's V		0.1887	

Chi-square test by PHU calculated:

Statistic	DF	Value	Prob
Chi-Square	39	10427.9279	<.0001
Likelihood Ratio Chi-Square	39	6110.2375	<.0001
Mantel-Haenszel Chi-Square	1	112.4856	<.0001
Phi Coefficient		0.1929	
Contingency Coefficient		0.1894	
Cramer's V		0.1929	

Sample Size = 280139

Confidence interval (95%) by Laboratory: (rejected samples / total samples)

<b>LAB</b>	<b>RISK</b>	<b>CI-Low</b>	<b>CI-High</b>
Windsor	0.0322	0.0236	0.0428
Timmins	0.0858	0.0758	0.0967
ThunderBay	0.0572	0.0518	0.0628
Peterborough	0.0583	0.0562	0.0603
Sudbury	0.0790	0.0739	0.0843
SStMarie	0.0214	0.0170	0.0267
Ottawa	0.0806	0.0779	0.0833
Orillia	0.0637	0.0614	0.0660
London	0.0977	0.0953	0.1002
Kingston	0.0408	0.0387	0.0430
Hamilton	0.1139	0.1092	0.1186
Central	0.0662	0.0622	0.0704

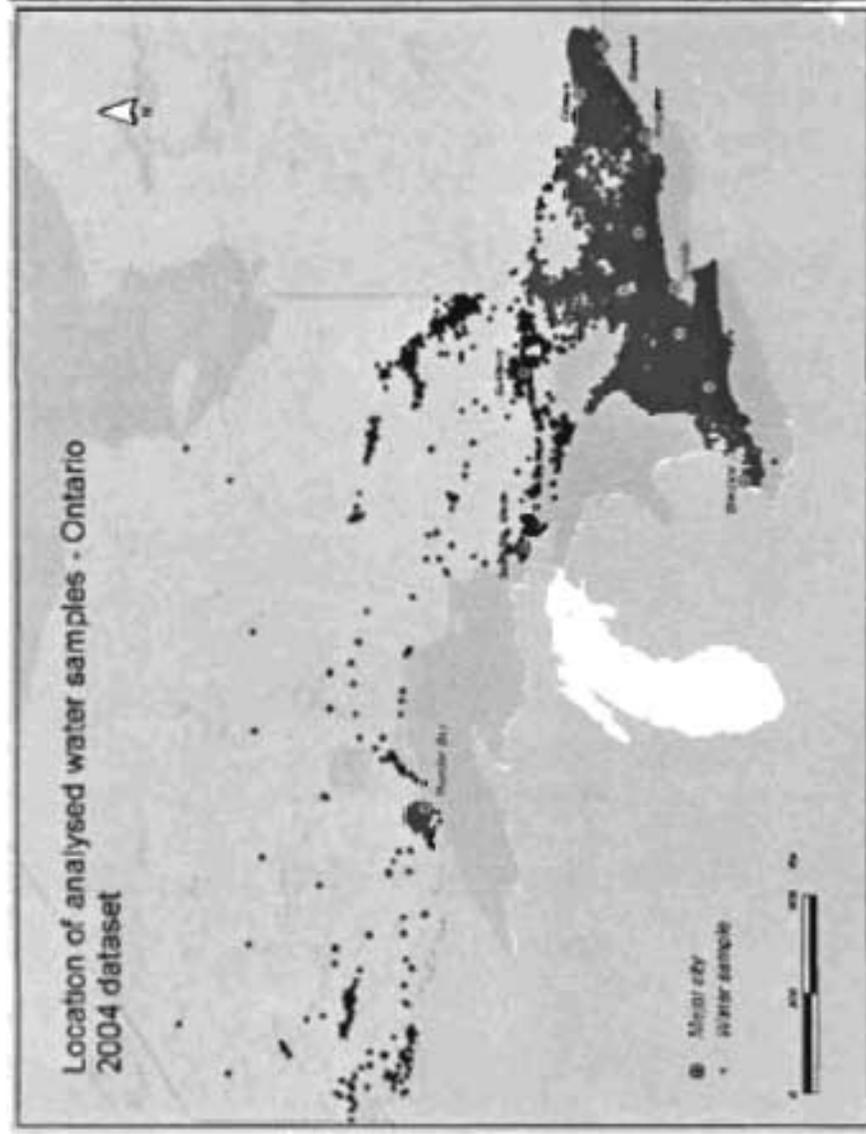
## Confidence interval (95%) by PHU: (rejected samples / total samples)

<b>PHU</b>	<b>RISK</b>	<b>CI-Low</b>	<b>CI-High</b>
0000	0.7624	0.7349	0.7884
2226	0.0227	0.0182	0.0281
2227	0.1082	0.0962	0.1212
2230	0.1081	0.1018	0.1147
2231	0.0929	0.0816	0.1052
2233	0.0977	0.0936	0.1019
2234	0.1048	0.0962	0.1139
2235	0.0497	0.0472	0.0523
2236	0.1142	0.1061	0.1227
2237	0.1020	0.0939	0.1107
2238	0.0633	0.0595	0.0673
2239	0.0758	0.0671	0.0852
2240	0.0300	0.0209	0.0417
2241	0.0208	0.0181	0.0236
2242	0.1161	0.0945	0.1406
2243	0.0412	0.0382	0.0444
2244	0.0981	0.0896	0.1071
2245	0.0653	0.0612	0.0696
2246	0.1093	0.0993	0.1200
2247	0.1104	0.1025	0.1188
2249	0.1083	0.0952	0.1226
2251	0.0900	0.0850	0.0952
2252	0.0925	0.0833	0.1023
2253	0.0782	0.0687	0.0886
2254	0.0962	0.0854	0.1079
2255	0.0436	0.0403	0.0471
2256	0.0447	0.0318	0.0608
2257	0.0795	0.0736	0.0856
2258	0.0650	0.0606	0.0695
2260	0.0605	0.0578	0.0633
2261	0.0408	0.0356	0.0466
2262	0.0361	0.0311	0.0416
2263	0.1002	0.0870	0.1147
2265	0.0951	0.0857	0.1051
2266	0.0890	0.0837	0.0945
2268	0.0605	0.0356	0.0951
2270	0.0549	0.0500	0.0600
3895	0.5000	0.3380	0.6620

Confidence interval (95%) by PHU calculated: (rejected samples / total samples)

<b>PHU</b>	<b>RISK</b>	<b>CI-Low</b>	<b>CI-High</b>
0000	0.9552	0.9364	0.9698
2222	0.0473	0.0448	0.0500
2226	0.0216	0.0171	0.0270
2227	0.1188	0.1049	0.1337
2230	0.1088	0.1025	0.1155
2231	0.0893	0.0789	0.1005
2233	0.0977	0.0935	0.1019
2234	0.1011	0.0930	0.1097
2235	0.1682	0.1423	0.1967
2236	0.1145	0.1065	0.1228
2237	0.1004	0.0920	0.1093
2238	0.0618	0.0582	0.0655
2239	0.0725	0.0642	0.0816
2240	0.0360	0.0261	0.0484
2241	0.0210	0.0185	0.0238
2242	0.1128	0.0915	0.1371
2243	0.0430	0.0398	0.0464
2244	0.1012	0.0928	0.1102
2245	0.0651	0.0612	0.0692
2246	0.1108	0.1014	0.1208
2247	0.1088	0.1007	0.1172
2249	0.1100	0.0966	0.1244
2251	0.0907	0.0857	0.0959
2252	0.0953	0.0861	0.1051
2253	0.0785	0.0687	0.0893
2254	0.0967	0.0855	0.1087
2255	0.0455	0.0422	0.0490
2256	0.0485	0.0349	0.0654
2257	0.0803	0.0742	0.0866
2258	0.0672	0.0628	0.0717
2260	0.0607	0.0579	0.0636
2261	0.0409	0.0356	0.0467
2262	0.0361	0.0310	0.0417
2263	0.1022	0.0887	0.1170
2265	0.0953	0.0860	0.1051
2266	0.0881	0.0828	0.0936
2268	0.0563	0.0337	0.0874
2270	0.0579	0.0529	0.0632
3895	1.0000	0.0250	1.0000

Appendix D : Resulting Map



**6.4. Appendix E : (electronic support) Geocoded Database (Access format)**

**Annexe 3**

Détail technique de l'algorithme permettant le lien entre les résultats de résistance antimicrobienne et les coordonnées géographiques

Document:

DEVELOPMENT AND TESTING OF AN ALGORITHM / Linkeage of antimicrobial-resistance results to geographic coordinates, version 2.1 / preliminary dataset – ARO study (Caroline Guénette, septembre 2005)

## DEVELOPMENT AND TESTING OF AN ALGORITHM

Linkeage of antimicrobial-resistance results to geographic coordinates

PRELIMINARY dataset – ARO Study

V. 2.1

Caroline Guénette, September 2005

<b>1. OVERVIEW.....</b>	<b>3</b>
1.1. DATA DESCRIPTION.....	3
1.2. TOOLS.....	3
1.3. DIAGRAM OF STEPS.....	3
<b>2. PRE-PROCESSING THE LOGBOOK INFORMATION.....</b>	<b>5</b>
2.1. CREATING A MAIN LOGBOOK.....	5
2.2. BUILDING-UP THE ACCESSION NUMBER.....	5
<b>3. PRE-PROCESSING THE ANTIMICROBIAL-RESISTANCE RESULTS.....</b>	<b>6</b>
3.1. ANTIMICROBIAL-RESISTANCE TESTING RESULTS SUMMARY.....	12
<b>4. LINKING THE LOGBOOK AND THE AMR RESULTS.....</b>	<b>13</b>
<b>5. GEOCODING A REFERENCE DATABASE.....</b>	<b>13</b>
5.1. MERGING REFERENCE DATABASES.....	14
5.2. PRE-PROCESSING RESULTS.....	14
5.3. GEOCODING RESULTS USING GEOPINPOINT.....	15
5.4. GEOCODING RESULTS FOR LOT/CONCESSION.....	15
5.5. SELECTED GEOCODING RESULTS.....	15
<b>6. LINKING AMR RESULTS TO THE GEOCODING RESULTS.....</b>	<b>16</b>
6.1. LINKING AMR- GEOCODING PERFORMANCE SUMMARY.....	17
<b>7. APPENDIX A: BREAKPOINTS TABLE.....</b>	<b>18</b>

## **1. OVERVIEW**

Antimicrobial-resistance (AMR) results needed to have a geographic location assigned to them in order to perform geospatial analysis. To obtain the information needed, three major steps had to be done: formatting a logbook, merging it to the previously cleaned AMR results and finally merging the AMR results with a previously geocoded reference database.

### ***1.1. Data description***

The antimicrobial-resistance (AMR) results are contained in a single database in Excel format. It contains results of two different tests performed: bacterial identification using API 20E and Antimicrobial-resistance testing using the NARMS susceptibility testing panel (National Antimicrobial Resistance Monitoring System, CDC).

The following fields are included in the AMR results database: AMR number; Isolate number; Final bacteria ID name; Minimal Inhibition Concentration (MIC) to the sixteen following antibiotics: Amikacin, Amoxicillin Clavulanic Acid, Ampicillin, Cefoxitin, Ceftiofur, Ceftriaxone, Cephalothin, Chloramphenicol, Ciprofloxacin, Gentamicin, Kanamycin, Nalidixic Acid, Streptomycin, Sulphamethoxazole, Tetracycline, Trimethoprim Sulphamethoxazole.

The AMR logbooks are provided by each of the two AMR testing laboratories. They contain various technical information regarding the performed analysis. Among other things, they state for each isolate the Isolate ID number that was submitted as well as its laboratory of origin and the AMR number that was assigned to it.

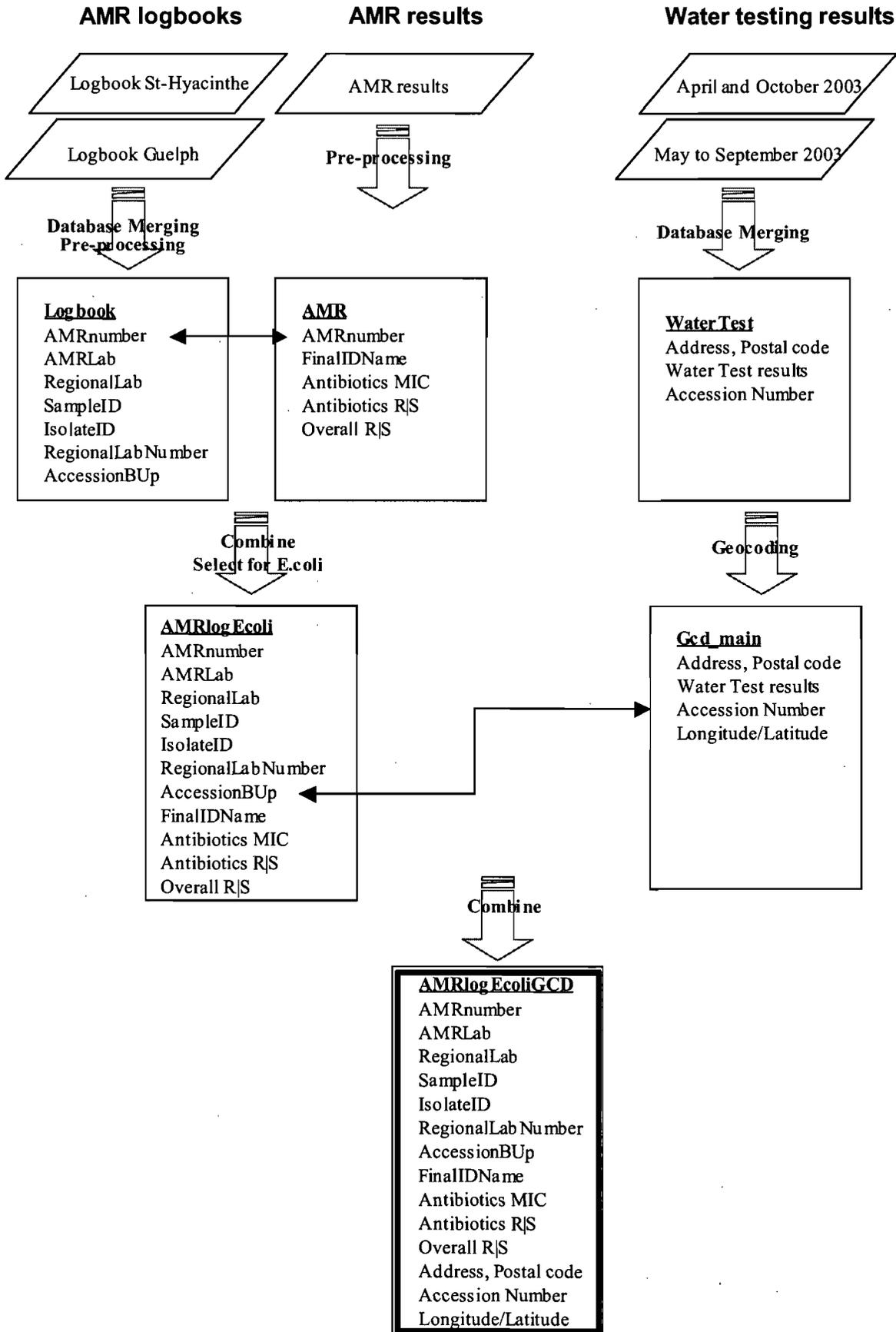
Geographical reference databases (2) are provided by the Food Safety and Safe Water Unit (SWU) of the Ministry of Health and Long Term Care of Ontario (MOHLTC). They are in Access format. They gather drinking water testing performed by the Central and Regional Public Health Laboratories of Ontario between May and September 2003 and for the months of April and October 2003.

### ***1.2. Tools***

Microsoft Access and GeopinPoint are the two software used to achieve the different steps detailed below.

### ***1.3. Diagram of steps***

The following diagram shows the different databases involved and their respective fields as well as the major steps to be achieved.



## 2. PRE-PROCESSING THE LOGBOOK INFORMATION

### 2.1. Creating a main logbook

The AMR analyses were performed at two distinct laboratories, Guelph and St-Hyacinthe. There was no standardized logging procedure, and therefore the two labs managed the samples numbers and logbooks different ways.

A main logsheet must be created to allow the link between the AMR number and the Sample ID number, also called in the geographical reference database the 'Accession number'.

Gather information from both logbooks and structure them in an Excel spreadsheet containing the following columns:

- AMRnumber (also called the Submission number );
- AMRlab (either St-Hyacinthe or Guelph);
- RegionalLab ( who submitted the sample to the study);
- SampleID;
- IsolateID;
- RegionLabNumber (called the Submitted sample number in Guelph logsheets).

Extract the SampleID number form the isolateID number when necessary.

### 2.2. Building-up the Accession number

The Accession number is a six digits number used by the Ministry of Health and Long-Term Care of Ontario (MOHLTC) to identify the water samples. It is the information that will be used to link the AMR data to the geocoded reference database.

When E. coli isolates were submitted to the study, the Accession number was not always written in full or had prefixes/suffixes added to it. It needs to be build and put in the standard MOHLTC format based on the IsolateID and/or RegionalLab Numbers that appeared on the logsheets.

In the Excel spreadsheet,

Copy the IsolateID column and rename it step1. Set cell format to text.

Use the find and replace function:

Column: Step 1  
Find What: -?  
Replace with: (blank)  
Replace all

Column: Step 1  
Find What: P  
Replace with: (blank)

**Replace all**

Insert a column called Step2. Set cell format to text. Fill '0000' for each line.

Insert a column named Step3. Set cell format to general.

Use the 'concatenate' function.

**Text1: Step2 column**

**Text2: Step1 column**

Copy the function to each line of the <logbook\_main> tab

Insert a column called Step4. Set cell format to general.

Use the 'right' function

**Text: Step3 column**

**Num chars: 6**

Copy the function to each line of the <logbook\_main> tab

Insert a column and name it AccessionBUP.

Copy the Step4 column in it by using the Copy \ Paste Special \ Paste values

Perform a few visual checks with the initial paper sheets provided with the samples to see if the Region \ SampleID \ IsolateID do fit.

Import in Access under the table name <logbook>.

### **3. PRE-PROCESSING THE ANTIMICROBIAL-RESISTANCE RESULTS**

The database containing the AMR testing results is in Excel format. The information needs to be modified to simplify reading prior to its importation in Access.

Using Excel:

In the Excel file containing the AMR results: change the column titles to avoid any special character or spaces and add the 'MIC' appellation to each of them. Delete the second title line containing the breakpoints. Set every cell format to text.

Import table in Access under the name <AMR>.

If problem to keep the text format while importing in Access, first export the Excel file to a text file and import the text file in the Access database.

Using Access:

Copy the content of the 16 'antibioticMIC' columns into new fields. Name them: AmikacinRS; AmoxicillinClavulanic AcidRS; AmpicillinRS; CefoxitinRS; CeftiofurRS; CeftriaxoneRS; CephalothinRS; ChloramphenicolRS; CiprofloxacinRS; GentamicinRS; KanamycinRS; Nalidixic AcidRS; StreptomycinRS; SulphamethoxazoleRS; TetracyclineRS; TrimethoprimSulphamethoxazoleRS.

Select the 'antibioticRS' columns and use the following find and replace functions on each of them to simplify and standardise the contained MIC results.

Edit menu \ Replace  
Find What: .0000 (dot and four zeros)  
Replace With: (blank)  
Match: Any part of the field  
Replace all

Find What: (space)  
Replace With: (blank)  
Match: Any part of the field  
Replace all

\*Repeat the process to get rid of double and multiples spaces

Find What: 0.25  
Replace With: 0.2500  
Match: whole field  
Replace all

Find What: 0.50  
Replace With: 0.5000  
Match: whole field  
Replace all

Find What: 0.5  
Replace With: 0.5000  
Match: whole field  
Replace all

Find What: 0.03  
Replace With: 0.0300  
Match: whole field  
Replace all

Find What: 0.06  
Replace With: 0.0600  
Match: whole field  
Replace all

Perform queries to assign either the letter 'R' or 'S' depending if the isolate was Resistant or Susceptible to that specific antibiotic, based on comparison between the observed concentration and the breakpoints provided. This process will simplify reading for future data users without losing any information as the original concentrations are preserved in the initial columns.

Query \ Design view \ Update query  
Field: AmikacinRS \ Table: AMRec  
Update To: 'R' \ Criteria: '>4' \ Run  
Save as <qr upd R amikacin>

Query \ Design view \ Update query  
Field: AmikacinRS \ Table: AMRec  
Update To: 'S' \ Criteria: '<=0.5000' Or '1' Or '2' Or '4' Or '=1' Or '=2' Or '=4'  
\ Run  
Save as <qr upd S amikacin>

Query \ Design view \ Update query  
Field: AmoxicillinClavulanicAcidRS \ Table: AMRec  
Update To: 'R' \ Criteria: '32' Or '>32' Or '=32' \ Run  
Save as <qr upd R AmoxicillinClavulanic>

Query \ Design view \ Update query  
Field: AmoxicillinClavulanicAcidRS \ Table: AMRec  
Update To: 'S' \ Criteria: '<=1' Or '1' Or '2' Or '4' Or '8' Or '16' Or '=1' Or '=2' Or  
'=4' Or '=8' Or '=16' \ Run  
Save as <qr upd S AmoxicillinClavulanic>

Query \ Design view \ Update query  
Field: AmpicillinRS \ Table: AMRec  
Update To: 'R' \ Criteria: '32' Or '>32' Or '=32' \ Run  
Save as <qr upd R Ampicillin>

Query \ Design view \ Update query  
Field: AmpicillinRS \ Table: AMRec  
Update To: 'S' \ Criteria: '<=1' Or '1' Or '2' Or '4' Or '8' Or '16' Or '=1' Or '=2' Or  
'=4' Or '=8' Or '=16' \ Run  
Save as <qr upd S Ampicillin>

Query \ Design view \ Update query  
Field: CefoxitinRS \ Table: AMRec  
Update To: 'R' \ Criteria: '>16' \ Run  
Save as <qr upd R Cefoxitin>

Query \ Design view \ Update query  
Field: CefoxitinRS \ Table: AMRec  
Update To: 'S' \ Criteria: '<= 1' Or '1' Or '2' Or '4' Or '8' Or '16' Or '=1' Or '=2' Or  
'=4' Or '=8' Or '=16' \ Run  
Save as <qr upd S Cefoxitin >

Query \ Design view \ Update query

Field: CefotiofurRS \ Table: AMRec  
Update To: 'R' \ Criteria: '8' Or '>8' Or '=8' \ Run  
Save as <qr upd R Cefotiofur >

Query \ Design view \ Update query  
Field: CefotiofurRS \ Table: AMRec  
Update To: 'S' \ Criteria: '<=0.1200' Or '0.2500' Or '0.5000' Or '1' Or '2' Or '4' Or  
'=0.2500' Or '=0.5000' Or '=1' Or '=2' Or '=4' \ Run  
Save as <qr upd S Cefotiofur >

Query \ Design view \ Update query  
Field: CeftriaxoneRS \ Table: AMRec  
Update To: 'R' \ Criteria: '64' Or '>64' Or '=64' \ Run  
Save as <qr upd R Ceftriaxone >

Query \ Design view \ Update query  
Field: CeftriaxoneRS \ Table: AMRec  
Update To: 'S' \ Criteria: '<=0.2500' Or '0.5000' Or '1' Or '2' Or '4' Or '8' Or '16'  
Or '32' Or '=0.5000' Or '=1' Or '=2' Or '=4' Or '=8' Or '=16'  
Or '=32' \ Run  
Save as <qr upd S Ceftriaxone >

Query \ Design view \ Update query  
Field: CephalothinRS \ Table: AMRec  
Update To: 'R' \ Criteria: '32' Or '>32' Or '=32' \ Run  
Save as <qr upd R Cephalothin >

Query \ Design view \ Update query  
Field: CephalothinRS \ Table: AMRec  
Update To: 'S' \ Criteria: '<=2' Or '4' Or '8' Or '16' Or '=4' Or '=8' Or '=16' \ Run  
Save as <qr upd S Cephalothin >

Query \ Design view \ Update query  
Field: ChloramphenicolRS \ Table: AMRec  
Update To: 'R' \ Criteria: '32' Or '>32' Or '=32' \ Run  
Save as <qr upd R Chloramphenicol >

Query \ Design view \ Update query  
Field: ChloramphenicolRS \ Table: AMRec  
Update To: 'S' \ Criteria: '<=2' Or '4' Or '8' Or '16' Or '=4' Or '=8' Or '=16' \ Run  
Save as <qr upd S Chloramphenicol >

Query \ Design view \ Update query  
Field: CiprofloxacinRS \ Table: AMRec  
Update To: 'R' \ Criteria: '4' Or '>4' Or '=4' \ Run  
Save as <qr upd R Ciprofloxacin >

```
Query \ Design view \ Update query
Field: CiprofloxacinRS \ Table: AMRec
Update To: 'S' \ Criteria: '0.12' Or '0.1200' Or '<=0.0150' Or '0.03' Or '0.2500'
Or '0.0300' Or '0.5000' Or '=0.12' Or '=0.1200' Or
'=0.5000' Or '=0.03' Or '=0.2500' Or '=0.0300' \ Run
Save as <qr upd S Ciprofloxacin>
```

```
Query \ Design view \ Update query
Field: GentamicinRS \ Table: AMRec
Update To: 'R' \ Criteria: '16' Or '>16' Or '=16' \ Run
Save as <qr upd R Gentamicin>
```

```
Query \ Design view \ Update query
Field: GentamicinRS \ Table: AMRec
Update To: 'S' \ Criteria: '<=0.2500' Or '0.5000' Or '1' Or '2' Or '4' Or '8' Or
'=0.5000' Or '=1' Or '=2' Or '=4' Or '=8' \ Run
Save as <qr upd S Gentamicin>
```

```
Query \ Design view \ Update query
Field: KanamycinRS \ Table: AMRec
Update To: 'R' \ Criteria: '64' Or '>64' Or '=64' \ Run
Save as <qr upd R Kanamycin>
```

```
Query \ Design view \ Update query
Field: KanamycinRS \ Table: AMRec
Update To: 'S' \ Criteria: '<=8' Or '16' Or '=16' \ Run
Save as <qr upd S Kanamycin>
```

```
Query \ Design view \ Update query
Field: NalidixicAcidRS \ Table: AMRec
Update To: 'R' \ Criteria: '32' Or '>32' Or '=32' \ Run
Save as <qr upd R NalidixicAcid>
```

```
Query \ Design view \ Update query
Field: NalidixicAcidRS \ Table: AMRec
Update To: 'S' \ Criteria: '<=0.5000' Or '1' Or '2' Or '4' Or '8' Or '16' Or '=1' Or
'=2' Or '=4' Or '=8' Or '=16' \ Run
Save as <qr upd S NalidixicAcid>
```

```
Query \ Design view \ Update query
Field: StreptomycinRS \ Table: AMRec
Update To: 'R' \ Criteria: '64' Or '>64' Or '=64' \ Run
Save as <qr upd R Streptomycin>
```

```
Query \ Design view \ Update query
```

Field: StreptomycinRS\ Table: AMRec  
Update To: 'S' \ Criteria: '<=32' \ Run  
Save as <qr upd S Streptomycin>

Query \ Design view \ Update query  
Field: SulphamethoxazoleRS\ Table: AMRec  
Update To: 'R' \ Criteria: '512' Or '>512' Or '=512' \ Run  
Save as <qr upd R Sulphamethoxazole>

Query \ Design view \ Update query  
Field: SulphamethoxazoleRS\ Table: AMRec  
Update To: 'S' \ Criteria: '<=16' Or '32' Or '64' Or '128' Or '256' Or '=32' Or '=64'  
Or '=128' Or '=256' \ Run  
Save as <qr upd S Sulphamethoxazole>

Query \ Design view \ Update query  
Field: TetracyclineRS\ Table: AMRec  
Update To: 'R' \ Criteria: '16' Or '>16' Or '32' Or '>32' '=16' Or '=32' \ Run  
Save as <qr upd R Tetracycline>

Query \ Design view \ Update query  
Field: TetracyclineRS\ Table: AMRec  
Update To: 'S' \ Criteria: '<=4' Or '8' Or '=8' \ Run  
Save as <qr upd S Tetracycline>

Query \ Design view \ Update query  
Field: TrimethoprimSulphamethoxazoleRS\ Table: AMRec  
Update To: 'R' \ Criteria: '4' Or '>4' Or '=4' \ Run  
Save as <qr upd R TrimethoprimSulpha>

Query \ Design view \ Update query  
Field: TrimethoprimSulphamethoxazoleRS\ Table: AMRec  
Update To: 'S' \ Criteria: '<=0.1200' Or '0.2500' Or '0.5000' Or '1' Or '2' Or  
'=0.2500' Or '=0.5000' Or '=1' Or '=2' \ Run  
Save as <qr upd S TrimethoprimSulpha>

Note: *The AMR plate used for testing antibiotic resistance to Amikacin does not make any distinction between concentrations over 4 ug/ml. The results classified as '>4' are considered as resistant even if the breakpoint is set at 64 ug/ml.*

*Also, the plate used for Cephoxitin testing does not make distinction between concentrations over 16 ug/ml. The results classified as '>16' are considered as resistant even if the breakpoint is set at 32 ug/ml.*

Add a field in the <AMR> table. Name it 'OverallRS'.

Use a query in two steps to assign the letter R or S to every tested isolate to mention if it is resistant to 'at least one antibiotic' or fully susceptible.

```
Query \ Select Query \ Design view
Field: AMRnumber \ Table: AMR
Field: AllRS: [AmikacinRS] & [AmoxicillinClavulanicAcidRS] &
[AmpicillinRS] & [CefoxitinRS] & [CeftiofurRS] & [CeftriaxoneRS] &
[CephalothinRS] & [ChloramphenicolRS] & [CiprofloxacinRS] &
[GentamicinRS] & [KanamycinRS] & [NalidixicAcidRS] &
[StreptomycinRS] & [SulphamethoxazoleRS] & [TetracyclineRS] &
[TrimethoprimSulphamethoxazoleRS]
Run.
Save query as <qr_overallRS_step1>
```

Link the <qr\_overallRS\_step1> query and <AMR> table by the commun 'AMRnumber' field.

```
Query \ Update query \ Design view
Field: OverallRS \ Table: AMR \ Update to: IIf(InStr([AllRS],"R")>0,"R","S")
Field: AllRS \ Table: qr_overallRS_step1 \ Criteria: is not null
Run.
Save query as <qr_overallRS_step2>
```

Rename the following columns in the <AMR> table to use the same appellations as in the <logbook> table.

Rename 'SubmissionNumber' as 'AMRnumber'

Rename 'SubmittedSampleNumber' as 'RegionLabNumber'

### ***3.1. Antimicrobial-resistance testing results summary***

In total, 606 isolates were send to either the Guelph or St-Hyacinthe labs for AMR testing.

- 190 isolates (31%) were rejected for various reasons
  - 39 (6%) were not E.coli
  - 69 (11%) did not score high enough to be sure they were E.coli
  - 82 (14%) did not show any growth
- 416 isolates (69%) were identified by API 20E as Eschericia coli

Of the 416 E. coli isolates,

- 358 (86%) were susceptible to all antibiotics tested
- 58 (14%) showed resistance to at least on antibiotic
  - ?? 6 showed resistance to Amikacin
  - ?? 4 showed resistance to Amoxicillin Clavulanic Acid
  - ?? 19 showed resistance to Ampicilin
  - ?? 3 showed resistance to Cefoxitin
  - ?? 1 showed resistance to Ceftiofur
  - ?? 0 showed resistance to Ceftriaxone
  - ?? 9 showed resistance to Cephalothin
  - ?? 3 showed resistance to Chloremphenicol
  - ?? 0 showed resistance to Ciprofloxacin
  - ?? 0 showed resistance to Gentamicine
  - ?? 4 showed resistance to Kanamycin
  - ?? 1 showed resistance to Nalidixic Acid
  - ?? 25 showed resistance to Streptomycin
  - ?? 22 showed resistance to Sulphamethoxazole
  - ?? 38 showed resistance to Tetracycline
  - ?? 10 showed resistance to TrimethoprimSulphamethoxazole

#### 4. LINKING THE LOGBOOK AND THE AMR RESULTS

Perform a query to append the information contained in the <logbook> table to the <AMR> table.

```
Link both tables by the 'AMRnumber' field
Query \ Make table Query \ Design view
Field: AMR* \ Table: AMR
Field: Logbook* \ Table: Logbook
Name new table <AMRlog> \ Run
Save query as <Or mkt AMRlog>
```

Create a sub-table containing only the E.coli isolates.

```
Query \ Design view \ Make Table query
Field: \ * use all fields in the table
Field: \ FinalIDName \ Table: AMRlog
Criteria: = 'Escherichia coli' \ Run
Name table <AMRlogecoli>
Save as <qr mkt amrecoli>
```

The resulting <AMRlogecoli> table contains 416 records.

#### 5. GEOCODING A REFERENCE DATABASE

### 5.1. merging reference databases

The reference database to be used for geocoding was provided in two separate files. One database covers the months of April and October 2003 while the other one covers the period from May to September 2003. Both databases contain the same fields of information and are structured the exact same way.

Merge both databases in order to create the main one that will contain all records from April 1<sup>st</sup> to October 31<sup>st</sup> 2003.

The database covering the months of April 2003 and October 2003 is containing 52,540 while the one covering from May to September 2003 represents 129,019 entries.

Make a copy of the <WaterTestMaySept> table \ Rename <WaterTest>

Use an append query to append the entries of the <WaterTestAprilOct> table.

```
Query \ Design view \ Query type: Append query
Field: WaterTestAprilOct*
Table: WaterTestAprilOct
Append To: WaterTest*
Save Query as: <qr_append_2DBs> \ Run
```

The <WaterTest> table now contains 181,559 entries.

Geocoding was performed as detailed in a previous report entitled '*Developing a geocoding algorithm and geocoding a database related to water testing V. 2.2*' – Caroline Guénette, March 2005.

### 5.2. Pre-processing results

The table called Wtd in the AMRgcd database contains the pre-processing results.

Of the initial 181,559 entries in the database :

- 1 was removed because it was a double;
- 776 were labelled 'POBOXONLY' in the gcdcode column as they contained PO Box alike information only in the address field;
- 1187 were labelled 'NOGEO' as they contained no usable spatial information;
- 181 were labeled 'NOGEO' as they contained lot/concession numbers as only spatial identifier but no municipality information;
- 14 520 were labelled 'NWT' as they did not contained any water test results;

167 038 records (92%) are potentially usable for geocoding and do have a water testing result.

Of the 181 558 distinct entries in the database, 57 310 (~32%) were modified to better match the reference databases at geocoding :

- 2732 were modified to remove the extra PO Box alike information;
- 583 were modified because they contained brackets in the address field;
- 39 were modified because they contained quotation marks in the address field;
- 54 were modified as they contained invalid Ontario postal codes;
- 53 902 were modified to standardize the lot/concession information.

### ***5.3. Geocoding results using GeoPinpoint***

- 37,454 (~20%) Have a successful geocoding result using the strict civic address
- 6,449 (~4%) Have a successful geocoding result using the relaxed civic address
- 23,586 (~13%) Have a successful geocoding result using six digits postal code
- 46,645 (~26%) Have a successful geocoding result using five digits postal code
- 148,536 (~82%) Have a successful geocoding result using tree digits postal code
- 138,411 (~76%) Have a successful geocoding result using municipality

As these numbers show, many entries had successful geocoding results using different information sources. In a further step, the most appropriate result is selected for use.

### ***5.4. Geocoding results for Lot/Concession***

- 14,211 (~8%) Have a successful geocoding result using the lot/concession numbers in the official municipality name
- 14,642 (~8%) Have a successful geocoding result using the lot/concession numbers in the city name appellation

### ***5.5. Selected geocoding results***

The columns <select\_x> and <select\_Y> of the <GCD\_main> table contain the geocoding results considered as the most appropriate to use.

Of the 181 558 entries in the initial database,

For 43,903 (~24%) entries, geocoding using the civic address was considered as the best option;

For 17,373(~10%) entries, geocoding using the six digits postal code was considered as the best option;

For 34,883 (~19%) entries, geocoding using the five digits postal code was considered as the best option;

For 30,340 (~17%) entries, geocoding using the three digits postal code was considered as the best option;

For 11,392 (~ 6%) entries, geocoding according to lot/concession numbers was considered as the best option;

For 34,681 (~ 19%) entries, geocoding to the centroid of their municipality was considered as the best option;

In total : 172,572 entries (~95%) were geocoded while 8,986 (~5%) were not.

## 6. LINKING AMR RESULTS TO THE GEOCODING RESULTS

Within the <AMRlogecoli> table, replace all records having 'Toronto' in their 'RegionLab' column by 'Central'. It will facilitate merges.

Use a query to link the <AMRlogecoli> table to the <GCD\_main> table previously geocoded.

Link the 'AccessionBU' field in the <AMRlogecoli> table with the 'Accession' field in the <GCD\_main> table AND link the 'RegionLab' field in the <AMRlogecoli> table to the 'Labo' field in the <GCD\_Main> table.

Use the join property 'Include all records from 'AMRlogecoli' and only those records from 'Gcd\_main' where the joined fields are equals' for both links.

```
Query \ Design view \ Make Table query
Field: * use all fields in the table \ Table: AMRlogecoli
Field: * use all fields in the table \ Table: Gcd_main
Name table <AMRlogecoliGCD> \ Run
Save as <qr_mkt_AMRlogecoliGCD>
```

Perform a query to find duplicates based on the 'AMRnumber' field. It represents records in the <AMRlogecoli> table that have duplicate records associated in the <Gcd\_main> database, meaning that the same Ontario regional laboratory used the same Accession number more than once in the time frame covered by this study.

When identified by the query, go back to the sample log sheet provided with the samples (paper format) and look for the 'collection date'. Manually delete the duplicates collected at another date than the one stated on the sample logsheet.

### ***6.1. Linking AMR- geocoding performance summary***

Out of the 416 E.coli isolates analysed for AMR,

- 34 do not have a link to the geographic reference database
- 382 have a link to the geographic reference database

Some accession numbers are in the database but the laboratory of origin does not fit or in some cases the Accession number is not found in the geographic reference database.

See <nofit\_error> table \ linkproblem column for more details

Out of the 382 AMR-analysis with a link to the geographic reference database:

- 10 do not have a successful geocoding
- 372 have a successful geocoding result

In total, 372 (~90%) of the initial 416 E. coli analysed can be associated with a longitude/latitude.

**7. APPENDIX A: BREAKPOINTS TABLE**

<b>Antibiotic</b>	<b>Breakpoint for resistance (MIC)</b>
Amikacin	$\geq 64$
Amoxicillin/Clavulanic Acid	$\geq 32$
Ampicillin	$\geq 32$
Cefoxitin	$\geq 32$
Ceftiofur	$\geq 8$
Ceftriaxone	$\geq 64$
Cephalothin	$\geq 32$
Chloramphenicol	$\geq 32$
Ciprofloxacin	$\geq 4$
Gentamicin	$\geq 16$
Kanamycin	$\geq 64$
Nalidixic Acid	$\geq 32$
Streptomycin	$\geq 64$
Sulphamethoxazole	$\geq 512$
Tetracycline	$\geq 16$
Trimethoprim/Sulphamethoxazole	$\geq 4$

**Annexe 4**

Résultats obtenus suite au lien entre les résultats de résistance antimicrobienne et les coordonnées géographiques

Document :

PERFORMANCE DESCRIPTION/ Linkage of antimicrobial-resistance results to geographic coordinates, version 1.1 / 2004 Dataset – ARO Study / par Caroline Guénette

**PERFORMANCE DESCRIPTION**

**Linkage of antimicrobial-resistance results to geographic coordinates**

**2004 dataset – ARO Study**

**Caroline Guénette, September 2005**

<b>1. OVERVIEW .....</b>	<b>3</b>
<b>2. ANTIMICROBIAL-RESISTANCE TESTING RESULTS SUMMARY.....</b>	<b>4</b>
<b>3. LINKING AMR- GEOCODING PERFORMANCE SUMMARY .....</b>	<b>5</b>
<b>4. APPENDIX A: BREAKPOINTS TABLE .....</b>	<b>6</b>

## 1. OVERVIEW

Antimicrobial-resistance (AMR) results needed to have a geographic location assigned to them in order to perform geospatial analysis. To obtain the information needed, three major steps had to be done: formatting a logbook, merging it to the previously cleaned AMR results and finally merging the AMR results with a previously geocoded reference database.

The steps to follow to perform the linkage and databases merging are described in a previous report entitled 'Development; testing and performance description of an algorithm for linkage of antimicrobial-resistance results to geographic coordinates – Caroline Guénette, V. 2.1, September 2005'.

The present document summarizes the performances when applying the algorithm to the 2004 dataset of the ARO Study.

Note: *The AMR plate used for testing antibiotic resistance to Amikacin does not make any distinction between concentrations over 4 ug/ml. The results classified as '>4' are considered as resistant even if the breakpoint is set at 64 ug/ml.*

*Also, the plate used for Cephoxitin testing does not make distinction between concentrations over 16 ug/ml. The results classified as '>16' are considered as resistant even if the breakpoint is set at 32 ug/ml.*

## 2. ANTIMICROBIAL-RESISTANCE TESTING RESULTS SUMMARY

In total, 1564 presumptive E.coli sample coming from distinct private drinking water sources in Ontario were send to the study for analysis.

The initial laboratory screening process (see SOP document) identified 1111 out of the 1564 samples (71%) as fully susceptible to all antibiotics involved. The screening process identified 453 samples (29%) as potentially resistant. Those 453 samples were send to the St-Hyacinthe lab for full AMR analysis.

Of the 453 potentially resistant presumptive E.coli coming from distinct Ontario private drinking water source:

- 72 isolates (16%) were rejected for various reasons
  - 51 (11%) were not E.coli
  - 21 (5%) did not score high enough to be sure they were E.coli
- 381 isolates (84%) were identified by API 20E as Eschericia coli

Of the 381 E. coli isolates,

- 143 (38%) were susceptible to all antibiotics tested
- 238 (62%) showed resistance to at least on antibiotic
  - 8 showed resistance to Amikacin
  - 14 showed resistance to Amoxicillin Clavulanic Acid
  - 76 showed resistance to Ampicilin
  - 16 showed resistance to Cefoxitin
  - 11 showed resistance to Ceftiofur
  - 0 showed resistance to Ceftriaxone
  - 54 showed resistance to Cephalothin
  - 23 showed resistance to Chloremphenicol
  - 4 showed resistance to Ciprofloxacin
  - 9 showed resistance to Gentamicine
  - 27 showed resistance to Kanamycin
  - 14 showed resistance to Nalidixic Acid
  - 94 showed resistance to Streptomycin
  - 107 showed resistance to Sulphamethoxazole
  - 168 showed resistance to Tetracycline
  - 56 showed resistance to TrimethoprimSulphamethoxazole

### 3. LINKING AMR- GEOCODING PERFORMANCE SUMMARY

Out of the 381 E.coli isolates analysed for AMR,

- 15 do not have a link to the geographic reference database
- 366 have a link to the geographic reference database

Some accession numbers are in the database but the laboratory of origin does not fit or in some cases the Accession number is not found in the geographic reference database.

See `<nofit_error>` table \linkproblem column for more details

Out of the 366 AMR-analysis with a link to the geographic reference database:

- 12 do not have a successful geocoding
- 354 have a successful geocoding result

In total, 354 (~93%) of the initial 381 E. coli analysed can be associated with a longitude/latitude.

#### 4. APPENDIX A: BREAKPOINTS TABLE

Antibiotic	Breakpoint for resistance (MIC)
Amikacin	$\geq 64$
Amoxicillin/Clavulanic Acid	$\geq 32$
Ampicillin	$\geq 32$
Cefoxitin	$\geq 32$
Ceftiofur	$\geq 8$
Ceftriaxone	$\geq 64$
Cephalothin	$\geq 32$
Chloramphenicol	$\geq 32$
Ciprofloxacin	$\geq 4$
Gentamicin	$\geq 16$
Kanamycin	$\geq 64$
Nalidixic Acid	$\geq 32$
Streptomycin	$\geq 64$
Sulphamethoxazole	$\geq 512$
Tetracycline	$\geq 16$
Trimethoprim/Sulphamethoxazole	$\geq 4$

## Annexe 5

### Tableaux des données

Tableau de fréquence de puits positifs par patron de résistance selon la SRU

Subdivision de recensement unifiée	Nb de puits échantillonnés	Fréquence absolue de puits positifs selon le patron de résistance					Fréquence relative de puits positifs selon le patron de résistance (%)				
		1	2	3	4	5	1	2	3	4	5
Adelaide Metcalfe	4	4	2	2	0	0	100	50	50	0.0	0.0
Adjala-Tosorontio	2	0	0	0	0	0	0	0	0	0.0	0.0
Alfred and Plantagenet	3	1	1	0	0	0	33.3	33.3	0	0.0	0.0
Amaranth	1	0	0	0	0	0	0	0	0	0.0	0.0
Armour	1	1	0	1	0	0	100	0	100	0.0	0.0
Asphodel-Norwood	1	1	1	0	0	0	100	100	0	0.0	0.0
Athens	1	1	0	1	0	0	100	0	100	0.0	0.0
Bayham	1	1	0	1	0	0	100	0	100	0.0	0.0
Beckwith	1	1	1	0	0	0	100	100	0	0.0	0.0
Belleville	3	2	1	1	0	0	66.7	33.3	33.3	0.0	0.0
Blandford-Blenheim	2	1	1	0	1	0	50	50	0	50.0	0.0
Bluewater	1	0	0	0	0	0	0	0	0	0.0	0.0
Bracebridge	3	1	1	0	1	0	33.3	33.3	0	33.3	0.0
Bradford West Gwillimbury	1	0	0	0	0	0	0	0	0	0.0	0.0
Brant	5	4	1	3	0	0	80	20	60	0.0	0.0
Brighton	2	1	1	0	0	0	50	50	0	0.0	0.0
Brockton	2	1	1	0	0	0	50	50	0	0.0	0.0
Burlington	3	2	2	0	0	1	66.7	66.7	0	0.0	33.3
Caledon	5	5	3	2	1	0	100	60	40	20.0	0.0
Campbellford/Seymour,Percy,Hastings	1	0	0	0	0	0	0	0	0	0.0	0.0
Central Elgin	1	1	1	0	0	0	100	100	0	0.0	0.0
Central Frontenac	1	1	0	1	0	0	100	0	100	0.0	0.0
Central Huron	1	1	0	1	0	0	100	0	100	0.0	0.0
Centre Hastings	1	1	1	0	0	0	100	100	0	0.0	0.0
Centre Wellington	3	2	0	2	0	0	66.7	0	66.7	0.0	0.0
Chatsworth	3	3	1	2	0	0	100	33.3	66.7	0.0	0.0
Clarington	3	2	0	2	0	0	66.7	0	66.7	0.0	0.0
Douro-Dummer	3	1	1	0	0	1	33.3	33.3	0	0.0	33.3
Dysart and Others	5	2	1	1	0	0	40	20	20	0.0	0.0
East Zorra-Tavistock	1	0	0	0	0	0	0	0	0	0.0	0.0
Edwardsburgh/Cardinal	1	1	0	1	0	0	100	0	100	0.0	0.0
Elizabethtown-Kitley	2	1	0	1	0	0	50	0	50	0.0	0.0
Essa	2	1	1	0	0	0	50	50	0	0.0	0.0
Fort Erie	4	3	2	1	0	0	75	50	25	0.0	0.0

Subdivision de recensement unifiée	Nb de puits échantillonnés	Fréquence absolue de puits positifs selon le patron de résistance					Fréquence relative de puits positifs selon le patron de résistance (%)				
		1	2	3	4	5	1	2	3	4	5
Galway-Cavendish and Harvey	2	1	1	0	0	0	50	50	0	0.0	0.0
Georgian Bluffs	5	4	2	2	0	1	80	40	40	0.0	20.0
Georgian Highlands	18	13	9	4	0	0	72.2	50	22.2	0.0	0.0
Georgina	7	3	2	1	0	0	42.9	28.6	14.3	0.0	0.0
Gravenhurst	3	2	1	1	0	0	66.7	33.3	33.3	0.0	0.0
Greater Napanee	2	2	1	1	0	0	100	50	50	0.0	0.0
Grey Highlands	5	4	3	1	0	0	80	60	20	0.0	0.0
Guelph/Eramosa	3	2	2	0	1	0	66.7	66.7	0	33.3	0.0
Haldimand	19	9	4	5	0	1	47.4	21.1	26.3	0.0	5.3
Halton Hills	5	3	2	1	0	0	60	40	20	0.0	0.0
Hamilton	9	7	5	2	0	0	100	100	0	0.0	0.0
Hamilton	2	2	2	0	0	2	77.8	55.6	22.2	0.0	22.2
Hastings Highlands	1	1	1	0	0	0	100	100	0	0.0	0.0
Havelock-Belmont-Methuen	1	0	0	0	0	0	0	0	0	0.0	0.0
Huntsville	1	0	0	0	0	0	0	0	0	0.0	0.0
Huron East	1	1	0	1	0	0	100	0	100	0.0	0.0
Innisfil	1	0	0	0	0	0	0	0	0	0.0	0.0
Kawartha Lakes	11	4	2	2	0	0	36.4	18.2	18.2	0.0	0.0
Kincardine	1	1	0	1	0	0	100	0	100	0.0	0.0
King	1	1	1	0	1	0	100	100	0	100.0	0.0
Kingston	3	3	2	1	0	1	100	66.7	33.3	0.0	33.3
Lambton Shores	1	1	1	0	0	0	100	100	0	0.0	0.0
Lanark Highlands	3	2	0	2	0	0	66.7	0	66.7	0.0	0.0
Leeds and the Thousand Islands	2	2	2	0	0	0	100	100	0	0.0	0.0
Lincoln	4	2	2	0	0	1	50	50	0	0.0	25.0
Loyalist	4	3	0	3	0	0	75	0	75	0.0	0.0
Madoc	1	1	1	0	0	0	100	100	0	0.0	0.0
Malahide	2	2	1	1	0	0	100	50	50	0.0	0.0
Mapleton	1	1	1	0	0	0	100	100	0	0.0	0.0
Marmora	1	1	1	0	0	0	100	100	0	0.0	0.0
McKellar	6	2	0	2	0	0	33.3	0	33.3	0.0	0.0
Melancthon	2	2	2	0	0	0	100	100	0	0.0	0.0
Middlesex Centre	1	1	1	0	0	0	100	100	0	0.0	0.0
Milton	7	2	1	1	0	0	28.6	14.3	14.3	0.0	0.0
Mississippi Mills	2	2	0	2	0	0	100	0	100	0.0	0.0
Mono	1	1	1	0	0	0	100	100	0	0.0	0.0
Montague	1	1	0	1	0	0	100	0	100	0.0	0.0
Muskoka Lakes	3	3	3	0	1	0	100	100	0	33.3	0.0
Niagara-on-the-Lake	3	1	1	0	0	0	33.3	33.3	0	0.0	0.0
Norfolk	4	1	0	1	0	0	25	0	25	0.0	0.0
North Dumfries	1	1	1	0	0	0	100	100	0	0.0	0.0
North Dundas	1	0	0	0	0	0	0	0	0	0.0	0.0

Subdivision de recensement unifiée	Nb de puits échantillonnés	Fréquence absolue de puits positifs selon le patron de résistance					Fréquence relative de puits positifs selon le patron de résistance (%)				
		1	2	3	4	5	1	2	3	4	5
North Frontenac	1	0	0	0	0	0	0	0	0	0.0	0.0
North Glengarry	6	1	1	0	0	0	16.7	16.7	0	0.0	0.0
North Huron	1	1	1	0	0	0	100	100	0	0.0	0.0
North Perth	1	1	0	1	0	0	100	0	100	0.0	0.0
North Stormont	1	0	0	0	0	0	0	0	0	0.0	0.0
Northern Bruce Peninsula	3	2	1	1	0	0	66.7	33.3	33.3	0.0	0.0
Oro-Medonte	7	2	1	1	0	0	28.6	14.3	14.3	0.0	0.0
Otonabee-South Monaghan	3	0	0	0	0	0	0	0	0	0.0	0.0
Ottawa	14	9	3	6	0	0	64.3	21.4	42.9	0.0	0.0
Parry Sound, Unorganized, Centre Part	1	0	0	0	0	0	0	0	0	0.0	0.0
Pelham	1	0	0	0	0	0	0	0	0	0.0	0.0
Port Colborne	5	5	2	3	0	1	100	40	60	0.0	20.0
Prince Edward	4	3	1	2	0	1	75	25	50	0.0	25.0
Puslinch	2	1	1	0	0	0	50	50	0	0.0	0.0
Quinte West	4	4	2	2	0	0	100	50	50	0.0	0.0
Ramara	2	1	0	1	0	0	50	0	50	0.0	0.0
Richmond Hill	2	1	0	1	0	0	50	0	50	0.0	0.0
Rideau Lakes	2	1	1	0	0	0	50	50	0	0.0	0.0
Scugog	2	2	1	1	0	0	100	50	50	0.0	0.0
Severn	3	1	0	1	0	0	33.3	0	33.3	0.0	0.0
Smith-Ennismore-Lakefield	1	0	0	0	0	0	0	0	0	0.0	0.0
South Dundas	1	0	0	0	0	0	0	0	0	0.0	0.0
South Frontenac	8	8	3	5	0	0	100	37.5	62.5	0.0	0.0
Springwater	1	0	0	0	0	0	0	0	0	0.0	0.0
Stone Mills	2	0	0	0	0	0	0	0	0	0.0	0.0
Strathroy-Caradoc	1	1	0	1	0	0	100	0	100	0.0	0.0
Tay	1	0	0	0	0	0	0	0	0	0.0	0.0
Thames Centre	2	2	2	0	0	0	100	100	0	0.0	0.0
The Nation Municipality	2	1	1	0	0	1	50	50	0	0.0	50.0
Thorold	1	0	0	0	0	0	0	0	0	0.0	0.0
Tiny	4	0	0	0	0	0	0	0	0	0.0	0.0
Tyendinaga	2	2	1	1	0	0	100	50	50	0.0	0.0
Wainfleet	4	2	1	1	0	0	50	25	25	0.0	0.0
Warwick	1	1	0	1	0	0	100	0	100	0.0	0.0
Wellesley	3	3	0	3	0	0	100	0	100	0.0	0.0
Wellington North	2	2	0	2	0	0	100	0	100	0.0	0.0
West Grey	4	3	2	1	0	0	75	50	25	0.0	0.0
West Lincoln	10	7	5	2	0	0	70	50	20	0.0	0.0
West Perth	2	2	2	0	0	0	100	100	0	0.0	0.0
Whitestone	1	1	1	0	0	0	100	100	0	0.0	0.0
Whitewater Region	2	1	0	1	0	0	50	0	50	0.0	0.0
Wilmot	4	3	2	1	1	0	75	50	25	25.0	0.0

Subdivision de recensement unifiée	Nb de puits échantillonnés	Fréquence absolue de puits positifs selon le patron de résistance					Fréquence relative de puits positifs selon le patron de résistance (%)				
		1	2	3	4	5	1	2	3	4	5
Woolwich	3	3	1	2	0	0	100	33.3	66.7	0.0	0.0
Zorra	1	1	1	0	0	0	100	100	0	0.0	0.0
Total	354	220	120	100	7	11					

Tableau de données des variables agricoles (densités et utilisations des terres) selon la SRU

Subdivision de recensement unifiée	DBo	DPo	DVo	DEq	DRu	UCult	UJach	UPatCE	UPatNat	UEpFS	UEpSI	UEpFLS	UEpFLI
Adelaide Metcalfe	0.27424	2.09118	17.49130	0.00782	0.11158	0.80330	0.00096	0.03779	0.04576	0.05784	0.03051	0.10226	.
Adjala-Tosorontio	0.23438	0.06781	8.74832	0.02945	0.07276	0.65310	0.01261	0.04102	0.09783	0.07687	.	0.02757	0.00000
Alfred and Plantagenet	0.46429	0.56900	7.33963	0.01064	0.07356	0.72665	0.00163	0.06248	0.08642	0.09626	.	0.05733	.
Amaranth	0.44324	0.53049	.	0.02474	0.20980	0.69318	0.01228	0.08779	0.08608	0.11810	.	0.04515	.
Armour	0.08323	.	0.19281	.	0.00000	0.20659	.	0.11198	0.23234	0.02814	0.00000	.	0.00000
Asphodel-Norwood	0.46041	0.04245	22.60522	0.01747	0.31771	0.54759	0.00416	0.06400	0.16971	0.12139	0.00000	0.01633	0.00000
Athens	0.50323	0.00000	0.02420	0.00610	0.01918	0.53621	.	0.08265	0.11402	0.12406	0.00000	0.03263	0.00000
Bayham	0.16001	0.00368	0.10356	0.01939	0.00866	0.73405	0.00539	0.02281	0.00695	0.06038	.	0.00503	0.00000
Beckwith	0.38971	0.00226	0.03669	0.01560	0.07306	0.40951	.	0.10179	0.15612	0.07618	0.00000	.	0.00000
Belleville	0.36689	0.00720	.	0.01965	0.04542	0.59119	.	0.03511	0.13588	0.09911	0.00000	0.01294	.
Blandford-Blenheim	0.46681	1.23493	5.50259	0.00837	0.07371	0.78414	0.00149	0.02822	0.03927	0.09307	0.01423	0.05192	0.00777
Bluewater	0.16119	1.22183	18.61948	0.00305	0.01753	0.88592	0.00000	0.01538	0.01533	0.04773	0.00884	0.03553	0.00750
Bracebridge	0.09065	0.01888	0.26073	0.02634	0.00000	0.12694	0.00000	0.03920	0.13068	0.01805	0.00000	.	0.00000
Bradford West Gwillimbury	0.17499	0.19836	.	0.01252	0.08362	0.83753	0.00288	0.02406	0.03847	0.04659	0.00000	0.00865	.
Brant	0.26074	0.34590	5.57455	0.01950	0.05118	0.81319	0.00357	0.02618	0.02675	0.07965	0.00438	0.01657	0.00410
Brighton	0.36014	0.26559	18.76969	0.00418	0.04748	0.61888	.	0.03411	0.08577	0.11988	0.00000	.	.
Brockton	0.80618	0.75830	3.87541	0.01168	0.10307	0.65975	0.00184	0.09754	0.07951	0.17959	.	0.02874	0.00899
Burlington	0.74266	.	.	0.05465	0.00000	0.77773	0.00734	.	0.03936	0.03405	0.00000	0.02529	.
Caledon	0.48686	.	.	0.05270	0.04335	0.74466	.	0.04279	0.09394	0.10942	.	0.01973	0.00191
Campbellford/Seymour,Percy.Hastings	0.44018	0.19770	0.42880	0.01106	0.09784	0.59145	0.00272	0.07009	0.14047	0.10264	0.00000	0.02060	.
Central Elgin	0.12494	0.23311	2.08820	0.01382	0.01022	0.83906	0.00071	0.01933	0.01184	0.04712	0.00895	0.01255	.
Central Frontenac	0.15840	0.00216	.	0.00415	0.02690	0.15818	.	0.04395	0.42639	0.05679	0.00000	0.00000	0.00000
Central Huron	0.30027	1.55506	60.22056	0.00482	0.13035	0.80305	0.00096	0.04200	0.02190	0.09671	0.01073	0.04497	.
Centre Hastings	0.34196	.	0.03765	0.00710	0.00321	0.47865	0.00370	0.05613	0.14097	0.07374	0.00000	0.02004	.
Centre Wellington	0.67764	1.43143	14.91098	0.02376	0.11417	0.80409	0.00059	0.04347	0.03972	0.14130	0.00177	0.11435	.
Chatsworth	0.68373	0.18577	1.80573	0.02314	0.13155	0.50054	0.00608	0.16045	0.11538	0.15471	.	0.01396	0.00000
Clarington	0.35518	0.29782	.	0.02479	0.09673	0.67587	0.00434	0.06993	0.11338	0.09766	0.00000	0.01665	.
Douro-Dummer	0.43266	0.01960	0.08900	0.01882	0.04654	0.39448	0.00413	0.12042	0.23083	0.08993	0.00000	.	.
Dysart and Others	0.05236	.	0.07791	0.01533	0.00000	0.11494	.	0.00958	0.36398	0.05172	0.00000	0.00000	0.00000
East Zorra-Tavistock	0.67232	2.86985	28.12576	0.01068	0.07280	0.88518	.	0.01632	0.01662	0.12455	0.01735	0.15301	0.00383
Edwardsburgh/Cardinal	0.34493	0.32868	.	0.02126	0.03946	0.56544	.	0.06956	0.07798	0.10324	.	0.02551	0.00000
Elizabethtown-Kitley	0.33633	0.06187	.	0.01074	0.08868	0.45229	0.00179	0.06125	0.16042	0.07957	.	0.02257	0.00337
Essa	0.34586	0.29743	0.05340	0.02264	0.04339	0.73680	0.01960	0.05092	0.05887	0.08539	0.00000	0.03732	0.00000
Fort Erie	0.19649	.	28.51945	0.04492	0.00176	0.69360	0.07202	0.05646	0.02785	0.06575	0.00000	.	0.00000
Galway-Cavendish and Harvey	0.24055	.	0.04154	0.01485	0.02415	0.18341	.	0.06809	0.35618	0.05924	0.00000	.	0.00000
Georgian Bluffs	0.61699	0.03444	0.05275	0.01147	0.20092	0.44234	.	0.18807	0.11414	0.14274	0.00000	0.01153	0.00000
Georgian Highlands	0.51829	0.05972	1.03867	0.01221	0.22077	0.50428	0.00151	0.12296	0.16692	0.12241	0.00000	0.01913	0.00000
Georgina	0.19753	0.39705	9.23284	0.03355	0.32937	0.69339	0.00198	0.03465	0.13896	0.05483	0.00000	0.01138	.
Gravenhurst	0.14104	.	0.52500	0.03868	0.09528	0.23962	0.00000	0.05330	0.24717	0.03349	0.00000	0.00000	0.00000
Greater Napanee	0.31353	.	.	0.01638	0.06995	0.54135	.	0.04703	0.22301	0.07978	.	0.02179	.
Grey Highlands	0.52774	0.13757	0.50967	0.01526	0.11581	0.52744	0.00306	0.13178	0.10494	0.14078	.	0.00788	0.00121
Guelph/Eramosa	0.42110	0.36523	2.88745	0.03202	0.11999	0.81274	0.00314	0.03412	0.02354	0.08681	0.00000	0.03785	.

Subdivision de recensement unifiée	DBo	DPo	DVo	DEq	DRu	UCult	UJach	UPatCE	UPatNat	UEpFS	UEpSI	UEpFLS	UEpFLI
Haldimand	0.28158	0.59112	13.89569	0.01223	0.08544	0.80651	0.00518	0.03720	0.03450	0.08152	0.00599	0.02287	0.00387
Halton Hills	0.24891	.	.	0.04083	0.02778	0.79777	0.00564	0.03342	0.04830	0.07598	0.00000	0.01940	.
Hamilton	0.32132	0.32116	35.33495	0.04149	0.09921	0.74859	0.00875	0.04583	0.04551	0.07509	0.00505	0.02116	0.00657
Hamilton	0.34419	0.31747	5.38476	0.02171	0.06567	0.61124	0.00192	0.07175	0.16467	0.07991	.	0.02048	0.00000
Hastings Highlands	0.08651	.	0.33557	0.00913	0.03224	0.18619	.	0.05884	0.13702	0.05535	.	.	0.00000
Havelock-Belmont-Methuen	0.28283	.	0.12921	0.01756	0.09024	0.29003	.	0.07374	0.29424	0.08743	0.00000	0.00000	0.00000
Huntsville	0.10705	0.02725	0.63994	0.05644	0.10627	0.33398	.	0.12145	0.10588	0.04632	0.00000	.	0.00000
Huron East	0.56310	4.06651	11.26277	0.00521	0.05887	0.84876	.	0.03057	0.01947	0.08746	0.01761	0.10636	0.02341
Innisfil	0.26036	1.43340	6.97791	0.02329	0.18186	0.80020	0.00283	0.03191	0.05162	0.07006	0.00000	0.03771	.
Kawartha Lakes	0.40059	0.08241	1.46967	0.01542	0.10879	0.44946	0.00277	0.09098	0.26268	0.08984	0.00043	0.00949	.
Kincardine	0.60040	0.45133	2.64441	0.01320	0.31208	0.65528	0.00050	0.15079	0.07595	0.12905	0.00000	0.03751	.
King	0.19566	0.04954	.	0.07472	0.13153	0.74881	0.01396	0.04873	0.06187	0.08036	.	0.00802	0.00000
Kingston	0.42768	0.00412	.	0.01878	0.05591	0.50259	0.00184	0.06315	0.20997	0.10579	0.00000	.	0.00000
Lambton Shores	0.15979	1.85767	.	0.00222	0.03542	0.87842	.	0.02277	0.02250	0.03410	0.02293	0.06763	0.00381
Lanark Highlands	0.20229	.	0.06960	0.00611	0.10031	0.18908	.	0.06694	0.21263	0.03822	0.00000	.	0.00000
Leeds and the Thousand Islands	0.38393	0.05514	.	0.01129	0.09810	0.50499	0.00219	0.07180	0.18377	0.09853	0.00000	0.01640	0.00530
Lincoln	0.10160	0.46202	54.06941	0.02025	0.00681	0.72473	0.01848	0.02812	0.02741	0.16341	.	0.01627	0.00000
Loyalist	0.24669	.	0.05898	0.01386	0.25219	0.47694	.	0.07031	0.21835	0.06414	.	0.01055	.
Madoc	0.29124	0.02240	0.04409	0.00750	0.11308	0.42271	0.00312	0.05578	0.21198	0.07604	0.00000	.	0.00000
Malahide	0.44840	0.92756	7.74442	0.01516	0.06573	0.82381	0.00138	0.02451	0.01338	0.09504	.	0.09480	0.01818
Mapleton	0.99157	2.19375	37.15702	0.01970	0.09919	0.85549	0.00053	0.03047	0.02324	0.19973	0.00992	0.13829	0.00367
Marmora	0.26443	.	0.16464	0.02356	0.06697	0.32080	0.00000	0.05056	0.20911	0.11117	0.00000	0.00000	0.00000
McKellar	0.15423	.	0.83744	0.00834	0.01099	0.29291	.	0.12050	0.18189	0.09739	0.00000	0.00000	0.00000
Melancthon	0.39289	0.04714	.	0.00868	0.11989	0.66229	0.00319	0.09313	0.10598	0.11550	0.00000	.	0.00000
Middlesex Centre	0.25447	0.78855	11.89803	0.01878	0.04203	0.85610	0.00073	0.02844	0.02371	0.05582	0.00968	0.05101	0.00132
Milton	0.20022	.	17.11909	0.10192	0.09270	0.69786	.	0.05991	0.06323	0.07789	0.00000	0.00700	.
Mississippi Mills	0.37503	0.00679	0.09250	0.01636	0.04572	0.44263	0.00233	0.08436	0.18255	0.08260	0.00000	0.00523	0.00000
Mono	0.58987	.	0.20585	0.05583	0.09419	0.55897	0.00206	0.08065	0.16915	0.08281	0.00000	.	0.00000
Montague	0.27661	.	0.04837	0.01827	0.07015	0.31690	.	0.12144	0.23773	0.05809	0.00000	0.00000	0.00000
Muskoka Lakes	0.18554	0.00898	0.12114	0.03636	0.00000	0.25411	0.00000	0.11676	0.17393	0.11260	0.00000	.	0.00000
Niagara-on-the-Lake	0.04161	.	47.16913	0.00897	0.03813	0.75348	0.01884	0.02905	0.03948	0.16420	0.00000	.	.
Norfolk	0.06760	0.24482	11.07769	0.00708	0.01940	0.75508	0.00723	0.01432	0.01979	0.07590	0.00231	0.00926	0.00089
North Dumfries	0.45195	.	38.73603	0.02504	0.07196	0.74135	.	0.05530	0.04968	0.11385	.	0.06230	.
North Dundas	0.48358	.	.	0.00843	0.09429	0.83046	0.00205	0.04598	0.04235	0.07513	0.02094	0.04583	0.00000
North Frontenac	0.04913	0.00000	.	0.01048	0.00000	0.09106	.	0.02326	0.14314	0.01670	0.00000	0.00000	0.00000
North Glengarry	0.44418	0.07958	.	0.00816	0.05223	0.62251	0.00247	0.06807	0.09633	0.07977	0.00300	0.05242	.
North Huron	0.81527	0.88009	.	0.00867	0.07656	0.70207	0.00063	0.06279	0.04957	0.15599	.	0.03545	0.00000
North Perth	0.77806	1.88846	29.62404	0.01261	0.04046	0.82991	.	0.03996	0.03975	0.16824	0.00866	0.12811	0.00380
North Stormont	0.44387	.	0.06055	0.00454	0.03181	0.73427	0.00090	0.03083	0.05814	0.05388	0.01152	0.05830	.
Northern Bruce Peninsula	0.52676	0.00000	.	0.00516	0.01162	0.28634	.	0.18070	0.26122	0.05922	0.00000	0.00000	0.00000
Oro-Medonte	0.39782	0.22497	0.10819	0.02625	0.26678	0.62377	0.00357	0.09166	0.08425	0.10110	0.00000	0.01274	0.00000
Otonabee-South Monaghan	0.42414	0.02672	.	0.01378	0.19349	0.57432	0.00838	0.11830	0.10941	0.09925	0.00000	0.01167	.
Ottawa	0.40782	0.06600	1.77190	0.02307	0.06446	0.66627	0.00642	0.06178	0.11278	0.06562	0.01251	0.03232	.
Parry Sound, Unorganized, Centre Part	0.09849	.	0.19610	0.00425	0.01045	0.24429	.	0.06554	0.10115	0.03012	0.00000	0.00000	0.00000

Subdivision de recensement unifiée	DBo	DPo	DVo	DEq	DRu	UCult	UJach	UPatCE	UPatNat	UEpFS	UEpSI	UEpFLS	UEpFLI
Pelham	0.13951	0.02214	31.05426	0.03415	0.00645	0.82090	.	0.01366	0.01581	0.08411	.	0.02846	0.00000
Port Colborne	0.09882	.	48.07573	0.01441	0.00000	0.80015	0.00425	0.02050	0.02789	0.05098	0.00000	0.00000	0.00000
Prince Edward	0.28018	0.01877	4.31403	0.01014	0.10499	0.64218	0.00531	0.05340	0.12050	0.06749	0.00086	0.01190	.
Puslinch	0.40012	0.77080	10.03082	0.07486	0.11553	0.59340	0.00521	0.07949	0.06964	0.11414	0.00000	0.03812	.
Quinte West	0.36389	0.16146	3.28457	0.01591	0.03662	0.66050	0.00249	0.04191	0.09453	0.07866	.	0.02107	0.00000
Ramara	0.36101	0.01545	0.08651	0.01367	0.07878	0.31593	0.00204	0.05739	0.49913	0.08820	0.00000	0.00403	0.00000
Richmond Hill	0.11319	0.00000	.	0.06020	0.00000	0.76316	.	0.03208	0.09084	0.05443	0.00000	0.00000	0.00000
Rideau Lakes	0.30492	0.00699	0.04532	0.00999	0.04710	0.36575	0.00144	0.05608	0.30645	0.06402	0.00000	0.01397	0.00000
Scugog	0.45569	0.19818	19.17842	0.03167	0.10157	0.68479	0.00188	0.05157	0.10441	0.12111	.	0.05302	0.00000
Severn	0.29011	0.10194	0.73663	0.02362	0.22756	0.53962	0.00762	0.08197	0.16745	0.07916	0.00000	0.00602	0.00000
Smith-Ennismore-Lakefield	0.45717	0.00609	3.60683	0.01536	0.03433	0.47527	0.00697	0.08281	0.23851	0.10299	0.00000	0.01804	0.00000
South Dundas	0.57098	0.21325	.	0.00425	0.04850	0.70196	0.00099	0.03891	0.05557	0.06334	0.01694	0.04153	0.00937
South Frontenac	0.30633	0.00415	0.06347	0.00719	0.05448	0.38861	0.00060	0.07270	0.28104	0.06828	0.00000	0.01455	0.00000
Springwater	0.45046	0.39689	4.52624	0.01332	0.07834	0.74488	0.00381	0.04053	0.05593	0.10537	0.00000	0.03741	0.00000
Stone Mills	0.37860	.	.	.	0.02931	0.37280	0.00305	0.07718	0.32585	0.08099	0.00000	0.01627	0.00000
Strathroy-Caradoc	0.20182	0.67299	7.34734	0.01093	0.02415	0.78632	0.00119	0.02989	0.03291	0.11435	.	0.03457	0.01633
Tay	0.50504	.	0.25787	0.02797	0.08554	0.50853	.	0.10220	0.13839	0.11762	0.00000	0.01789	0.00000
Thames Centre	0.42314	1.02517	6.27126	0.02203	0.03372	0.81117	0.00253	0.03958	0.03751	0.08345	0.00589	0.09061	.
The Nation Municipality	0.53855	0.08651	17.34404	0.00613	0.04025	0.81144	0.00211	0.04226	0.04945	0.06381	0.04161	0.09740	.
Thorold	0.17948	.	.	0.01135	0.00450	0.88370	.	0.01007	0.02078	0.06661	0.00000	0.00000	0.00000
Tiny	0.37423	.	.	0.02807	0.06222	0.63373	0.00332	0.04819	0.10676	0.08543	0.00000	0.02365	0.00000
Tyendinaga	0.31822	0.05786	0.02253	0.01149	0.01555	0.40495	0.00349	0.05803	0.31639	0.07364	0.00000	0.00343	0.00000
Wainfleet	0.19120	0.47086	35.44859	0.01277	0.05563	0.85567	0.00265	0.01733	0.02042	0.05971	0.00826	0.03843	.
Warwick	0.18877	2.85489	18.17307	0.00231	0.03409	0.85822	0.00291	0.01782	0.01521	0.05346	0.01831	0.11911	0.00995
Wellesley	1.71784	3.32185	16.70675	0.08433	0.07439	0.74988	0.00090	0.07050	0.05921	0.30364	0.02295	0.05569	.
Wellington North	0.69265	0.78988	22.46055	0.01931	0.09217	0.76991	0.00138	0.06026	0.03445	0.16813	0.00422	0.03825	.
West Grey	0.65441	0.24895	17.93144	0.02664	0.21322	0.56547	0.00411	0.11229	0.08347	0.17496	0.00000	0.02098	0.00125
West Lincoln	0.33812	0.81257	67.37567	0.01163	0.11833	0.77789	0.00751	0.04249	0.03137	0.13972	0.00954	0.03324	0.00762
West Perth	0.38492	3.62342	7.13853	0.00617	0.03374	0.87871	.	0.02538	0.01447	0.08229	0.01531	0.11787	0.01536
Whitestone	0.10017	.	0.07534	0.01221	0.16919	0.17593	0.00000	0.08249	0.21549	0.02146	0.00000	0.00000	0.00000
Whitewater Region	0.59882	0.03984	0.09387	0.00674	0.05010	0.54760	0.00136	0.10786	0.11093	0.11880	0.00000	0.02502	0.00000
Wilmot	0.60249	1.10437	38.26441	0.01216	0.05953	0.83156	0.00326	0.03020	0.02534	0.14467	0.02249	0.07830	0.00624
Woolwich	1.16360	1.40930	36.45214	0.03887	0.04512	0.82945	0.00679	0.02472	0.02539	0.21608	.	0.07317	0.00506
Zorra	0.52021	2.74014	16.88390	0.01216	0.06990	0.85819	0.00030	0.03060	0.01656	0.09168	0.02351	0.13619	0.01543

Tableau de données des variables agricoles (rapport population humaine/population animale) selon la SRU

Subdivision de recensement unifiée	BoHu	PoHu	EqHu	VoHu	RuHu
Adelaide Metcalfe	2.27215	17.32582	0.06478	144.91839	0.92442
Adjala-Tosorontio	0.44604	0.12904	0.05604	16.64888	0.13846
Alfred and Plantagenet	1.09228	1.33865	0.02502	17.26731	0.17305
Amaranth	1.92440	2.30318	0.10743	.	0.91088
Armour	0.06134	.	.	0.14210	0.00000
Asphodel-Norwood	1.41531	0.13049	0.05370	69.48908	0.97666
Athens	0.91942	0.00000	0.01114	0.04422	0.03505
Bayham	0.48424	0.01114	0.05867	0.31341	0.02620
Beckwith	0.23941	0.00139	0.00958	0.02254	0.04488
Belleville	0.08202	0.00161	0.00439	.	0.01016
Blandford-Blenheim	0.38382	1.01538	0.00688	4.52434	0.06060
Bluewater	0.86935	6.58983	0.01648	100.42246	0.09452
Bracebridge	0.02624	0.00547	0.00763	0.07549	0.00000
Bradford West Gwillimbury	0.10374	0.11760	0.00742	.	0.04958
Brant	0.14180	0.18812	0.01060	3.03171	0.02784
Brighton	0.45592	0.33623	0.00529	23.76156	0.06011
Brockton	3.75844	3.53520	0.05446	18.06730	0.48053
Burlington	0.02415	.	0.00178	.	0.00000
Caledon	0.34675	.	0.03753	.	0.03087
Campbellford/Seymour,Percy,Hastings	1.10820	0.49773	0.02785	1.07956	0.24632
Central Elgin	0.07316	0.13650	0.00809	1.22280	0.00599
Central Frontenac	0.62804	0.00856	0.01646	.	0.10665
Central Huron	0.72797	3.77008	0.01168	145.99870	0.31603
Centre Hastings	0.83176	.	0.01727	0.09158	0.00781
Centre Wellington	0.80515	1.70078	0.02824	17.71678	0.13566
Chatsworth	3.54809	0.96401	0.12006	9.37054	0.68264
Clarington	0.17477	0.14655	0.01220	.	0.04760
Douro-Dummer	1.33734	0.06058	0.05818	0.27511	0.14387
Dysart and Others	0.00839	.	0.00246	0.01248	0.00000
East Zorra-Tavistock	2.15750	9.20945	0.03426	90.25656	0.23363
Edwardsburgh/Cardinal	0.60773	0.57911	0.03746	.	0.06952
Elizabethtown-Kitley	0.25715	0.04730	0.00821	.	0.06780
Essa	0.33906	0.29159	0.02219	0.05236	0.04254
Fort Erie	0.02782	.	0.00636	4.03830	0.00025
Galway-Cavendish and Harvey	0.24616	.	0.01519	0.04251	0.02471
Georgian Bluffs	1.80191	0.10057	0.03349	0.15406	0.58678
Georgian Highlands	0.47356	0.05457	0.01116	0.94904	0.20172
Georgina	0.06854	0.13776	0.01164	3.20350	0.11428
Gravenhurst	0.02743	.	0.00752	0.10212	0.01853
Greater Napanee	0.62378	.	0.03258	.	0.13918
Grey Highlands	2.70313	0.70465	0.07819	2.61059	0.59319
Guelph/Eramosa	0.08003	0.06941	0.00608	0.54875	0.02280
Haldimand	0.57954	1.21663	0.02518	28.59980	0.17586
Halton Hills	0.10171	.	0.01669	.	0.01135
Hamilton	0.03683	0.03682	0.00476	4.05063	0.01137
Hamilton	0.15992	0.14751	0.01009	2.50199	0.03051
Hastings Highlands	0.08066	.	0.00852	0.31288	0.03006
Havelock-Belmont-Methuen	0.35968	.	0.02233	0.16432	0.11476
Huntsville	0.01586	0.00404	0.00836	0.09482	0.01575
Huron East	3.66291	26.45207	0.03388	73.26271	0.38295
Innisfil	0.14428	0.79436	0.01291	3.86699	0.10078
Kawartha Lakes	0.84524	0.17388	0.03254	3.10097	0.22955
Kincardine	1.86789	1.40412	0.04107	8.22695	0.97089
King	0.18157	0.04597	0.06934	.	0.12205
Kingston	0.06719	0.00065	0.00295	.	0.00878
Lambton Shores	0.45615	5.30309	0.00634	.	0.10113
Lanark Highlands	0.80792	.	0.02440	0.27800	0.40063
Leeds and the Thousand Islands	0.70237	0.10087	0.02065	.	0.17947
Lincoln	0.05574	0.25349	0.01111	29.66578	0.00374
Loyalist	0.30130	.	0.01693	0.07204	0.30802
Madoc	1.59638	0.12280	0.04110	0.24168	0.61986

Subdivision de recensement unifiée	BoHu	PoHu	EqHu	VoHu	RuHu
Malahide	0.91484	1.89244	0.03094	15.80038	0.13411
Mapleton	5.59959	12.38858	0.11125	209.83349	0.56014
Marmora	0.25069	.	0.02233	0.15609	0.06349
McKellar	0.03506	.	0.00190	0.19039	0.00250
Melancthon	1.08673	0.13038	0.02400	.	0.33160
Middlese Centre	0.93238	2.88920	0.06881	43.59381	0.15398
Milton	0.08633	.	0.04395	7.38159	0.03997
Mississippi Mills	0.78741	0.01425	0.03434	0.19421	0.09599
Mono	0.18688	.	0.01769	0.06522	0.02984
Montague	0.18437	.	0.01218	0.03224	0.04676
Muskoka Lakes	0.10307	0.00499	0.02020	0.06729	0.00000
Niagara-on-the-Lake	0.02681	.	0.00578	30.38948	0.02457
Norfolk	0.13159	0.47661	0.01379	21.56534	0.03777
North Dumfries	0.52286	.	0.02897	44.81435	0.08325
North Dundas	1.75504	.	0.03060	.	0.34220
North Frontenac	0.08329	0.00000	0.01777	.	0.00000
North Glengarry	1.57815	0.28275	0.02899	.	0.18557
North Huron	2.33949	2.52548	0.02488	.	0.21970
North Perth	3.07209	7.45641	0.04977	116.96765	0.15977
North Stormont	2.30999	.	0.02363	0.31510	0.16557
Northern Bruce Peninsula	2.94832	0.00000	0.02890	.	0.06502
Oro-Medonte	0.55430	0.31346	0.03658	0.15075	0.37172
Otonabee-South Monaghan	1.53374	0.09661	0.04981	.	0.69968
Ottawa	0.06346	0.01027	0.00359	0.27572	0.01003
Parry Sound, Unorganized, Centre Part	0.20130	.	0.00869	0.40080	0.02136
Pelham	0.07222	0.01146	0.01768	16.07615	0.00334
Port Colborne	0.02900	.	0.00423	14.10743	0.00000
Prince Edward	0.65214	0.04369	0.02361	10.04140	0.24437
Puslinch	0.58675	1.13033	0.10977	14.70960	0.16941
Quinte West	0.19773	0.08773	0.00865	1.78478	0.01990
Ramara	0.92325	0.03951	0.03495	0.22123	0.20148
Richmond Hill	0.00132	0.00000	0.00070	.	0.00000
Rideau Lakes	0.96284	0.02206	0.03155	0.14312	0.14873
Scugog	0.63533	0.27631	0.04416	26.73922	0.14161
Severn	0.09459	0.03324	0.00770	0.24019	0.07420
Smith-Ennismore-Lakefield	0.09387	0.00125	0.00315	0.74058	0.00705
South Dundas	1.65622	0.61857	0.01233	.	0.14068
South Frontenac	0.62071	0.00841	0.01456	0.12860	0.11039
Springwater	0.09489	0.08360	0.00281	0.95342	0.01650
Stone Mills	1.50361	.	.	.	0.11640
Strathroy-Caradoc	0.24830	0.82798	0.01345	9.03945	0.02972
Tay	0.07290	.	0.00404	0.03722	0.01235
Thames Centre	1.09966	2.66423	0.05724	16.29784	0.08763
The Nation Municipality	1.64616	0.26442	0.01873	53.01429	0.12303
Thorold	0.04643	.	0.00294	.	0.00116
Tiny	0.35455	.	0.02660	.	0.05895
Tyendinaga	1.00018	0.18185	0.03612	0.07080	0.04888
Wainfleet	0.49537	1.21988	0.03308	91.83893	0.14414
Warwick	1.23950	18.74584	0.01516	119.32820	0.22385
Wellesley	4.48308	8.66909	0.22007	43.59989	0.19413
Wellington North	2.31004	2.63432	0.06440	74.90756	0.30739
West Grey	1.74653	0.66443	0.07109	47.85701	0.56905
West Lincoln	0.74861	1.79907	0.02576	149.17330	0.26198
West Perth	2.41845	22.76613	0.03878	44.85179	0.21196
Whitestone	0.27902	.	0.03400	0.20985	0.47128
Whitewater Region	3.10552	0.20660	0.03497	0.48681	0.25982
Wilmot	0.88302	1.61859	0.01783	56.08133	0.08725
Woolwich	1.99648	2.41805	0.06670	62.54404	0.07741
Zorra	3.02322	15.92462	0.07067	98.12245	0.40623

Tableau des données des variables socio-économiques, météorologiques et pédologique selon la SRU

Subdivision de recensement unifiée	LogeA	Rev	Scol	DHu	PrTo	NMarAv05	FonN	Drain
Adelaide Metcalfe	38	53083	0.37472	0.09484	1027.4	0	16	Pauvre
Adjala-Tosorontio	29	66969	0.28971	0.27007	846.2	0	15	Bien
Alfred and Plantagenet	31	45689	0.38362	0.21560	974.1	12	15	Imparfait
Amaranth	31	79095	0.31834	0.14213	859	0	10	Bien
Armour	40.5	30244	0.47382	0.12412	945	45	45	Bien
Asphodel-Norwood	41	43245	0.38622	0.24155	948.1	1	31	Bien
Athens	38	52675	0.32035	0.23853	1109.2	0	15	Bien
Bayham	37	42206	0.52356	0.25939	1031.3	0	17	Bien
Beckwith	29	59458	0.31352	0.57296	875.2	0	5	Bien
Belleville	38	41175	0.34794	1.89427	1082.35	0	14.5	Pauvre
Blandford-Blenheim	38.5	54237	0.38348	0.95187	967.7	0	18	Bien
Bluewater	44	50562	0.37375	0.16488	1004.8	0	18	Bien
Bracebridge	29.5	45987	0.32679	0.11458	1028.2	10	27	.
Bradford West Gwillimbury	24	67136	0.35524	1.10287	809.4	0	10	Bien
Brant	38.5	51682	0.38609	1.07542	802.2	0	20	Pauvre
Brighton	31	44116	0.36628	0.42697	1006.8	0	9	Bien
Brockton	45	42516	0.39418	0.16944	1122	0	33	Bien
Burlington	27	68372	0.23900	8.09579	757.3	0	21	Imparfait
Caledon	23	84381	0.27858	0.72897	901.0167	2.3333	13.6667	Imparfait
Campbellford/Seymour,Percy,Hastings	41	38559	0.39932	0.23725	1002.6	0	21	Bien
Central Elgin	38.5	52528.5	0.34350	1.44100	1017.2	0	24	Bien
Central Frontenac	36	38193	0.43590	0.04120	1032.8	0	18	.
Central Huron	43.5	45147	0.35904	0.32981	1177.5	0	16	Pauvre
Centre Hastings	37	41747	0.39852	0.18105	1037.8	0	21	Bien
Centre Wellington	31	60783	0.31089	0.58367	958.175	0	21.25	Imparfait
Chatsworth	41	44318	0.39535	0.10467	1153.3	0	19	Bien
Clarington	24	68428	0.30623	1.14019	775.6	0	6	Bien
Douro-Dummer	35	48459	0.34807	0.14517	948.1	1	31	Bien
Dysart and Others	30.3333	35545.6667	0.42875	0.03446	1013.35	38	38	.
East Zorra-Tavistock	41	57729	0.39328	0.29180	941.2	0	18	Bien
Edwardsburgh/Cardinal	40	49312	0.36985	0.21146	923.4	0	13	Bien
Elizabethtown-Kitley	38.5	47709	0.32618	0.54102	1161.2	0	8	Bien
Essa	31	60015	0.34394	0.59923	745.95	0	16.5	Bien
Fort Erie	40	43893	0.37182	1.72726	1227.6	0	10	Pauvre
Galway-Cavendish and Harvey	29	36560	0.38948	0.03757	948.1	1	31	Très pauvre
Georgian Bluffs	33	53136	0.32467	0.17145	1073.6	0	14	Très pauvre
Georgian Highlands	40.5	40004.5	0.36063	0.51924	1009.7	0	13	Imparfait
Georgina	29	56149	0.36744	1.29652	811.05	0	11.5	Pauvre
Gravenhurst	35	38109	0.37411	0.19655	937.25	10	17	.
Greater Napanee	39	40978	0.37015	0.32408	1005.6	0	11	Pauvre
Grey Highlands	39	46239	0.37331	0.10310	1222.4	12	15	Bien
Guelp/Eramosa	30.5	63838.5	0.28389	3.09082	794.5	0	24	Imparfait
Haldimand	37	53319	0.36724	0.34137	920.4	0	13	Pauvre
Halton Hills	30	74921	0.29020	1.73822	722.5	0	13	Imparfait
Hamilton	37	47855	0.35516	4.37293	898.2667	0	20	Pauvre
Hamilton	32.5	53306	0.34284	1.00392	1022.7	0	6	Bien
Hastings Highlands	32	33632	0.41604	0.03885	937.9	15	33	.
Havelock-Belmont-Methuen	38	35122	0.45752	0.07633	929.4	1	31	Imparfait
Huntsville	30	46212	0.35108	0.22522	1098.1	45	45	.
Huron East	48	43936	0.43103	0.14445	1205.9	0	16	Pauvre
Innisfil	26	57975	0.35653	1.00745	841.5	0	10	Bien
Kawartha Lakes	34	46156	0.39843	0.22410	952.8	0	13	Bien
Kincardine	36	52911	0.32283	0.20252	965.6	0	33	Imparfait
King	33	82122	0.27294	0.55260	863.5	0	10	Bien
Kingston	33	46076	0.26694	2.52057	989.2	0	12	Bien
Lambton Shores	36	46019	0.37386	0.30746	1043.1	0	18	Pauvre
Lanark Highlands	35	42534	0.40102	0.04401	912.6	0	27	.
Leeds and the Thousand Islands	41	47059.5	0.35754	0.23649	1047.2	0	15	Imparfait
Lincoln	32	59837	0.30697	1.26308	940.3	0	14	Pauvre

Subdivision de recensement unifiée	LogeA	Rev	Scol	DHu	PrTo	NMarAv05	FonN	Drain
Loyalist	33	51870	0.29355	0.48635	1007.1	0	21	Très pauvre
Madoc	39	39906	0.47284	0.07299	1030.4	0	12	.
Malahide	38.5	48681	0.48131	0.39671	1083.9	0	17	Bien
Mapleton	40	60711	0.54747	0.17157	1044.1	0	16	Imparfait
Marmora	37	31692	0.49842	0.06895	1030.4	0	12	.
McKellar	29.1667	37829.5	0.39165	0.10890	859.75	15.5	15.5	.
Melancthon	35	50307.5	0.38446	0.21802	998.4	5	25	Bien
Middlese Centre	36	73252	0.26201	0.24214	1041.3	0	10	Bien
Milton	32	74279	0.27761	0.85496	702.2	0	13	Bien
Mississippi Mills	35	60499	0.29350	0.22120	875.2	0	5	.
Mono	27	70050	0.30088	1.09244	859	0	10	Bien
Montague	42	42488	0.40714	0.43986	912.6	0	27	Bien
Muskoka Lakes	27.3333	26351	0.36216	0.05486	1092.85	45	45	.
Niagara-on-the-Lake	35	56956	0.29607	1.02397	1038	0	14	Imparfait
Norfolk	40	46845	0.42521	0.37984	861.5	0	19	Pauvre
North Dumfries	31	75364	0.30240	0.46266	937.1	0	18	Bien
North Dundas	39	52070	0.33998	0.21729	919.8	2	11	Très pauvre
North Frontenac	35	29773	0.47284	0.01417	1079.5	22	22	Bien
North Glengarry	40	38397	0.41038	0.16212	912.8	2	12	Pauvre
North Huron	48	45487	0.38423	0.27672	1171.7	0	33	Pauvre
North Perth	42	50074	0.41471	0.24432	906.8	0	11	Pauvre
North Stormont	42	48004	0.36301	0.13203	986.7	2	18	Pauvre
Northern Bruce Peninsula	32	36467	0.36601	0.04539	1055.9	0	29	Bien
Oro-Medonte	29	60169	0.28752	0.31203	915.9999	6	22.5	Bien
Otonabee-South Monaghan	30.5	45715.5	0.32232	0.19484	995.5	0	9	Bien
Ottawa	30	62130	0.21698	2.76129	900.1083	0.5	17	Bien
Parry Sound, Unorganized, Centre Part	22.8	5274.8	0.59615	0.00901	1129.5	30	30	.
Pelham	32	65220	0.24285	1.20549	1037.2	0	16	Pauvre
Port Colborne	45	40770	0.41730	1.50472	1028	0	20	Pauvre
Prince Edward	39	44516	0.35179	0.22559	1128.75	0	11	Bien
Puslinch	32	76996	0.27396	0.27139	723.8	0	18	Bien
Quinte West	35	46696	0.36597	0.82790	1043.1	0	9	Bien
Ramara	22	44071	0.38395	0.20986	682.7	0	32	Imparfait
Richmond Hill	20	77637	0.23980	12.53644	865.05	0	17.5	Bien
Rideau Lakes	40.5	41670.5	0.31268	0.12787	1103.2	0	15	Bien
Scugog	23.5	32445.5	0.34626	0.42059	982	4	21	Pauvre
Severn	34	45122.5	0.37201	0.69593	1070.1	0	32	.
Smith-Ennismore-Lakefield	31.6667	40834	0.33392	2.30606	1202.2	0	11	Bien
South Dundas	41	42712	0.41458	0.20584	1067.2	0	13	Pauvre
South Frontenac	31	55319	0.30752	0.16753	1007.1	0	21	.
Springwater	23.6667	54388.6667	0.31176	1.96344	896.5	0	12	Bien
Stone Mills	39	48771	0.29668	0.09888	1005.6	0	21	Bien
Strathroy-Caradoc	32	51795	0.38823	0.55156	1027.4	0	16	Imparfait
Tay	35	43170.3333	0.39502	1.80204	872.1	0	12	Bien
Thames Centre	34	66675	0.31311	0.28701	1087.5	0	24	Bien
The Nation Municipality	29	53110	0.35142	0.20129	1003.4	2	18	Imparfait
Thorold	37	49776	0.35436	2.10178	925.3	0	16	Pauvre
Tiny	22.5	34258	0.35242	0.24216	1127.3	0	12	Bien
Tyendinaga	42	41265	0.42417	0.14302	1037.8	0	21	Bien
Wainfleet	40	51949	0.34234	0.28693	1028	0	24	Pauvre
Warwick	44	50447	0.42645	0.13855	1027.4	0	16	Imparfait
Wellesley	37	63323	0.55058	0.33621	1004.5	0	24	Bien
Wellington North	38	43344	0.47261	0.21469	886	0	11	Imparfait
West Grey	40.5	42153	0.41987	0.20811	1150.1167	5	14.5	Bien
West Lincoln	34	59322	0.34004	0.31614	1060.8	0	14	Pauvre
West Perth	44	55257	0.40449	0.15750	1068.2	0	17	Bien
Whitestone	37	39402	0.38411	0.00826	1049.1	30	30	.
Whitewater Region	37	42850	0.34513	0.12057	774.9	0	32	Imparfait
Wilmot	33	61364	0.34131	0.55884	937.1	0	18	Bien
Woolwich	37	63020	0.42403	0.55800	775	0	18	Imparfait
Zorra	44	57574	0.33991	0.15116	1087.5	0	24	Bien