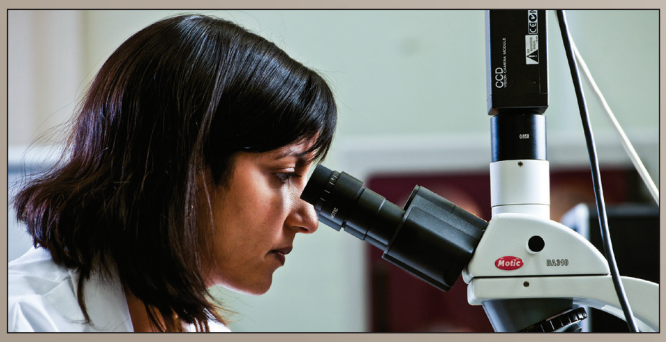
A SYSTEMATIC INVESTIGATION OF MANITOBA'S PROVINCIAL LABORATORY DATA

DECEMBER 2012

MANITOBA CENTRE FOR HEALTH POLICY
DEPARTMENT OF COMMUNITY HEALTH SCIENCES
FACULTY OF MEDICINE, UNIVERSITY OF MANITOBA



Authors: Lisa M. Lix, PhD P.Stat.
Mark Smith, MSc
Mahmoud Azimaee, BSc
Matthew Dahl, BSc
Patrick Nicol, BComm (Hons)
Charles Burchill, MSc
Elaine Burland, PhD
Chun Yan Goh, MSc
Jennifer Schultz, MA
Angela Bailly, MA

University of Manitoba | Manitoba Centre for Health Policy

This report is produced and published by the Manitoba Centre for Health Policy (MCHP).
It is also available in PDF format on our website at:
http://mchp-appserv.cpe.umanitoba.ca/deliverablesList.html

Information concerning this report or any other report produced by MCHP can be obtained by contacting:

Manitoba Centre for Health Policy
Dept. of Community Health Sciences
Faculty of Medicine, University of Manitoba
4th Floor, Room 408
727 McDermot Avenue
Winnipeg, Manitoba, Canada
R3E 3P5

Email: reports@cpe.umanitoba.ca
Phone: (204) 789-3819
Fax: (204) 789-3910

How to cite this report:
Lix LM, Smith M, Azimaee M, Dahl M, Nicol P, Burchill C, Burland E, Goh CY, Schultz J, Bailly A. A Systematic Investigation of Manitoba's Provincial Laboratory Data. Winnipeg, MB. Manitoba Centre for Health Policy, December 2012.

Legal Deposit:
Manitoba Legislative Library
National Library of Canada

ISBN 978-1-896489-68-1

1st printing (December 2012)

This work was supported through funding provided by the Department of Health of the Province of Manitoba to the University of Manitoba (HIPC#2010/2011-12). The results and conclusions are those of the authors and no official endorsement by Manitoba Health was intended or should be inferred. Data used in this study are from the Population Health Research Data Repository housed at the Manitoba Centre for Health Policy, University of Manitoba and were derived from data provided Manitoba Health.

ABOUT THE MANITOBA CENTRE FOR HEALTH POLICY

The Manitoba Centre for Health Policy (MCHP) is located within the Department of Community Health Sciences, Faculty of Medicine, University of Manitoba. The mission of MCHP is to provide accurate and timely information to healthcare decision–makers, analysts and providers, so they can offer services which are effective and efficient in maintaining and improving the health of Manitobans. Our researchers rely upon the unique Population Health Research Data Repository (Repository) to describe and explain patterns of care and profiles of illness, and to explore other factors that influence health, including income, education, employment, and social status. This Repository is unique in terms of its comprehensiveness, degree of integration, and orientation around an anonymized population registry.

Members of MCHP consult extensively with government officials, healthcare administrators, and clinicians to develop a research agenda that is topical and relevant. This strength, along with its rigorous academic standards, enables MCHP to contribute to the health policy process. MCHP undertakes several major research projects, such as this one, every year under contract to Manitoba Health. In addition, our researchers secure external funding by competing for research grants. We are widely published and internationally recognized. Further, our researchers collaborate with a number of highly respected scientists from Canada, the United States, Europe, and Australia.

We thank the University of Manitoba, Faculty of Medicine, Health Research Ethics Board for their review of this project. MCHP complies with all legislative acts and regulations governing the protection and use of sensitive information. We implement strict policies and procedures to protect the privacy and security of anonymized data used to produce this report and we keep the provincial Health Information Privacy Committee informed of all work undertaken for Manitoba Health.

Manitoba Centre for Health Policy | University of Manitoba

ACKNOWLEDGEMENTS

The authors wish to express appreciation for the contribution of the individuals who assisted in the development of this report:

The Advisory Group for their advice and insight into the process for the project: Laurel Biluk (Continuous Improvement Manager, Interlake Regional Health Authority), Dr. Bernie Chodirker (Professor, Pediatrics & Child Health and Biochemistry & Medical Genetics, University of Manitoba), Dr. Brenda Hemmelgarn (Associate Professor, Departments of Medicine and Community Health Sciences, University of Calgary), Catherine Hynes (Regional Manager, Decision Support, NOR–MAN Regional Health Authority), Julene Reimer (Manager, MIS, Data Management and Development, Manitoba Health), and Dr. Paul Van Caeseele (Medical Director, Cadham Provincial Laboratory). Thanks to the staff at the Cadham Provincial Laboratory for their assistance with various queries about the data.

We wish to extend a sincere thanks to our external reviewers Dr. Laurel Thorlacius (Director, Clinical Biochemistry and Genetics, Diagnostic Services of Manitoba) and Karey Iron (Director, Data Partnerships and Development, ICES) and to our senior reader Dr. Greg Finlayson (Research Scientist, MCHP, University of Manitoba).

We send our appreciation to many who have provided valuable feedback about specific components of this report: To Beth Ann Davies for her feedback on the coding of sexually transmitted infections. To Deborah Malazdrewicz (Executive Director, Health Information, Manitoba Health) for her feedback on the data management process. To Debbie Nowicki and Myrna Dyck for their feedback on the analysis plan. To Colette Raymond for her consultations related to identification of infectious disease populations. To Carol Schoenfeld for information on the management of personal health information numbers.

A special thanks to Dr. Alan Katz for his guidance and input throughout the life of the project.

We acknowledge the financial support of Manitoba Health for this report. The results and conclusions are those of the authors and no official endorsement by the Manitoba Government is intended or should be inferred. This report was prepared at the request of Manitoba Health, as part of the contract between the University of Manitoba and Manitoba Health.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

CIHI – Canadian Institute for Health Information

CMV – Cytomegalovirus

CPL – Cadham Provincial Laboratory

DA – Dissemination Area

DSM - Diagnostic Services of Manitoba

EA – Enumeration Area

GP – General Practitioner

HIPC – Health Information Privacy Committee

HIV – Human Immunodeficiency Virus

HSV – Herpes Simplex Virus

ICD – International Classification of Diseases

ICD–9–CM – International Classification of Disease, 9th Revision, Clinical Modification

ICD–10–ICA – International Classification of Disease, 10th Revision, Canadian Version

LIS – Laboratory Information System

LTC – Long–Term Care

MCHP – Manitoba Centre for Health Policy

PHAC – Public Health Agency of Canada

PHIN – Personal Health Identification Number

RHA – Regional Health Authority

SPDS – Scalable Performance Data Server

STI – Sexually Transmitted Infection

TB – Tuberculosis

VIMO – Valid, Invalid, Missing, Outlier

VZV – Varicella Zoster Virus

EXECUTIVE SUMMARY

Background

This study examines Cadham Provincial Laboratory (CPL) archived data from 1992 to 2010, which were recently acquired into the Population Health Research Data Repository (Repository) at Manitoba Centre for Health Policy (MCHP). These data contain population–based screening and test requisitions and results for notifiable diseases from clinical microbiology, virology, parasitology, and serology departments within CPL. Notifiable disease data have traditionally been used for infectious disease surveillance, disease control, and outbreak detection. However, when these data are linked with other administrative health databases, they can also be used for a variety of studies about the health and health service use of infectious disease populations.

Given the large quantity of CPL records acquired into the Repository, it was timely for MCHP data acquisition and management staff to formalize their data management and evaluation processes. Specifically, the MCHP Executive was interested in a framework for acquiring new administrative health databases and conducting evaluations of their quality.

Objectives

The study objectives were:

- Develop a data management template to apply to new datasets acquired into the Repository, including the CPL data
- Develop a data quality framework for evaluating administrative databases in the Repository
- Investigate key components of CPL data quality, including accuracy, consistency, and completeness
- Investigate the role of the CPL data for identifying infectious disease populations in Manitoba

Methods

A six–step data management template was created with input from a variety of individuals, including MCHP researchers and data analysts, representatives from Manitoba Health and the province's regional health authorities. The MCHP Data Quality Framework, which encompasses both database–specific and project–specific quality, was constructed following a scoping review of provincial, national, and international quality evaluation frameworks for secondary data sources and with consideration for data privacy legislation in Manitoba.

The accuracy and consistency of the CPL data were evaluated using descriptive statistics, including percentages of valid, invalid, and missing observations, and graphical analyses of temporal consistency in screenings and tests for clinical microbiology, virology, parasitology, and serology sections. An assessment of the completeness of coverage for the entire Manitoba population and the prenatal population was conducted by such variables as age, sex, and residence location. Case studies, in which the CPL data were linked with hospital abstracts and physician billing claims, were conducted to explore methods to identify HIV, tuberculosis (TB), and sexually transmitted infection (STI) populations.

Key Findings

The six–step data management process that arose from this research emphasizes the iterative nature of the data acquisition process. The MCHP Data Quality Framework encompassess constructs of accuracy, validity, timeliness, and interpretability, which are also found in data quality frameworks produced by other organizations, but the MCHP framework emphasizes the role of data privacy legislation in conducting database–specific quality and project–specific quality evaluations . Macros were developed to routinely produce a Data Quality Report. The data management process and data quality reporting mechanisms can be generalized to other administrative databases acquired into the Repository, such as health, education, and justice databases.

The quality evaluation of the CPL data revealed that there were few invalid or missing observations in the data fields. Temporal consistency analyses revealed some substantial variations over time, although an overall increasing trend was noted for most sections. Assessment of the completeness of coverage of the Manitoba population demonstrated the potential for incomplete coverage in some years for southwestern Manitoba regional health authorities.

CPL data for identifing individuals with HIV tests only contain linkable Personal Health Information Numbers (PHINs) beginning in 2006/07 fiscal year, which limits opportunities to study the health and healthcare use of HIV–infected individuals. Identification of TB and STI populations that rely solely on the CPL data will also result in an incomplete picture of the total number of cases in the population.

Conclusions

Notifiable disease data have many potential uses beyond surveillance and outbreak detection. When the data are anonymously linked with other administrative health databases, they can be used to construct population–based cohorts for investigations of health outcomes and health services utilization. Linkage with health databases that contain diagnostic information can also be used to produce comprehensive population estimates of communicable disease prevalence and incidence. As well, the data can be used to evaluate the effectiveness of population–based disease prevention or management programs by investigating changes over time in testing rates for different population groups. Comparisons of differences in testing rates between geographic areas or income groups can be used to assess disparities in the utilization of public health services.

However, a systematic process for database management and quality evaluation is essential to ensure that the data can be fully utilized for population health and health services research. In particular, acquisition of historical documentation about the contents and organization of the data are critical to ensure that study results can be correctly interpreted.

The following recommendations arise from this study:

1. Link notifiable disease data to other administrative databases to explore the full potential of the CPL data for population health and health services research.
2. Add other sources of disease tests to the Repository to improve the comprehensiveness of the Repository for the investigation of notifiable diseases.
3. Apply the Data Quality Framework to all administrative databases in the Repository. Explore the use of case studies to promote best practices in data quality evaluation.
4. Develop a framework and tools to evaluate the quality of administrative database documentation.
5. Conduct studies about the validity of disease cases ascertained from notifiable disease and diagnostic information in administrative data.

CHAPTER 1: INTRODUCTION

Background

The **Manitoba Centre for Health Policy (MCHP)**[1] recently acquired archived data from the Cadham Provincial Laboratory (CPL) mainframe system into the **Population Health Research Data Repository** (Repository) housed at MCHP. CPL provides public health laboratory services in Manitoba as well as to other provinces; it maintains testing data about **notifiable diseases**. CPL services encompass the areas of microbiology, virology, parasitology, serology, and **newborn screening and public health chemistry**.

The acquisition of the CPL archived data provided an opportunity for MCHP to expand its capability for policy–relevant population health and health services research. In other jurisdictions, notifiable disease data from public health laboratories have been used to conduct studies about quality and accessibility of care, investigate at–risk populations, develop methods and tools to predict patient outcomes, and evaluate diagnostic information contained in hospital and physician administrative data (Emons, 2001).

The acquisition of such a large amount of **administrative health data** also presented a timely opportunity for MCHP to formalize its procedures and processes for acquiring data into the Repository and evaluating their quality. This involves identifying the steps and key participants in the **data management process** and developing techniques to routinely evaluate **data quality**. Data quality evaluation is important because many administrative databases were not originally collected for the purpose of conducting research. Instead, these data are often collected for health system management and provider remuneration. Assessment of data quality includes the evaluation of such characteristics as completeness, **accuracy**, **validity**, and **timeliness**.

Public Health Laboratories in Canada and Manitoba

Manitoba, like other Canadian provinces, is home to both public and private laboratory service providers. Diagnostic Services of Manitoba (DSM), a not–for–profit corporation established in 2002, is responsible for all of Manitoba's public laboratory services. DSM oversees services in 77 laboratory facilities located in hospitals, health centres, and clinics.

The CPL, which is operated by **Manitoba Health**, is the province's only public health laboratory. The CPL is a member of the Canadian Public Health Laboratory Network, and is responsible for the provision of a core set of services related to preventing, detecting, and monitoring human disease and providing related education to healthcare professionals and the public. The focus of the Network is on testing for notifiable diseases, which are required by law to be reported by government authorities. Table 1.1, which contains information about notifiable diseases in Canada, provides an indication of the range of diseases for which Canadian public health laboratories, like CPL offer testing services (Public Health Agency of Canada, 2005).

1 Terms in **bold** typeface are defined in the glossary located at the end of the report.

Table 1.1: Nationally Notifiable Diseases and Year(s) Positive Reports were Recorded in Canada

Disease	First Positive Case (Year)
Acute Flaccid Paralysis	2000 -
AIDS	1986 -
Amoebiasis	1927 - 1999
Anthrax	2002 -
Botulism	1933, 1940 -
Brucellosis	1928 -
Campylobacteriosis	1986 -
Chancroid	1979 - 1999
Chickenpox	1924 to 1959, 1986 -
Chlamydia, Genital	1990 -
Cholera	1974 -
Creutzfeldt-Jakob Disease	2000 -
Cryptosporidiosis	2000 -
Cyclosporiasis	2000 -
Diphtheria	1924 -
Giardiasis	1983 -
Gonorrhea	1924 -
Gonococcal Ophthalmia Neonatorum	1979 - 1999
Group B Streptococcal Disease of the Newborn	2000 -
Hantavirus Pulmonary Syndrome	2000 -
Hepatitis A	1927 to 1958, 1969 -
Hepatitis B	1969 -
Hepatitis C	1991 -
Hepatitis Non-A, Non-B	1983 - 1999
Human Immonodeficiency Virus	2000 -
Influenza,Laboratory-Confirmed	2000 -
Invasive *Haemophilus influenzae* type b Disease	1979 -
Invasive Group A Streptococcal Disease	2000 -
Invasive Meningococcal Disease	1924 -
Invasive Pneumococcal Disease	2000 -
Legionellosis	1986 -
Leprosy	1925 -
Listeriosis (all types)	1990 - 1999
Malaria	1929 to 1978, 1983 -
Measles	1924 -
Meningitis, Pneumococcal	1979 - 1999
Meningitis, Other Bacterial	1979 - 1999
Meningitis, Viral	1952 - 1999
Mumps	1924 to 1959, 1986 -
Paratyphoid	1924 to 1952, 1969 - 1999
Pertussis	1924 -
Plague*	
Poliomyelitis	1924 -
Rabies	1927 -
Rubella	1924 -
Rubella, Congenital	1979 -
Salmonellosis	1958 -
Shigellosis	1924 -
Smallpox	2002 -
Syphilis, All	1924 - 1978

Disease	First Positive Case (Year)
Syphilis, Congenital	1992 -
Syphilis, Early Latent	1992 -
Syphilis, Early Symptomatic (Primary and Secondary)	1979 -
Syphilis, Other	1924 -
Tetanus	1957 -
Tuberculosis	1924 -
Tularemia	2002 -
Trichinosis	1929 - 1999
Typhoid	1924 to 1952, 1969 -
Verotoxigenic *E. coli*	1990 -
Viral Hemorrhagic Fevers (Crimean Congo, Ebola, Lassa, Margurg)	2002 -
West Nile Virus Asymptomatic Infection	2003 -
West Nile Virus Fever	2003 -
West Nile Virus Neurological Syndromes	2003 -
West Nile Virus Unclassified/Unspecified	2003 -
Yellow Fever*	

* The notifiable disease database has never received a report of plague or yellow fever.
PHAC website: http://dsol-smed.phac-aspc.gc.ca/dsol-smed/ndis/list-eng.php.

History and Functions of the Cadham Provincial Laboratory

The CPL has served as Manitoba's public health laboratory for more than 100 years. It was established in 1897 as the Provincial Board of Health Laboratory under the direction of the Provincial Bacteriologist, Dr. Graham Bell. The Board of Health Laboratory mainly provided services for the identification of patients who were carriers of **infectious disease**. It also provided water and milk testing.

The CPL has several sections, including clinical microbiology, serology and parasitology, virus detection, and newborn screening and public health chemistry. The CPL identifies and monitors virus outbreaks, including outbreaks of notifiable respiratory and enteric diseases (Appendix A contains a listing of notifiable diseases in Manitoba). The CPL staff conduct research and develop new techniques, such as nucleic acid detection and characterization and antigen detection for timely detection of microbial pathogens—typing pathogens to establish source and the etiology of outbreaks. As well, the CPL provides technical support, information management, and administration services.

Within the CPL, the **clinical microbiology section** is responsible for investigating human bacterial enteric pathogens and food–borne illness (e.g., salmonella), as well as **sexually transmitted infections (STIs)** such as **chlamydia** and **gonorrhea**. Screening for mothers and babies at risk for Group B hemolytic streptococcal colonization and infections is also provided. Screening protocols for detection of antibiotic resistance include methicillin–resistant staphylococcus aureus (MRSA) and vancomycin–resistant enterococci. Testing for respiratory organisms such as microbacterium tuberculosis, pertussis, Legionella species, diphtheria, and streptococcus is also conducted. Staff are involved in both identifying and typing organisms.

The serology and **parasitology section** provides surveillance and diagnosis for public health programs, such as prenatal screening for **syphilis**, **hepatitis B**, and rubella. This section monitors Western Equine Encephalitis and West Nile infections (since 2003). **Serodiagnostic testing** is done for viral and bacterial diseases, and parasitology testing is also conducted. **Human immunodeficiency virus (HIV)** serology was established in 1990; **hepatitis C** serology was established in 1991.

The virology section of the University of Manitoba joined the CPL in 1963. This section supplies viral strain information and primary isolates to the **Public Health Agency of Canada (PHAC)** for national viral surveillance. Viral culture testing, including influenza, other respiratory viruses, Herpes Simplex Virus (HSV), Varicella Zoster Virus (VZV), Cytomegalovirus (CMV), and enteroviruses was established in the 1980s. Molecular detection was introduced in the late 1990s. CPL monitors viral isolates to ensure vaccine–preventable diseases are not spreading and detects viral infections of immune–compromised persons (i.e., transplant recipients).

During the 1960s, newborn screening for inherited metabolic diseases was added as a laboratory service. This includes phenylketonuria (PKU), galactosemia, congenital hypothyroidism, biotinidase deficiency, congenital adrenal hyperplasia, and screening of male infants for Duchenne Muscular Dystrophy. In 1988, **maternal serum screening program** services were incorporated into the laboratory, including screening of specimens of Alpha–Fetoprotein for prenatal detection of neural open–tube defects and testing for Down Syndrome.

The **Cadham Provincial Laboratory (CPL) database** was installed in 1982; it initially captured data on virus detection and clinical microbiology **requisitions** and tests. Serology and parasitology requisitions and tests were added in 1992. Newborn screening, maternal serum screening, and selected other chemistry tests were never included in the mainframe database, but were instead managed via multiple other databases. Before moving from the mainframe database to a new Laboratory Information Management (LIM) System in 2010, more than 1,000 records were entered daily into the mainframe database. Approximately 4.5 million records were added to the mainframe database between 1992 and 2010 and about 13.5 million test results were accumulated. These are the archived data that were transferred to MCHP in the fall of 2010 and that were analysed for this study. At present, it is not known if the data contained in the LIM will be incorporated into the Repository on an on–going basis.

Purpose and Objectives

The study purpose was to conduct a systematic investigation of the CPL archived data. The objectives were:

1. Develop a data management template to apply to new datasets acquired into the Repository, including the CPL data
2. Develop a **data quality framework** for evaluating administrative databases in the Repository
3. Investigate key components of CPL data quality, including accuracy, consistency, and completeness
4. Investigate the role of the CPL data for identifying infectious disease populations in Manitoba

Report Organization

This report is divided into two sections. The first section, which is found in Chapters 2 and 3, focuses on frameworks for data management and data quality evaluation. The data management framework was developed using the process of acquiring the CPL data into the Repository as a guide. The data quality framework was developed after a review of existing data quality frameworks and literature, with a particular emphasis on their relevance to administrative databases.

The second section, which encompasses Chapters 4 through 6, examines the characteristics of the CPL data. In Chapters 4 and 5, various aspects of the quality of the CPL data are described using the data quality framework presented in Chapter 3. Chapter 4 primarily investigates quality of unlinked CPL data, while Chater 5 focuses on the quality of the data after linkage to other administrative databases in the Repository. In Chapter 6, the CPL data are investigated for identifying cases of HIV, **tuberculosis (TB)**, and sexually transmitted infections (STIs) via linkage to other administrative data sources.

Chapter 7 concludes the report with a summary of the study's key findings and a discussion of potential opportunities to use these data for population health and health services research. Recommendations arising from this study, which include recommendations about further enhancements of the data quality framework, are provided.

CHAPTER 2: THE DATA MANAGEMENT PROCESS AT MCHP

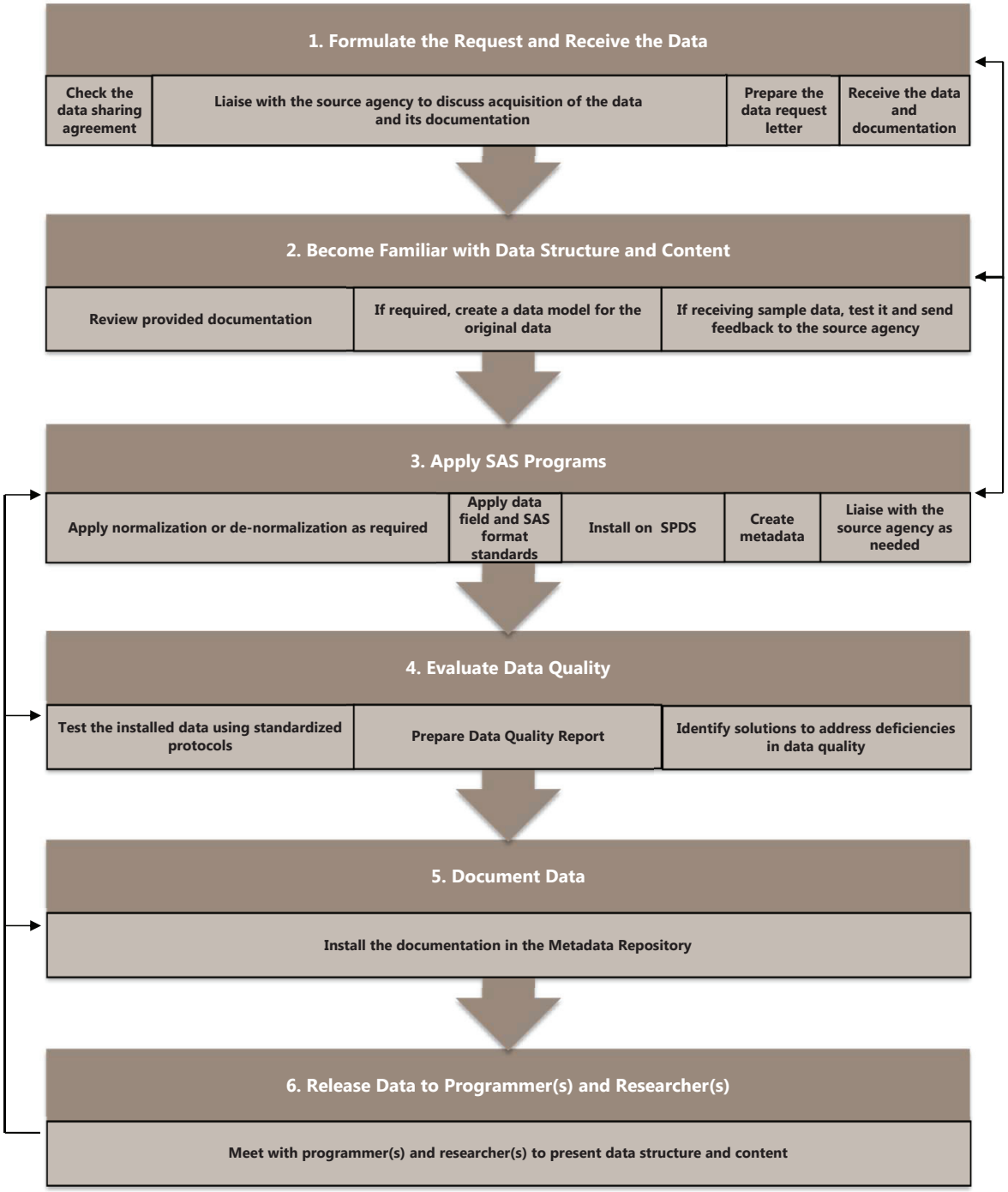
This chapter focuses on the process used to acquire the CPL data into the Repository housed at MCHP, which is then generalized to encompass any of the databases acquired into the Repository. A six–step process for acquiring, documenting, evaluating, and installing data into the Repository was formalized during the process of acquiring the CPL data (Figure 2.1). In the following sections, the steps and terminology in the data management process are described. The individuals responsible for executing each of these steps are identified. As Figure 2.1 illustrates, the data management process is iterative; individuals may return to previous steps in the process as needed in order to ensure that the data acquired into the Repository are complete and well documented for research purposes.

Step 1: Formulate the Request and Receive the Data

A data sharing agreement must be in place before any data can be received from the source agency. MCHP works in consultation with the University of Manitoba's Office of Legal Counsel and the source agency to develop an agreement. The data sharing agreement defines policies and practices about data confidentiality, privacy, legislative and regulatory requirements, data transfer, and ongoing use of the data for research purposes. Data sharing agreements are of two types: (a) agreements for data added to the Repository at regular intervals (typically annually) and (b) agreements for data provided for a single research project. For data added to the Repository at regular intervals, MCHP assumes responsibility for overseeing the use of the data. This involves ensuring that appropriate policies and procedures governing use are established, documented, and enforced.

Once a data sharing agreement is established, a data management analyst is assigned to work with the source agency to facilitate the transfer. Initially this involves meeting with representatives from the source agency to acquire background information, documentation, data model diagrams, data dictionaries, documentation about historical changes in the data (including changes in program scope, content, structure, and format), existing data quality reports, and other information relevant to the description or use of the data. This background information is used to: (a) develop a formal data request; (b) enhance the MCHP **Metadata Repository**, which contains database documentation; and (c) prepare the Data Quality Report. The data management analyst will also ask the source agency for reports or publications (e.g., annual reports) that document the number of entities in the data, such as people, places, events, or activities. This information is used to assess the accuracy and validity of the data files that are installed in the Repository. Financial data, such as annual budgets and total expenditures for specific programs, are also requested if they are available.

Figure 2.1: The Six–Step Data Management Process



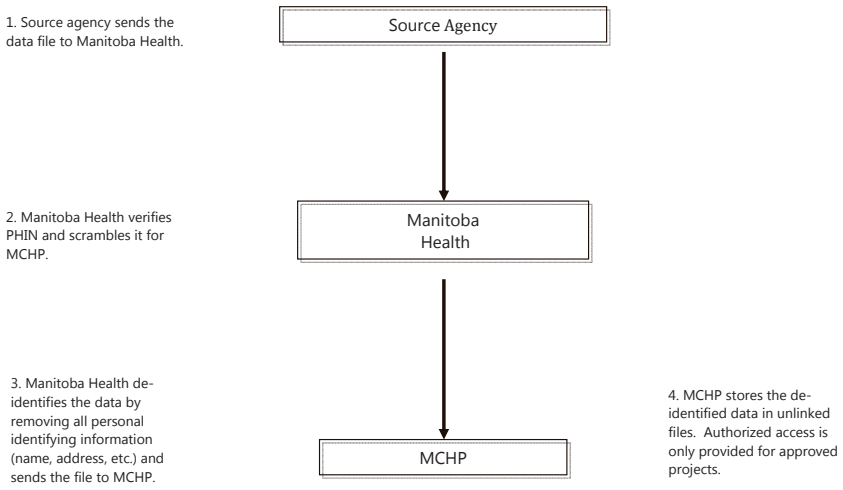
The initial data request encompasses historical documentation. This is information that may have gone through multiple revisions over time, particularly in response to health system changes. The initial data request may in fact be a series of requests, one for each generation of the source data that has existed. Future requests for updates to the data may refer to the most recent generation only. All of the changes in coding methods, program constraints, and accounting measures should be documented and incorporated into the MCHP Metadata Repository.

A sample data file is often prepared by the source agency and transferred to MCHP at the same time as the initial transfer of documentation. Ideally, the sample data file consists of a random sample of the original data. If the sample data file is sent directly to MCHP, it must be anonymized, which involves the source agency removing the **personal health information number (PHIN)** and other identifiable information such as name, address, and telephone number from each record. Any data elements containing comments that might identify individuals must also be removed.

Once the documentation and sample data file have been evaluated, a formal data request is prepared and sent to the source agency. At this time, the amount of programmer/analyst time and the cost associated with extracting and preparing the data for shipping to MCHP or Manitoba Health are identified.

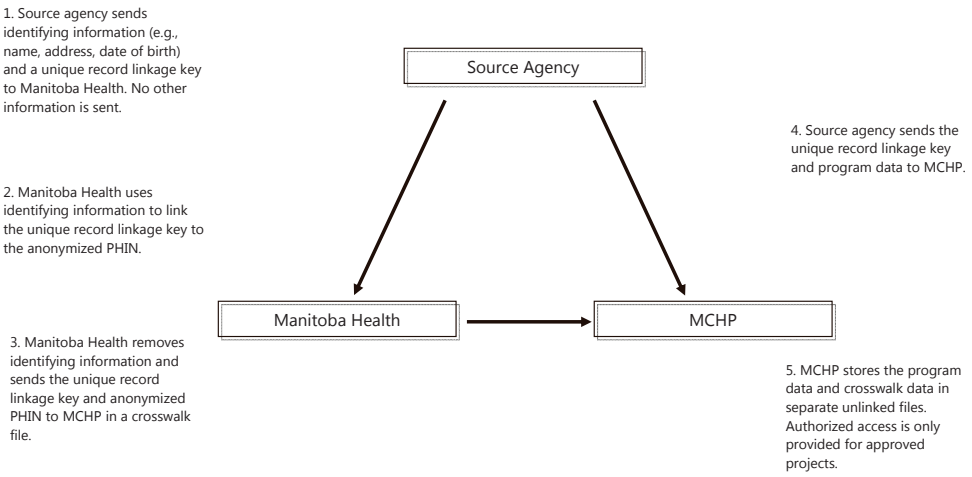
Before the data file(s) can be shipped to MCHP, a de–identification and linkage process is undertaken to strip the data of any identifying information and prepare it for linkage to the other files in the Repository. If the source agency is a health–related agency such as Healthy Child Manitoba or a **Regional Health Authority (RHA)**, then the data are sent to Manitoba Health and subjected to the process depicted in Figure 2.2 prior to being sent to MCHP. Manitoba Health uses identifying information such as name, address, date of birth, sex, and **postal code** to confirm the validity of the PHIN that is provided on a record or to provide a PHIN if it is missing from the data. Identifiable information in the data file is linked to the same information in the **Provincial Health Insurance Registry** using probabilistic linkage techniques (Blakely & Salmond, 2002) to impute (i.e., fill in) a missing PHIN. Once a PHIN is verified or imputed for each data record, identifiable information is stripped from the record. The PHIN is then anonymized (i.e., scrambled) to protect the confidentiality of individuals. The anonymized data files are sent to MCHP. The same anonymization algorithm is used for all data files contained in the Repository to facilitate record linkage.

Figure 2.2: The De–Identification Process for Health Data



If the source agency does not have health–related data then an alternate de–identification process is used. Some examples of source agencies for which this alternate de–identification process is used include the provincial departments of Housing and Community Development, Justice, Education, and Family Services. The data are split by the source agency into two parts. The first part contains program–specific data and a unique (i.e., person–specific) record linkage key, but does not contain identifying information. This record linkage key should not be an identifier normally accessible to end users of the source agency, and it must be unique to an individual over time. The second part contains identifying information, the same unique record linkage key, but no program–specific information. The first part is sent to MCHP, while the second part is sent to Manitoba Health where probabilistic linkage with the Provincial Health Insurance Registry is used to impute a PHIN. This probabilistic linkage is based on name, address, sex, and birth date. Manitoba Health produces a **crosswalk** file that contains the PHIN and unique record linkage key. The PHIN is anonymized using the algorithm described previously. The crosswalk file is sent to MCHP, where it is only used for approved research projects.

Figure 2.3: The De–Identification and Linkage Process for Health–Related Data



Step 2: Become Familiar with the Data Structure and Content

Once the data are received at MCHP, a data management analyst will review the data documentation and the organization of the files and structures. While data in the Repository are usually organized to reflect the structure of the original source data, sometimes the files must be reorganized to permit researchers to address questions about the different units of analysis that comprise the data, including persons, places, objects, events, and event dates.

Tasks that the data management analyst undertakes in the process of becoming familiar with the data structure and content include:

1. Standardizing unique record identifiers and the anonymized PHIN. If the PHIN is missing, then a value is imputed by MCHP analysts
2. Standardizing dates of events and correcting incomplete dates, where possible
3. Standardizing frequently used demographic data elements, including sex and postal code
4. Identifying and restricting access to data elements not normally made available to researchers without special permission. Examples include registration numbers and hospital chart numbers
5. Re–organizing and converting files to a different file format, if necessary

Step 3: Apply SAS Programs

MCHP uses **SAS®** for analysis, which performs optimally with data files that have been denormalized (SAS Institute Inc., 2006). Denormalization is a process of removing redundant information from a data file in order to reduce the processing time required to read the data. Standardized formats will be applied to selected fields, such as date fields. SAS can also be used to create a summary of the contents of the data files.

Once a datafile has been prepared for research use, it is installed in the SAS Scalable Performance Data Server (SPDS) using sort order indices and other design elements appropriate for the most commonly used research applications. During this process, standard naming conventions for data files are applied.

Step 4: Evaluate Data Quality

A Data Quality Report can be produced for each dataset in the Repository. This report is housed in the Metadata Repository, which provides a single point of access for all documentation concerning a data file. The structure and contents of the Report, and the framework that guided the development of the report, are described in Chapter 3.

Step 5: Document the Data

Data dictionaries, which contain information about the name, contents, and format of each field, are created and stored in the Metadata Repository. The data dictionaries can be used to conduct an initial review of data quality; a cursory review of the data dictionaries can reveal problems with missing data, completeness of labels and descriptors for categorical data values, ranges in numeric values, and/or integrity of data linkage keys.

Before the data management analyst completes the data installation in the Repository, the data dictionaries are subjected to an initial assessment of accuracy and completeness. If deficiencies are identified, the analyst will investigate them through further contacts with the source agency, Manitoba Health, and/or MCHP staff.

Step 6: Release the Data

If the data files and documentation appear ready, then the data can be released for use. Release may occur via an informal process by notifying data analysts at MCHP that the new data and documentation are available for use. A more formal process may also be used that involves presentations to data analysts and researchers. The latter process is useful when a new data source is intended to be used in multiple research projects or if substantial changes have occurred to an existing data source. The latter arises, for example, when the source agency has introduced a new data capture process or system.

Participants in Data Management

The MCHP Director and the Associate Director of Administration are responsible for data sharing agreements for all deliverables and MCHP–initiated projects. However, for project–specific data, a data sharing agreement may be developed amongst the principal investigator, source agency, and the Associate Director of Administration, with confirmation being sent to the Associate Director of the Repository and the Associate Director of Access and Use.

Data management tasks will primarily be undertaken by data management analysts, one or more representatives from the source agency, and one or more Manitoba Health representatives. Usually the process of adding new types of data to the Repository will be guided by a content expert (e.g., principal investigator) and an Advisory Group formed by the principal investigator. The Advisory Group is usually composed of potential data users, researchers, and clinicians with relevant expertise. Advisory Groups have been formed for deliverable data

acquisition projects as well as for data acquisition projects that have been funded by the Canadian Foundation for Innovation. The principal investigator and steering committee members are responsible for guiding many of the data management tasks identified in this chapter.

Project–specific data are not identified as being a component of the Repository and accordingly, MCHP takes limited responsibility for their preparation and content. For project–specific data, MCHP ensures a copy of the data, as provided, is made available to the team conducting the research.

Conclusions

The six–step data management process developed by MCHP follows standards and practices suggested in similar initiatives and follows recommendations developed by other organizations that maintain repositories of anonymized personal health information for research purposes (Daas, Arends–Tóth, Schouten, Kuijvenhoven, & Statistics Netherlands, 2008; Holman, Bass, Rouse, & Hobbs, 1999; Lyman, Scully, & Harrison, 2008). However, the data management process at MCHP also reflects unique aspects of the environment in which MCHP operates, including relationships with source agencies, the software platform on which the Repository is maintained, and provincial health privacy legislation.

CHAPTER 3: A FRAMEWORK FOR DATA QUALITY EVALUATION AT MCHP

The collection and maintenance of administrative data is usually the responsibility of the source agency and is therefore generally beyond the control of the researcher(s) who will use the data. Data quality evaluations are therefore critical to ensure that the data will satisfy the objectives of the proposed project(s).

Quality evaluations conducted by the source agency, which are collected as part of the data management process, can provide useful contextual information. However, these evaluations are unlikely to encompass all relevant aspects of the research uses of administrative data, such as record linkage with the MCHP **Research Registry**[2] or other administrative databases. As well, data quality evaluations conducted by different source agencies may not be comparable because of the lack of standardization of data quality definitions. These factors provided the motivation to develop a Data Quality Framework and evaluation process for MCHP. The framework and formalized process of data quality evaluation provide analysts and researchers with access to indicators and procedures to conduct standardized, automated evaluations on a regular and timely basis.

In this chapter, we describe the MCHP Data Quality Framework and its features. We identify dimensions of quality that are important for administrative data, measures of quality encompassed by the Framework, and tools (i.e., **macros**) that have been developed to operationalize key components of the Framework.

Overview of MCHP Data Quality Framework

The MCHP Data Quality Framework (Figure 3.1) was developed after a scoping review of existing frameworks, including those developed by the **Canadian Institute for Health Information (CIHI)**, Public Health Agency of Canada (PHAC), **Statistics Canada**, and Australian Bureau of Statistics. As well, QuAAD, the Quality Assessment of Administrative Data framework developed by the Institute for Clinical Evaluative Sciences (Iron & Manuel, 2007), was a useful reference. A description of selected data quality frameworks is provided in Appendix B.

2 The MCHP Research Registry is commonly referred to as the Research Registry; the term "Research Registry" will be used throughout the remainder of this document. The Research Registry captures information on dates of coverage for beneficiaries of the Manitoba Health Services Insurance Plan as well as date of birth, sex, and RHA of residence based on municipal and/or postal codes.

Data quality is a broad concept that is both relative and multidimensional (Statistics Canada, 2009). One comprehensive definition is "the totality of features and characteristics of a data set that bear on its ability to satisfy the needs that result from the intended use of the data" (Arts, De Keizer, & Scheffer, 2002). The multidimensional nature of data quality is evident in all of the frameworks (Appendix B). For example, CIHI's data quality framework encompasses concepts of accuracy, timeliness, comparability, usability, and relevance. PHAC's framework encompasses similar concepts, including accuracy, timeliness, serviceability, usability, and relevance. The Statistics Canada framework includes the concepts of relevance, accuracy, timeliness, accessibility, **interpretability**, and coherence. The latter concept, which is unique across these frameworks, refers to the ability to bring together data from different sources; for example, coherence can be achieved by using common methods across surveys or standardized coding systems across time or geography. While the ICES QuAAD model encompasses concepts similar to those found in other frameworks, including correctness, reliability, completeness and usability, the developers emphasize that these concepts do not necessarily represent distinct criteria. Correctness and reliability, for example, are interrelated, and may also be described using terms such as accuracy, reproducibility, and validity. The ICES framework introduces a broad perspective on data quality, recognizing the importance of defining the target audience, the political environment, and the purpose of the data quality evaluation prior to initiating the process. Furthermore, the opportunities to use data quality assessments for improvement and planning for change in the system are also recognized within QuAAD.

Wang, Storey, and Firth (1995) noted that there is a lack of consensus about what constitutes an optimal set of data quality concepts and measures. However, in their review of data quality frameworks, Arts et al. (2002) found that accuracy and completeness were the most frequently cited concepts or measures. Wang and Strong (1996) emphasized that investigating multiple quality concepts can be helpful for identifying the root causes of deficiencies in data quality.

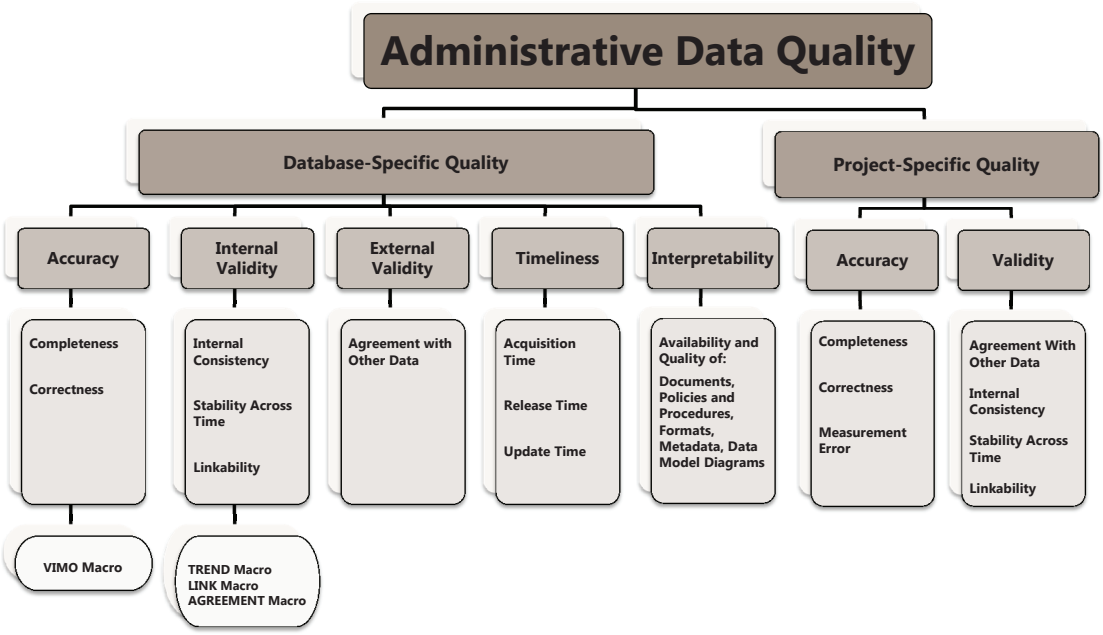
The MCHP Data Quality Framework distinguishes database–specific and **project–specific quality** (Figure 3.1). The environment of data access approvals in Manitoba was a key factor in understanding how data quality evaluations can be conducted. Another important factor was the intended use of the data.

In Manitoba, the types of quality evaluations that can be conducted on a database are primarily determined by the data access approvals received by the researcher. **Database–specific quality** encompasses concepts that can be evaluated without project–specific approval by the provincial Health Information Privacy Committee (HIPC). For example, evaluations requiring limited linkage to the Research Registry, such as linkages to assess the quality of demographic characteristics of age or sex, can be conducted without project–specific approval. On the other hand, data quality evaluations that require extensive linkage with other administrative databases must have both HIPC and Health Research Ethics Board (HREB) approvals.

Database–specific quality encompasses concepts of accuracy, internal and **external validity**, timeliness, and interpretability. Each of these quality concepts is useful for assessing the usability of a database in a different way and is measured by one or more indicators using quantitative or qualitative methods. Some of the quality indicators are linked to macros, written in SAS and developed as part of this study. These macros can be used to generate summary data for a Data Quality Report.

Project–specific quality focuses on concepts of accuracy and validity. These concepts have the same meaning as the corresponding database–specific quality concepts, but indicators of these concepts are applied to a specific cohort, region, or time period that is the focus of the project. Accordingly, data quality evaluation results may vary for database–specific and project–specific analyses. For example, completeness of a field may be very high for the entire population, but if analysis is limited to a specific segment of the population (e.g., to study participants of a program), completeness may be much lower because of the characteristics of the program administration or population participation in the program.

Figure 3.1: MCHP Data Quality Framework



Dimensions of Database–Specific Quality

Accuracy is the degree to which the data correctly describe the phenomenon they were designed to measure (Arts et al., 2002) or the degree to which the data reflect the truth (Iron & Manuel, 2007). Measures of accuracy include completeness or comprehensiveness and correctness (Iron & Manuel, 2007). Completeness can be measured by the percentage of records that contain non–missing values. However, completeness or comprehensiveness can also be measured by investigating database exclusions. For example, Schmidtmann and Blettner (2009) describe completeness of cancer registries in terms of their coverage of the target population. If selected sub–groups are missing from a database because of exclusions based on age, stage/type of disease, or geography, then the databases will result in incomplete estimates of the target outcome (e.g., incidence or prevalence).

Correctness is measured by the percentage of valid values, that is, values within the domain of possible or plausible values. Values may be invalid because they violate physical, logical, or metadata–based constraints. An assessment of validity of data values requires documentation about plausible values as well as knowledge gained through exploratory analyses of the data.

MCHP data management analysts constructed the VIMO macro to provide summary information about completeness and correctness. VIMO is the acronym for Valid, Invalid, Missing, Outlier, and is based on a similar data quality assessment conducted by the United Kingdom's National Health Service (http://www.ic.nhs.uk/services/independent–sector–information–programme/data–quality–assessment–reports) . The VIMO macro produces the percentage of valid, invalid, and missing values for each field in a data file. The sum of the percentages of valid, invalid, and missing values is 100 for each field. The macro produces other descriptive statistics for each field, including the mean, median, standard deviation, minimum, and maximum values, as well as the percentage of outliers. This macro is described in Appendix C.

Validity refers to whether the data makes sense. In the MCHP Data Quality Framework, **internal validity** focuses on the relationship of the data in one field to the data in another field. It encompasses measures of internal consistency, temporal consistency, and linkability. Internal validity also encompasses the coherence between subject–specific identifiers in the database and the identifiers in the Research Registry or another patient or provider registry.

Internal consistency is measured by the numeric agreement between fields or the logical relationships between fields. Numeric agreement is often quantified using the Pearson correlation coefficient for continuous and normally distributed data or the kappa statistic, polychoric correlation coefficient, or tetrachoric correlation coefficient for categorical data. Internal consistency can also be evaluated using logical rules. For example, a 70–year–old woman would not have a baby, a man would not be scheduled for a caesarean section, a four–year–old child would not be recorded as having an occupation, and a hospital employing more than 50 nurses would not likely report an annual salary budget of less than one million dollars. A list of logical rules developed by data analysts and researchers can be used to identify internal inconsistencies in administrative databases.

Temporal consistency is measured by the degree to which a set of time–related observations conforms to a smooth line or curve over time and the percentage of observations that are classified as outliers from that line or curve. Temporal consistency can be assessed using trend analysis, which involves fitting different types of lines or curves to a set of data and applying graphic or inferential techniques to compare observed values with expected values. Process control charts might also been used to investigate temporal consistency (Omar, 2010).

The TREND macro was developed by MCHP data analysts to assess temporal consistency (Appendix C). This macro fits a series of smooth lines or curves to a set of observations and can compute the mean square error (MSE) for the statistical model associated with a line or curve; this information can be used to identify the best–fit line for a set of data. The macro estimates studentized residuals, standardized differences between observed and predicted values. Studentized residuals that are statistically significant (i.e., larger or smaller than expected) are identified. The macro also identifies repeated observations with the exact same value (indicating no change over time) and will flag these as potential coding errors.

Linkability measures the ability to connect one data file to another data file using a unique subject–specific identifier (Iron & Manuel, 2007). In the MCHP Data Quality Framework, linkability is defined as the percentage of records that have common identifiers in two or more administrative databases. This is an important data quality indicator because linkability determines the extent to which different databases can be used in project–specific analyses. A record is considered linkable if its subject–specific identifier corresponds to a valid personal health information number (PHIN) in the Research Registry. A record is not linkable if its subject–specific identifier does not correspond to a valid PHIN in the Research Registry.

Besides the TREND macro, other macros that were developed by MCHP analysts to measure internal validity include the LINK and AGREEMENT macros (Appendix C). The LINK macro produces the number and percentage of linkable records and linkable individuals (e.g., patients, providers) in a data file. The AGREEMENT macro produces values of Cohen's kappa statistic for sex and date of birth for linkable individuals; these are two of the most important demographic variables in a data file. The Research Registry is used as the comparator data source for the production of agreement statistics.

External validity refers to the relationship between the values in a data file and an external source of information. For example, if caesarian section rates computed from a data file are much higher or lower than values published in a report this might suggest a lack of external validity. External validity of data can sometimes be quantified by comparison with a "gold standard", that is, an external data source that contains error–free information about the measure or construct under investigation. Sensitivity, specificity, positive and negative predictive values, and likelihood ratio statistics are used to quantify validity. In the absence of a gold standard or when the gold standard contains measurement error, validity can be quantified using specialized statistical models such as latent class models (Bernatsky et al., 2005).

Timeliness refers to how current the data are. Indicators of timeliness include: (1) time to acquisition, (b) time to release, and (c) recency of the data. The first indicator is the number of days between the date of a data–sharing agreement and the date the file was acquired into the Repository. The second indicator is the number of days between the date the file was acquired and the date the file was released to users. Recency of the data is the number of days between the last reference date in a file and the date the file was released to users.

Finally, the concept of interpretability focuses on the documentation for a data file, including historical and concurrent documentation. The former refers to documentation that is maintained over time, while the latter is developed as the database is examined for inclusion in the Repository. Changes in program inclusion criteria, data collection methods, or reporting criteria may confound an analyst's or researcher's ability to identify data quality problems. For example, changes over time in program eligibility criteria may result in large increases or decreases in the total number of patient or client records contained in a data file. These increases or decreases might inadvertently be flagged as data quality problems if the analyst does not have comprehensive documentation to accompany the data file. As well, changes in the methods used by program staff to code data values might also result in erroneous detection of invalid or out–of–range values. Thus, documentation is important for establishing that a data quality problem does or does not exist.

Dimensions of Project–Specific Data Quality

The majority of data quality frameworks that were reviewed for this report produce general assessments of data quality. Project–specific data quality assessments are intended to examine fitness for use of administrative databases for specific populations, time periods, or geographic regions. Completeness of a database may, for example, vary with age or sex of subjects. Regional variations in data collection mechanisms or program delivery constraints may result in missing data for specific locales. General assessments of data quality may mask these limitations of the data and stratification of the data by all study variables is not possible for a database–specific quality evaluation. Detailed, project–specific evaluations are conducted once a specific project has been identified and approved.

Accuracy and validity are the key concepts to be investigated when assessing data quality for specific research projects. As noted previously, these concepts can be measured in the same way as for a database–specific data quality evaluation. The macros developed by data management analysts at MCHP can also be used to prepare a project–specific data quality report. However, additional objectives for data quality assessment may be developed by the researcher, in consultation with clinicians, analysts, or representatives of the source agency who are familiar with the database contents.

MCHP Data Quality Report

The MCHP Data Quality Report focuses on database–specific indicators of quality. It contains:

a. Summary information about the administrative database, including information about the data provider, rationale for creation of database, and overview of its contents
b. A listing of the data files that comprise the database and the number of fields and records contained in each data file
c. A summary report about database–specific quality indicators
d. Detailed information about each of the database–specific indicators of quality

The CONTENTS macro is used to prepare the Data Quality Report. It produces a summary of the contents of the data files and generates an overview table (Appendix C).

The summary report uses qualitative criteria to describe the percentages of invalid and missing values for each field in a data file: (a) minimal or none: less than 2.0%, (b) moderate: 2.0 to 5.0%, and (c) significant: greater than 5.0%. These criteria are similar to those used by CIHI in its data quality reporting (2009). Similarly, measures of agreement based on the kappa statistic are also given qualitative ratings: (a) very good agreement: 0.90 or higher, (b) good agreement: 0.70 to 0.89, and (c) moderate agreement: below 0.70 (Landis & Koch, 1977).

The Data Quality Report is intended to be used by:

1. MCHP data management staff, for quality assurance processes
2. Associate Director of the Repository, as an accountability mechanism to the MCHP Executive Committee and Advisory Board
3. Users of the Research Data Repository such as data analysts and researchers, to increase efficiency and accuracy of the research
4. Data providers, to improve the quality of ongoing data requests and generate discussions about mechanisms for improving data quality

CHAPTER 4: DATA QUALITY REPORT FOR THE CADHAM PROVINCIAL LABORATORY DATA

The MCHP Data Quality Framework, which was introduced in Chapter 3, was applied to the CPL data to produce a Data Quality Report that focuses on database–specific quality. This chapter begins with a summary of the data quality assessment. Following that, data–base specific quality is described for the clinical microbiology, serology and parasitology, and virus detection sections. While serology and parasitology comprise a single section of CPL, the results of the data quality assessment are usually discussed separately. Appendix D contains additional tables containing data quality information.

Data Quality Report Summary

- There were 28 data files received from CPL that cover the period from fiscal years 1992/93 to 2009/10. The data files contained more than 12 million records and 575 fields.
- For all sections—clinical microbiology, serology and parasitology, virology—the CPL patient number, requisition number, report date, and receive date (i.e., date a specimen was received for testing) were always provided.
- For the clinical microbiology requisitions data file, the fields containing information on sex and date of birth were always complete. The field containing the RHA of the client, which was based on a municipal code or postal code, was almost always (99%) complete.
- For the clinical microbiology results/organisms data file, the record sequence and record type were always complete. The field containing **referring facility** information was almost always complete (98%) as was the specimen date (95%). There were few invalid codes in this section, based on a comparison with the documentation provided by CPL.
- For the parasitology test data file, the field containing test results was always complete. The field containing codes for the interpretation of results was almost always incomplete; this does not necessarily signal a data quality problem because not all results may require an interpretation. The field containing specimen date was complete 90% of the time.
- Results for the serology and virology requisition and test data files showed that information about tests were coded reliably and contained few errors.
- For the virology requisition and test data files, the fields containing information on the referring facility and specimen date were complete between 93% and 95% of the time. Free–form fields that contained comments about the tests were rarely completed.
- For the provider/physician data file, fields containing information on the provider identification number (scrambled), unit office number, region, and municipal code were always complete. The postal code was provided in virtually all (97%) of the records.
- For all requisitions, the agreement between client date of birth and sex in the CPL data and compared to the Research Registry was very good (99% or higher).

- The internal consistency of requisition and testing dates showed good results.
- Results of the analysis of stability over time reveal some large changes in the frequency of requisitions and tests. There are many cases of large increases and decreases in the number of records in the sections, which could indicate changes in testing policies or practices. Information about these changes will need to be incorporated into database documentation.

Data Quality Report Details

Table 4.1 provides an overview of the CPL database. A total of 28 data files comprise the database, with eight of the data files containing requisitions and tests or results. The largest number of records is contained in the data files for serology results and clinical microbiology results and requisitions. The data file containing the parasitology section requisitions has the largest number of fields, followed by the clinical microbiology section organisms file. The auxiliary data files contain small numbers of records; these files are primarily used for interpretation of the codes contained in the requisition and results data files.

Table 4.1: Overview of Data Files, 1992/93–2009/10

Number	Name	Label	Number of Records	Number of Fields
1	MHCPL_SPSEROTESTS_19922010	CPL Serology Section Tests 19922010	4051042	35
2	MHCPL_CMRESULTS_19922010	CPL Clinical Microbiology Section Results 19922010	2366194	29
3	MHCPL_CMSECTION_19922010	CPL Clinical Microbiology Section Requisitions 19922010	2114577	60
4	MHCPL_SPSECTION_19922010	CPL Serology Parasitology Section Requisitions 19922010	2094537	87
5	MHCPL_SPPARATESTS_19922010	CPL Parasitology Section tests 19922010	517459	39
6	MHCPL_VIRUSTESTS_19922010	CPL Virus Detection Section Tests 19922010	365340	29
7	MHCPL_CMORGANISM_19922010	CPL Clinical Microbiology Section Organisms 19922010	283835	74
8	MHCPL_VIRUSSECTION_19922010	CPL Virus Detection Section Requisitions 19922010	232286	49
9	MHCPL_PROVIDER_19922010	CPL Physician - Provider 19922010	3964	10
10	MHCPL_REFTYPE09_19922010	CPL 1992-2010 Auxiliary Type 09	1616	7
11	MHCPL_REFTYPE02_19922010	CPL 1992-2010 Auxiliary Type 02	1169	8
12	MHCPL_REFFACIL_19922010	CPL Referring Facility Table 19922010	949	16
13	MHCPL_REFTYPE08_19922010	CPL 1992-2010 Auxiliary Type 08	861	8
14	MHCPL_REFTYPE15AA_19922010	CPL 1992-2010 Auxiliary Type 15	786	20
15	MHCPL_REFTYPE01_19922010	CPL 1992-2010 Auxiliary Type 01	257	8
16	MHCPL_REFTYPE10_19922010	CPL 1992-2010 Auxiliary Type 10	169	8
17	MHCPL_REFTYPE14_19922010	CPL 1992-2010 Auxiliary Type 14	129	6
18	MHCPL_REFTYPE06_19922010	CPL 1992-2010 Auxiliary Type 06	91	6
19	MHCPL_REFTYPE12_19922010	CPL 1992-2010 Auxiliary Type 12	55	9
20	MHCPL_REFTYPE11_19922010	CPL 1992-2010 Auxiliary Type 11	51	6
21	MHCPL_REFTYPE04_19922010	CPL 1992-2010 Auxiliary Type 04	46	9
22	MHCPL_REFTYPE17_19922010	CPL 1992-2010 Auxiliary Type 17	36	9
23	MHCPL_REFTYPE07_19922010	CPL 1992-2010 Auxiliary Type 07	34	6
24	MHCPL_REFTYPE05_19922010	CPL 1992-2010 Auxiliary Type 05	23	6
25	MHCPL_REFTYPE18_19922010	CPL 1992-2010 Auxiliary Type 18	21	6
26	MHCPL_REFTYPE16_19922010	CPL 1992-2010 Auxiliary Type 16	20	12
27	MHCPL_REFTYPE03_19922010	CPL 1992-2010 Auxiliary Type 03	15	6
28	MHCPL_REFTYPE13_19922010	CPL 1992-2010 Auxiliary Type 13	10	7

Accuracy

The Data Quality Report contains tables that capture information on accuracy (i.e., completeness and correctness) of each of the fields contained in each of the data files enumerated in Table 4.1. Table 4.2 provides this information for the data file that contains information about clinical microbiology organisms identified during testing. When interpreting the results in this table, bold values represent a significant percentage of invalid or missing data (greater than 5%), while values in italics represent a moderate percentage of invalid or missing data (2% to 5%). Values in regular font indicate either no missing or invalid data or minimal missing or invalid data (less than 2%). This table reveals that fields pertaining to the presence of a resistant or susceptible antibiotic detected during testing often do not contain data, as do the fields that identify whether this antibiotic was of a resistant or susceptible type. While the presence of missing data does not signal that the data are of poor quality, a lack of agreement in the percentages of missing data between the antibiotic fields and the corresponding resistant/susceptible field could possibly signal a problem. However, as Table 4.2 reveals, the percentages of valid values are the same in both sets of fields. Almost all other fields in the clinical microbiology organisms section are complete or nearly complete. Similar tables for other clinical microbiology, serology, parasitology, and virus detection sections are provided in Appendix D. As well, detailed descriptions of each of the data files are found in Appendix D.

Tables 4.3 and 4.4 provide information about the accuracy (i.e., completeness and correctness) of the physician/provider and referring facility data files. This information is useful for identifying who submitted a requisition for a laboratory test and where the requisition originated. As the data in these two tables reveal, almost all of the fields were complete and contained valid data.

Table 4.2: Valid, Invalid, and Missing Data for Clinical Microbiology Organisms Data File, 1992/93–2009/10

Type	Variable Name	Variable Label	% Valid	% Invalid	% Missing
Identification	FILEPHIN	MH Scrambled PHIN	100.00	0.00	0.00
	SCRPHIN	MCHP Scrambled PHIN	100.00	0.00	0.00
Numeric	RECPOSN	Position of this record in requisition	100.00	0.00	0.00
Character	ANTIBIOTIC01	CM type 2 -Antibiotic 01	32.88	0.00	**67.12**
	ANTIBIOTIC02	CM type 2 -Antibiotic 02	15.27	0.00	**84.73**
	ANTIBIOTIC03	CM type 2 -Antibiotic 03	30.40	0.00	**69.60**
	ANTIBIOTIC04	CM type 2 -Antibiotic 04	3.64	0.00	**96.36**
	ANTIBIOTIC05	CM type 2 -Antibiotic 05	24.81	0.00	**75.19**
	ANTIBIOTIC06	CM type 2 -Antibiotic 06	43.24	0.00	**56.76**
	ANTIBIOTIC07	CM type 2 -Antibiotic 07	0.28	0.00	**99.72**
	ANTIBIOTIC08	CM type 2 -Antibiotic 08	13.28	0.00	**86.72**
	ANTIBIOTIC09	CM type 2 -Antibiotic 09	12.97	0.00	**87.03**
	ANTIBIOTIC10	CM type 2 -Antibiotic 10	6.40	**7.45**	**86.15**
	ANTIBIOTIC11	CM type 2 -Antibiotic 11	14.51	0.00	**85.49**
	ANTIBIOTIC12	CM type 2 -Antibiotic 12	3.90	0.00	**96.10**
	ANTIBIOTIC13	CM type 2 -Antibiotic 13	0.28	0.00	**99.72**
	ANTIBIOTIC14	CM type 2 -Antibiotic 14	3.97	0.00	**96.03**
	ANTIBIOTIC15	CM type 2 -Antibiotic 15	0.77	0.00	**99.23**
	ANTIBIOTIC16	CM type 2 -Antibiotic 16	0.99	0.00	**99.01**
	ANTIBIOTIC17	CM type 2 -Antibiotic 17	17.64	0.00	**82.36**
	ANTIBIOTIC18	CM type 2 -Antibiotic 18	0.70	0.00	**99.30**
	ANTIBIOTIC19	CM type 2 -Antibiotic 19	0.52	0.00	**99.48**
	ANTIBIOTIC20	CM type 2 -Antibiotic 20	9.20	0.00	**90.80**
	ANTIBIOTIC21	CM type 2 -Antibiotic 21	3.87	0.00	**96.13**
	ANTIBIOTIC22	CM type 2 -Antibiotic 22	0.00	0.00	**100.00**
	ANTIBIOTIC23	CM type 2 -Antibiotic 23	0.00	0.00	**100.00**
	ANTIBIOTIC24	CM type 2 -Antibiotic 24	0.00	0.00	**100.00**
	ANTIBIOTIC25	CM type 2 -Antibiotic 25	0.00	0.00	**100.00**
	CMTESTTYPE	CM Test Type	100.00	0.00	0.00
	CPLPATIENTNUMBER	CPL Patient Number	100.00	0.00	0.00
	FILEPHINTYPE	Method to determine FILEPHIN	100.00	0.00	0.00
	ORGANISM	CM type 2 - Organism	100.00	0.00	0.00
	POSNEG	Positive-Negative	100.00	0.00	0.00
	RECSEQUENCE	Record Sequence	100.00	0.00	0.00
	RECTYPE	Record Type	100.00	0.00	0.00
	REFERFACIL	Referring Facility	98.12	0.00	1.88
	REFEROUT	Referred Out	0.00	0.00	**100.00**
	REQUISITIONNUMBER	Requisition Number	100.00	0.00	0.00
	RESISTSUSCEPT01	CM type 2 -Resistant or Susceptible 01	32.88	0.00	**67.12**
	RESISTSUSCEPT02	CM type 2 -Resistant or Susceptible 02	15.27	0.00	**84.73**
	RESISTSUSCEPT03	CM type 2 -Resistant or Susceptible 03	30.40	0.00	**69.60**
	RESISTSUSCEPT04	CM type 2 -Resistant or Susceptible 04	3.64	0.00	**96.36**
	RESISTSUSCEPT05	CM type 2 -Resistant or Susceptible 05	24.81	0.00	**75.19**
	RESISTSUSCEPT06	CM type 2 -Resistant or Susceptible 06	43.24	0.00	**56.76**
	RESISTSUSCEPT07	CM type 2 -Resistant or Susceptible 07	0.28	0.00	**99.72**
	RESISTSUSCEPT08	CM type 2 -Resistant or Susceptible 08	13.28	0.00	**86.72**
	RESISTSUSCEPT09	CM type 2 -Resistant or Susceptible 09	12.97	0.00	**87.03**
	RESISTSUSCEPT10	CM type 2 -Resistant or Susceptible 10	13.85	0.00	**86.15**
	RESISTSUSCEPT11	CM type 2 -Resistant or Susceptible 11	14.51	0.00	**85.49**
	RESISTSUSCEPT12	CM type 2 -Resistant or Susceptible 12	3.90	0.00	**96.10**

Table 4.2 - Continued

Type	Variable Name	Variable Label	% Valid	% Invalid	% Missing
	RESISTSUSCEPT13	CM type 2 -Resistant or Susceptible 13	0.28	0.00	**99.72**
	RESISTSUSCEPT14	CM type 2 -Resistant or Susceptible 14	3.97	0.00	**96.03**
	RESISTSUSCEPT15	CM type 2 -Resistant or Susceptible 15	0.77	0.00	**99.23**
	RESISTSUSCEPT16	CM type 2 -Resistant or Susceptible 16	0.99	0.00	**99.01**
	RESISTSUSCEPT17	CM type 2 -Resistant or Susceptible 17	17.64	0.00	**82.36**
	RESISTSUSCEPT18	CM type 2 -Resistant or Susceptible 18	0.70	0.00	**99.30**
	RESISTSUSCEPT19	CM type 2 -Resistant or Susceptible 19	0.52	0.00	**99.48**
	RESISTSUSCEPT20	CM type 2 -Resistant or Susceptible 20	9.20	0.00	**90.80**
	RESISTSUSCEPT21	CM type 2 -Resistant or Susceptible 21	3.87	0.00	**96.13**
	RESISTSUSCEPT22	CM type 2 -Resistant or Susceptible 22	0.00	0.00	**100.00**
	RESISTSUSCEPT23	CM type 2 -Resistant or Susceptible 23	0.00	0.00	**100.00**
	RESISTSUSCEPT24	CM type 2 -Resistant or Susceptible 24	0.00	0.00	**100.00**
	RESISTSUSCEPT25	CM type 2 -Resistant or Susceptible 25	0.00	0.00	**100.00**
	SCRPHINTYPE	Method to determine SCRPHIN	100.00	0.00	0.00
	SECTION	Section	100.00	0.00	0.00
	SPECIMENSOURCE	Specimen Source	100.00	0.00	0.00
	STATUS	Status	100.00	0.00	0.00
	TECHINIT	Technician Initials	100.00	0.00	0.00
	TESTSUBSECTION	Test Subsection	1.68	0.00	**98.32**
	VERIFIED	Verified	100.00	0.00	0.00
Date	ACQDT	Date record was acquired by MCHP	100.00	0.00	0.00
	RECEIVEDDT	Received Date	100.00	0.00	0.00
	REPORTDT	Report Date	100.00	0.00	0.00
	SPECIMENDT	Specimen Date	95.38	0.00	*4.62*

Legend for Invalid and Missing Columns:
Regular font: None or Minimal; *Italics: Moderate* **; Bold: Significant**

Table 4.3: Valid, Invalid, and Missing Data for Physician/Provider Data File, 1992/93–2009/10

Type	Variable Name	Variable Label	% Valid	% Invalid	% Missing
Character	HEALTHUNITOFFICE	Health Unit - Office	100.00	0.00	0.00
	HEALTHUNITREGION	Health Unit- Region	100.00	0.00	0.00
	MHREGION	MH Region	99.90	0.10	0.00
	MUNCODE	Municipal Code	99.90	0.10	0.00
	PHYSICIANNUMBER	Physician Number	100.00	0.00	0.00
	POSTAL	Postal Code	96.70	0.00	*3.30*
	RHA	Regional Health Authority	99.90	0.10	0.00
Date	ACQDT	Date record was acquired by MCHP	100.00	0.00	0.00
	DELETEDT	Delete Date	82.27	0.00	**17.73**
	STARTDT	Start Date	100.00	0.00	0.00

Legend for Invalid and Missing Columns:
Regular font: None or Minimal; *Italics: Moderate* **; Bold: Significant**

Table 4.4: Valid, Invalid, and Missing Data for Referring Facility Data File, 1992/93–2009/10

Type	Variable Name	Variable Label	% Valid	% Invalid	% Missing
Character	FACILADDR1	Facility Address #1	100.00	0.00	0.00
	FACILADDR2	Facility Address #2	46.26	0.00	**53.74**
	FACILADDR3	Facility Address #3	96.73	0.00	*3.27*
	FACILADDR4	Facility Address #4	93.68	0.00	**6.32**
	FACILNAME	Facility Name	100.00	0.00	0.00
	FACILPOSTAL	Facility Postal Code	99.47	0.00	0.53
	HEALTHUNITOFFICE	Health Unit - Office	100.00	0.00	0.00
	HEALTHUNITREGION	Health Unit- Region	100.00	0.00	0.00
	MESSAGECD	Message Code	2.21	0.00	**97.79**
	MHREGION	Manitoba Health Region	98.95	0.00	1.05
	MUNCODE	Municipal Code	100.00	0.00	0.00
	REFERFACIL	Referring Facility #	100.00	0.00	0.00
	RHA	Regional Health Authority	100.00	0.00	0.00
Date	ACQDT		100.00	0.00	0.00
	DELETEDT	Delete Date	100.00	0.00	0.00
	STARTDT	Start Date	100.00	0.00	0.00

Legend for Invalid and Missing Columns:
Regular font: None or Minimal; *Italics: Moderate* **; Bold: Significant**

Internal Validity

The results of the internal validity assessment are provided next. Table 4.5 provides information on the agreement of sex and date of birth fields contained in the CPL data files with the Research Registry. This evaluation is conducted only for those data files that contain information on both variables. As this table reveals, the agreement is extremely high.

Table 4.5: Agreement of Sex and Date of Birth Fields in CPL Files with Population Registry, 1992/93–2009/10

Name	Label	Sex: CPL Agreement with Registry (Kappa Statistic)	Date of Birth: CPL Agreement with Registry (Kappa Statistic)
MHCPL_CMSECTION_19922010	CPL Clinical Microbiology Section Requisitions 19922010	0.9997	0.9951
MHCPL_SPSECTION_19922010	CPL Serology Parasitology Section Requisitions 19922010	0.9996	0.9952
MHCPL_VIRUSSECTION_19922010	CPL Virus Detection Section Requisitions 19922010	0.9993	0.9954

An assessment of linkability of the data files was also conducted. Table 4.6 provides information on the number and percentage of linkable records and the number of linkable individuals for the data files that contain a personal health information number (PHIN). Linkability was high (i.e., above 80%) for all data files except for those containing requisitions for a serology or parasitology test. Internal validity was also assessed by conducting **descriptive analyses** of the frequency of the number of records over time. Specifically, we conducted trend analyses of the number of records that contained a receive date in each fiscal year; this date was recorded in 100% of the records in each of the requisition and test results data files that comprise the CPL sections. In the figures that follow, only a linear trend line has been fit to the annual frequencies, although the TREND macro will produce the best–fit trend line or curve for a set of data. Darkened circles are used to denote values that are outliers from this linear trend line.

Table 4.6: Linkability of CPL Data Files with Population Registry, 1992/93 – 2009/10

Data File		Total Number of Records	Number of Linkable Records	Percentage of Linkable Records	Number of Linkable Individuals
MHCPL_CMORGANISM_19922010	CPL Clinical Microbiology Section Organisms 19922010	283,835	265,942	93.70	103,264
MHCPL_CMRESULTS_19922010	CPL Clinical Microbiology Section Results 19922010	2,366,194	2,242,674	94.78	542,488
MHCPL_CMSECTION_19922010	CPL Clinical Microbiology Section Requisitions 19922010	2,114,577	2,009,005	95.01	557,748
MHCPL_SPPARATESTS_19922010	CPL Parasitology Section Tests 19922010	517,459	491,307	94.95	83,068
MHCPL_SPSECTION_19922010	CPL Serology Parasitology Section Requisitions 19922010	2,094,537	1,579,243	75.40	606,009
MHCPL_SPSEROTESTS_19922010	CPL Serology Section Tests 19922010	4,051,042	3,401,205	83.96	566,781
MHCPL_VIRUSSECTION_19922010	CPL Virus Detection Section Requisitions 19922010	232,286	199,134	85.73	122,345
MHCPL_VIRUSTESTS_19922010	CPL Virus Detection Section Tests 19922010	365,340	313,230	85.74	122,345

Note that the trend analysis is conducted for fiscal years 1992/93 to 2008/09 only. The analyses showed a strong downward trend for 2009/10 for all data files indicating that the data received for the final year was incomplete.

Overall, these data quality assessments reveal a lack of consistency in the frequency of requisitions and tests over time. However, the annual values that are classified as outliers changes from one data file to the next, suggesting that there is no single year which warrants a more detailed data quality assessment.

Figure 4.1: Trend Analysis for Frequency of Records in the Clinical Microbiology Organisms Data File, 1992/93–2008/09

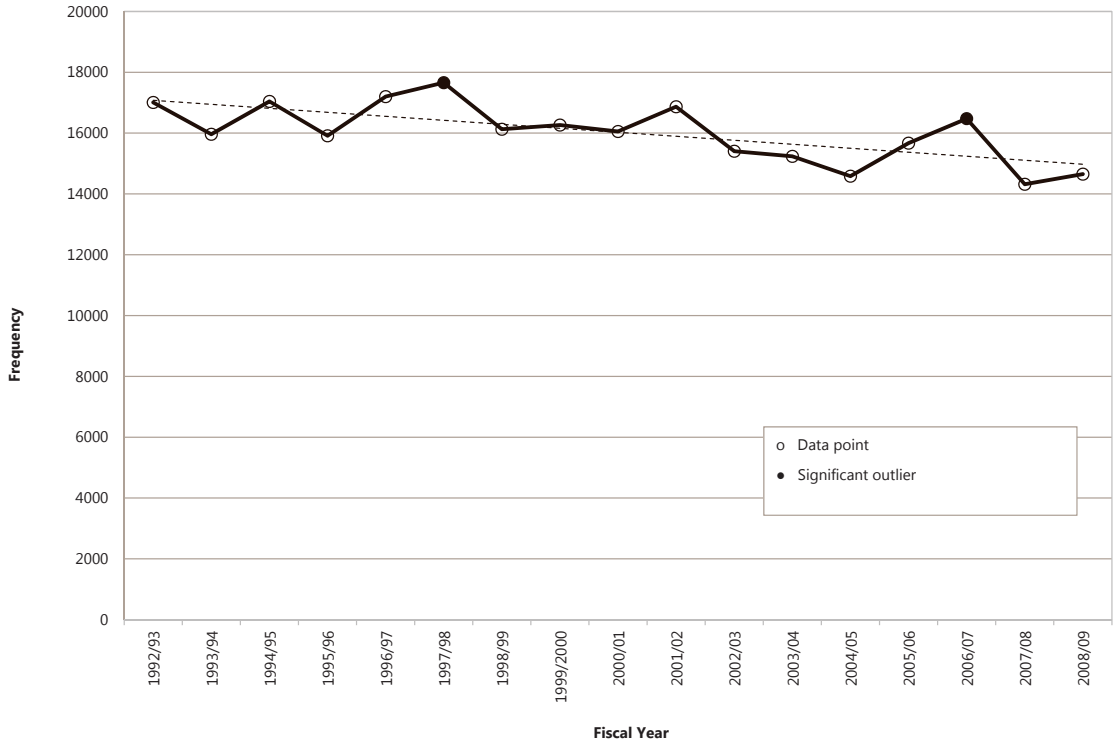


Figure 4.2: Trend Analysis for Frequency of Records in the Clinical Microbiology Requistions Data File, 1992/93–2008/09

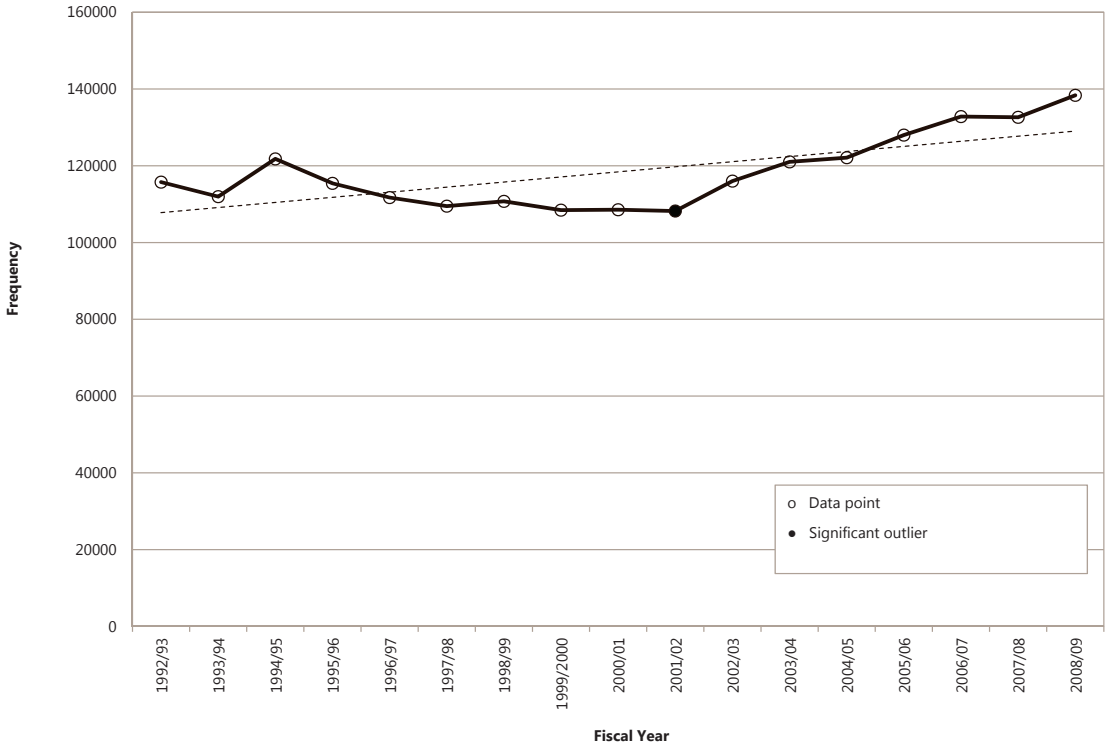


Figure 4.3: Trend Analysis for Frequency of Records in the Clinical Microbiology Results Data File, 1992/93–2008/09

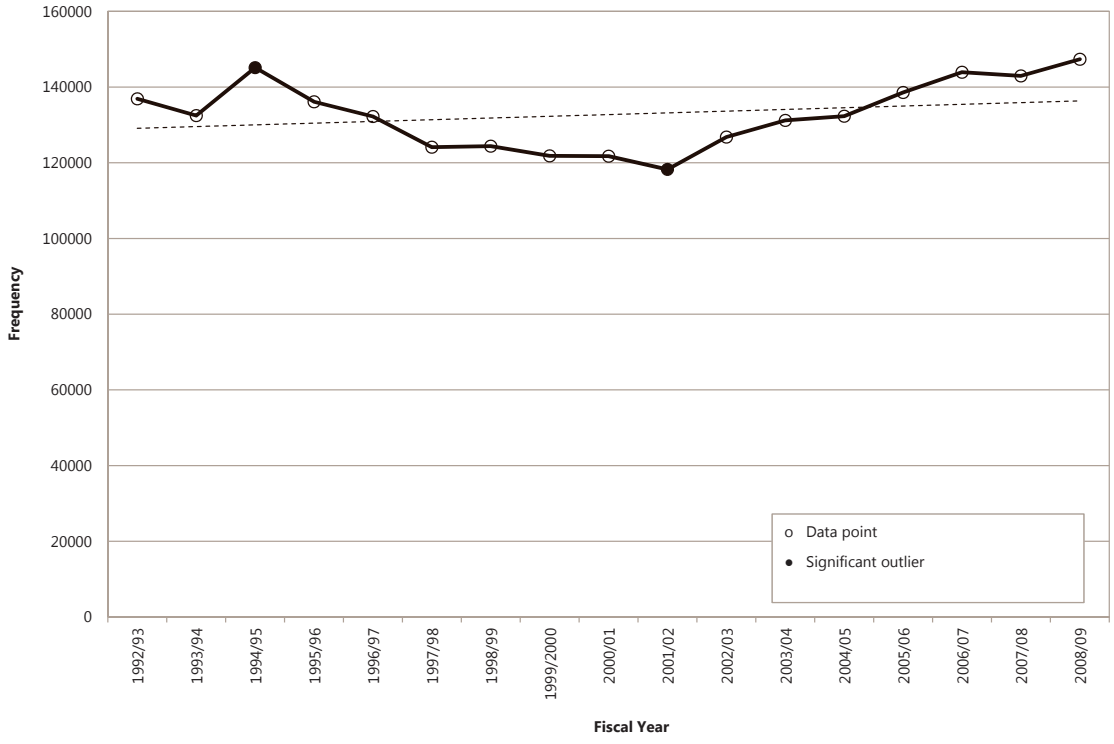


Figure 4.4: Trend Analysis for Frequency of Records in the Parasitology and Serology Requisitions Data File, 1992/93–2008/09

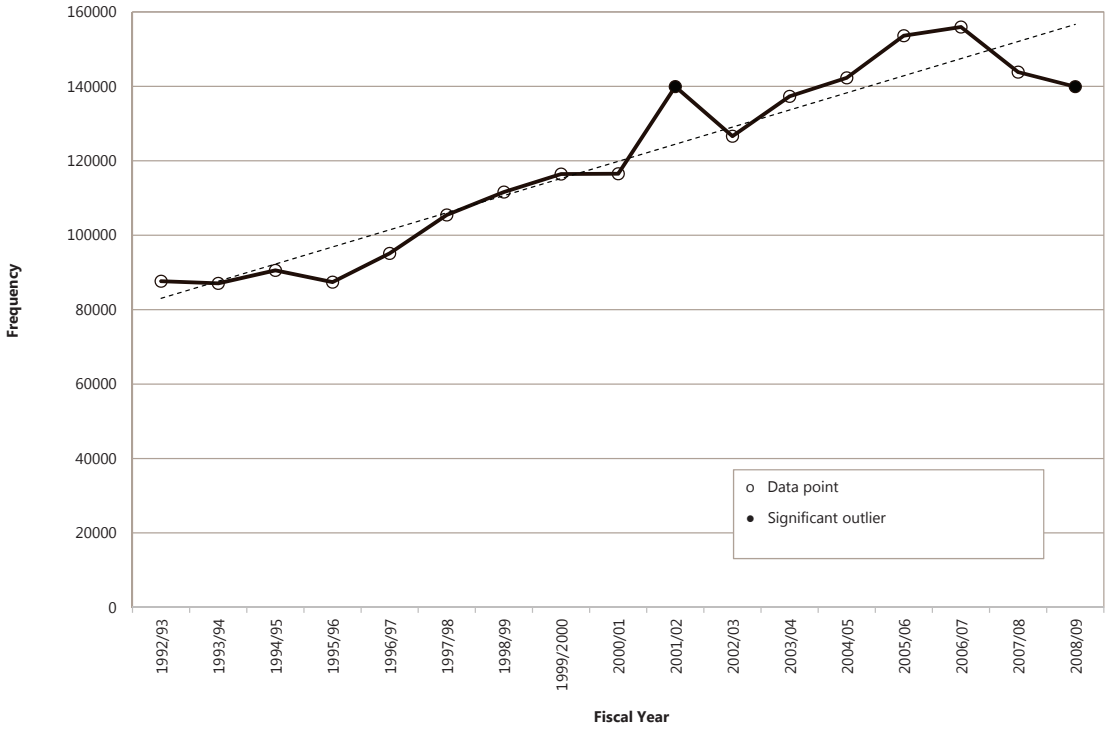


Figure 4.5: Trend Analysis for Frequency of Records in the Parasitology Results Data File, 1992/93–2008/09

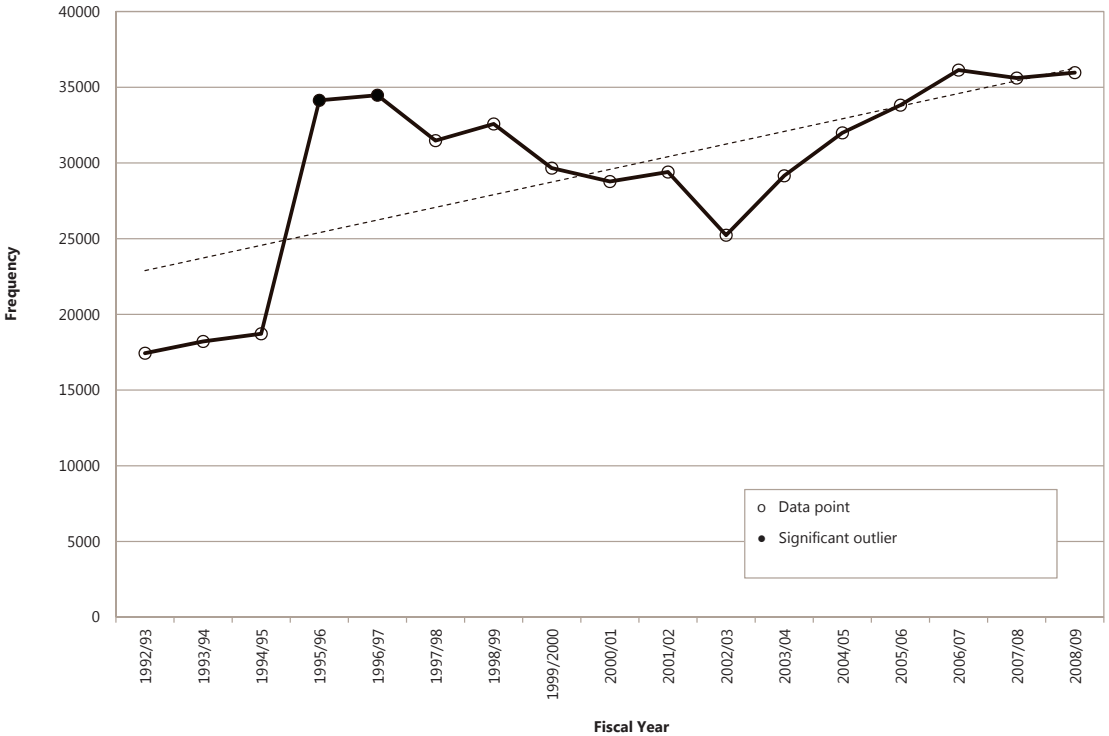


Figure 4.6: Trend Analysis for Frequency of Records in the Serology Results Data File, 1992/93–2008/09

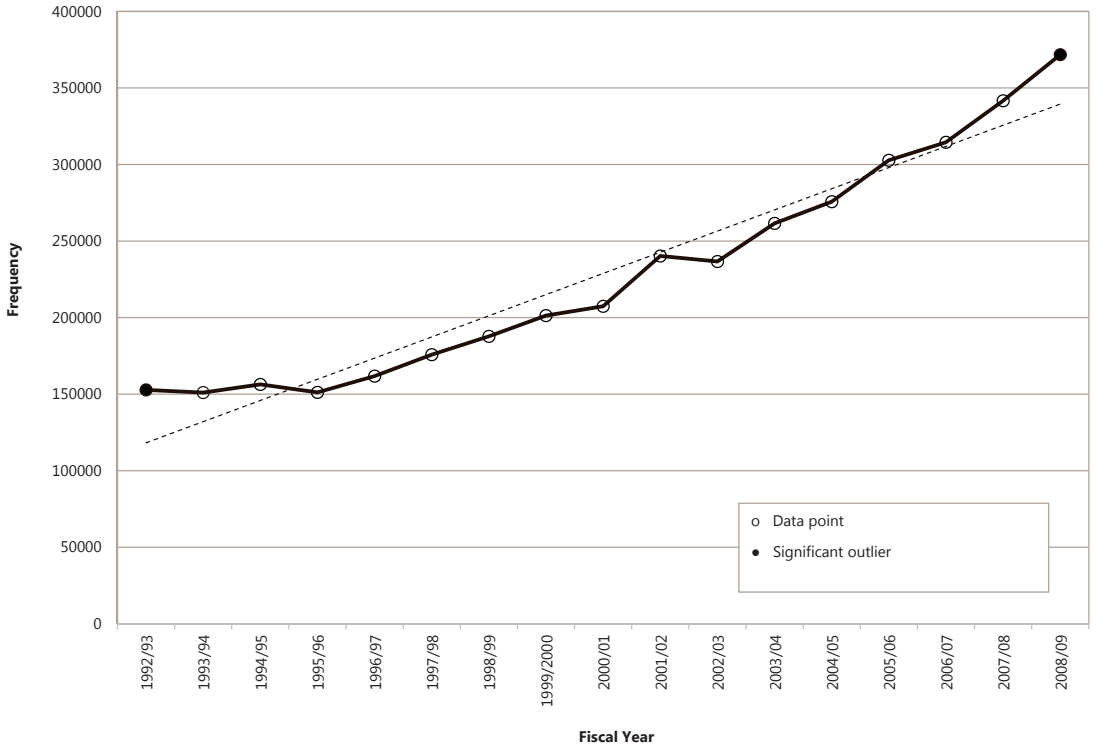


Figure 4.7: Trend Analysis for Frequency of Records in the Virus Detection Requisitions Data File, 1992/93–2008/09

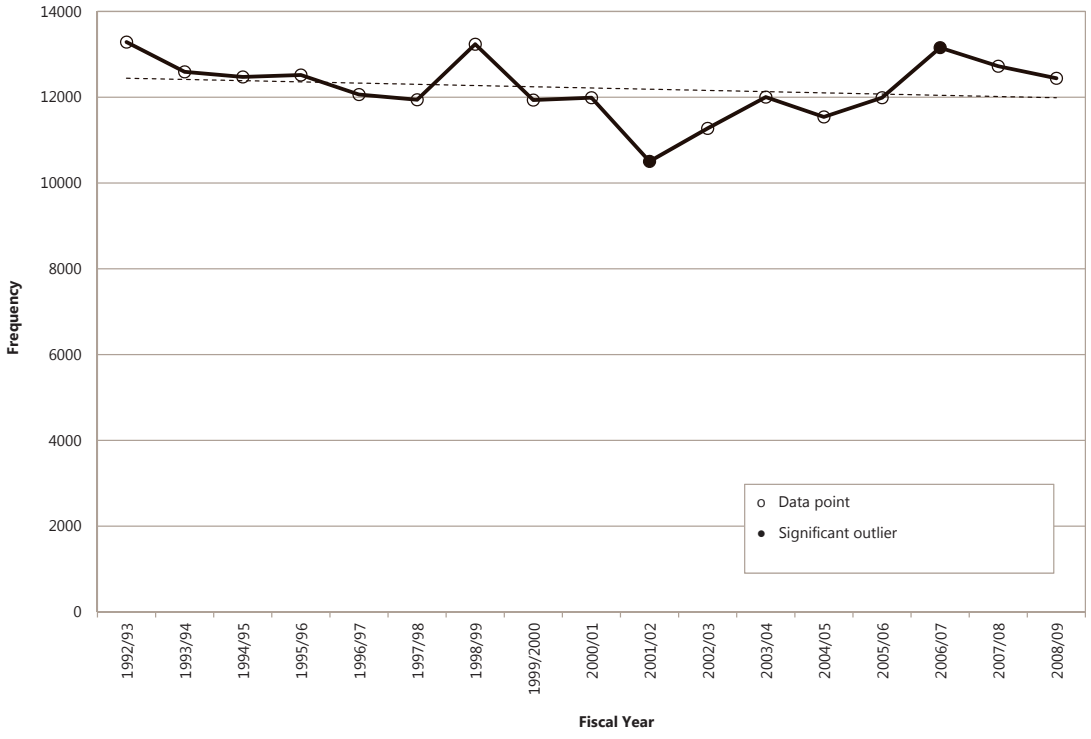
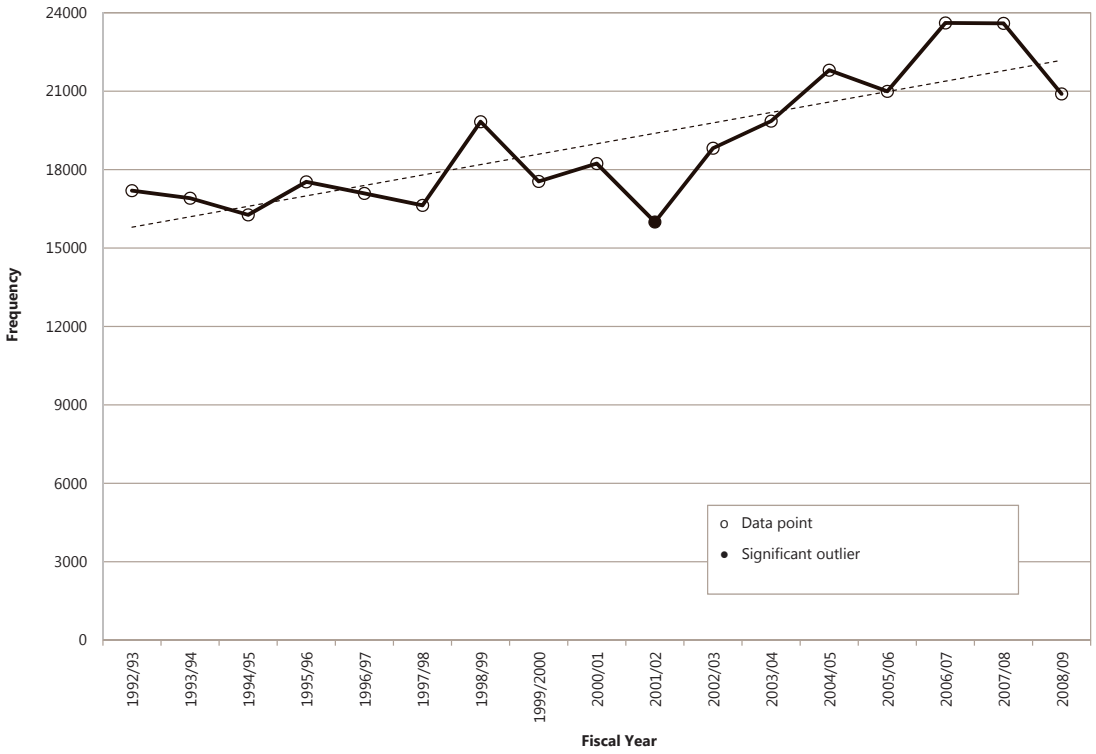


Figure 4.8: Trend Analysis for Frequency of Records in the Virus Detection Results Data File, 1992/93–2008/09



Timeliness

The CPL data sharing agreement was established on February 4, 2010 and the data were acquired on June 16, 2010, giving a total of 132 days to acquisition. The data were subsequently installed on the SPD Server on August 26, 2010, giving a total of 71 days to release. The latest date on a record contained in the data file is August 10, 2010, giving a recency of 16 days. There are currently no guidelines for best practices around data timeliness at MCHP. Compilation of these statistical across multiple Repository databases could be used to define benchmarks based on average or median values.

CHAPTER 5: A PROFILE OF CADHAM PROVINCIAL LABORATORY REQUISITIONS AND TESTS IN THE MANITOBA POPULATION

This chapter focuses on measures of data quality that are relevant to specific projects that may be undertaken with the CPL data, by focusing on the characteristics of requisitions and tests for healthcare providers, facilities, and Manitoba residents. It examines the distribution of requisitions by characteristics of the referring provider and facility, including the type of provider or facility and their RHA location. The distribution of tests and requisitions by selected characteristics of the client population, including residence location, age, sex, and **income quintile**, are also examined. Separate analyses are conducted for the entire Manitoba population and the Manitoba **prenatal population**. The latter was included because pregnant women are routinely tested for hepatitis B, syphilis, and rubella and, therefore, represent a large and important population for future research projects that use the CPL data.

The primary motivation for the descriptive analyses presented in this chapter is the examination of trends over time in the source of requisitions and the recipients of tests. Changes in annual trends may provide an indication of changes in data collection practices or program characteristics that may either limit or enhance the usefulness of the CPL data for population health and health services research.

Methods

The Repository data sources used for these analyses were the CPL database, Research Registry, physician resource file, hospital discharge abstracts, and physician billing claims. The physician resource file captures characteristics of physicians licensed to practice in Manitoba, including specialty and clinic or service location. A hospital discharge abstract is completed when a patient is discharged from an **acute care** facility. Diagnoses in hospital abstracts are recorded using the **International Classification of Diseases, 9th revision, Clinical Modification (ICD–9–CM)** up to and including the 2003/04 **fiscal year** and the **International Classification of Diseases, 10th revision, Canadian version (ICD–10–CA)** for subsequent years. As many as 16 ICD–9–CM codes and 25 ICD–10–CA codes are recorded on each abstract. Physicians who are paid on a fee–for–service basis submit billing claims to the provincial health ministry; these claims capture almost all outpatient services, including those provided in hospital emergency departments and outpatient departments. **Physician claims** contain a single ICD–9–CM code.

For the referring provider and facility analyses, the study cohort consisted of all healthcare providers and facilities with an identification number recorded on a CPL requisition for the study period from fiscal year 1992/93 to 2009/10. The CPL referring provider number was linked to the physician resource file to capture information on provider characteristics. Specifically, the physician resource file was used to distinguish **general practitioners (GPs)/family practitioners (FP)** from specialists. The latter category encompasses specializations in psychiatry, pediatrics, obstetrics and gynecology, medical specialties (e.g., internal medicine), general or specialty surgery, anaesthesiology, radiology, pathology, and non–medical specialties (e.g., dentistry, optometry, chiropractic). Alphanumeric numbers in the CPL physician/provider data file were used to identify other types of providers, including health unit directors, public health units, midwives, pathologists, and veterinary services.

Some referring provider numbers in the CPL data files were not associated with a corresponding number in the physician resource file. These numbers were assigned to an unlinked category. For some requisitions, a referring provider number was missing. However, a referring facility was listed on all requisitions and was used in the analyses.

For all records on which a CPL referring facility number was recorded, the number was linked to a crosswalk file of facility identification numbers developed by MCHP analysts. The following categorization of facilities was developed: acute care, long–term care, nursing station, medical group, laboratory, physician office, other, and missing. Acute care facilities include all hospitals except Deer Lodge and Riverview Health Centre. Long–term care facilities include personal care homes, Deer Lodge, and Riverview Health Centre. Nursing stations include all facilities for which a nursing station name was provided. Medical groups encompass groups of physicians or other health professionals, including clinics, medical centres, health centres, medical groups, family practices, family medical clinics, medical buildings, health stations, and medical co–operatives. Laboratories include facilities for which a laboratory name was provided. Physician offices include all referring facilities listed as physician's office or for which a physician's name was listed as the referring facility. The 'other' category includes referring facilities that could not be assigned to one of the previous categories. Requisitions that did not have a referring facility number were included in the missing category.

For the client analyses, the study cohort was comprised of all Manitoba residents who met the following inclusion criteria for the period from fiscal year 1992/93 to 2009/10: (a) residents with complete health insurance coverage for the entire fiscal year, (b) residents who were born in the fiscal year and had complete coverage from birth until the end of the fiscal year, and (c) residents who died in the fiscal year and had complete coverage from the start of the fiscal year until the date of death. Postal code and municipal code (if provided) was identified from the Research Registry, and used to assign the RHA of residence. For the income quintile assignment, postal code was assigned to an **enumeration area (EA)**, the smallest geographic area for which **Census** data are available prior to 2001, or a **dissemination area (DA)**, which replaced the EA as the Census unit of geography in 2001. Income ranges were determined such that the Manitoba population was divided into five approximately equal groups. Residents were assigned an income quintile according to their EA or DA average household income. Some residents could not be assigned to a quintile; for example postal codes in which more than 90% of the residents are in long–term care facilities are excluded because the Census does not collect information on income for institutionalized persons. Other postal codes that are not included are those belonging to the public trustee office, prisons, and mental health institutions. The income quintile methodology was applied separately to data for rural and urban regions of Manitoba. Urban regions included Winnipeg, while all other regions were classified as rural.

For the prenatal analyses, the study cohort was comprised of all females 10 years of age or older in a fiscal year who met the following inclusion criteria:

a. hospital discharge abstract with an admission date between April 1, 1992 and March 31, 2010 and a diagnosis code for a delivery outcome (i.e., live or still born) of ICD–9–CM V27 or ICD–10–CA Z37OR

b. pregnancy episode, defined as two or more physician visits within a 60–day period with a service date between January 31, 1992 and March 31, 2010 AND
   (i) a physician billing (i.e., tariff) code indicating a prenatal service (8400 or 8401) OR
   (ii) a physician billing code (8501, 8507, 8509, 8540) in conjunction with a diagnosis code of 640 to 648 (complications mainly related to pregnancy), 650 to 659 (normal delivery and other indicators for pregnancy, labour, and delivery), 660 to 669 (complications during labour and delivery) OR
   (iii) a diagnosis code of V22 (normal pregnancy) or V23 (supervision of high–risk pregnancy) OR
c. hospital discharge abstract with an admission date between April 1, 1992 and March 31, 2010 and a diagnosis code for abnormal pregnancy outcomes including ectopic and molar pregnancies and intrauterine deaths or a procedure code related to an abnormal pregnancy outcome procedure (ICD–9–CM 630–637, 656.4; ICD–10–CA O00, O01, O02.1, O03 to O07, O36.4, D39.2; ICD–9–CM procedure codes 66.62, 69.01, 69.51, 74.3, 74.91, 75.0; CCI codes 5.CA.88, 5.CA.89, 5.CA.90, 5.CA.93, 5.MD.5, 5.MD.60)

For hospitalizations with a live or still born delivery outcome, gestational age (in weeks) on either the cohort study member's record or the linked infant's record was used to establish the prenatal period. For individuals identified from physician claims data, the prenatal period was defined as the interval extending from 30 days prior to the first prenatal care visit to 30 days after the last prenatal care visit within the pregnancy episode. For hospitalizations for other outcomes, the prenatal period began 280 days prior to the admission date.

For the referring provider analyses, frequencies and percentages of requisitions were calculated by fiscal year, type of provider, and RHA of the provider. For the referring facility analyses, frequencies and percentages of requisitions were calculated by fiscal year, type of facility, and RHA of the facility. Frequencies and percentages of requisitions and tests for the remaining analyses were calculated by fiscal year, age group, sex, RHA of residence, Winnipeg/non–Winnipeg residence, and income quintile.

Overview of Notifiable Disease Requisitions and Tests

Table 5.1 provides information about the total number of requisitions and tests by fiscal year. It is apparent that there is a substantial decline in the number of requisitions and tests in the 2009/10 fiscal year, which is when the transition to the new CPL Laboratory Information Management System began. Given the anomalous results for this fiscal year, it is excluded from further analyses in which the data are stratified by fiscal year.

The number of requisitions increased by 35.2% between fiscal years 1992/93 and 2008/09 while the number of tests increased by 73.0% during this period (Table 5.1). Accordingly, the ratio of tests to requisitions increased from 1.60 to 2.05.

The serology and clinical microbiology sections accounted for the vast majority of the tests (Figure 5.1). However, while the percentage of all tests associated with the **serology section** increased over time, from 44.8% in 1992/93 to 62.9% in 2008/09, the percentage of tests associated with the clinical microbiology section decreased, from 45.1% in 1992/93 to 27.4% in 2008/09. The percentage of parasitology section tests increased slightly, from 5.1% to 6.1%, while the percentage of tests associated with the **virus detection section** decreased slightly from 5.0% to 3.5%.

An investigation of seasonal variations in requisitions (Figure 5.2) revealed that while there was no consistent trend for the years selected for investigation (i.e., 1992/93, 2000/01, and 2008/09), the frequency of requisitions tended to be highest in or around October and lowest in December and July or August. When a detailed investigation was conducted by section for the most recent fiscal year 2008/09 (see Figure 5.3), the month with the highest number of requisitions was October for the clinical microbiology, serology, and parasitology sections and March for virus detection section.

Table 5.1: Frequency of Tests and Requisitions, 1992/93–2009/10

Fiscal Year	Total Requisitions	Total Tests	Tests/ Requisition	Tests: Virus Detection		Tests: Parasitology		Tests: Serology		Tests: Clinical Microbiology	
	Freq	Freq	Average	Freq	%	Freq	%	Freq	%	Freq	%
1992/93	213343	341345	1.60	17195	5.04	17434	5.11	152797	44.76	153919	45.09
1993/94	207920	334594	1.61	16907	5.05	18214	5.44	151025	45.14	148448	44.37
1994/95	221359	353541	1.60	16268	4.60	18719	5.29	156382	44.23	162172	45.87
1995/96	211569	354860	1.68	17533	4.94	34137	9.62	151146	42.59	152044	42.85
1996/97	214892	362775	1.69	17091	4.71	34474	9.50	161783	44.60	149427	41.19
1997/98	222947	365723	1.64	16632	4.55	31477	8.61	175828	48.08	141786	38.77
1998/99	231269	380676	1.65	19833	5.21	32576	8.56	187747	49.32	140520	36.91
1999/2000	233168	386661	1.66	17549	4.54	29662	7.67	201358	52.08	138092	35.71
2000/01	233618	392198	1.68	18234	4.65	28773	7.34	207385	52.88	137806	35.14
2001/02	255090	420755	1.65	15995	3.80	29401	6.99	240244	57.10	135115	32.11
2002/03	251084	422903	1.68	18821	4.45	25228	5.97	236660	55.96	142194	33.62
2003/04	267860	456991	1.71	19855	4.34	29152	6.38	261540	57.23	146444	32.05
2004/05	273211	476393	1.74	21805	4.58	31987	6.71	275718	57.88	146883	30.83
2005/06	291269	511836	1.76	20996	4.10	33820	6.61	302770	59.15	154250	30.14
2006/07	299093	534675	1.79	23612	4.42	36137	6.76	314520	58.82	160406	30.00
2007/08	286834	558141	1.95	23594	4.23	35613	6.38	341653	61.21	157281	28.18
2008/09	288465	590607	2.05	20897	3.54	35965	6.09	371736	62.94	162009	27.43
2009/10	180576	336069	1.86	41782	12.43	14685	4.37	160707	47.82	118895	35.38

Figure 5.1: Percent of Requisitions by Section, 1992/93–2008/09

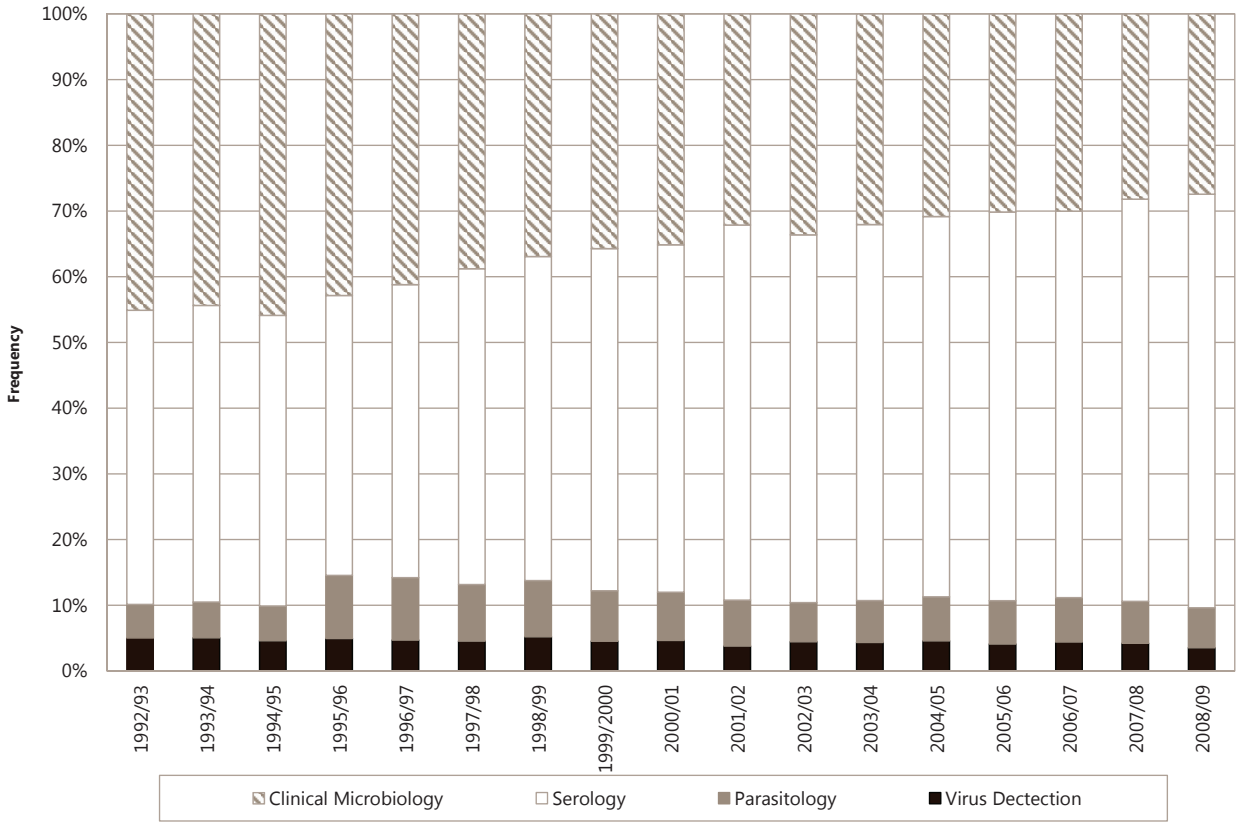


Figure 5.2: Frequency of Requisitions by Month, All Sections,1992/93, 2000/01, and 2008/09

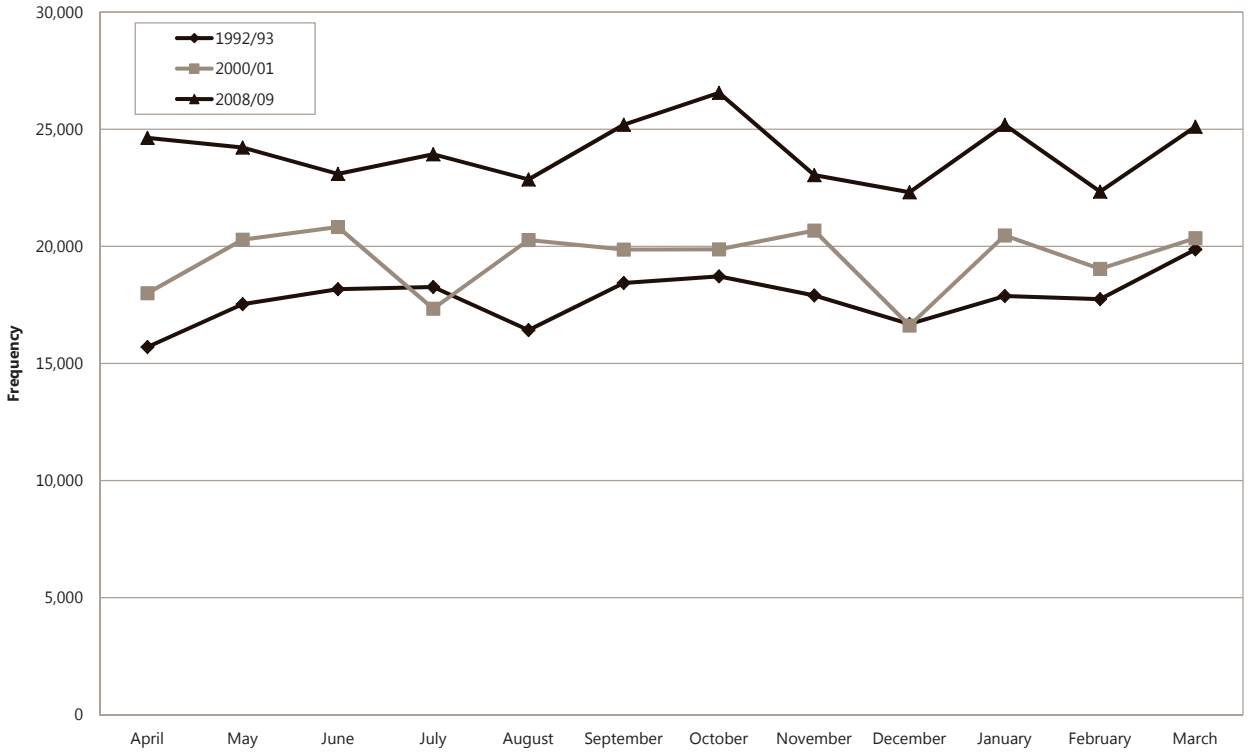
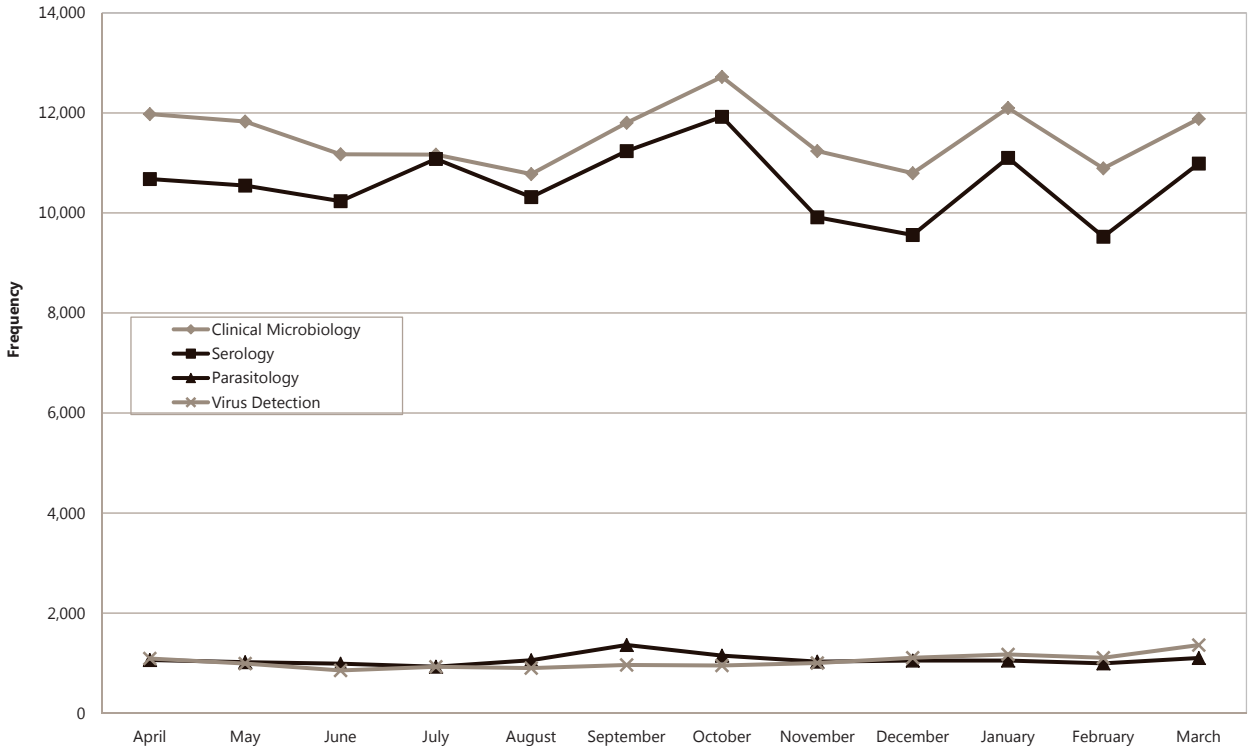


Figure 5.3: Frequency of Requisitions by Month and Section, 2008/09



Requisitions by Referring Provider and Facility

The majority of requisitions were made by general practitioners (GPs) in each fiscal year (Figure 5.4). The number of requisitions from GPs increased by 32.2% between 1992/93 and 2008/09. However, requisitions by specialist physicians showed a larger percentage increase (51.6%) between these two study years. At the same time, the percentage of all requisitions made by GPs remained close to 68.0% annually and the percentage of all requisitions made by specialists was close to 20.0% in each study year. The percentage of all requisitions with a missing referring physician/provider number and therefore could not be linked to the physician resource file varied between 8.6% in 2008/09 and 12.8% in 1996/97. The 'other' category saw an increase in requisitions attributed to midwives; the number of requisitions from this group increased from none in 1992/93 to almost 1,400 in 2008/09. For other referring providers, there was little change over time. The analysis of RHA of the **provider type** revealed that for Winnipeg RHA providers, 69.3% of requisitions were from GPs; while for non–Winnipeg RHAs providers, 88.4% of requisitions were from GPs.

Figure 5.4: Frequency of Requisitions by Provider Type and Fiscal Year

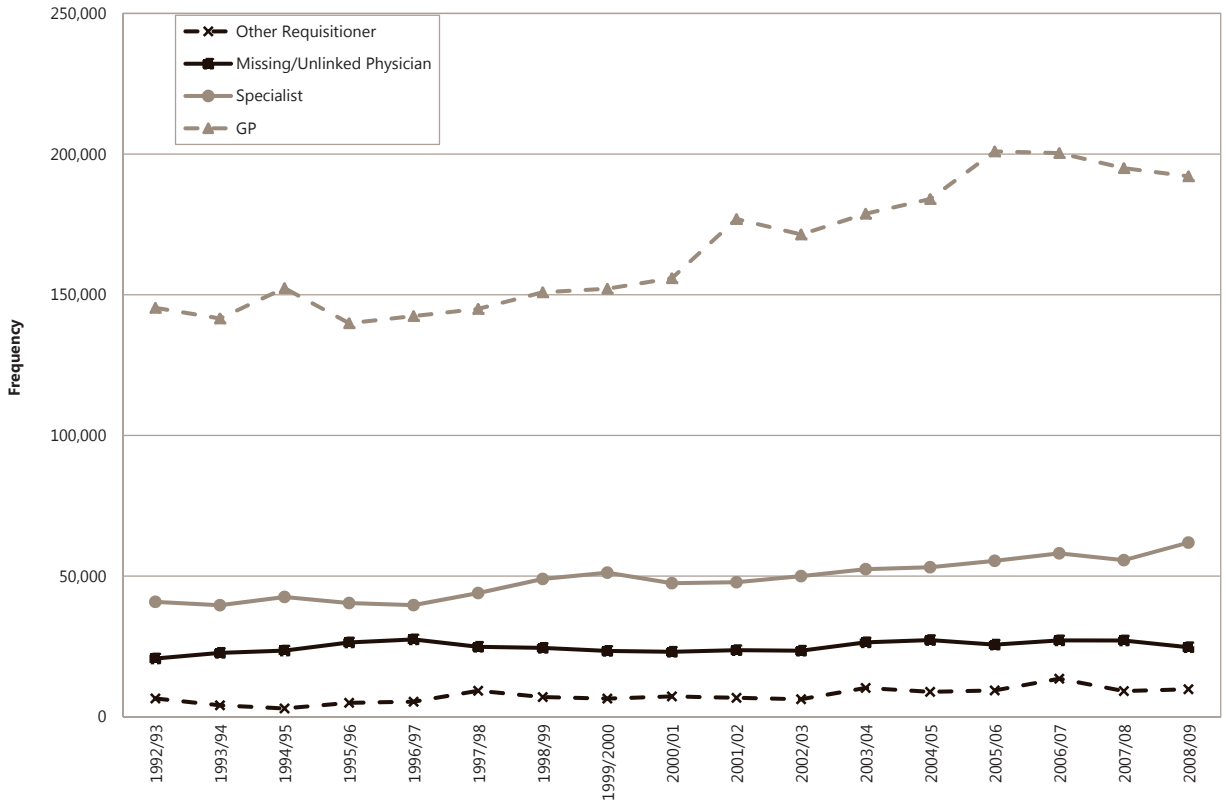


Figure 5.5 reveals that for specialists and other providers, the majority of requisitions were for the serology section. For GPs, half of the requisitions were for the clinical microbiology section. As expected, the vast majority of requisitions (63.1%) were made by referring providers from Winnipeg RHA. Amongst the non–Winnipeg RHA providers, the greatest numbers of requisitions were from Central RHA providers (19.5%) followed by Interlake RHA providers (15.2%).

Figure 5.5: Percent of Requisitions by Referring Provider and Section, 1992/93–2009/10

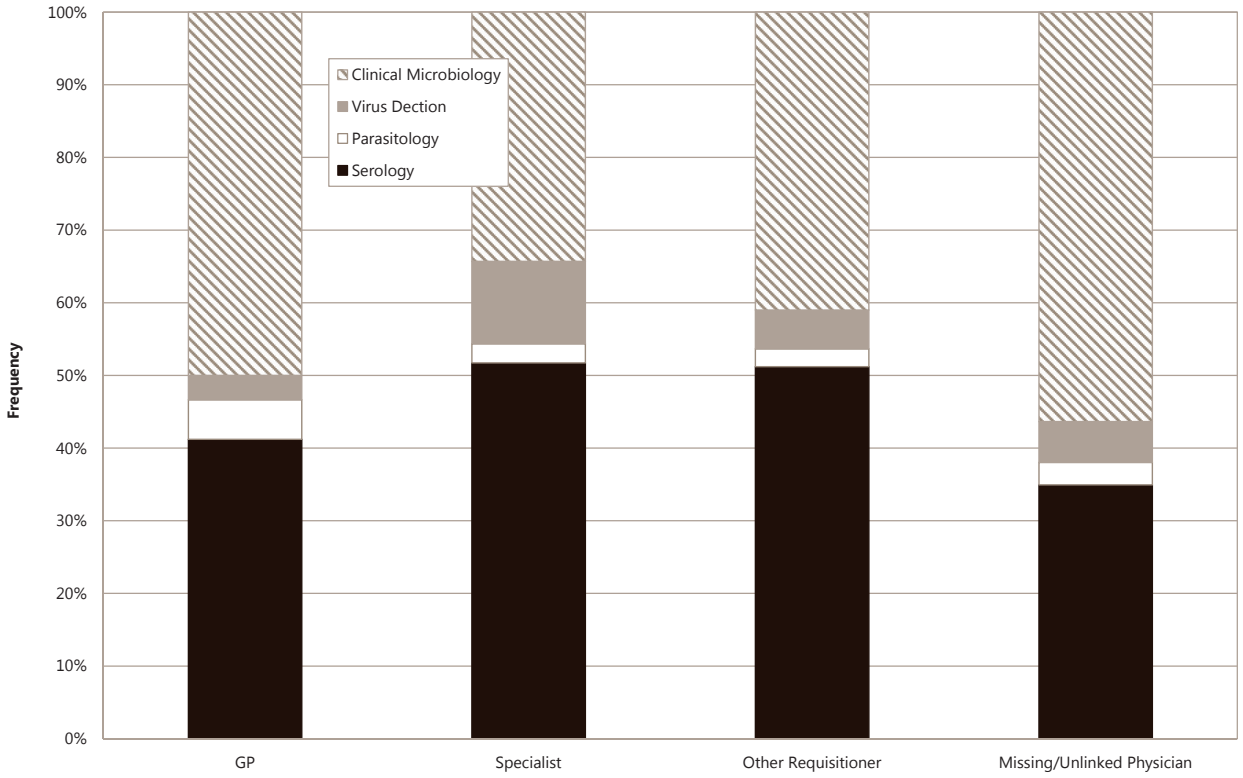


Figure 5.6: Frequency of Requisitions by RHA of Referring Provider, 1992/93–2009/10

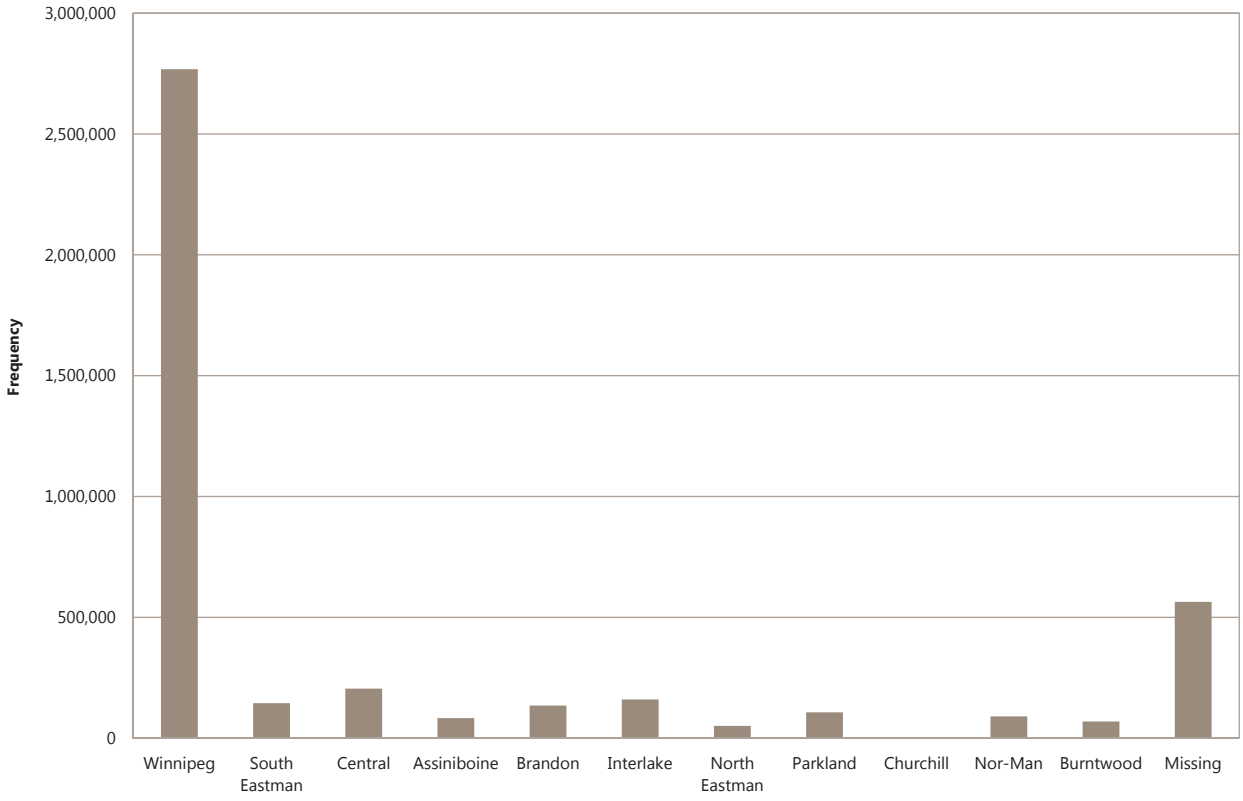


Figure 5.7: Frequency of Requisitions by Referring Facility, 1992/93–2009/10

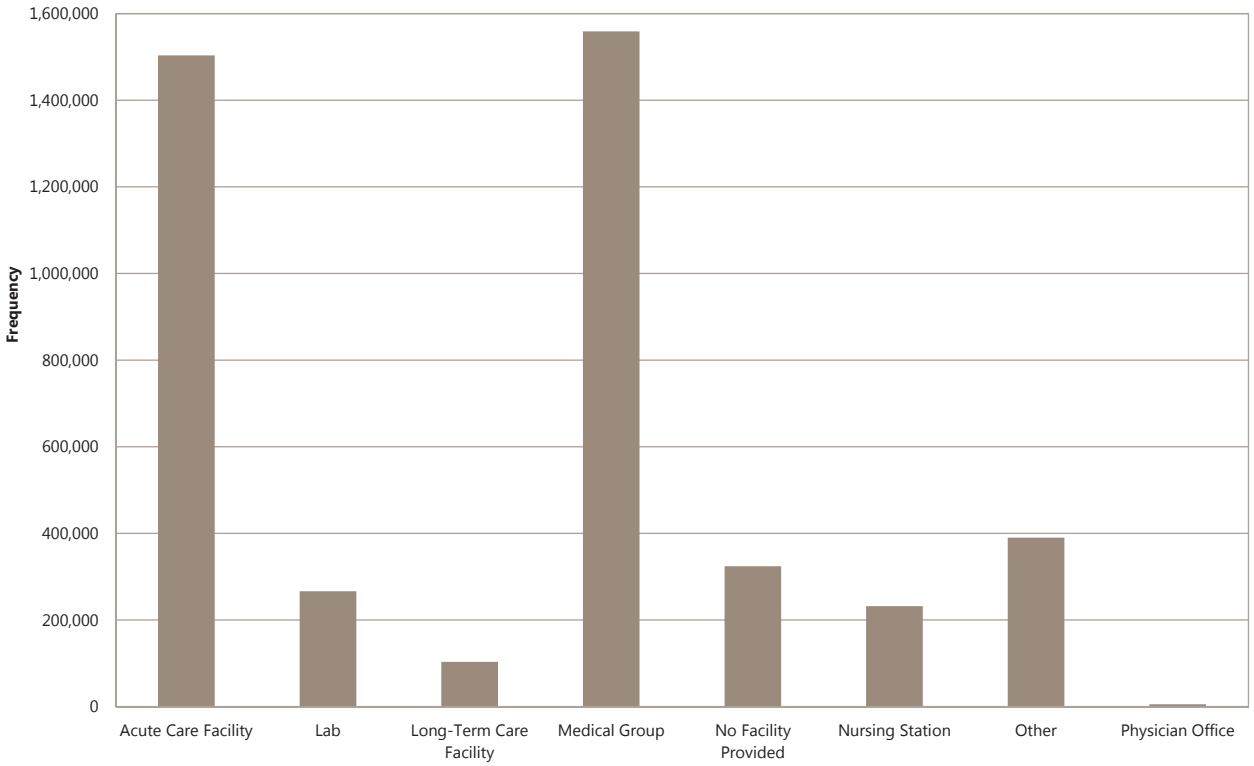
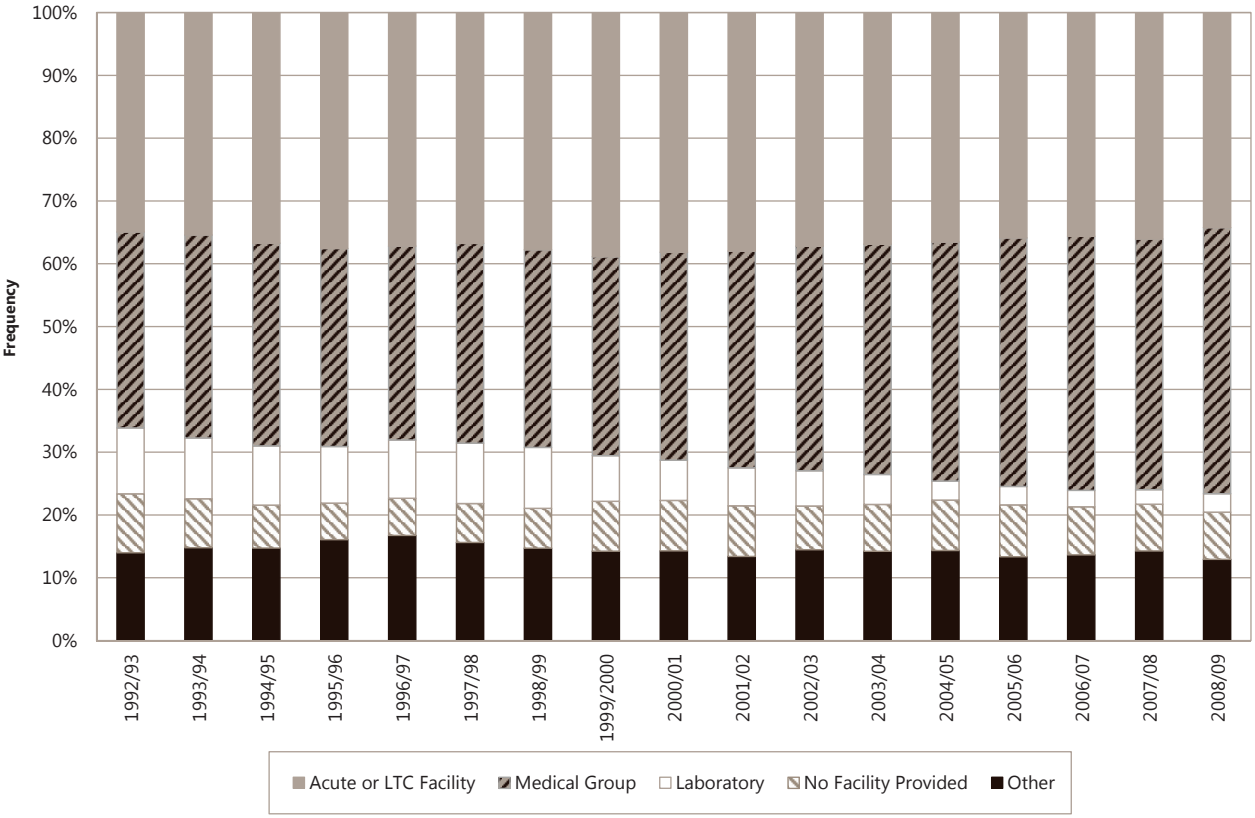


Figure 5.8: Percent of Requisitions by Type of Referring Facility



The analyses by referring facility (Figure 5.7) reveal that the majority of requisitions were from medical groups and acute care facilities. The percentage of all requisitions made from a medical group increased over time (Figure 5.8), from 31.2% in 1992/93 to 42.4% in 2008/09, while the percentage of all requisitions made from an acute care or long–term care (LTC) facility remained relatively constant, at about 33.0%.

The analyses of the RHA of the referring facility revealed that 38.7% of requisitions from acute care facilities were for Winnipeg RHA facilities and 85.0% of requisitions from medical groups were also for Winnipeg RHA facilities. When the requisition was from a lab, the vast majority (92.5%) were from Winnipeg RHA facilities.

Tests in the Manitoba Population

Figure 5.9 reveals that the percentage of Manitoba residents with at least one test in the CPL data files in each year increased from 8.9% in 1992/93 to 11.9% in 2008/09. The anomalous increase in tests in 2001/02 (Figure 5.9) will be examined further in subsequent analyses.

The percentage of Manitoba residents having at least one test in the clinical microbiology section rose from 6.0% to 7.7% between 1992/93 and 2008/09 (Figure 5.10). The corresponding percentages for the serology section were 4.2% and 7.0%, with a large one–year increase occurring in 2001/02. This one–year increase is attributed to an increased number of tests for HIV, Hepatitis B, and Hepatitis C. However, the frequency of tests for Hepatitis B and Hepatitis C remained high in subsequent years. For the virus detection and parasitology sections, the percentages remained largely unchanged over the study period.

In terms of the results by sex (Figure 5.11), testing rates were higher for females than for males in all study years. However, for both sexes there was an increase over time, from 4.8% to 7.2% for males and from 13.0% to 16.5% for females between 1992/93 and 2008/09.

An investigation of the results for children and youth (Figure 5.12) shows that rates were highest for the 10 to 19 years age group at the beginning and end of the study periods. However, there was substantial variability, particularly for newborns and the 1 to 9 years age group in the earliest study years (i.e., 1994/95), which may reflect changes in policies or practices around testing. For the adult population (Figure 5.13), the rates were substantially higher for the 20 to 29 years age group than for other age groups. The rates were similar for the 45 to 64 and 65+ age groups. For the 20 to 29 years age group, the percentage of the population having at least one CPL test increased from 20.2% to 26.3% between the first and last years of the study period. For the oldest age group there was also a small increase, from 5.3% to 7.7%, although the highest value of 11.3% was observed in 2001/02.

Figure 5.14 demonstrates that overall, the rates of testing were similar for urban and rural RHAs. However, there was substantial variation across Manitoba's RHAs (Figures 5.15 and 5.16). The percentage of the population having at least one CPL test increased substantially for residents of Churchill RHA from 13.9% in 1992/93 to 21.9% in 2008/09. For Burntwood RHA, rates increased between 1992/93 and 1998/99, from 18.6% to 21.6%, and then declined slightly to 18.9% in 2008/09. In contrast, the rates for North Eastman were relatively constant and remained around 10.0% for the duration of the study period. For Winnipeg RHA (Figure 5.16), the percentage increased slightly from 9.2% to 12.1%; while for Brandon, Assiniboine, and Parkland RHAs, there was a substantial increase in the percentages around the 2001/02 or 2002/03 fiscal years. Accordingly, the rates from Brandon RHA rose from 4.6% in 1992/93 to 12.1% in 2008/09, those in Assiniboine rose from 4.0% to 7.2%, and those in Parkland rose from 6.4% to 10.0% between the first and last study years.

The percentages of the population having at least one test were also compared for rural and urban income quintiles. Figure 5.17 contains the results for the lowest and highest quintiles in each of these regions. As expected, rates were highest for the lowest urban and rural quintiles and lowest for the highest urban and rural quintiles. The population in the highest urban income quintile consistently had the lowest rates for all study years. However, for all income groups, there was an increase in the rates of testing over time. For example, in the lowest urban income quintile, the percentage increased from 12.4% to 15.7%; and for the highest urban income quintile, the percentage rose from 7.4% to 10.0%.

Figure 5.9: Percent of Manitoba Population Having at Least One CPL Test by Fiscal Year

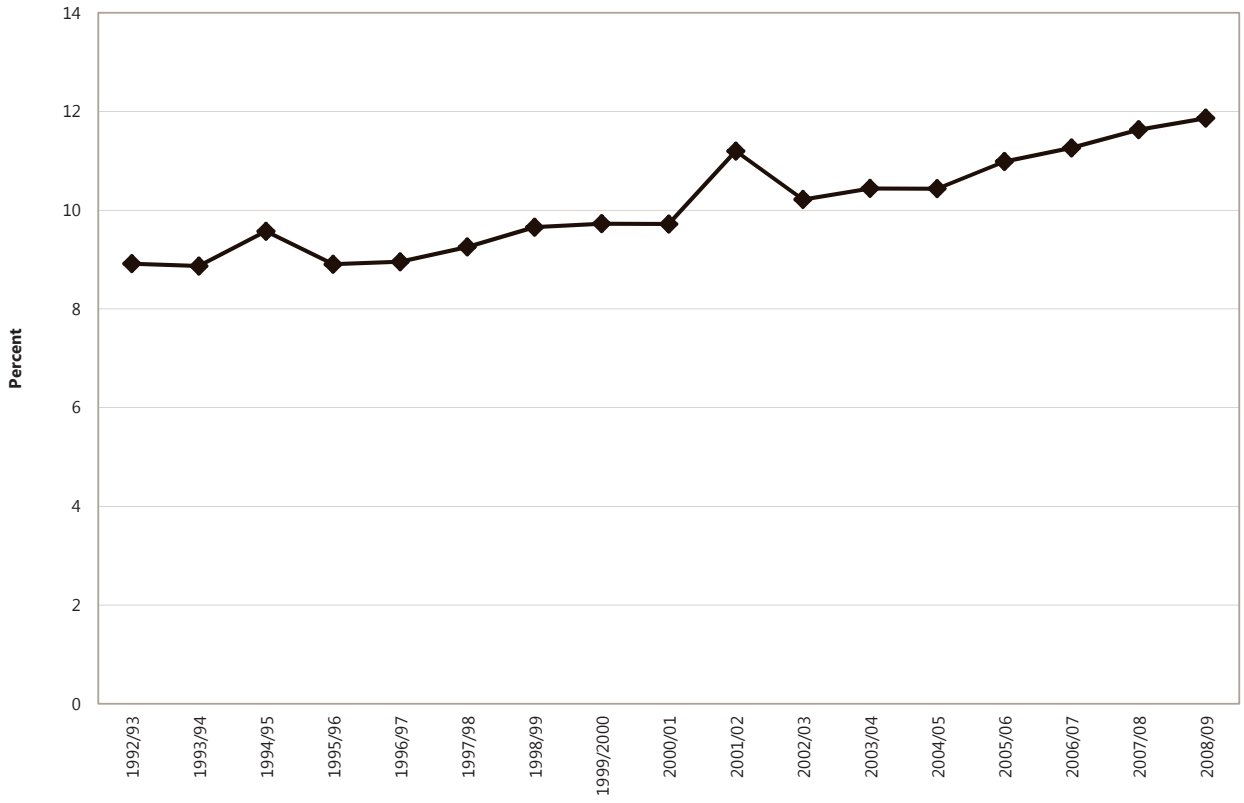


Figure 5.10: Percent of Manitoba Population Having at Least One Test by Section and Fiscal Year

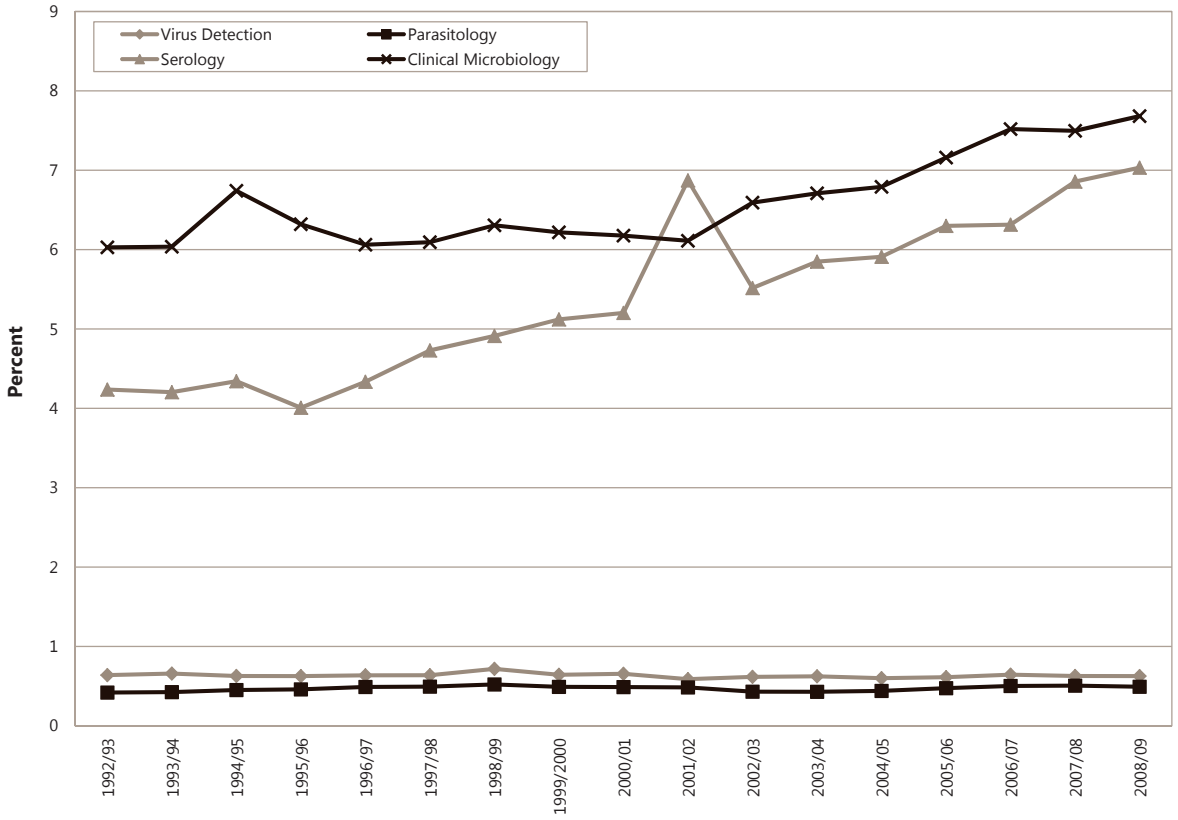


Figure 5.11: Percent of Manitoba Population Having at Least One Test by Sex and Fiscal Year

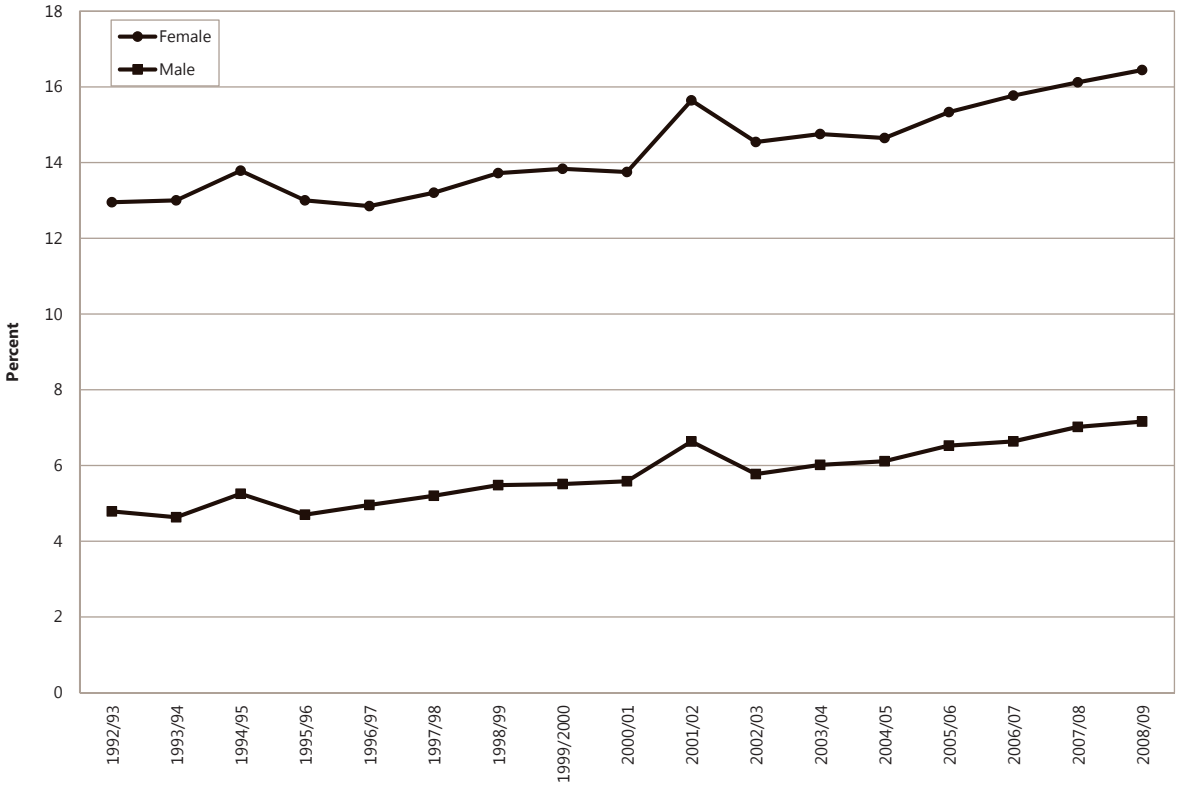


Figure 5.12: Percent of Manitoba Population, 0–19 Years, Having at Least One Test by Age Group and Fiscal Year

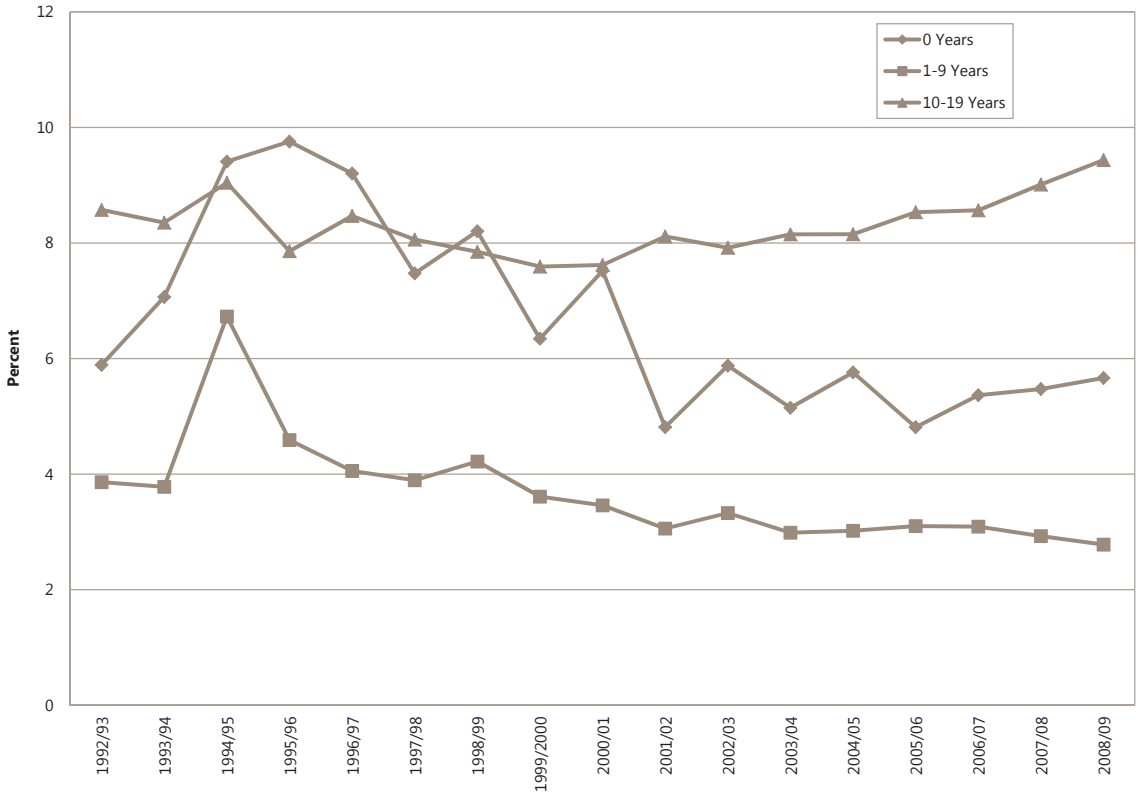


Figure 5.13: Percent of Manitoba Population, 20+ Years, Having at Least One Test by Age Group and Fiscal Year

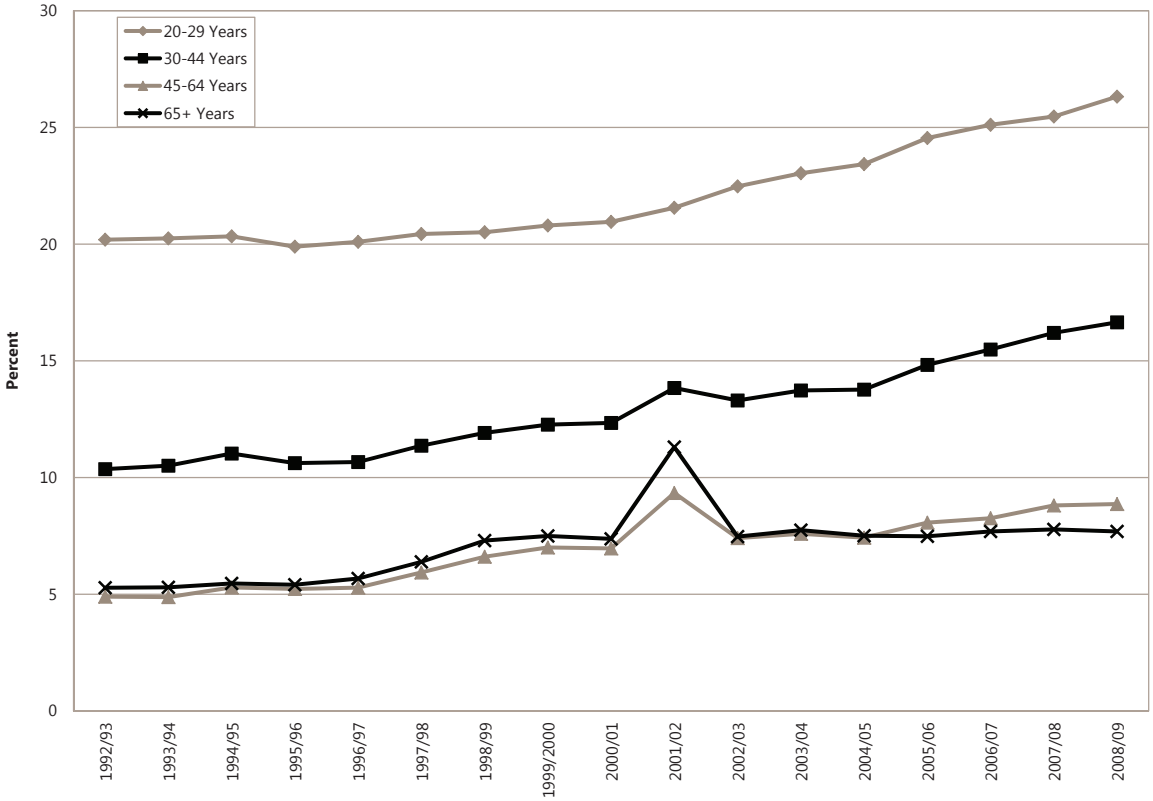


Figure 5.14: Percent of Manitoba Population Having at Least One Test by Urban or Rural Area and Fiscal Year

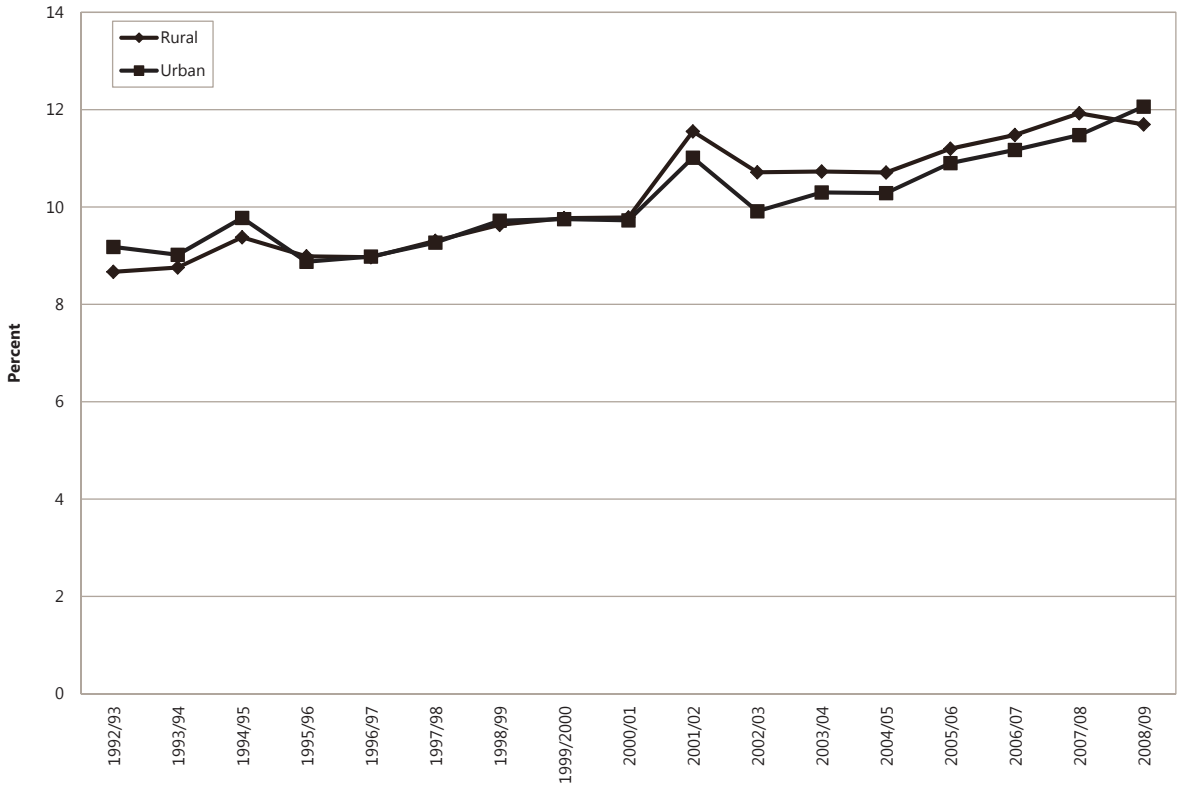


Figure 5.15: Percent of Manitoba Population Having at Least One Test by Selected RHAs and Fiscal Year, Northern RHAs

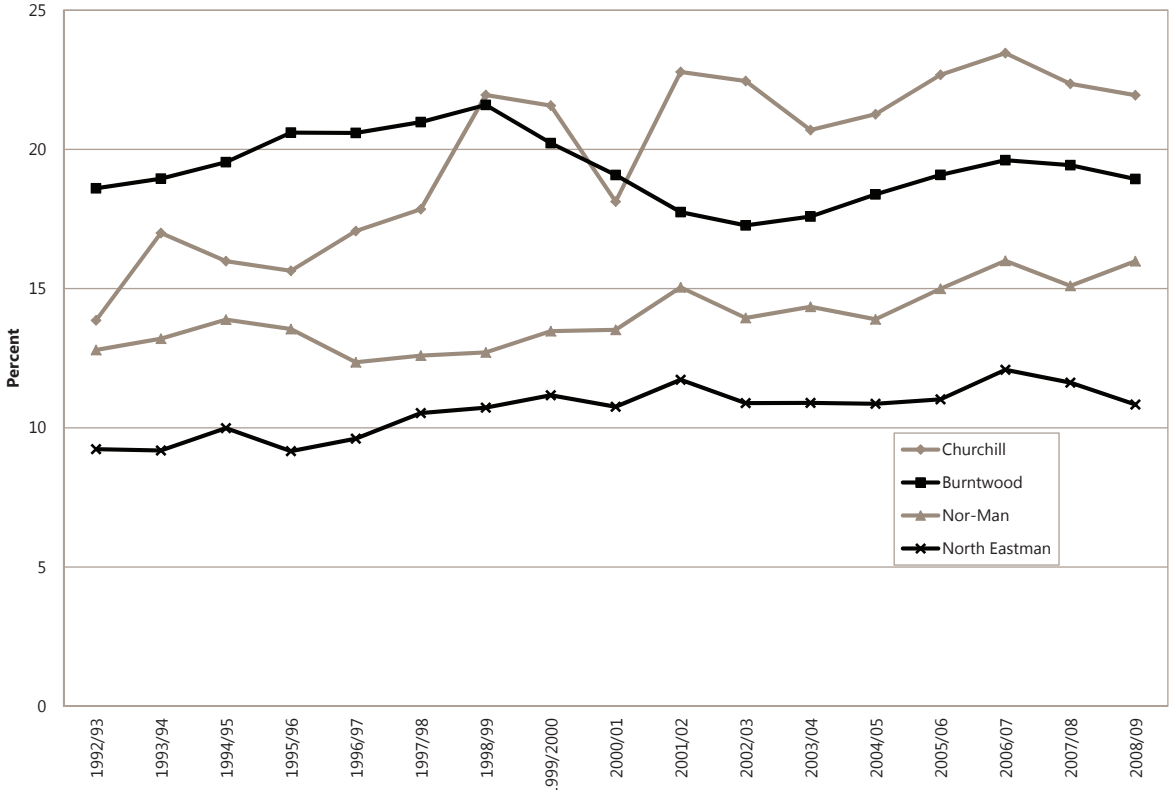


Figure 5.16: Percent of Manitoba Population Having at Least One Test by Selected RHAs and Fiscal Year, Southern RHAs

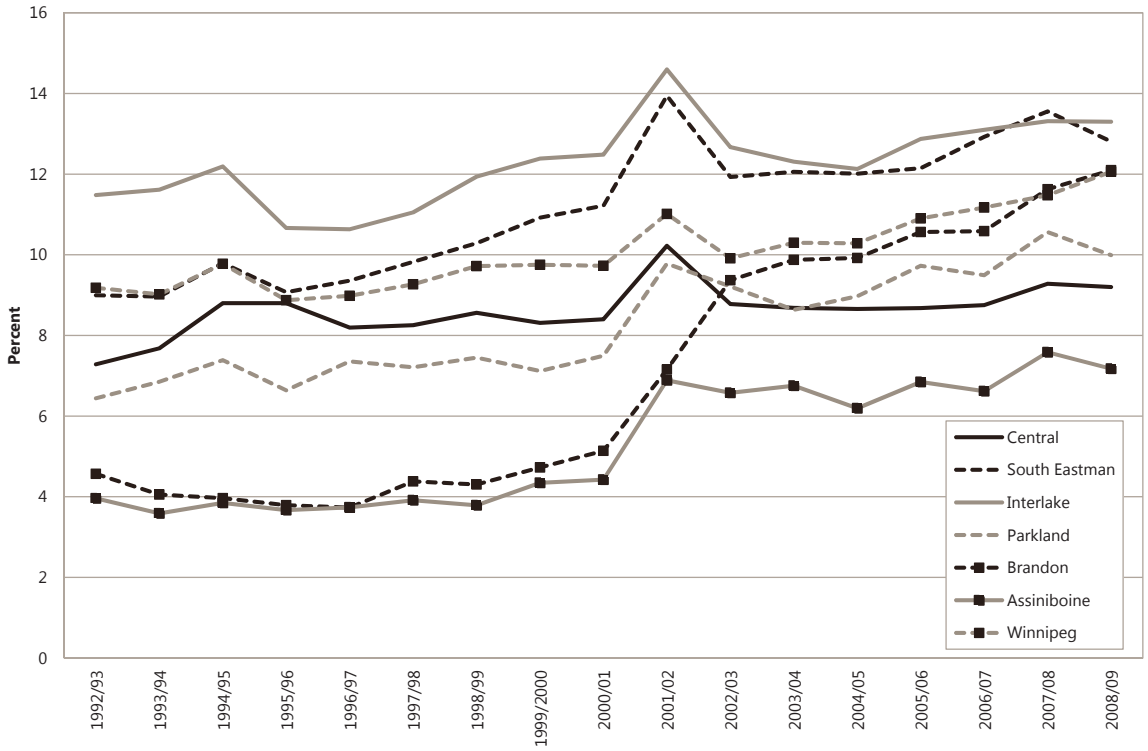
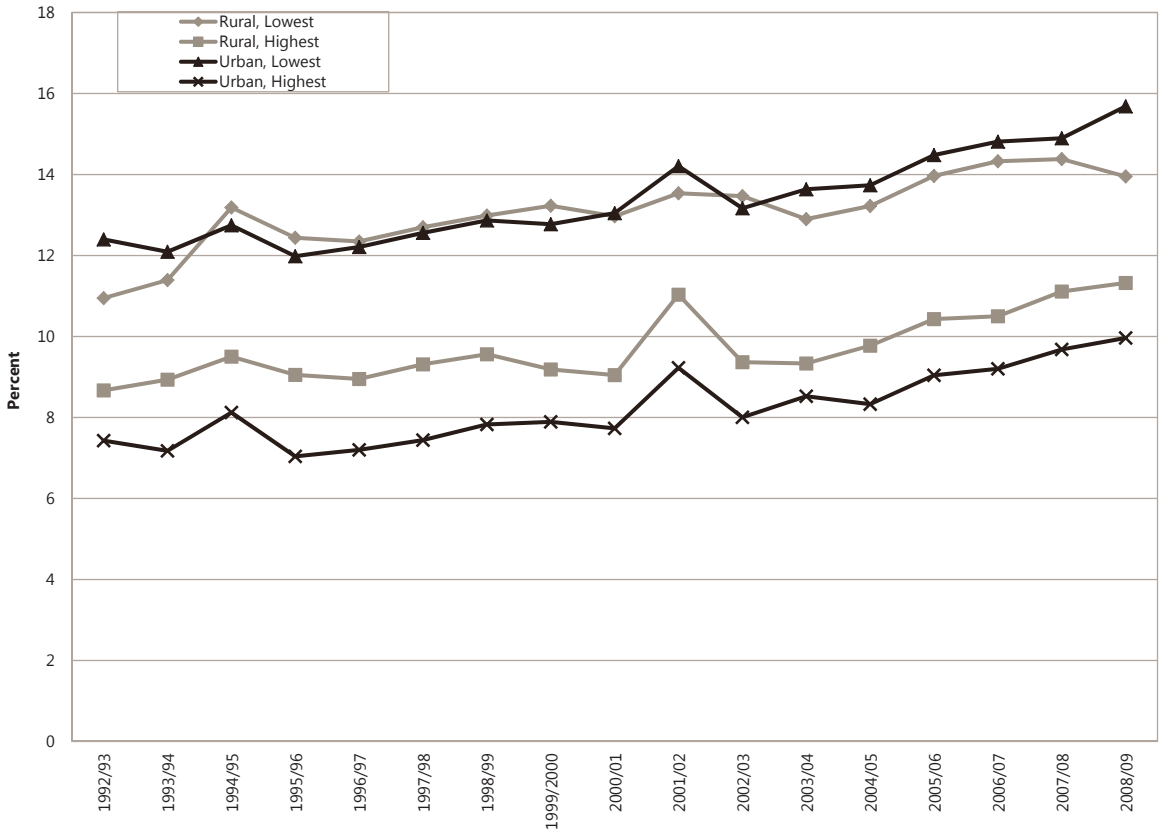


Figure 5.17: Percent of Manitoba Population Having at Least One Test by Income Qunitile and Fiscal Year



Tests in the Manitoba Prenatal Population

A similar analysis to that reported in the previous section, which focused on the entire Manitoba population (including the prenatal population) was conducted just for the prenatal (i.e., pregnant women) population in Manitoba. Routine screening for infectious diseases (hepatitis B, syphilis, and rubella) in prenatal populations is typically done early in pregnancy . We first focused on the percentage of the Manitoba prenatal population having at least one CPL test in any section (Figure 5.18). This percentage was high throughout the study period but increased slightly from 89.4% in 1993/94 to 94.1% in 2008/09. Note that this analysis starts in 1993/94 to give a one–year 'look back' for the prenatal period. Furthermore, as Figure 5.19 reveals, for the three routine screening tests for pregnant women, the percentages of women having at least one of these tests were consistently high across the study period. Finally, given that Hepatitis C incidence is increasing in Canada and children born to women with Hepatitis C are at risk of being infected with the virus, we also investigated tests for this notifiable disease in pregnant women. Testing rates increased, from less than 1% of pregnant women to almost one third (32.7%) of women between 1993/94 and 2008/09 (Figure 5.19).

An examination of testing rates across age groups reveals that testing rates were highest amongst pregnant women under 40 years of age (Figure 5.20). For those between 40 and 49 years, the percentage of women having at least one test was lowest in 1993/94 (78.1%) and highest in 2008/09 (89.7%). The testing percentages were similar for pregnant women living in rural and urban RHAs (Figure 5.21). Subsequent analyses for the individual RHAs revealed less variability than for the general population (Figures 5.22 and 5.23). Percentages of pregnant women having at least one test were above 80.0% in all years and for all RHAs. For example, the average percent of the prenatal population that had at least one rubella test, averaged over the study period, ranged from 82.1% in NOR–MAN RHA to 88.8% in South Eastman. There was also little variation by income quintile (Figure 5.24).

Figure 5.18: Percent of Manitoba Prenatal Population Having at Least One Test in Any Section by Fiscal Year

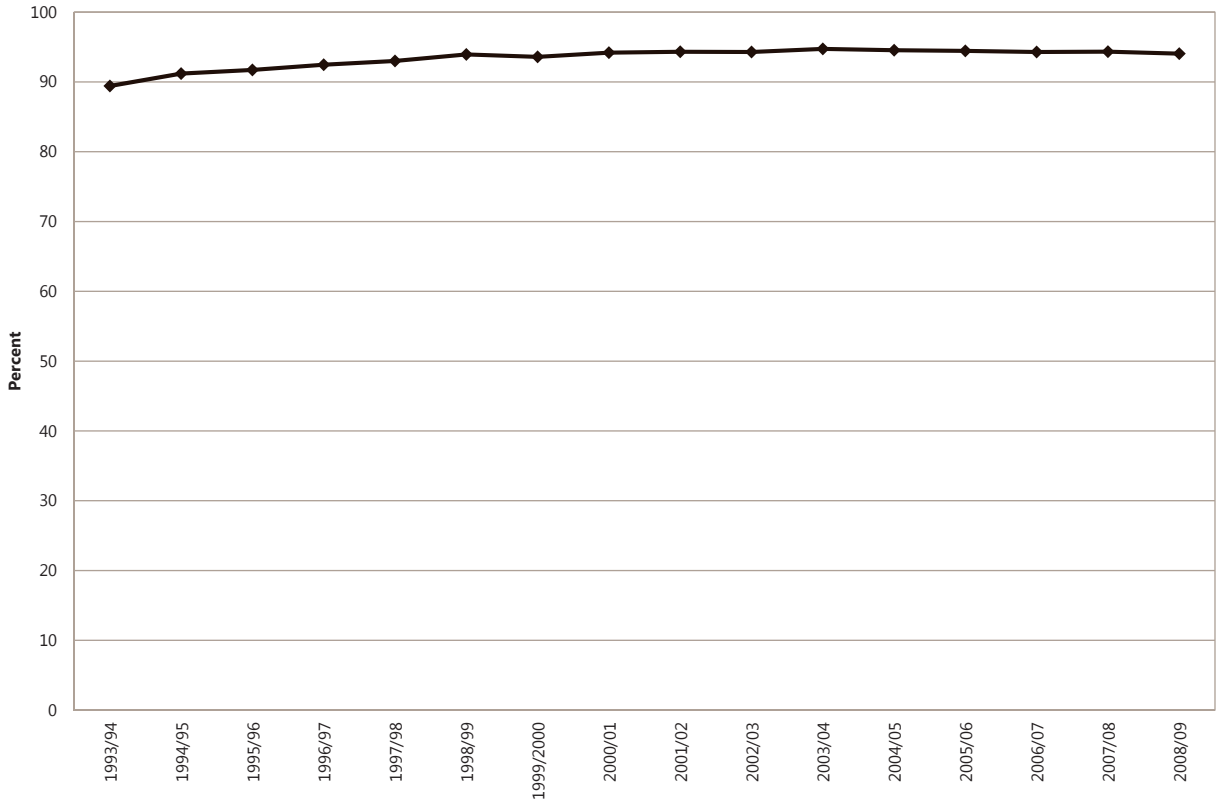


Figure 5.19: Percent of Manitoba Prenatal Population Having at Least One Test by Type of Test and Fiscal Year

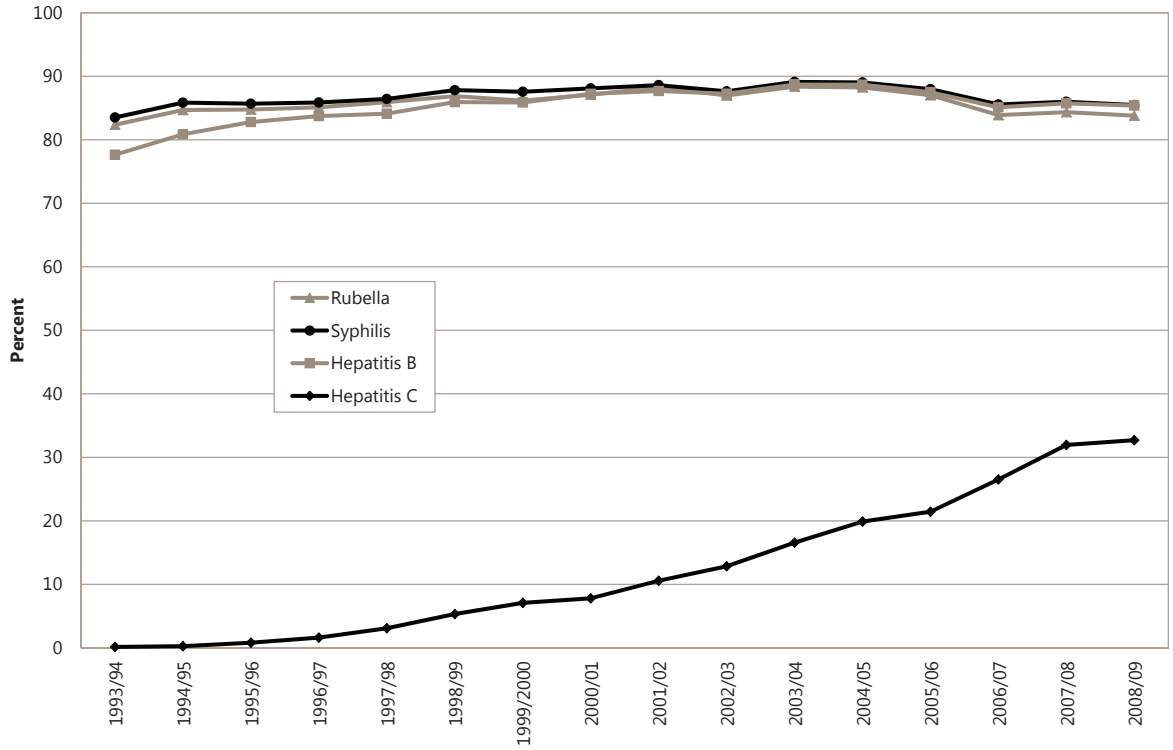


Figure 5.20: Percent of Manitoba Prenatal Population Having at Least One Test in Any Section by Age Group and Fiscal Year

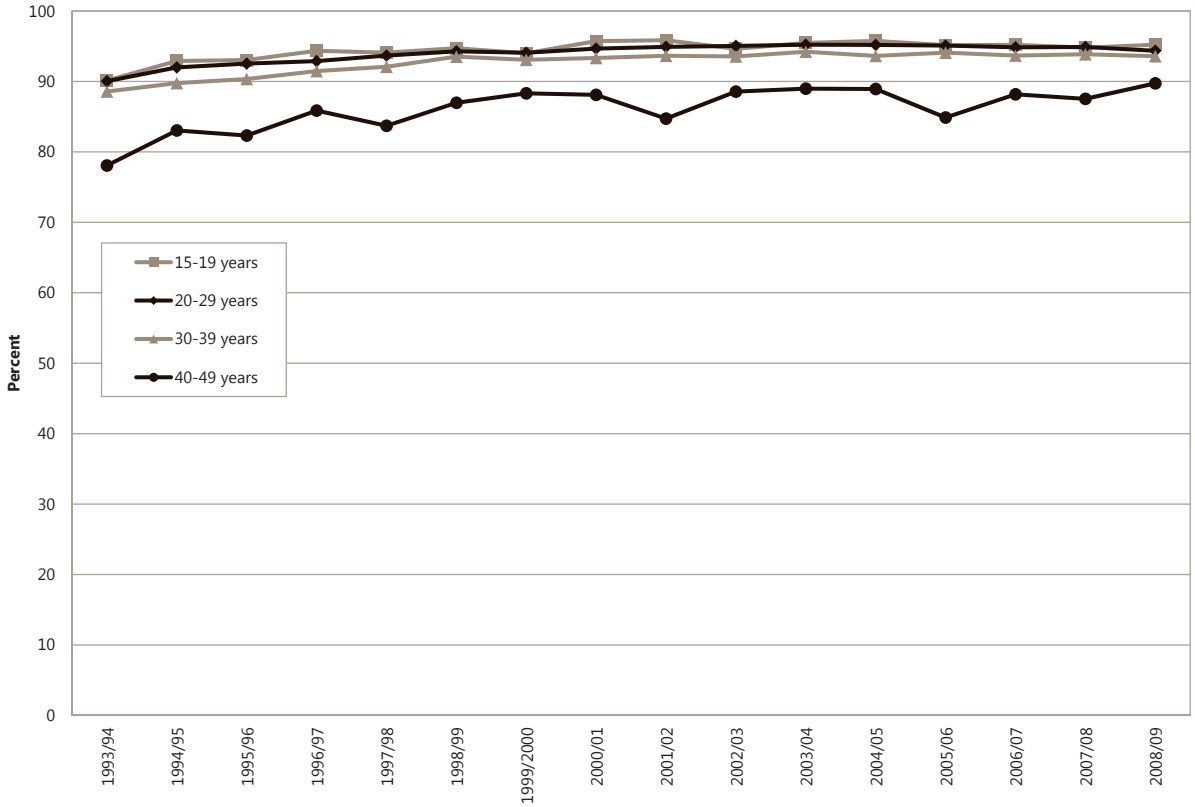


Figure 5.21: Percent of Manitoba Prenatal Population Having at Least One Test in Any Section by Urban/Rural RHA and Fiscal Year

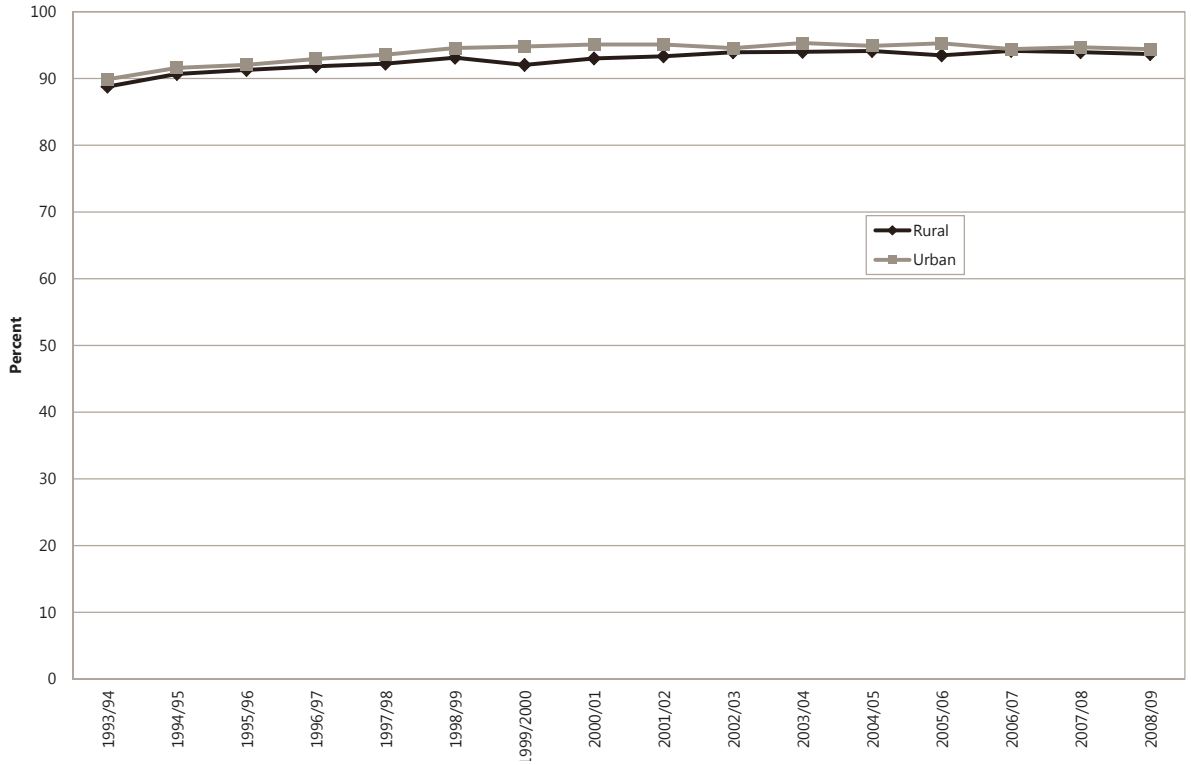


Figure 5.22: Percent of Manitoba Prenatal Population Having at Least One Test in Any Section by Selected RHAs and Fiscal Year, Northern RHAs

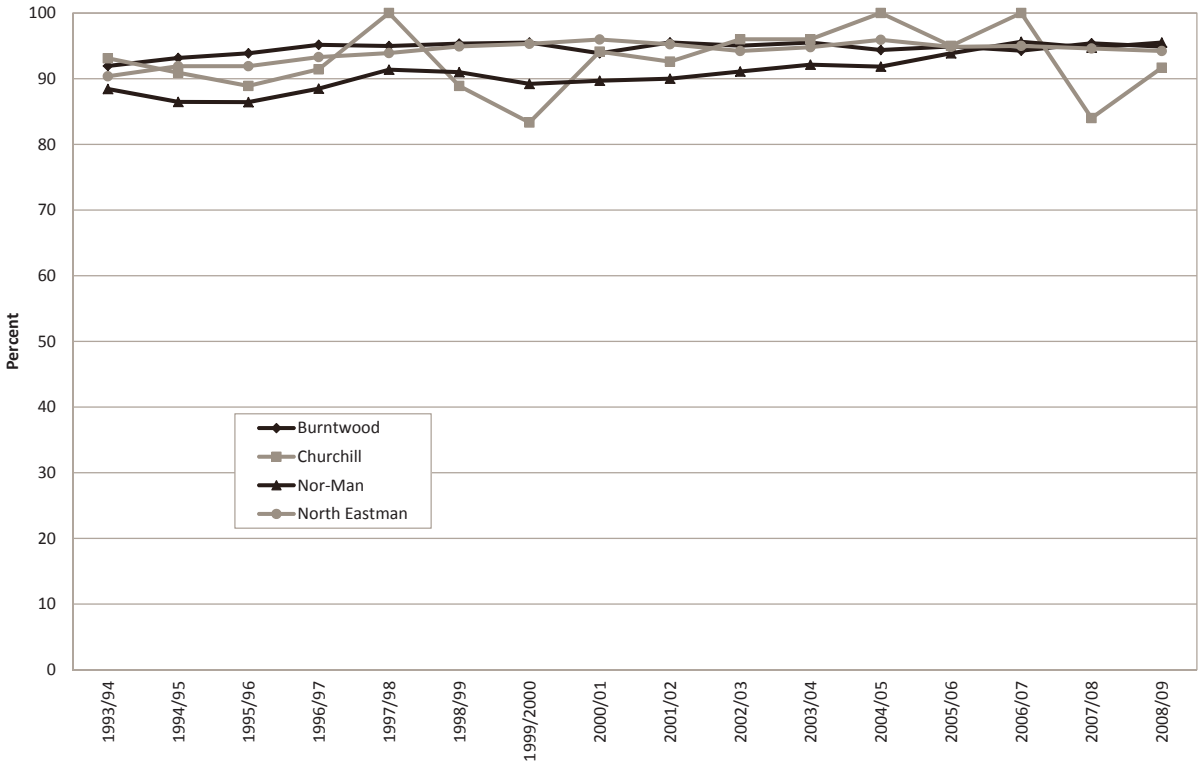


Figure 5.23: Percent of Manitoba Prenatal Population Having at Least One Test in Any Section by Selected RHAs and Fiscal Year, Southern RHAs

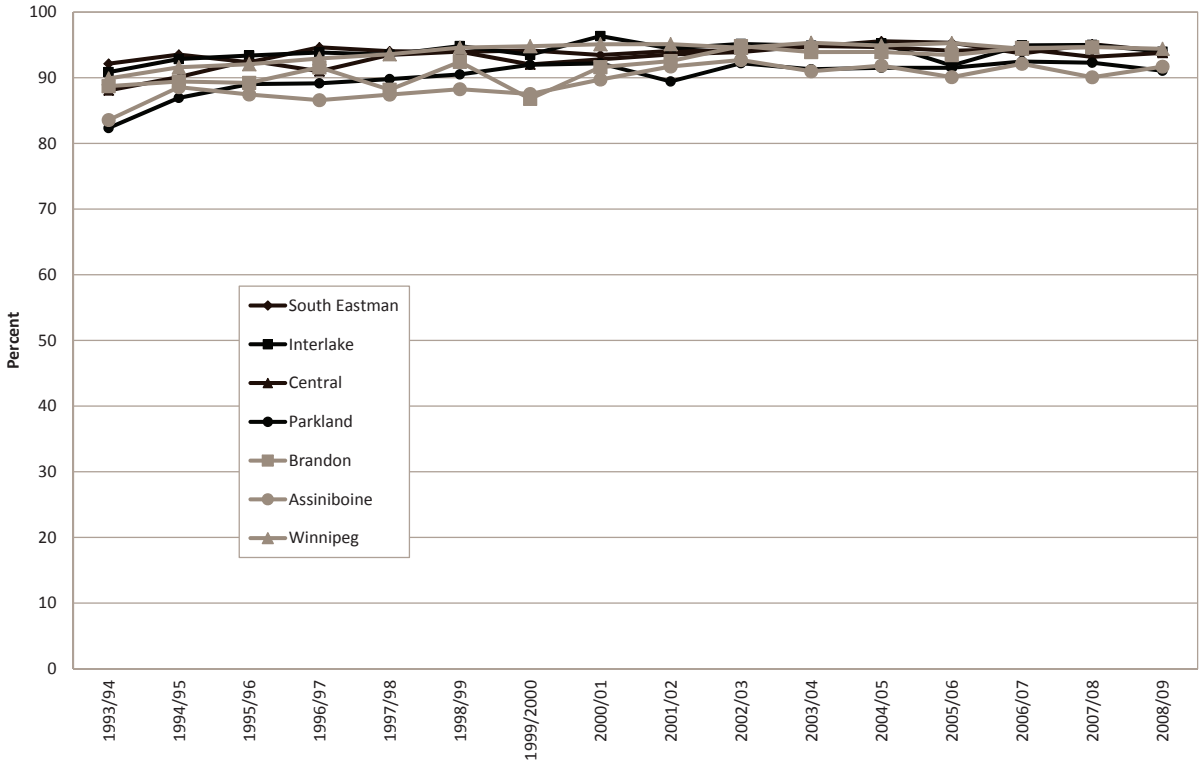
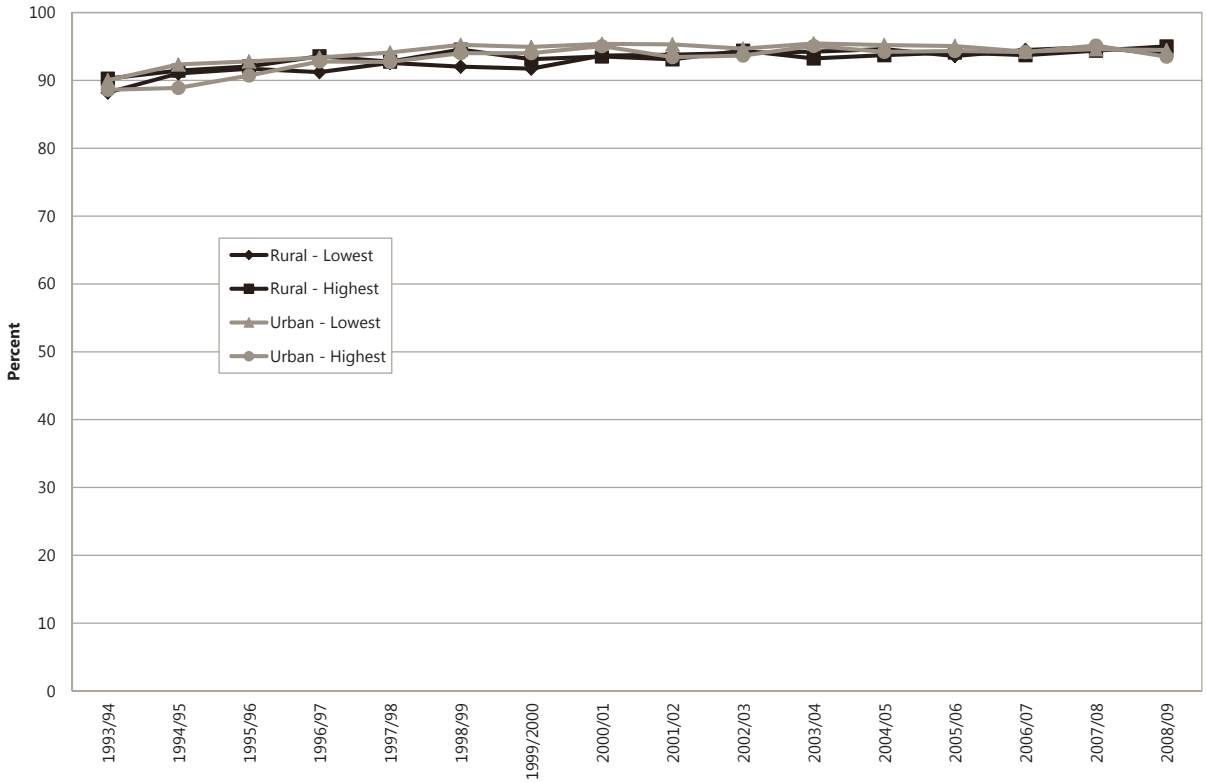


Figure 5.24: Percent of Manitoba Prenatal Population Having at Least One Test in Any Section by Income Quintile and Fiscal Year



Conclusions

The CPL data files contain a wealth of information about requisitions and tests for notifiable diseases in Manitoba. This chapter focused on investigating the quality of the data by characteristics of the requesting provider and facility, the general client population, and the prenatal client population. The results suggest that the data are generally of high quality; there is limited missing information; coverage of the population appears to be high. However, there are some variations in geographic coverage and section coverage that suggest changes in program delivery or data capture over time. Specifically, a large increase in the serology section tests was observed in 2001/02. As well, there were large increases in testing for residents of some of the southern RHAs around 2001/02 and a decrease in testing for residents of Burntwood RHA between 1998/99 and 2005/06.

CHAPTER 6: IDENTIFYING DISEASE POPULATIONS USING LINKED ADMINISTRATIVE HEALTH DATABASES

This chapter examines the role of the CPL data for identifying specific infectious disease populations in Manitoba. This chapter builds on the previous one, which examined overall issues of data quality for tests and requisitions by socio–demographic and regional variables, by focusing on the use of the CPL data for measuring incidence and prevalence of infectious diseases, identifying disease populations for case–control or cohort studies, and investigating healthcare utilization in these populations. Throughout this chapter, linkage of the CPL to other administrative health databases is emphasized.

The CPL data are only one source of information to identify infectious disease populations for surveillance and research in Manitoba. Additional sources include test data from other private and public laboratories such as DSM, diagnoses in hospital records and physician billing claims, and records of vital events. The prescription drug data are another potential source of case ascertainment for infectious diseases treated with specific types of medications. Record linkage can be used to assess the completeness of the CPL data, by comparing the percentage and characteristics of cases identified from different administrative data sources. This information, coupled with database documentation, can be used to identify gaps in the CPL data. Similar types of studies using descriptive methods or capture–recapture models have been undertaken to investigate the completeness of population–based data sources, including laboratory testing data, for infectious disease research surveillance (Doyle, Glynn, & Groseclose, 2002; Christensen et al., 2012).

This chapter is organized as a series of case studies. The first study focuses on the linkage of records for HIV tests to records for other administrative databases in the Repository. The second case study investigates TB tests in the CPL data and their concordance with diagnoses for TB in hospital abstracts and physician billing claims. The third case study examines the completeness of the CPL data for identifying cases of sexually transmitted infections (STIs). We focus on these infectious diseases primarily because of their importance and relevance for the Manitoba healthcare system (e.g., Orr, 2011), but also because they highlight the strengths and limitations of the CPL for infectious disease case ascertainment.

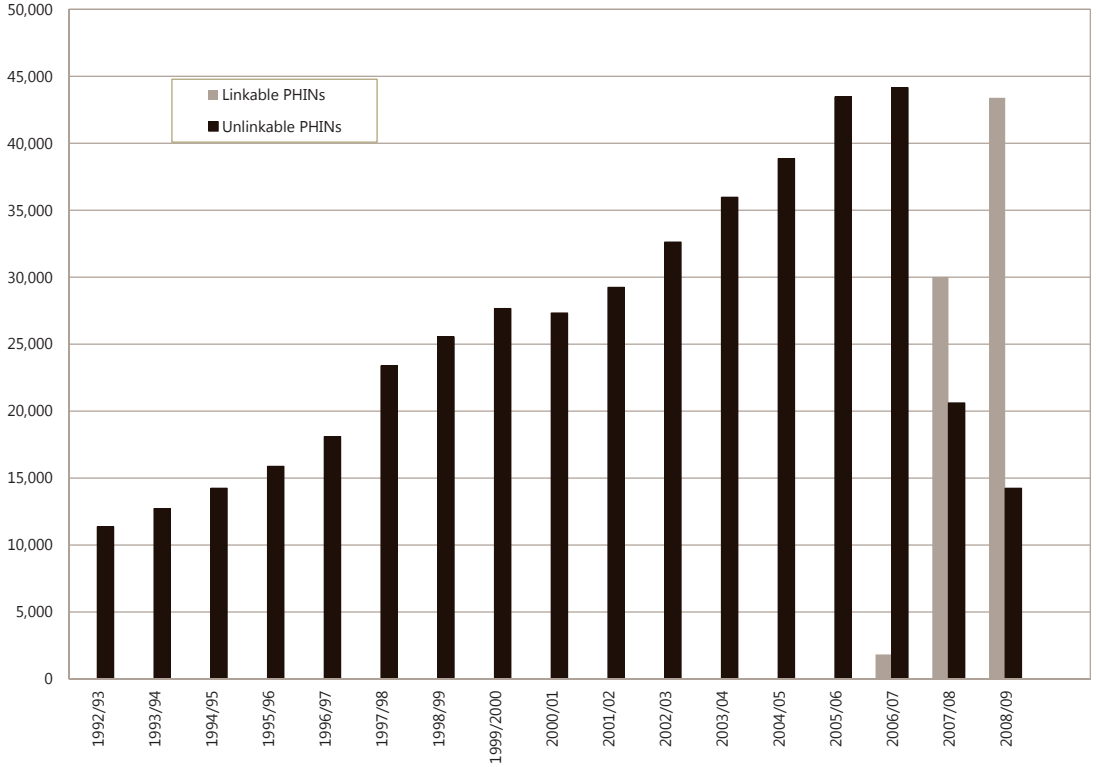
Case Study #1: Record Linkage for HIV Tests

In this first case study, we examine the frequency of records for HIV tests with a linkable PHIN; that is, a PHIN that can be used to link the CPL data to other administrative health databases in the Repository. The frequency of tests with either a positive or negative result was compiled by fiscal year. Frequencies of tests that did and did not contain a linkable PHIN were calculated. The identification of HIV tests in the CPL data files was conducted in consultation with CPL staff.

As well, we examined the agreement between positive and negative HIV tests in the CPL data and diagnoses for HIV in hospital separation records and physician billing claims. In the two–year period from fiscal year 2007/08 to 2008/09, we identified all incident HIV cases ascertained from diagnosis codes (ICD–9–CM 042; ICD–10–CA: B20–B24) in these two data sources. Incident cases were defined as those individuals who did not have another HIV diagnosis in the five–year period prior to their HIV index date (i.e., the date of diagnosis in the period from 2007/08 to 2008/09). Then we examined 30–, 90–, and 180–day observation windows from the HIV index date to ascertain the frequency of agreement between diagnosis–confirmed cases and laboratory–confirmed cases.

As Figure 6.1 reveals, prior to 2006/07, none of the tests in the CPL data contained a linkable PHIN. In 2006/07, a small number of tests had a linkable PHIN. However, this number increased sharply in 2007/08 and continued to rise to the end of the observation period. Accordingly, studies that require record linkage to study the health and healthcare use of the HIV population are not feasible prior to 2007/08 in Manitoba.

Figure 6.1: Frequency of HIV Tests by Type of PHIN



A total of 203 new (i.e., incident) diagnosed cases of HIV were identified from hospital records and physician billing claims in the two–year period from 2007/08 to 2008/09. Among these cases, 33.0% had a positive HIV test within 30 days before or after the date of diagnosis. Using a 180–day period before or after the date of diagnosis, 45.8% of diagnosed cases of HIV had a positive HIV test in the CPL data. Thus, these data indicate that only a moderate association exists between diagnosis–confirmed HIV cases and laboratory–confirmed HIV cases.

Case Study #2: Tuberculosis Tests in CPL Data and Tuberculosis Diagnoses in Hospital and Physician Billing Records

We began the second case study by identifying the total number of TB tests, number of individuals with at least one TB test, and number of positive tests (i.e., laboratory–confirmed cases) of TB in the CPL data. As Figure 6.2 reveals, tests were conducted in all years from 1992/93 to 2000/01 but not in subsequent years. Documentation provided by CPL for this study revealed that in these subsequent years, all TB testing was conducted by DSM; while prior to 2001/02, about half of all TB tests were conducted by CPL.

The annual number of positive tests ranged from 17 in 1992/93 to 52 in 1999/00. Overall, in the period from 1992/93 to 2001/02 a total of 283 positive TB tests were identified in the CPL data amongst 211 individuals.

Figure 6.2: Frequency of TB Tests by Fiscal Year

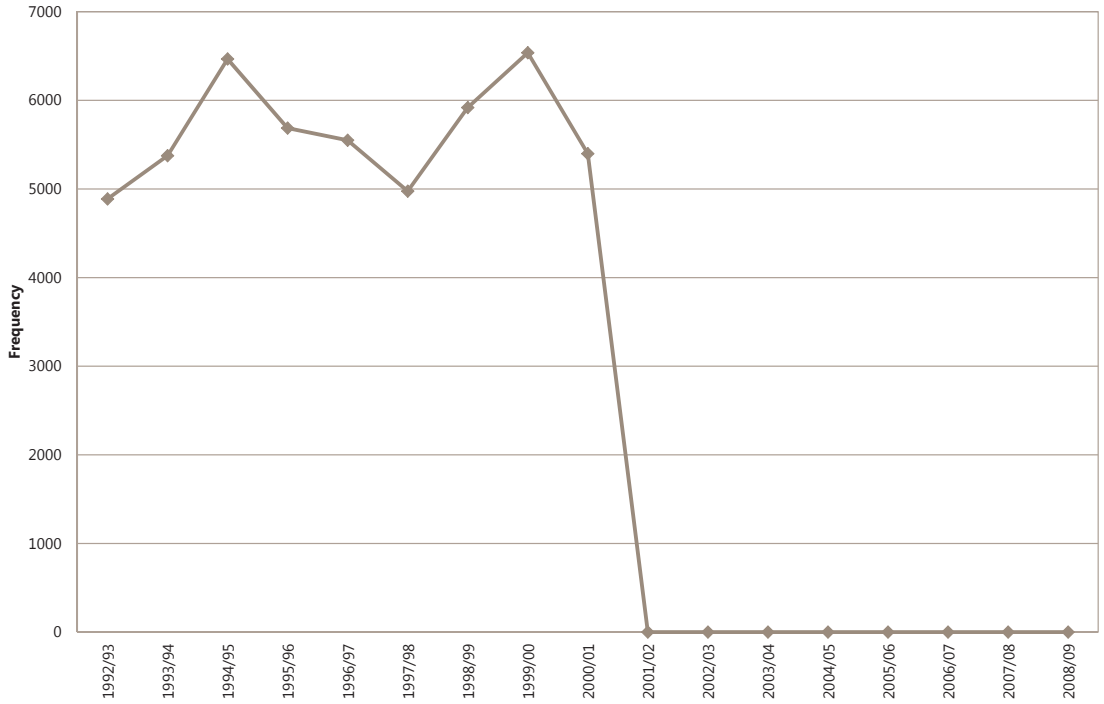
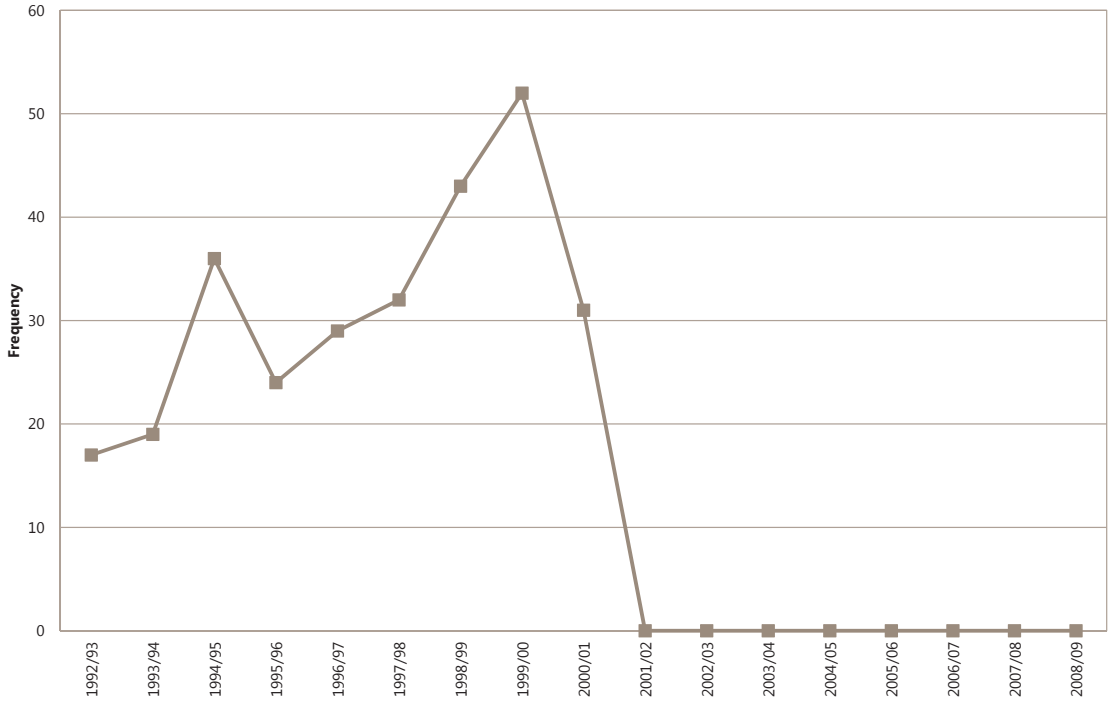


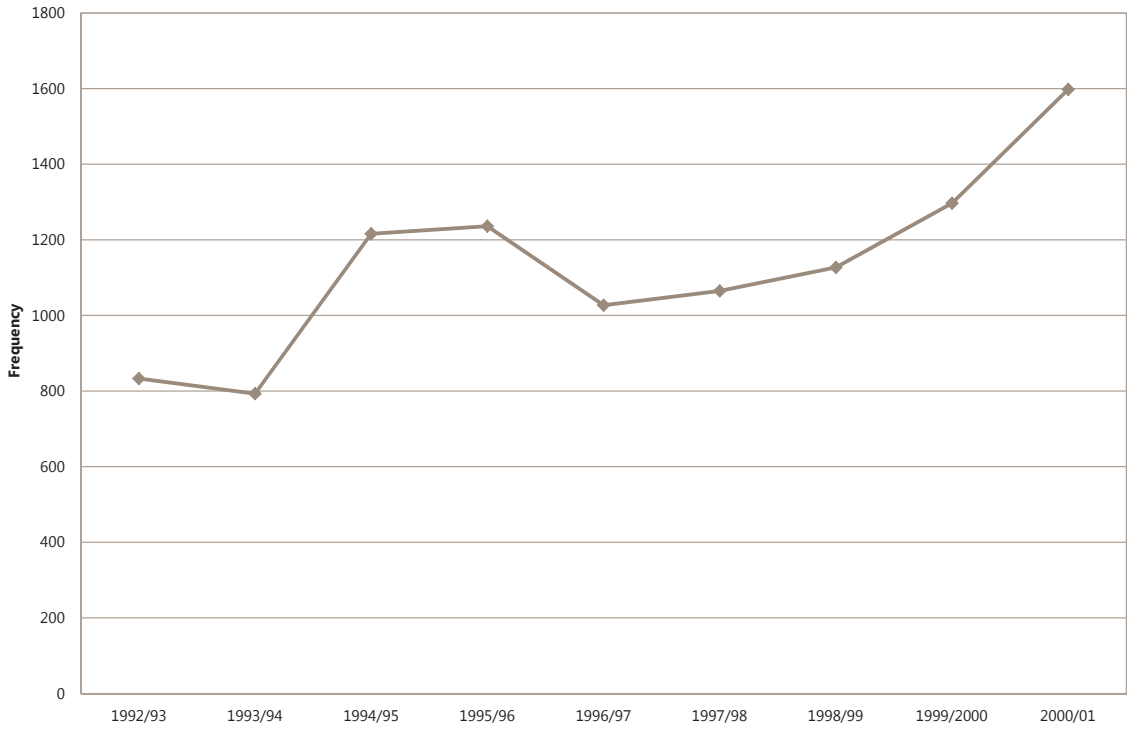
Figure 6.3: Frequency of Positive TB Tests by Fiscal Year



The frequency of individuals with a TB diagnosis in hospital separations or physician billing claims is reported in Figure 6.4 and ranged from 793 individuals in 1993/94 to 1598 individuals in 2000/01. When we restricted our analysis to new (i.e., incident) TB cases, the frequency was still substantially higher than in the CPL data. For example, in 2000/01, there were 946 incident TB cases identified; these were identified using a five–year washout period, meaning that a TB case was identified as incident if there was no other TB diagnosis in the five–year period prior to the diagnosis date in 2000/01.

Given these differences between the frequencies of cases with a positive TB test from the CPL data and a diagnosis in hospital records or physician billing claims, the agreement between the two data sources was further examined. We first established the index date for each individual with at least one positive TB test in the CPL data. The index date was the date that a positive TB test was first recorded for that individual. Then the hospital records and physician billing claims databases for each indivdiual were searched for a TB diagnosis both before and after the index date.

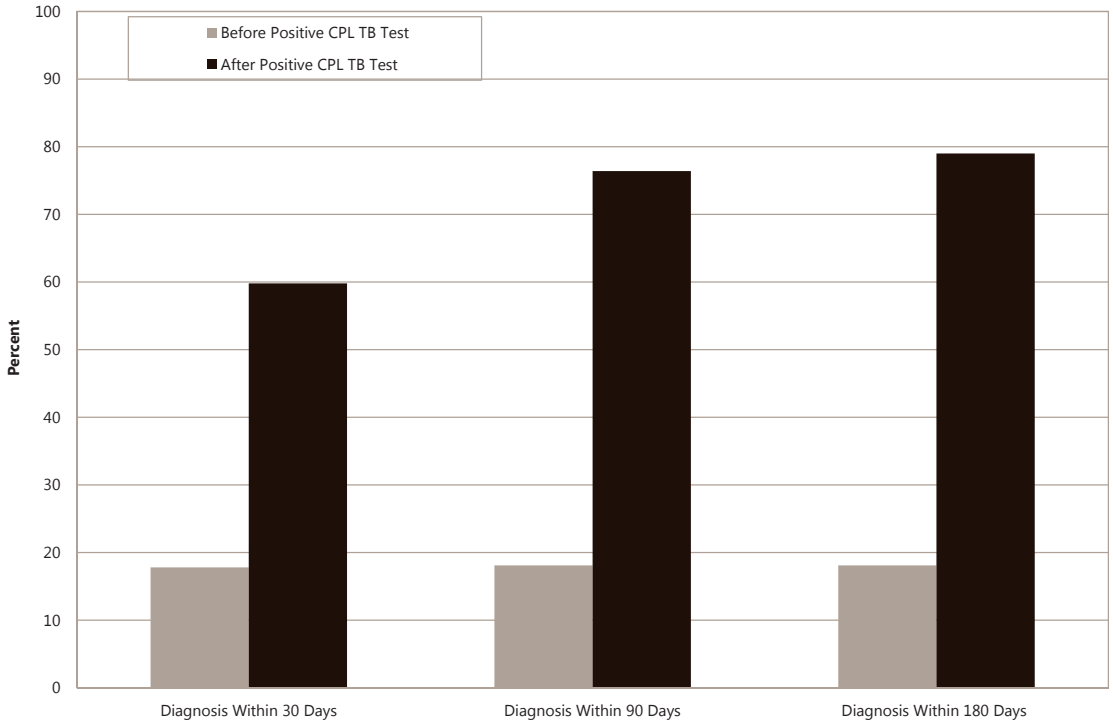
Figure 6.4: Frequency of Individuals with a TB Diagnosis in Hospital Data or Physician Billing Records



As Figure 6.5 reveals, in the 30 days after the index date for a positive TB test in the CPL data, more than half (59.8%) of individuals had a diagnosis for TB recorded in hospital records or physician billing claims data. This percentage increased to 79.0% if a 180–day period after the index date for a positive TB test was examined. For 30–, 90–, and 180–day windows prior to the positive TB test index date, the percentage of individuals with a diagnosis in hospital records or physician billing claims data remained constant, at approximately 18.0%.

To further investigate the relationship between the CPL data and hospital and physician administrative health databases for TB, we identified a cohort of individuals who had a TB diagnosis between April 1, 1998 and March 31, 2001 (i.e., a diagnosis in fiscal years 1998/99 to 2000/01) in hospital records or physician billing claims. This TB cohort was stratified into: (a) incident TB cases, who did not have a prior diagnosis of TB in either hospital records or

Figure 6.5: Agreement Between a Positive TB Test and Diagnosis in Hospital Data or Physician Billing Records



physician billing claims during the period from April 1, 1984 to March 31, 1998 and (b) non–incident TB cases, who did have a prior diagnosis of TB in either hospital or physician claims data in this observation period. A total of 2,320 individuals were identified in the incident cohort and 428 individuals were identified in the non–incident cohort.

For each of the sub–groups within the cohort, the frequency of CPL tests, by section and overall, was recorded. This investigation was conducted using the index date of the TB diagnosis as the reference date. In this analysis, the index date was the date that a TB diagnosis was recorded in hospital or physician billing claims data during the cohort definition period.

The percentage of cohort members with at least one CPL test by section and for any section is reported in Table 6.1. The results demonstrate that a higher percentage of incident TB cohort members than non–incident cohort members had tests in each CPL section. This finding was observed regardless of the size of the observation window before or after the index date. However, the results were similar to those observed for the general population, in that most of the tests were conducted for the clinical microbiology section and few tests were conducted for the parasitology and virus detection sections.

In conclusion, this case study demonstrates that it is important to recognize that the CPL data files do not contain population–based tests for all infectious diseases. This is the case for TB, where about only half of all tests were performed by CPL prior to 2001/02 and all tests were performed by DSM beginning in this year. There is a moderate concordance between positive test results and diagnoses in hospital records and physician billing claims, which suggests that sole reliance on either source is not likely to provide a complete picture of the total number of cases in the population. However, individuals who have an incident diagnosis in hospital records or physician billing claims are likely to have higher rates of tests in CPL data, suggesting that they may be heavier users of public health services than individuals who were previously diagnosed with TB.

Table 6.1: Percentage of Diagnosed TB Cases Having a CPL Test by Section

Clinical Microbiology Section (Results)	Incident TB Cohort (N=2,320)	Non-Incident TB Cohort (N=428)
Within 30 days prior to diagnosis index date	6.6	1.6
Within 90 days prior to diagnosis index date	10.5	6.5
Within 180 days prior to diagnosis index date	14.1	11.2
Within 30 days after to diagnosis index date	5.3	2.6
Within 90 days after to diagnosis index date	8.2	4.7
Within 180 days after to diagnosis index date	11.5	8.2
Parasitology Section		
Within 30 days prior to diagnosis index date	s	s
Within 90 days prior to diagnosis index date	0.6	s
Within 180 days prior to diagnosis index date	1.2	s
Within 30 days after to diagnosis index date	0.4	s
Within 90 days after to diagnosis index date	0.7	s
Within 180 days after to diagnosis index date	1.0	s
Serelogy Section		
Within 30 days prior to diagnosis index date	6.4	1.6
Within 90 days prior to diagnosis index date	11.5	5.6
Within 180 days prior to diagnosis index date	15.0	9.1
Within 30 days after to diagnosis index date	6.7	4.4
Within 90 days after to diagnosis index date	9.8	5.8
Within 180 days after to diagnosis index date	12.8	7.7
Virus Detection Section		
Within 30 days prior to diagnosis index date	1.1	0.0
Within 90 days prior to diagnosis index date	1.8	s
Within 180 days prior to diagnosis index date	2.5	s
Within 30 days after to diagnosis index date	2.0	s
Within 90 days after to diagnosis index date	2.8	s
Within 180 days after to diagnosis index date	3.0	1.6
Any Section		
Within 30 days prior to diagnosis index date	12.5	3.3
Within 90 days prior to diagnosis index date	20.8	11.2
Within 180 days prior to diagnosis index date	26.5	17.5
Within 30 days after to diagnosis index date	12.6	6.8
Within 90 days after to diagnosis index date	17.8	10.1
Within 180 days after to diagnosis index date	23.2	14.7

Note: 's' means that a result has been supressed due to small numbers

Case Study #3: Sexually Transmitted Infection Tests in CPL Data and Sexually Transmitted Infection Diagnoses in Hospital and Physician Billing Records

This last case study examined the CPL data and hospital records and physician billing claims for chlamydia, gonorrhea, and syphilis. CPL performs virtually all tests for these STIs in Manitoba, but as previous research has demonstrated, laboratory data may not capture all STI cases in a jurisdiction (Doyle, Glynn, & Groseclose, 2002). Thus, our aim with this case study was to explore the relationship between the CPL data and other administrative data sources.

The total number of tests for each of these STIs, as well as the number of positive tests, was identified in each fiscal year and analyzed using descriptive techniques, such as frequencies and percentages. In addition, the CPL data were linked to hospital records and physician billing claims; diagnoses for STIs recorded in these administrative health databases were compared to STI test results. The index date for each individual having a test for chlamydia, gonorrhea, or syphilis in the CPL data was established. The index date was the date that a test for one of these STIs was first recorded in the CPL data. Then, the hospital records and physician billing claims were searched for an STI diagnosis both before and after the index date, using 30–, 90–, and 180–day time period These analyses were conducted for both **positive and negative tests**.

As the results in Figure 6.6 reveal, approximately 5% of all chlamydia tests were positive in each of the study years. This percentage was lower for gonorrhea and syphilis and was, on average, less than 3% (Figures 6.7 to 6.8).

Figure 6.6: Frequency of Tests and Positive Tests for Chlamydia

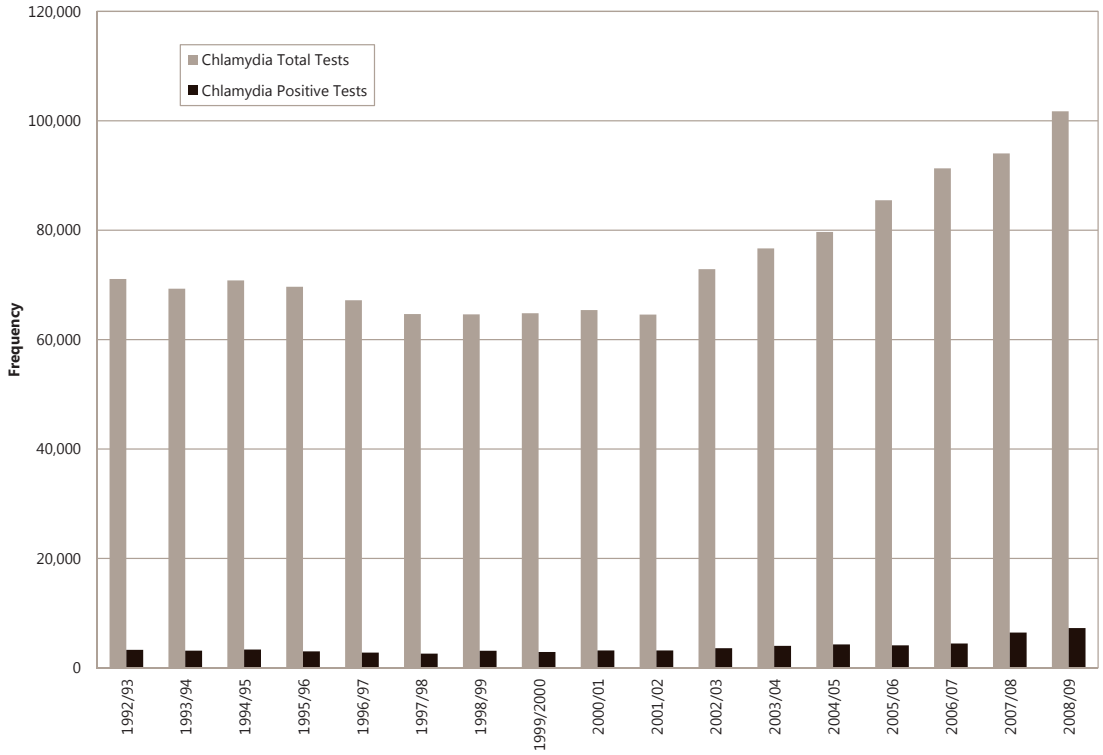


Figure 6.7: Frequency of Tests and Positive Tests for Gonorrhea

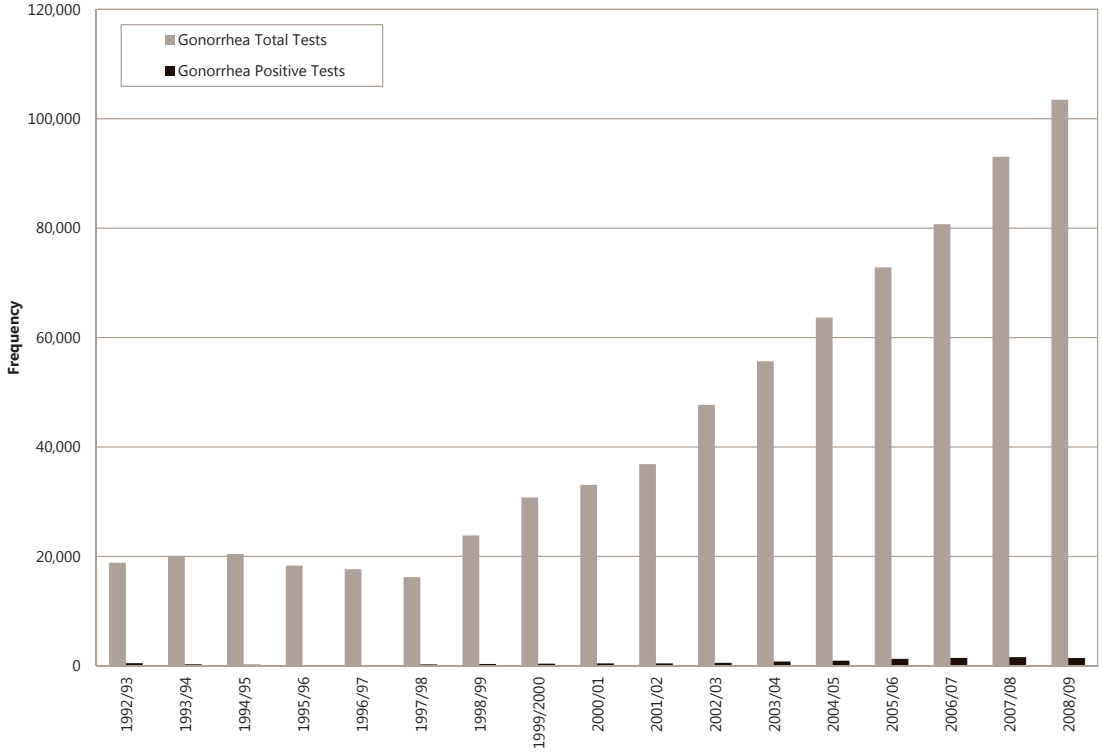
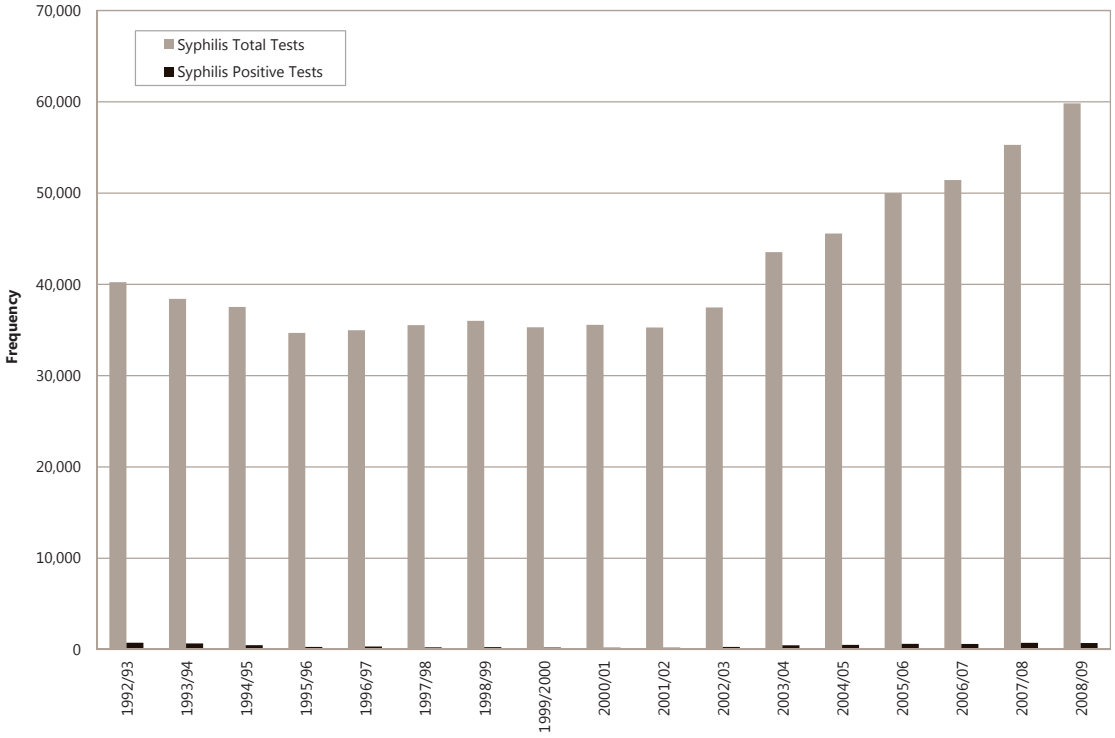


Figure 6.8: Frequency of Tests and Positive Tests for Syphilis



As the results in Table 6.2 reveal, only a small proportion of positive tests for chlamydia and gonorrhea had a corresponding diagnosis in hospital or physician claims data, regardless of the size of the reference window applied to the data. However, for syphilis a substantial portion of positive tests were associated with a diagnosis in hospital and physician billing claims data. In fact, almost one–third of positive tests had a corresponding diagnosis if a 180–day period before and after the index date was investigated. When a parallel analysis was conducted for negative tests, the percentages were much smaller, particularly for gonorrhea and syphilis.

Table 6.2: Comparison of Test Results for Sexually Transmitted Infections in CPL Databases and STI Diagnoses in Hospital Data and Physician Billing Records

	Positive Test in CPL Data					
	Chlamydia (N=69,403)		Gonorrhea (N=11,844)		Syphilis (N=7,775)	
	N	%	N	%	N	%
Diagnosis Within 30 Days of Test	2,186	3.1	591	5.0	1,580	20.3
Diagnosis Within 90 Days of Test	2,787	4.0	655	5.5	2,026	26.1
Diagnosis Within 180 Days of Test	3,410	4.9	706	6.0	2,390	30.7
	Negative Test in CPL Data					
	Chlamydia (N=1,277,087)		Gonorrhea (N=814,817)		Syphilis (N=723,784)	
	N	%	N	%	N	%
Diagnosis Within 30 Days of Test	5,889	0.5	1,359	0.2	868	0.1
Diagnosis Within 90 Days of Test	12,861	1.0	1,928	0.2	1,167	0.2
Diagnosis Within 180 Days of Test	22,176	1.7	2,656	0.3	1,445	0.2

CHAPTER 7: SUMMARY AND CONCLUSIONS

This study investigated a new administrative health database, the CPL database, in the Repository housed at MCHP. The CPL database in the Repository contains population–based notifiable disease requisitions and tests from 1992 to 2010. The quality of the data and its potential usefulness for population health and public health services research were examined. Data management and data quality frameworks were developed as part of this study to facilitate the ongoing acquisition and evaluation of administrative health databases into the Repository.

Key Findings

A key finding of this study is that the CPL data are generally of high quality. Using the data–base specific dimensions of quality found in the MCHP Data Quality Framework to guide the evaluation, we found that the data were largely complete; they covered the majority of the Manitoba population across study years. Correctness, as measured by the percentage of invalid and missing values, was also high; only small amounts of invalid or missing values were identified. The data were not, however, always consistent over time; the linear trend analyses for the annual frequencies of results and requisitions frequently had one or more outlying values. However, this may not necessarily reflect a data quality problem, but rather the nature of the CPL environment. The demand for disease testing will vary as a function of such factors as the number of research studies on which CPL scientists are working, the number of disease outbreaks that require confirmation, and the initiation of disease prevention/health promotion campaigns by the provincial ministry of health. In terms of linkability, most of the records in the CPL data can be linked to other administrative health databases using a unique, anonymized PHIN. However, interpretability of the data may be limited by gaps in historical documentation about the data.

Another key finding is that notifiable disease testing data may not always have a high degree of agreement with diagnostic data contained in hospital records and physician billing claims. This concurs with the finding of Yiannakoulias and Svenson (2009), who compared these two data sources for surveillance of gastrointestinal illness. Testing data and diagnosis data arise from two different administrative processes and therefore are more likely to be complementary, rather than concordant, sources of information about notifiable diseases. In terms of the analyses that we conducted for HIV, TB, and STIs, we found that the agreement varied with the disease under investigation. For example, for syphilis there was higher rate of concordance between the two data sources than for other STIs. While we explored agreement for only a small number of diseases, previous research also confirms that completeness of case capture from laboratory testing results will vary across diseases (Doyle, Glynn, & Groseclose, 2002).

Conclusions

Notifiable disease data have many potential uses beyond surveillance when they can be anonymously linked to other administrative databases. These data can be used to construct population–based cohorts, by identifying individuals with positive or negative test results, for investigations of health outcomes and health services utilization. Linkage with administrative health databases that contain diagnostic codes can be used to produce comprehensive population estimates of disease prevalence and incidence. As well, notifiable disease data can be used to evaluate the effectiveness of population–based disease prevention programs by investigating changes over time in testing rates for different population groups. Comparisons of differences in testing rates between geographic areas or income quintiles can be used to assess disparities in the utilization of public health services.

However, the use of laboratory data in research projects requires careful evaluation of both data–base specific quality and project–specific quality. Changes over time in program delivery and testing protocols may affect some characteristics of the data, including completeness of population coverage, and temporal consistency. Variations are dependent on the notifiable disease or population under investigation. For example, HIV test results in Manitoba cannot be linked to other administrative health databases prior to 2006/07 fiscal year. RHAs in southwestern Manitoba have experienced substantial changes in requisition and testing rates for some notifiable diseases over time, which suggest the existence of some gaps in the CPL data.

To date, only a small number of Canadian studies have linked notifiable disease data to other administrative health databases in order to conduct population health and public health services research. Yiannakoulias and Svenson (2009) compared spatial and temporal variations in Alberta's rates of gastrointestinal illnesses using test results from notifiable disease data and diagnostic codes in hospital and physician billing claims databases. They found that illness rates estimated from these two sources differed substantially over time and across health regions and, accordingly, recommended that both data sources be used in order to comprehensively ascertain notifiable disease cases. Kwong et al. (2010) reported on the burden of infectious diseases in Ontario, including bacterial infections, viral hepatitis, sexually transmitted infections, viral respiratory infections, intestinal infections, vaccine–preventable diseases, tuberculosis, and various other infections. Through the anonymized linkage of notifiable disease data from an integrated Public Health Information System with other administrative health databases, the authors estimated years of life lost and year–equivalents of reduced functioning for these diseases.

Given the limited research that has been conducted to date in Canada using linked notifiable disease data, there are opportunities for further research. Mak and Watkins (2008) used linkage techniques to combine notifiable disease data on sexually transmitted infections, Hepatitis B, and Hepatitis C with other administrative health databases to produce accurate estimates of disease rates in the Aboriginal and non–Aboriginal populations of Western Australia. The authors found that rates for Aboriginal populations were substantially overestimated if the error–prone ethnicity identifier found in the notifiable disease database was used instead of the less–biased identifier found in other administrative data sources. Vrijens et al. (2010) linked laboratory reports of hospital–acquired bloodstream infections with hospital administrative health data to estimate the incremental costs and lengths of stay associated with bloodstream infections in hospital patients. The study was conducted using administrative health databases from Belgium. Faustini et al. (2010) produced an accurate estimate of prevalence of viral Hepatitis C in one region of Italy by linking notifiable disease laboratory data with hospital records and dialysis registry data.

While the acquisition of the CPL data into the Repository affords an opportunity to expand population health and health services research using administrative health databases, this study provides a number of other benefits to researchers, data managers, and data analysts. The MCHP Data Quality Framework and an accompanying set of tools to evaluate database–specific quality was proposed. The Framework and analytic techniques can be applied to new databases that are acquired into the Repository, as well as to existing databases that require ongoing evaluation. Data quality is a multidimensional construct and, therefore, requires multiple indicators and strategies for its evaluation. The framework distinguishes database–specific quality evaluation, which can be routinely

conducted on any administrative health database acquired into the Repository, from project–specific quality evaluation, which focuses on the unique needs for quality evaluation once the objectives of a research project have been defined. In both types of evaluations, concepts of accuracy and validity are represented. In addition, concepts of timeliness and interpretability are incorporated into database–specific quality evaluations.

This framework was constructed using other data quality frameworks as a guide, but recognizes the unique evaluative requirements of administrative health databases. Similarly, Iron and Manuel (2007) proposed that the domains of correctness, reliability, completeness and usability should be evaluated for administrative health databases, although they did not develop specific tools to conduct these evaluations. In the current study, macros were constructed for the routine evaluation of some components of database–specific quality. However, macros were not developed for all components of data quality, in part because appropriate quantitative measures have not yet been defined in the research literature. For example, it is not clear how interpretability of the data should be quantified. Moreover, external validity is difficult to routinely assess because it requires the identification of an appropriate, error free data source that can be linked to the database in question.

Poor data quality can have a number of adverse effects on the progress of science. Perhaps most importantly, it can lead to erroneous research conclusions. Poor quality data can reduce the efficiency of data users by impacting on their productivity. It affects the ability of researchers to produce high–impact studies that have the potential to influence clinical and population health policy and program delivery. Therefore evaluations of data quality are critical to the timely and valid conduct of population health and health services research.

Recommendations

The following recommendations arise from this study:

Recommendation #1: Link notifiable disease data to other administrative databases to explore the full potential of the CPL data for population health and health services research.

While a number of studies have used the CPL for surveillance purposes, their full potential to contribute to population health and health services research has not yet been explored. MacDonald et al. (2007) identified that a limited amount of public health services research has been conducted within Canada. The CPL data provide one important source of information about the use of public health services. MCHP could assume a leadership role in Canada on the conduct of public health services research as these data become increasingly used in MCHP deliverables and other research projects.

Recommendation #2: Add other sources of disease tests to the Repository to improve the comprehensiveness of the Repository for the investigation of notifiable diseases.

CPL captures all or virtually all of the tests conducted in Manitoba for many notifiable diseases. However, the data have some gaps that could be addressed by acquiring test data from other sources, including DSM. Consultations with Manitoba Health and other laboratory service providers in the province will be useful to identify opportunities to expand the Repository at MCHP to include comprehensive data on notifiable diseases for the province.

Recommendation #3: Apply the Data Quality Framework to all administrative databases in the Repository at MCHP. Explore the use of case studies to promote best practices in data quality evaluation.

MCHP's Data Quality Framework provides the conceptual framework for the evaluation of data quality, but the routine application of data quality evaluation tools to each database in the Repository can contribute to the efficient use of the data for population health and health services research. Roos et al. (2005) in their synthesis of the data quality literature in Canada found that while many studies have been conducted about the validity of chronic disease diagnoses in administrative data, there have been few, if any studies of other dimensions of quality. As well, data quality studies have primarily focused on hospital and physician administrative databases; few studies have examined other sources, including population registry and prescription drug databases.

The MCHP Concept Dictionary affords a unique opportunity to share information on project–specific quality. Indeed, details about validation methodologies are already incorporated into the Concept Dictionary. However, concepts devoted to each of the components of the Data Quality Framework could help to enhance the visibility of this research in Canada and internationally.

Recommendation #4: Develop a framework for evaluating the quality of database documentation.

Administrative databases frequently have limited documentation about their contents and data collection processes. It can be a time–consuming process for researchers to learn about the characteristics of the data that may affect the research process and outputs. High–quality documentation about the contents of data fields, changes in data collection methods and data recording techniques, and reasons for missing data can expedite the research process and ensure accuracy and validity of research outputs.

A framework for evaluating data quality is an important component of the Metadata Repository that has been developed at MCHP. It can be used to identify gaps in both the content and usability of documentation for administrative health data. Ultimately, this framework will help to improve researchers' abilities to correctly interpret administrative health data.

However, a key challenge in developing a framework for quality of data documentation quantifying the adequacy of documentation. Concepts of data quality developed in the Data Quality Framework also apply to documentation, including accuracy, completeness, and interpretability. However, measures of these concepts have, to the best of our knowledge, not been described in the literature.

Recommendation #5: Conduct studies about the validity of cases ascertained from notifiable disease tests and diagnostic information in administrative data.

This study found that the agreement between laboratory–confirmed cases and diagnosis–based cases of notifiable diseases could be low for some diseases, emphasizing the importance of using more than one source of data to comprehensively ascertain disease cases. However, the validity of different data sources for ascertaining cases of notifiable diseases provides another opportunity for further research and evaluation. Sensitivity and specificity of laboratory tests may vary over time, as new clinical procedures are introduced. Similarly, the validity of diagnoses codes may not be constant across time. While Roos et al. (2005) identified a number of Canadian validation studies about diagnosis codes for ascertaining cases of chronic disease, there has been few validation studies about infectious diseases.

REFERENCE LIST

Arts DGT, de Keizer NF, Scheffer GJ. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. *Journal of the American Medical Informatics Association.* 2002;9(6):600–611.

Bernatsky S, Joseph L, Bélisle P, et al. Bayesian modelling of imperfect ascertainment methods in cancer studies. *Statist Med.* 2005;24(15):2365–2379.

Blakely T, Salmond C. Probabilistic record linkage and a method to calculate the positive predictive value. *International Journal of Epidemiology.* 2002;31(6):1246–1252.

Canadian Institute for Health Information. The CIHI data quality framework. Canadian Institute for Health Information. 2009. http://www.cihi.ca/CIHI–ext–portal/pdf/internet/DATA_QUALITY_FRAMEWORK_2009_EN. Accessed January 26, 2012.

Christensen PB, Hay G, Jepsen P, Omland LH, Just SA, Krarup HB, Weis N, Obel N, Cowan S. Hepatitis C prevalence in Denmakr – an estimate based on multiple national registers. *BMC Infectious Diseases* 2012;12:178.

Daas PJH, Arends–Tóth J, Schouten B, Kuijvenhoven L, Statistics Netherlands. Quality framework for the evaluation of administrative data. 2008. http://www.pietdaas.nl/beta/pubs/pubs/21Daas.pdf. Accessed January 26, 2012.

Doyle TJ, Glynn MK, Groseclose SL. Completeness of notifiable infectious disease reporting in the United States: An analytical literature review. *Am J Epidemiol* 2002;155:866–874.

Emons MF. Integrated patient data for optimal patient management: the value of laboratory data in quality improvement. *Clin Chem.* 2001;47(8):1516–1520.

Faustini A, Colais P, Fabrizi E, et al. Hepatic and extra–hepatic sequelae, and prevalence of viral hepatitis C infection estimated from routine data in at–risk groups. *BMC Infectious Diseases.* 2010;10(1):97.

Holman CDJ, Bass AJ, Rouse IL, Hobbs MST. Population–based linkage of health records in Western Australia: development of a health services research linked database. *Australian and New Zealand Journal of Public Health.* 1999;23(5):453–459.

Iron K, Manuel DG. Quality assessment of administrative data (QuAAD): an opportunity for enhancing Ontario's health data. Institute for Clinical Evaluative Sciences. 2007. http://www.ices.on.ca/file/QuAAD_Jul–07.pdf. Accessed January 26, 2012.

Karr AF, Sanil AP, Banks DL. Data quality: a statistical perspective. *Statistical Methodology.* 2006;3(2):137–173.

Kwong JC, Crowcroft NS, Campitelli MA, et al. Ontario burden of infectious disease study (ONBOIDS): an OAHPP/ICES report. Toronto: Ontario Agency for Health Protection and Promotion, Institute for Clinical Evaluative Sciences. 2010. http://www.ontla.on.ca/library/repository/mon/24012/306590.pdf.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159–174.

Lyman JA, Scully K, Harrison JH. The development of health care data warehouses to support data mining. *Clinics in Laboratory Medicine.* 2008;28(1):55–71.

MacDonald M, Hancock T, Strosher HW, Core Public Health Functions Research Team. Developing a research agenda on the implementation and impact of Core Public Health Functions in British Columbia. Core Public Health Functions Research Priority Think Tank. 2007. http://www.uvic.ca/research/groups/cphfri/assets/docs/Core_Functions_Research_Think_Tank_report.pdf. Accessed January 26, 2012.

Mak DB, Watkins RE. Improving the accuracy of Aboriginal and non–Aboriginal disease notification rates using data linkage. *BMC Health Services Research.* 2008;8(1):118.

Omar MH. Statistical process control charts for measuring and monitoring temporal consistency of ratings. *Journal of Educational Measurement.* 2010;47(1):18–35.

Orr P. Adherence to tuberculosis care in Canadian Aboriginal populations. Part I: definition, measurement, responsibility, barriers. *International Journal of Circumpolar Health.* 2011;70(2):113–127.

Roos LL, Gupta S, Soodeen R, Jebamani L. Data quality in an information–rich environment: Canada as an example. *Canadian Journal on Aging/Revue canadienne du vieillissement.* 2005;24(1):153–170.

SAS Institute Inc. SAS data integration studio 3.3: user's guide. SAS Institute Inc. 2006. http://support.sas.com/documentation/onlinedoc/etls/usage33.pdf. Accessed January 26, 2012.

Schmidtmann I, Blettner M. How do cancer registries in Europe estimate completeness of registration? *Methods of Information in Medicine.* 2009;48(3):267–271.

Statistics Canada. *Statistics Canada Quality Guidelines.* 5th ed. Ottawa, ON; Statistics Canada; 2009.

Vrijens F, Hulstaert F, Van de Sande S, Devriese S, Morales I, Parmentier Y. Hospital–acquired, laboratory–confirmed bloodstream infections: linking national surveillance data to clinical and financial hospital data to estimate increased length of stay and healthcare costs. *Journal of Hospital Infection.* 2010;75(3):158–62.

Wang RY, Storey VC, Firth CP. A framework for analysis of data quality research. *IEEE Transactions on Knowledge and Data Engineering.* 1995;7(4):623–640.

Wang RY, Strong DM. Beyond accuracy: what data quality means to data consumers. *Journal of Management Information Systems.* 1996;12(4):5–33.

Yiannakoulias N, Svenson LW. Differences between notifiable and administrative health information in the spatial–temporal surveillance of enteric infections. *Int J Med Inform.* 2009;78(6):417–424.

GLOSSARY

Accuracy

This term is used in the MCHP Data Quality Framework. "... is the degree to which the information correctly describes the phenomena it was designed to measure. It is usually characterized in terms of error in statistical estimates and is traditionally decomposed into bias (systematic error) and variance (random error) components. It may also be described in terms of the major sources of error that potentially cause inaccuracy (e.g., coverage, sampling, nonresponse, response)"

Note that for a **database–specific quality** assessment, measures of accuracy are applied to the entire database; whereas for a project–specific assessment, they are applied to the cohort, region, or time period that is the focus of the project.

Statistics Canada. http://www.statcan.gc.ca/pub/12–539–x/4147797–eng.htm. Accessed April 20, 2012.

Acute Care

Hospital stays with a length of stay between 1 and 59 days. Also known as Short Stay Inpatient Care or Short Stay Care. Or services provided within an acute care hospital.

Administrative Health Data

Refers to information collected "usually by government, for some administrative purpose (e.g., keeping track of the population eligible for certain benefits, paying doctors or hospitals), but not primarily for research or surveillance purposes" (Spasoff, 1999). MCHP's research uses administrative data from hospital discharge summaries, physician billing claims, claims for prescription drugs, and other health related data. Using these data, researchers can study the utilization of health resources over time and the variations in rates within and across the provinces.

Spasoff, RA. Epidemiologic Methods for Health Policy. New York, NY: Oxford University Press; 1999.

Cadham Provincial Laboratory (CPL) database

An administrative health database containing information about the services provided by the Cadham Provincial Laboratory (CPL), including public health laboratory services (microbiology, serology, parasitology, and virology) and reference services for identification and typing of microorganisms. Request for these services (from health practitioners) are captured in this database, as well as the results of the requests. Patient information and clinical information are also provided.

Canadian Institute for Health Information (CIHI)

An independent, not–for–profit organization that provides essential data and analysis on Canada's health system and the health of Canadians.

Clinical Microbiology Section

Service provider at the Cadham Provincial Laboratory. Clinical Microbiology services involve the detection, isolation and epidemiological characterization of bacterial or fungal pathogens or toxins from clinical specimens.

Government of Manitoba. http://www.gov.mb.ca/health/publichealth/cpl/docs/guide.pdf. Accessed July 17, 2012

Census

Official count of a population, often including demographic information such as age, sex, employment and income. **Statistics Canada** conducts a Census every five years. It takes account of all persons living in Canada, including any individuals residing in Canada on a temporary basis. The Census also includes Canadians abroad on military missions or on merchant vessels that are registered in Canada (See the Statistics Canada Dictionary and Concepts page http://www.statcan.gc.ca/concepts/index–eng.htm).

Chlamydia

Sexually transmitted infection that, if left untreated, may increase the risk of infertility, epididymitis in males, or ectopic pregnancy and pelvic inflammatory disease in females, and of contracting HIV (particularly for chlamydia infections with ulcerations) (Chin, 2000; Dickerson et al., 1996).

Chin J. Control of communicable diseases manual. 17th ed. Washington, DC: American Public Health Association; 2000.
Dickerson M, Johnson J, Delea T, White A, Andrews E. The casual role for genital ulcer disease as a risk factor for transmission of human immunodeficiency virus: An application of the Bradford Hill criteria. Sexually Transmitted Diseases. 1996;23:429–440.

Crosswalk

"A crosswalk is a specification for mapping one metadata standard to another. Crosswalks provide the ability to make the contents of elements defined in one metadata standard available to communities using related metadata standards." (Note: The terms "crosswalking" and "mapping" are sometimes used interchangeably.)

National Information Standards Organization. http://www.niso.org/publications/white_papers/crosswalk/. Accessed January 26, 2012.
American Medical Association. http://www.ama–assn.org/ama1/pub/upload/mm/399/crosswalking–between–icd–9–and–icd–10.pdf. Accessed January 26, 2012.

Data Management Process (MCHP)

Method to acquire and incorporate databases into the Repository housed at MCHP. The steps are: 1) formulate the request and receive the data, 2) become familiar with the data structure and content, 3) apply the data to SAS®, 4) evaluate the quality of the data, 5) document information about the data, and 6) release the data for use by data analysts and researchers.

Data Quality

The quality of data is measured according to its reliability and **validity**; the completeness and **accuracy** of a data set. It is usually measured by comparing the data set to another data set identified as the "gold standard" and assessing the level of agreement. For example, linking with a data source such as **census** data, vital statistics, or surveys to determine missing information.

Data Quality Framework (MCHP)

Part of MCHP's formalized process of **data quality** evaluation for data housed at the **Population Health Research Data Repository**. The framework, developed by Lix et al. (2012), depicts measures of database– and project–specific data quality.

Database–Specific Quality

Type of quality evaluation and a component in the MCHP **Data Quality Framework**. This component encompasses concepts of **accuracy**, **internal validity**, **external validity**, **timeliness**, and **interpretability**, all used to assess the general usability of a database. **Database–specific quality** is distinct from **project–specific quality**, the other type of quality evaluation in the MCHP Data Quality Framework, which focuses on concepts relevant to the quality of a database to address research questions associated with a project or study

Descriptive Analyses

Descriptive analyses are used for summarizing, organizing, graphing, and, in general, describing quantitative information. Often contrasted with inferential analyses, which are used to make inferences.

Dissemination Area (DA)

"A small, relatively stable geographic unit composed of one or more blocks. It is the smallest standard geographic area for which all **census** data are disseminated. DAs cover all the territory of Canada." As of 2001, the DA replaces the **Enumeration Area** (EA) as a basic unit for dissemination (Statistics Canada, 2007).

Statistics Canada. http://www12.statcan.ca/english/census01/Products/Reference/dict/geo021.htm. Accessed July 31, 2007.

Enumeration Area (EA) – see Dissemination Area (DA)

External Validity

This term is used in the MCHP Data Quality Framework. The relationship between the values in a data file or results of a project and an external source of similar or identical information that is assumed to be error free. For example, if caesarian section rates computed from a data file are much higher or lower than values published in a report that are based on another data source, this might suggest a lack of external validity.

Fiscal Year (FY)

For most Canadian government agencies and healthcare institutions, the fiscal year is defined as starting April 1 and ending the following year at March 31. For example, the 2005/06 fiscal year would be April 1, 2005 to March 31, 2006, inclusive and may also be denoted as FY 2005.

General Practitioner (GP)/Family Practitioner (FP)

A physician who operates a general or family practice and is not certified in another specialty in Manitoba.

Gonorrhea

Sexually transmitted infection that, if left untreated, may increase the risk of pelvic inflammatory disease, ectopic pregnancy, and infertility in females; urethritis, epididymitis, and gonococcal arthritis in males (Berkow & Fletcher, 1992); and of contracting HIV (Chin, 2000; Dickerson et al., 1996). It also may cause pharyngeal and anorectal infections in females and homosexual males (Chin, 2000). Symptoms may be absent, or may appear within 1–4 weeks after infection. Symptoms include (but are not limited to) a feeling of burning or pain while urinating, increased urination, penile or vaginal discharge, or tender or swollen genitals (PubMed Health, 2011).

Berkow R, Fletcher AJ, eds. The Merck manual of diagnosis and therapy. 16th ed. Rahway, NJ:Merck Research Laboratories; 1992.
Chin J. Control of communicable diseases manual. 17th ed. Washington, DC: American Public Health Association; 2000.
Dickerson M, Johnson J, Delea T, White A, Andrews E. The casual role for genital ulcer disease as a risk factor for transmission of human immunodeficiency virus: An application of the Bradford Hill criteria. Sexually Transmitted Diseases. 1996;23:429–440.
PubMed Health. 2011. http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0004526/. Accessed October 15, 2012.

Hepatitis B (HBV)

Hepatitis B virus (HBV) causes Hepatitis B, a vaccine–preventable infection of the liver. Symptoms are similar to the flu and appear in about half of those infected. The infection is transmitted through sexual contact or exposure to the blood or body fluids of an infected person. A newborn may contract HBV from an infected mother at birth. Most infected adults clear the virus after several months; approximately 10% of those infected become chronically infected. Treatment is available for HBV.

Hepatitis C (HCV)

Hepatitis C virus (HCV) causes Hepatitis C, a disease of the liver. Most people infected with HCV are asymptomatic for years. The infection is transmitted through sexual contact or exposure to the blood or body fluids of an infected person. A newborn may contract HBV from an infected mother at birth. Treatment is available for HCV.

Human Immunodeficiency Virus (HIV)

"A virus that attacks the immune system, resulting in a chronic, progressive illness that leaves people vulnerable to opportunistic infections and cancers. When the body can no longer fight infection, the disease is known as AIDS, which stands for Acquired Immunodeficiency Syndrome. On average, it takes more than 10 years to progress from initial HIV infection to AIDS… In order to be infected, the virus must enter a person's bloodstream (HIV cannot survive outside the body). HIV is transmitted from one person to another through: unprotected sexual intercourse (vaginal, anal or oral); shared needles or equipment for injecting drugs; unsterilized needles for tattooing, skin piercing or acupuncture; pregnancy, delivery and breast feeding (i.e., from an HIV–infected mother to her infant); occupational exposure in health care settings"

Public Health Agency of Canada. 2008. http://www.phac–aspc.gc.ca/aids–sida/info/index–eng.php. Accessed October 15, 2012.

Income Quintile

A method to measure the average (mean) household income of residents, ranking them from poorest to wealthiest, and then grouping them into five income quintiles (1 being poorest and 5 being wealthiest). Each quintile contains approximately 20% of the population. The income quintile measure is derived from Statistics Canada **Census** data by aggregating household income to the **dissemination area** (as of 2001 Census data, dissemination area replaces **enumeration area** as a basic unit for dissemination) and then ranking neighbourhoods by income quintile. Income quintiles are available for both urban and rural populations. Income quintiles are often used as a proxy measure of socio–economic status.

Infectious Disease

A group of illnesses that include pneumonia, influenza, hepatitis, tuberculosis, **sexually transmitted infections (STIs)**, pelvic inflammatory disease, and AIDS.

Internal Validity (Data Quality Framework)

The strength of the relationship of data in one field to the data of another field. It is measured by internal consistency (numeric agreement between fields or the logical relationships between fields), temporal consistency (stability of data fields across time), and linkability (connections of one data file to another using a unique subject–specific identifier).

International Classification of Diseases (ICD)

A classification system of diseases, health conditions, and procedures developed by the World Health Organization (WHO), which represents the international standard for the labeling and numeric coding of diseases and health related problems. Within this system, all diseases/conditions are assigned numbers in hierarchical order. There are several versions of the ICD coding system, including ICD–8, ICD–9, ICD–9–CM (Clinical Modifications), ICD–O (Oncology), ICD–10, and ICD–10–CA (Canadian Enhancements).

International Classification of Diseases, 9th Revision, with Clinical Modifications (ICD–9–CM)

The 9th version of the ICD (International Classification of Disease) coding system (with Clinical Modifications) developed by the World Health Organization (WHO). It is used to classify diseases, health conditions, and procedures. This version was used extensively in Canadian hospitals.

As of April 1, 2004, Manitoba hospitals replaced ICD–9–CM with ICD–10–CA for coding diagnoses and the Canadian Classification of Health Interventions (CCI) for coding procedures/interventions.

International Classification of Diseases, 10th Revision, with Canadian Enhancements (ICD–10–CA)

The 10th version of the ICD (International Classification of Disease) coding system developed by the World Health Organization (WHO). It is used to classify diseases and related health problems (morbidity), but includes enhancements developed by **Canadian Institute for Health Information (CIHI)** for use in Canadian hospitals and other medical facilities. The Canadian Classification of Health Interventions (CCI) is the companion classification system to ICD–10–CA for coding procedures in Canada. ICD–10–CA and CCI are being used on Manitoba hospital abstracts beginning April 1, 2004.

NOTE: ICD–10–CA and ICD–10 are similar to the 4th digit, but they are not the same.
For more information on ICD–10–CA, please visit the CIHI Web Site at: http://www.cihi.ca/cihi–ext–portal/internet/en/document/standards+and+data+submission/standards/classification+and+coding/codingclass_icd10.

Interpretability

This term is used in the MCHP Data Quality Framework. The extent to which a dataset may be understood. This is measured by the availability and quality of metadata, including documentation, policies and procedures relevant to the creation and maintenance of the dataset, variable formats and data model diagrams.

Macro

Abbreviated command that represents a series of instructions.

Wikipedia. http://en.wikipedia.org/wiki/Macro. Accessed May 17, 2012.
Dictionary.com. http://dictionary.reference.com/browse/macro. Accessed May 17, 2012.
ITBusinessEdge. http://www.webopedia.com/TERM/M/macro.html. Accessed May 17, 2012.
The Free Dictionary by Farlex. http://www.thefreedictionary.com/macro. Accessed May 17, 2012.

Manitoba Centre for Health Policy (MCHP)

A unit within the Department of Community Health Sciences, Faculty of Medicine, University of Manitoba. MCHP is active in health services research, evaluation, and policy analysis by concentrating on using the Manitoba **Population Health Research Data Repository** (Repository) to describe and explain patterns of care and profiles of health and illness.

Manitoba Health

A provincial government department responsible for providing healthcare services in Manitoba. From April 1, 2008 to November 2009, this department was part of a larger department called **Manitoba Health** and Healthy Living (MHHL).

Manitoba Health Insurance Registry

The Manitoba Health Insurance Registry (also known as the Master Registry and the Manitoba Health Services Insurance Plan (MHSIP) Registration File) is a longitudinal population–based registry of all individuals who have been registered with **Manitoba Health** at some point since 1970. It includes date fields for registration, birth, entry into province, migration in/out of province, and death. It provides the needed follow–up information to track residents for longitudinal and intergenerational analyses. Primary identification is achieved by two numbers: every family in Manitoba is assigned a family registration number and every individual is assigned a unique **Personal Health Identification Number (PHIN)** by the Ministry of Health. These components are also included in the Manitoba Health Insurance Registry. The PHIN is encrypted in the registry data received by MCHP so that individuals cannot be identified. Individuals moving into the province and not yet eligible for coverage, families of military personnel (insured federally), and members of the RCMP (insured federally) are not included in the registry. "Snapshot files" of the Manitoba Health Insurance Registry data, received semi–annually at MCHP from Manitoba Health, are used to create and maintain information in the MCHP Population Registry.

Maternal Serum Screening Program

This program is funded by the province and offers all pregnant women in Manitoba a blood test that screens for chromosome abnormalities or birth defects.

Metadata Repository

File containing documentation about the databases from the **Population Health Research Data Repository** held at the **Manitoba Centre for Health Policy**.

Newborn Screening and Public Health Chemistry Section

Service provider at the Cadham Provincial Laboratory. Responsibilities include Newborn Screening Program and Maternal Serum Screening. The Newborn Screening Program is guided by the Manitoba Perinatal Screening Committee and screens all newborns using a blood test taken between 24 hours and five days after the birth of a baby, used to identify the risk for a number of treatable disorders. The Maternal Serum Screening is completed in collaboration with the Department of Human Genetics of the University of Manitoba: this test is optional to pregnant women during prenatal care and provides an estimation of the risk for fetal open neural tube defects, Down Syndrome, Trisomy 18, and Smith–Lemli–Opitz syndrome.

Government of Manitoba. http://www.gov.mb.ca/health/publichealth/cpl/docs/guide.pdf. Accessed May 17, 2012.
Government of Manitoba. http://www.gov.mb.ca/health/publichealth/cpl/baby.html. Accessed May 17, 2012.

Notifiable Disease

"A disease deemed of sufficient importance to public health to require that its occurrence be reported to public health officials. The reporting of notifiable diseases is mandated by the provinces and territories; notifiable diseases may vary from province to province. Reporting by the provinces and territories to the federal level is voluntary; however, agreement is reached by consensus of the Advisory Committee on Epidemiology (ACE), which comprises representatives from all provinces and territories."

Public Health Agency of Canada. http://dsol–smed.phac–aspc.gc.ca/dsol–smed/ndis/glossa–eng.php. Accessed May 17, 2012.

Parasitology Section

Service provider at the Cadham Provincial Laboratory. Responsibilities include **screening** for parasites and eggs (ova) that may cause or carry a disease. A parasite is a biological agent that lives on or within the host, simultaneously causing illness or disease.

Personal Health Identification Number (PHIN)

A unique numeric identifier assigned by **Manitoba Health** to every person registered for health insurance in Manitoba and non–residents who are treated at facilities which submit claims electronically. Introduced as a linkage key in 1984, it was issued to the public in 1994 as the basic access identifier for the Pharmacare/Drug Programs Information Network (DPIN).

At MCHP, the PHIN is either a scrambled (encyrpted) version of the **Manitoba Health** PHIN or an alphanumeric identifier assigned via the **Research Registry** to individuals who do not have scrambled numeric PHINs.

Physician Claims

Claims (billings) for payment that are submitted to the provincial government by individual physicians for services they provide. Fee–for–service physicians receive payment based on these claims, while those submitted by physicians on alternate payment plans (APP) are for administrative purposes only. The physician claims are collected and stored in the Medical Services Database, which is part of the **Population Health Research Data Repository**.

Population Health Research Data Repository

The Population Health Research Data Repository is a comprehensive collection of administrative, registry, survey, and other databases primarily comprised of residents of Manitoba. This repository is housed at the **Manitoba Centre for Health Policy (MCHP)**. It was developed to describe and explain patterns of healthcare and profiles of health and illness, facilitating inter–sectoral research in areas such as healthcare, education, and social services. The administrative health database, for example, holds records for virtually all contacts with the provincial healthcare system, the Manitoba Health Services Insurance Plan (including physicians, hospitals, personal care homes, home care, and pharmaceutical prescriptions), for all registered individuals. MCHP acts as a trustee or steward of the information in the Repository for agencies such as **Manitoba Health**.

NOTE: This term Population Health Research Data Repository is commonly referred to as the Repository.

Positive and Negative Tests

Results of tests in determining the presence of a disease. When the disease has been identified, the test result is 'positive'; when the disease is absent, the test result is 'negative'.

Postal Code

A six digit code defining postal areas within Canada. Postal code is used to define groups based on their location in the province of Manitoba.

Prenatal Population

Pregnant females aged 10 years or older.

Project–Specific Quality

Type of quality evaluation and a component in the MCHP **Data Quality Framework**. This component encompasses concepts of **accuracy** and **validity**, both used to determine the usability of data for project–specific analyses. As opposed to measures of **database–specific quality**, which are applied to the entire database, measures of **project–specific quality** are applied to the cohort, region, or time period that is the focus of a project.

Provider Type

The practitioner who ordered the test.

Provincial Health Insurance Registry – see Manitoba Health Insurance Registry

Public Health Agency of Canada (PHAC)

The Government of Canada agency responsible for public health whose "primary goal is to strengthen Canada's capacity to protect and improve the health of Canadians and to help reduce pressures on the health–care system".

Research Registry

Also known as MCHP Research Registry. A longitudinal population based research registry that is derived from data in the **Manitoba Health Insurance Registry** and other data files in the MCHP Data Repository. "Snapshot files" of the Manitoba Health Insurance Registry data, received semi–annually at MCHP from **Manitoba Health**, are integrated with historical registry data at MCHP to maintain the MCHP Research Registry. Consistent programming efforts are applied to the repository data files in order to provide value–added data from the MCHP Research Registry. MCHP Research Registry is a key resource for the research conducted at the Centre and is central to the use of the **Population Health Research Data Repository**.

Referring Facility

The hospital, area of the hospital (e.g. Emergency department at Health Sciences Centre), or clinic where the patient was seen by the practitioner.

Regional Health Authority (RHA)

Regional governance structure set up by the province to be responsible for the delivery and administration of health services in specified areas. In Manitoba, from July 1, 2002 to April 17, 2012, there were 11 RHAs: Winnipeg, Brandon, South Eastman, Assiniboine, Central, Parkland, North Eastman, Interlake, Burntwood, NOR–MAN, and Churchill.

Requisition

The form that must be completed for all tests requested of the Cadham Provincial Laboratory.

SAS®

A statistical software package for analyzing data. Originally called Statistical Analysis System, SAS® is also referred to as Statistical Analysis Software.

Screening

A process (tests, examinations, or other procedures) to distinguish between well individuals who probably have (or are likely to develop) a particular disease from those who probably do not have it. This is also considered the secondary level of preventive care, involving the early detection of illness.

Serodiagnostic Testing – See Serology Section

Serology Section

Service provider at the Cadham Provincial Laboratory. Responsibilities include: Detection and determination of antigens or antibodies; **screening** and diagnosis of infections due to viral, bacterial, fungal, or parasitic agents; evaluating response to immunization; screening for donor and transplant selections; and viral load and genotyping for patient management and surveillance.

Sexually Transmitted Infection (STI)

"Infections that are transmitted through sexual contact (oral, vaginal or anal) with an infected individual. Blood–borne infections are transmitted by blood. Some infections (HIV/AIDS, **hepatitis B**, **hepatitis C** and **syphilis**) are capable of being transmitted through both sexual and blood–borne transmission routes."

Government of Manitoba. http://www.gov.mb.ca/health/publichealth/cdc/sti/index.html. Accessed July 4, 2012.

Statistics Canada

Statistics Canada (or Stats Can) is a federal government agency commissioned with producing statistics to help better understand Canada's population, resources, economy, society, and culture.

Syphilis

Infectious disease that may be transmitted through sexual contact, contaminated needles, or may be transmitted in utero. Symptoms can occur within a few weeks or a couple of months after infection. The first symptom may be a painless, open sore or ulcer (where the bacteria first entered the body). Later symptoms include patchy hair loss, a rash on soles of the feet or palms of the hands; fever; swollen glands, and muscle and joint pain. Symptoms usually disappear without treatment. If left untreated, syphilis can affect the brain, blood vessels, heart and bones, and can eventually lead to death.

Health Canada. 2006. http://www.hc–sc.gc.ca/hl–vs/iyh–vsv/diseases–maladies/syphilis–eng.php. Accessed October 22, 2012.

Timeliness

This term is used in the MCHP Data Quality Framework. How current the data are in a dataset. This is indicated by a) time until a dataset is acquired, b) time until the data is released to MCHP, and c) time until updates to the data are in place.

Tuberculosis (TB)

Disease that is acquired through an infection from a bacterium called Mycobacterium tuberculosis. TB is highly contagious: it is spread through the air by individuals with infected lungs or throats when they cough, sneeze, or talk. An individual with TB will become sick; and if left untreated, the individual may die.

Validity

This term is used in the MCHP Data Quality Framework. A measure of **data quality** to indicate whether the data make sense. See glossary terms **internal validity** and **external validity** for more information.

Note that as a **database–specific quality** assessment, measures of validity are applied to the entire database; whereas for a project–specific assessment, they are applied to the cohort, region, or time period that is the focus of the project.

Virus Detection Section

Service provider at the Cadham Provincial Laboratory. Clinical virology services involve the isolation or detection and identification of human viral pathogens from clinical specimens using established procedures such as: cell culture – many viruses are grown and identified in established cell lines; rapid diagnostics – results within hours to aid in patient management; and viral strain identification – subtyping for epidemiological and public health purposes (i.e., outbreak management, etc.)

Government of Manitoba. http://www.gov.mb.ca/health/publichealth/cpl/docs/guide.pdf. Accessed July 17, 2012.

APPENDIX A: NOTIFIABLE DISEASES IN MANITOBA

Common name	Scientific or technical name of disease or its infectious agent
AIDS	Acquired Immunodeficiency Syndrome
Amoebiasis	Entamoeba histolytica
Anthrax	Bacillus anthrasis
Avian Influenza	Influenza A virus, select Hemaglutinin and Neuraminidase types
Blastomycosis	Blastomyces dermatitidis
Botulism	Clostridium botulinum
Brucellosis	Brucella species
Campylobacter	Camplylobacter species
Cancer or malignant neoplasm	Cancer or malignant neoplasm
Chancroid	Haemophilus ducreyi
Chlamydia	Chlamydia trachomatis
Cholera	Vibrio cholerae, typable
Clostridium difficile toxin	Clostridium difficile
Clostridium perfringens (except wound specimens)	Clostridium perfringens
Congenital Rubella Infection/Syndrome	Rubella virus
Cryptosporidium	Cryptosporidium parvum
Cyclospora	Cyclospora cayetanensis
Creutzfeldt–Jakob Disease	Creutzfeldt–Jakob disease prion
Dengue Fever	Dengue virus
Diphtheria (Cases and Carriers)	Toxigenic Corynebacterium diptheriae (all subspecies)
Encephalitis	Encephalitis
Fish Tapeworm	Diphyllobothrium latum (Dibothriocephalus latus)
Food poisoning caused by Bacillus cereus	Bacillus cereus
Giardia	Giardia lamblia
Gonorrhea	Neisseria gonorrhoaea
Hantavirus	Hantavirus
Haemophilus influenza invasive disease from type-able Haemophilus organisms	Haemophilus influenzae
Hemolytic Uremic Syndrome (HUS)	Hemolytic Uremic Syndrome
Hepatitis A	Hepatitis A virus
Hepatitis B	Hepatitis B virus
Hepatitis C	Hepatitis C virus
Hepatitis, Viral (Other)	Hepatitis viruses other than A, B or C
HIV	Human immunodeficiency virus
Influenza A	Influenza A viruses
Influenza B	Influenza B viruses
Legionellosis	Legionella pneumophilia
Leprosy	Mycobacterium leprae
LGV	Lymphogranuloma venereum (Chlamydia trachomatis)
Listeriosis invasive disease	Listeria monocytogenes in normally sterile tissue
Lyme Disease	Borrelia burgdorferi
Malaria	Plasmodium falciparum, Plasmodium vivax, Plasmodium malariae, Plasmodium ovale
Measles	Rubeola virus
Meningococcal invasive disease	Neisseria meningitidis
Methicillin Resistant Staphylococcus aureus	Staphylococcus aureus with Methicillin resistance
Mumps	Mumps virus
Parapertussis	Bordetella parapertussis

Appendix A – Continued

Common name	Scientific or technical name of disease or its infectious agent
Parasitic Diseases other than amoebiasis, cryptosporidium, cyclospora, fish tapeworm, giardia, malaria, strongyloidiasis, toxoplasmosis, trichinosis and trypanosomiasis	
Parrot Fever (Psittacosis)	Chlamydophilia psittaci
Penicillin resistant pneumococci	Streptococcus pneumoniae with penicillin resistance
Pertussis	Bordetella pertussis
Plague	Yersinia pestis
Pneumococcal invasive disease (any normally sterile body site)	Streptococcus pneumoniae
Polio	Poliovirus
Q fever	Coxiella burnetii
Rabies	Rabies virus
Relapsing Fever	Borrelia recurrentis, Borrelia duttoni
Rickettsial Diseases other than Rocky Mountain Spotted Fever, Q-fever and typhus	
Rocky Mountain Spotted Fever	Rickettsia rickettsii
Rubella	Rubella virus
Salmonella	Salmonella species
Severe Acute Respiratory Syndrome (SARS)	SARS coronavirus
Severe Respiratory Illness (SRI)	Severe Respiratory Illness
Shigella	Shigella species
Smallpox	Variola major virus, Variola minor virus
Staphylococcal Food Poisoning	Staphylococcus aureus
Staphylococcal Toxic Shock Syndrome	Staphylococcus aureus in blood or normally sterile tissue in association with Toxic Shock Syndrome
Beta Hemolytic Streptococcal invasive disease, typable	Beta Hemolytic Streptococcal typable species in blood or normally sterile tissue. (Includes all samples of Strep. Group A, B, C, D, E, F or G found in blood, sterile tissue or internal aspirates — not in skin or wounds.)
Streptococcal Necrotizing Fasciitis	Streptococcus species in blood or normally sterile tissue in association with Necrotizing Fasciitis. (Includes all samples of Strep. Group A, B, C, D, E, F or G found in tissue or wounds that are accompanied by a clinical assessment of NF.)
Streptococcal Necrotizing Myositis	Streptococcus species in blood or normally sterile tissue in association with Necrotizing Myositis. (Includes all samples of Strep. Group A, B, C, D, E, F or G found in tissue or wounds that are accompanied by a clinical assessment of NM.)
Streptococcal Toxic Shock Syndrome	Streptococcus species in blood or normally sterile tissue in association with Toxic Shock Syndrome. (Includes all samples of Strep. Group A, B, C, D, E, F or G found in blood that are accompanied by a clinical assessment of TSS.)
Strongyloidiasis	Strongyloides stercoralis
Syphilis	Treponema pallidum pallidum
Tetanus	Clostridium tetani
Toxoplasmosis	Toxoplasma gondii
Trichinosis	Trichinella spiralis
Trypanosomiasis	Trypanosoma species

Appendix A – Continued

Common name	Scientific or technical name of disease or its infectious agent
Tuberculosis — respiratory	Mycobacterium tuberculosis, Mycobacterium africanum, Mycobacterium canetti, Mycobacterium bovis
Tuberculosis — other	Mycobacterium species (non-tuberculosis)
Tularemia	Francisella tularensis
Typhoid Fever	Salmonella typhi
Typhus	Rickettsia species
Vancomycin Resistant Enterococci (VRE)	Enterococcus species with vancomycin resistance
Vancomycin Resistant Staphylococcus aureus (VRSA)	Staphylococcus aureus with vancomycin resistance
Verotoxin-producing organisms	Verotoxin-producing organisms
Vibrio parahaemolyticus	Vibrio parahaemolyticus
Viral Hemorrhagic Fever	Viral Hemorrhagic Fever
West Nile Virus (WNV)	West Nile virus
Western Equine Encephalitis	Western Equine Encephalitis virus
Yellow Fever	Yellow fever virus
Yersinia infections	Yersinia pseudotuberculosis, Yersinia enterocolitica

Government of Manitoba. Public Health Act. 2009. http://web2.gov.mb.ca/laws/regs/pdf/p210–037.09.pdf. Accessed January 26, 2012.

APPENDIX B: SELECTED NATIONAL AND INTERNATIONAL DATA QUALITY FRAMEWORKS

Canadian Institute for Health Information (CIHI) Data Quality Framework

The CIHI Data Quality Framework (see Figure B1.1) encompasses the following domains:

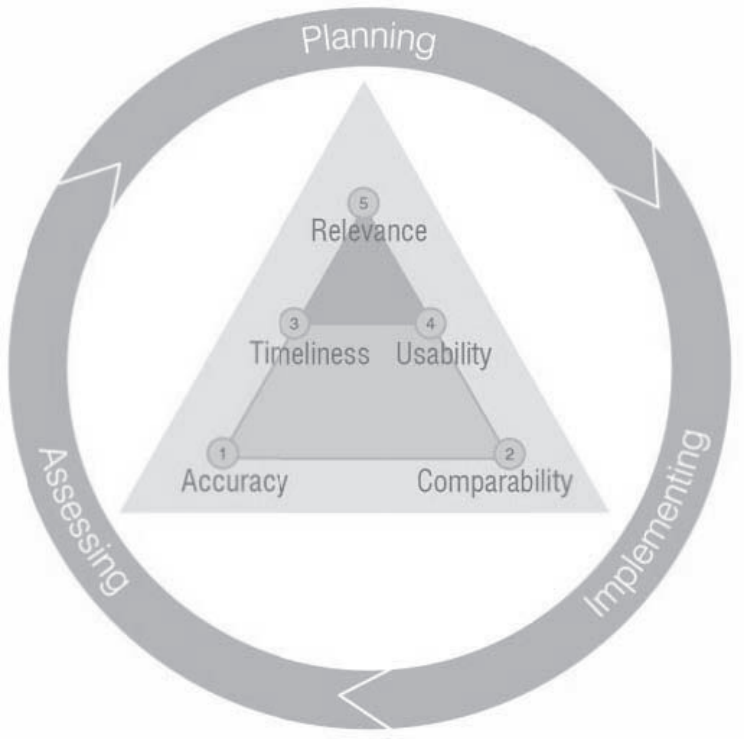
1. Accuracy: how well information in or derived from the data holding reflects the reality it was designed to measure.
2. Timeliness: how current or up to date the data is at the time of release, by measuring the gap between the end of the reference period to which the data pertains and the date on which the data becomes available to users.
3. Comparability: the extent to which databases are consistent over time and use standard conventions (such as data elements or reporting periods), making them comparable to other databases.
4. Usability: the ease with which a data holding's data may be understood and accessed.
5. Relevance**:** the degree to which a data holding meets the current and potential future needs of users.

The Framework is embedded within a data quality work cycle composed of three types of activities:

1. Planning: the activities necessary to prepare and prioritize the processes required for a data holding, as well as the design of any changes that are needed.
2. Implementing: developing the processes needed and applying them to the data holding (such as collecting data, monitoring incoming records and releasing written reports). The results of implementation activities for one process can be useful in the planning of similar future processes.
3. Assessing: evaluating the quality of data holding and determining if any changes to the processes are needed.

Canadian Institute for Health Information. The CIHI data quality framework, 2009. CIHI. 2009. Available from http://www.cihi.ca/CIHI–ext–portal/pdf/internet/DATA_QUALITY_FRAMEWORK_2009_EN. Accessed January 26, 2012.

Appendix Figure B.1: The CIHI Data Quality Framework



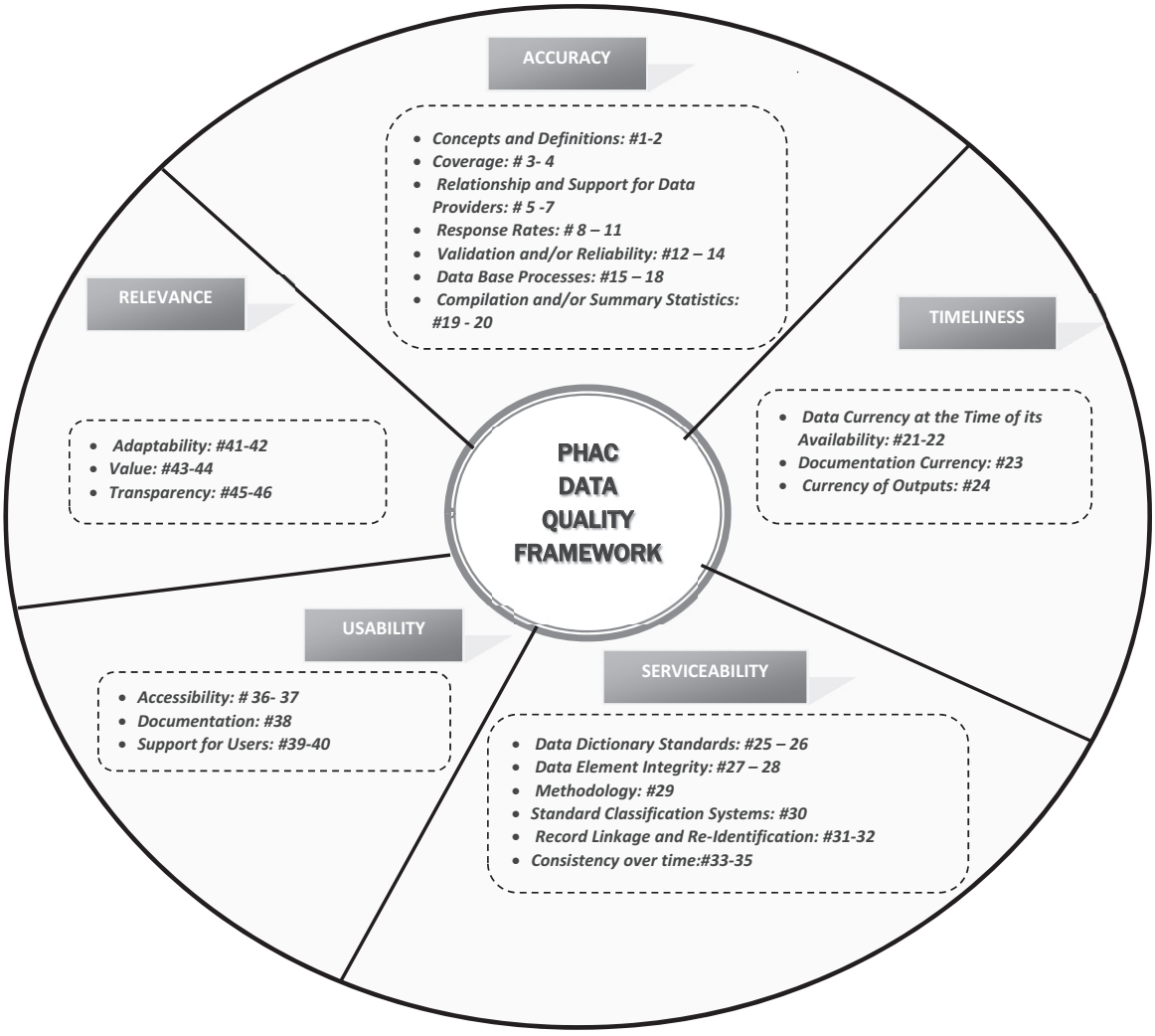
Source: CIHI, 2009

Public Health Agency of Canada (PHAC) Data Quality Framework

The PHAC Data Quality Framework consists of dimensions, characteristics, and criteria. Dimensions are the distinct components that encompass the broader definition of data quality. Each dimension is divided into related characteristics, and each characteristic is composed of several criteria. There are five dimensions, 22 characteristics, and 46 criteria. Descriptions of these three components are provided in Table B1.1.

Public Health Agency of Canada. PHAC data quality framework. Ottawa, ON: Public Health Agency of Canada. March 2009.

Appendix Figure B.2: PHAC Data Quality Framework



Source: PHAC, 2009

Appendix Table B.1: Components of the PHAC Data Quality Framework

Dimension & Characteristic	***Criteria***
ACCURACY	
1. Concepts and Definitions	1- Operational definitions provide reasonable representation of required subject matter concepts.
	2- The population of reference is explicitly documented.
2. Coverage	3- Routine assessment and monitoring is done to validate the list of all units providing data for the population of reference.
	4- Sources of under- and/or over-coverage are known and documented.
3. Relationship and Support for Data Providers	5- Practices exist that support data providers.
	6- Data capture and/or submission procedures are standardized.
	7-Follow-up is routinely done according to requirements.
4. Response Rates	8- Routine validation of the number of records sent and received is done.
	9- The magnitude of the record level response rates is explicitly documented.
	10 -Routine validation of response rates for key data elements is done.
	11- Procedures exist to follow-up on unexpected response rates for record and key data elements.
5. Validation and/or Reliability	12- Sources of bias and measurement (i.e., non-sampling) errors are known and documented.
	13- Routine checking (edit procedures) of data elements is performed for validity, consistency and logical relationships of related data items.
	14- Routine edit reports result from criterion 13.
6. Database Processes	15 Unmodified raw data are saved in a secure location.
	16- Data cleaning and/or imputation results from criterion 13.
	17 All data processes and systems procedures are documented and kept up to date.
	18 When database production procedures are modified, testing is done to ensure that the modification was done correctly.
7. Compilation and/or Summary Statistics	19- Procedures and techniques used to compile record level data into summary values are statistically sound.
	20- Routine assessment of summary statistics (and main intermediate results where needed) is performed.

Appendix Table B.1 – Continued

Dimension & Characteristic	***Criteria***
TIMELINESS	
8. Data Currency at the Time of its Availability	21- The difference between the reference period and the date of availability is reasonably brief and explicitly documented.
	22- Database procedures are regularly reviewed for timeliness.
9. Documentation Currency	23- Data quality documentation is available at the time the data is ready for use.
10. Currency of Outputs	24- Major outputs, including reports, summary tables and web releases are available on schedule.
SERVICEABILITY	
11. Data Dictionary Standards	25- Database dictionary exists and data elements conform to it.
	26- Data elements conform to PHAC Data Dictionary if and when it becomes available.
12. Data Element Integrity	27- Data are collected at the finest level of detail as is reasonable.
	28- For any derived data element, the original data element is also maintained on the main database.
13. Methodology	29- Appropriate methodology is used according to recognized best practices.
14. Standard Classification Systems	30- Standard classification coding systems are used according to recognized best practices.
15. Record Linkage and Re-Identification	31- Data elements that could be used to link other databases are identified.
	32- Data elements that could be used for re-identification purposes are identified.
16. Consistency over Time	33- Data are collected using a consistent collection period.
	34- All relevant changes in the database are evaluated, including changes to key data elements and methodology.
	35- Historical changes to the database are explicitly documented.
USABILITY	
17. Accessibility	36- Procedures exist for publicizing availability of major outputs to existing and potential users.
	37- Procedures exist for authorized access to micro-data.
18. Documentation	38- Up-to-date and pertinent documentation is available to users.
19. Support for Users	39- Different level of technical information is provided, according to the needs of the users.
	40- Feedback mechanisms exist for users to provide feedback and receive prompt and knowledgeable replies.

Appendix Table B.1 – Continued

Dimension & Characteristic	***Criteria***
RELEVANCE	
20. Adaptability	41- There are established processes in place for adapting to change, as needed.
	42- There are established processes in place to solicit input from stakeholders and major clients for possible improvements to the database.
21. Value	43- The objective of the database, filling a public health information need, is regularly assessed.
	44- The level of usage and satisfaction of users is regularly monitored.
22. Transparency	45- Outputs of the database are easily identifiable as such.
	46- Advance notice is given of major changes to the database.

Source: Public Health Agency of Canada, 2009

Statistics Canada Quality Assurance Framework

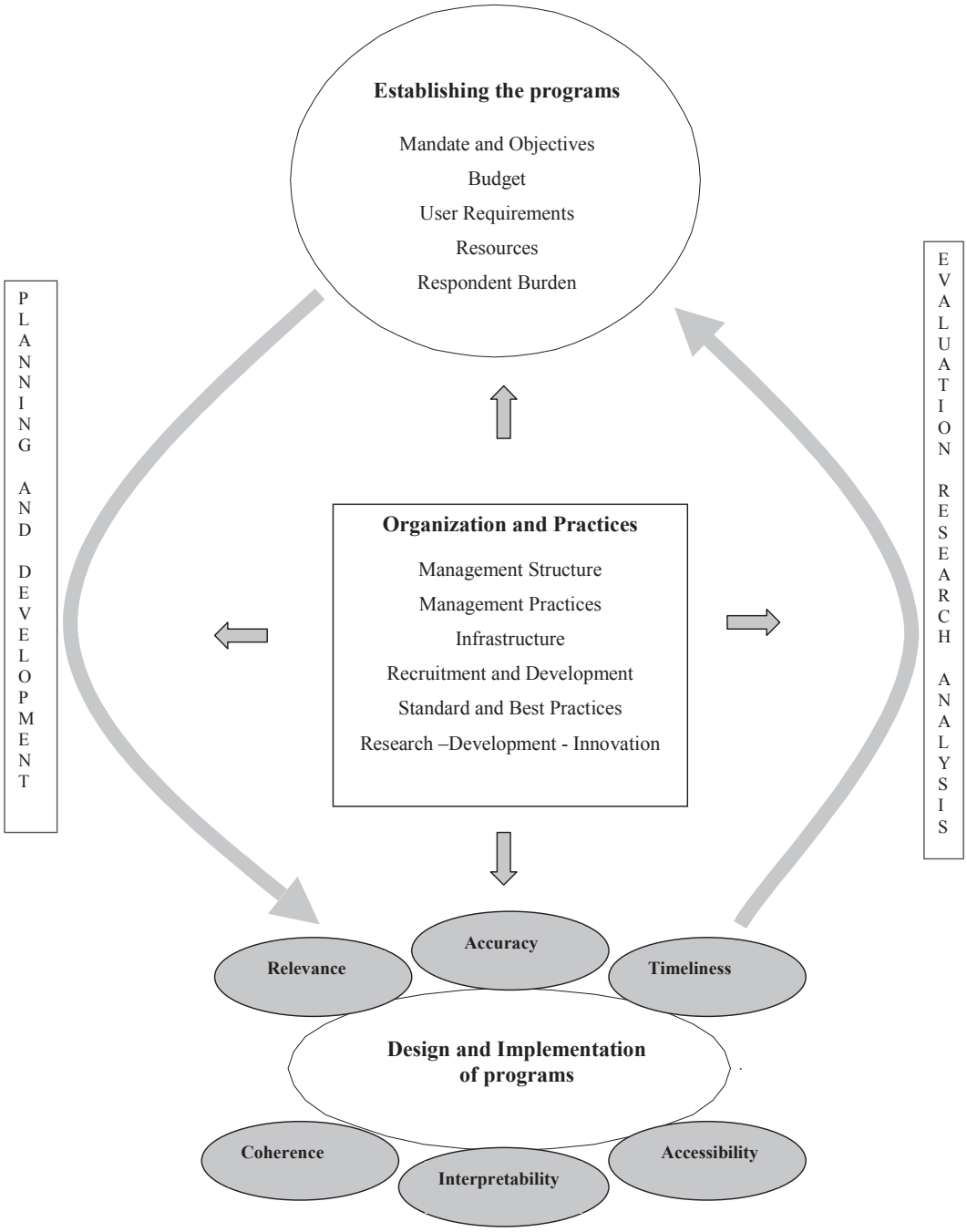
Statistics Canada defines information quality in terms of its fitness for use by its users. Quality is a multidimensional concept that includes the relevance of information to users' needs and characteristics of the information, such as accuracy, timeliness, accessibility, interpretability, and coherence that affects how it can be used.

Six dimensions of information quality are identified in the framework:

1. Relevance: reflects the degree to which the information meets the real needs of clients. It is concerned with whether the available information sheds light on the issues of most importance to users. Assessing relevance is a subjective matter dependent upon the varying needs of users.
2. Accuracy: the degree to which the information currently describes the phenomena it was designed to measure. It is usually characerized in terms of error in statistical estimates and is traditionally decomposed into bias (systematic error) and variance (random error) components. It may also be described in terms of the major sources of error that potentially cause inaccuracy (e.g., coverage, sampling, nonresponse, response).
3. Timeliness: refers to the delay between the reference point (or the end of the reference period) to which information pertains, and the date on which the information becomes available. It is typically involved in a trade–off against accuracy. The timeliness of information will influence its relevance.
4. Accessibility: the ease with which it can be obtained from the Agency. This includes the ease with which the existence of information can be ascertained, as well as the suitability of the form or medium through which the information can be accessed. The cost of the information may also be an aspect of accessibility for some users.
5. Interpretability: reflects the availability of the supplementary information and metadata necessary to interpret and use it appropriately. The information normally covers the underlying concepts, variables and classification used, the methodology of data collection and processing, and indications of the accuracy of the statistical information.
6. Coherence: the degree to which the information can be successfully brought together with other statistical information within a broad analytic framework and over time. The use of standard concepts, classifications, and target populations promotes coherence as does the use of common methodology across surveys. Coherence does not necessarily imply full numerical consistency.

Statistics Canada. Statistics Canada's quality assurance framework, 2002. Statistics Canada. 2002. http://publications.gc.ca/Collection/Statcan/12–586–XIE/12–586–XIE2002001.pdf. Accessed January 26, 2012.

Appendix Figure B.3: Statistics Canada Quality Assurance Framework



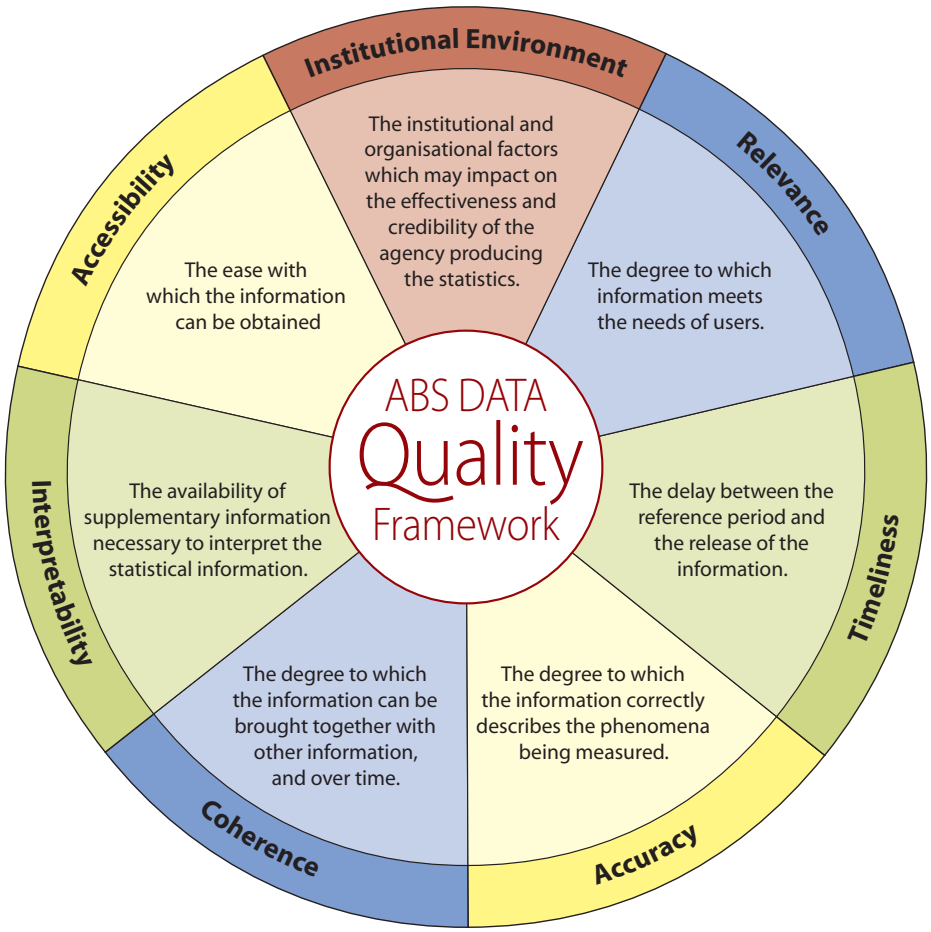
Source: Statistics Canada, 2002

Australian Bureau of Statistics DATAfitness: A Guide to Keeping Your Data in Good Shape

This Data Quality Framework contains seven dimensions, which encompass a broad range of concepts and assessment strategies. The dimensions are not necessarily equally weighted; importance will vary depending on the data source and its context.

Australian Bureau of Statistics. Datafitness: a guide to keeping your data in good shape. Australian National Statistical Service. 2009. Available from http://www.nss.gov.au/nss/home.nsf/0/c8805e7ccc865da3ca2575b4002024ed/$FILE/DataFitness%20A4%20Brochure%20single%20pages.pdf. Accessed January 26, 2012.

Appendix Figure B.4: Australian Bureau of Statistics Data Quality Framework



APPENDIX C: DESCRIPTION OF MACROS FOR DATA QUALITY EVALUATION

All macros were developed using SAS software (version 9.1). A description of each macro is provided below, along with the macro parameters (i.e., inputs and outputs) and some example code to implement the macros. All macros are available in the Manitoba Centre for Health Policy (MCHP) Concept Dictionary. http://umanitoba.ca/faculties/medicine/units/community_health_sciences/departmental_units/mchp/resources/concept_dictionary.html

VIMO Macro

Syntax: %VIMO (DS=);

Description: For a specified data file, this macro generates a table of valid, invalid, missing, and outlier (VIMO) observations. The table is in Excel format.

Parameters:
DS= Name of data file, which is in SAS format. This could be a temporary or permanent SAS dataset
INVALID= Option to turn invalid checks on or off (Default value=ON)
MEMNUM= List number of cluster members to include in the VIMO table. This parameter is not specified if there are no clusters.
MUNCODES= List of variables containing municipal codes, separated by blanks.
POSTALS= List of variables containing postal codes, separated by blanks.

Examples:

```
%VIMO (health.MHCPL_virustests_19922010);
%VIMO (DS = health.MHCPL_virustests_19922010,  INVALIDS = OFF);
%VIMO (DS = HEALTH.mhmed_1997apr,  MEMNUM = 23 24 25);
%VIMO (DS = social.hcm_edi_2006jan,  MEMNUM = 3,
        POSTALS = POSTAL p_code_original CL_POSTAL P_CODE P_CODE_E
        POSTAL_CODE POSTAL_CODE_HCM );
```

LINK Macro

Syntax: %LINK (DOMAIN=, DB=, PHIN=);

Description: For a series of data files, this macro creates a table (or members of a cluster) that calculates the linkability of individual data files based on the personal health information number (PHIN) in the Research Registry. The output is shown on screen and also saved in an Excel file. The macro will also generate a frequency table for PHIN types.

Parameters:
DOMAIN= Database domain
DB= Database prefix (or full name of cluster)
PHIN= Name of PHIN variable (Default=SCRPHIN)
TYPE= Name of PHINTYPE variable (Default=SCRPHINTYPE)

Example:

```
%LINK (health,MHCPL);
%LINK (health,MHCPL, PHIN=filephin);
%LINK (DOMAIN=social, DB=hcm_edi_2006jan, PHIN=FILEPHIN, TYPE=FILEPHINTYPE);
```

AGREEMENT Macro

Syntax: %AGREEMENT (DS=, REGYR=, SEX=, M=, F=, BIRTHDT=);

Description: This macro measures the agreement between a dataset and the Research Registry and produces kappa statistics for sex and date of birth.

Parameters:
DS= Name of dataset
REGYR= Latest available registry file (Default=2010)
PHIN= Variable containing PHIN (Default=SCRPHIN)
SEX= Variable containing sex (Default=SEX)
M= Numeric value assigned to males (Default=1)
F= Numeric value assigned to females (Default=2)
BIRTHDT= Variable containing date of birth (Default=BIRTHDT)

Example:
%AGREEMENT (DS=health.MHCPL_SPSECTION_19922010);
%AGREEMENT (DS=health.MHCPL_SPSECTION_19922010, REGYR=2009);

TREND Macro

Syntax: %TREND (DS=, STARTYR=, ENDYR=, BYDATE=, BYVAR=, BYFMT=, BYMONTH=);

Description: This macro conducts a trend analysis for a specified period of time. The results are summarized in a graphical format. The graph(s) are shown on screen and also saved in a PNG file.

Parameters:
DS= Name of dataset
STARTYR= Beginning fiscal year (1st part, 4–digit)
ENDYR= Ending fiscal year (1st part, 4–digit)
BYDATE= Desired date variable (Must be a SAS date)
BYVAR= An optional categorical variable to conduct stratified analyses. If omitted only one analysis is conducted for all records in the dataset.
BYFMT= An optional format for BYVAR.
BYMONTH= An optional parameter that will produce the analyses by month, instead of year, if assigned a value of YES (default value = NO).

Example:

```
%TREND (DS=health.wrha_ccic_med_2003mar,
        STARTYR=2003,
        ENDYR=2010,
        BYDATE=admit_dt,
        BYVAR=HOSP);
%TREND (DS=health.MHCPL_virustests_19922010,
        STARTYR=1992,
        ENDYR=2009,
        BYDATE=RECEIVEDDT);
```

```
%TREND (DS=health.MHCPL_virustests_19922010,
        STARTYR=1992,
        ENDYR=2009,
        BYDATE=RECEIVEDDT
        BYFMT=$HOSPFMTL.);
%TREND (DS=health.MHCPL_virustests_19922010,
        STARTYR=1992,
        ENDYR=2009,
        BYDATE=RECEIVEDDT
        BYMONTH=YES);
```

CONTENTS Macro

Syntax: %CONTENTS (DOMAIN=, DB);

Description: This macro runs PROC CONTENTS for a series of tables within a specified Domain and Database and generates a single overview table. Please refer to Table 4.1 for an example of the format of the overview table.

Parameters:
DOMAIN= Database domain on SPDS
DB= Database prefix

Example:
%CONTENTS (HEALTH,BMD);

APPENDIX D: SUPPLEMENTARY DATA QUALITY TABLES

Appendix Table D.1: Valid, Invalid, and Missing Data for Clinical Microbiology Requisitions Data File, 1992/93–2009/10

Type	Variable Name	Variable Label	% Valid	% Invalid	% Missing
Identification	FILEPHIN	MH Scrambled PHIN	100.00	0.00	0.00
	SCRPHIN	MCHP Scrambled PHIN	100.00	0.00	0.00
Numeric	NCMORGANISMS	N CM organisms sections this requisition	100.00	0.00	0.00
	NCMRESULTS	N CM result sections this requisition	100.00	0.00	0.00
Character	ACUTECONVAL	Acute/Convalescence	0.00	0.00	**100.00**
	CHANGEDDEMODATACD	Changed Demo Data Code	12.60	0.00	**87.40**
	COMPLETEDIND	Completed Indicator	100.00	0.00	0.00
	CPLPATIENTNUMBER	CPL Patient Number	100.00	0.00	0.00
	FILEPHINTYPE	Method to determine FILEPHIN	100.00	0.00	0.00
	FREEFORMCOMMENT	Freeform Comment	5.66	0.00	**94.34**
	HOSPITALCLINIC	Hospital/Clinic	51.38	0.00	**48.62**
	INSUFFICIENTINFOCD	Insufficient Information Code	0.00	0.00	**100.00**
	LABNO	Laboratory Number	100.00	0.00	0.00
	LASTUSEDSEQ	Last Used Sequence	100.00	0.00	0.00
	MATCHREQNUMBER	Match Requisition Number	0.00	0.00	**100.00**
	MHREGION	MH Region Code	99.62	0.00	0.38
	MORETHAN18TESTS	More Than 18 Tests	0.00	0.00	**100.00**
	MUNCODE	Municipal Code	94.63	0.00	**5.37**
	PHYSICIANNUMBER	Physician Number	87.74	0.00	**12.26**
	POSTAL	Patient Postal Code	94.62	0.00	**5.37**
	QCREQUISITION	Quality Control Requisition	0.00	0.00	**100.00**
	RECSEQUENCE	Record Sequence	100.00	0.00	0.00
	REFERFACIL	Referring Facility	94.18	0.00	**5.82**
	REPORTCOMMENTCD1	Report Comment Code 1st	5.50	0.00	**94.50**
	REPORTCOMMENTCD2	Report Comment Code 2nd	0.22	0.00	**99.78**
	REPORTCOMMENTCD3	Report Comment Code 3rd	0.02	0.00	**99.98**
	REPORTCOMMENTSTATUS1	Report Comment Status 1st	5.49	0.01	**94.50**
	REPORTCOMMENTSTATUS2	Report Comment Status 2nd	0.22	0.00	**99.78**
	REPORTCOMMENTSTATUS3	Report Comment Status 3rd	0.02	0.00	**99.98**
	REQCOMMENT	Requisition Comment	3.96	0.00	**96.04**
	REQCOMMENTSTATUS	Requisition Comment Status	100.00	0.00	0.00
	REQUISITIONNUMBER	Requisition Number	100.00	0.00	0.00
	RHAMUNCODE	Regional Health Authority of MUNCODE	99.86	0.00	0.14
	RHAPOSTAL	RHA of Postal Code	99.86	0.00	0.14
	SCRPHINTYPE	Method to determine SCRPHIN	100.00	0.00	0.00
	SECTION	Section	100.00	0.00	0.00
	SEX	Sex of Patient	100.00	0.00	0.00
	SPECIALSTUDYCD	Special Study Code	0.00	1.08	**98.92**
	STATREQUISITION	Stat Requisition	0.01	0.00	**99.99**
	SUBSECTION01	Subsection 01	35.40	0.00	**64.60**
	SUBSECTION02	Subsection 02	0.00	0.00	**100.00**
	SUBSECTION03	Subsection 03	0.00	0.00	**100.00**
	SUBSECTION04	Subsection 04	0.00	0.00	**100.00**
	SUBSECTION05	Subsection 05	0.00	0.00	**100.00**
	SUBSECTION06	Subsection 06	0.00	0.00	**100.00**
	SUBSECTION07	Subsection 07	0.00	0.00	**100.00**
	SUBSECTION08	Subsection 08	0.00	0.00	**100.00**
	SUBSECTION09	Subsection 09	0.00	0.00	**100.00**
	SUBSECTION10	Subsection 10	0.00	0.00	**100.00**
	TESTCOUNT	Test Count	100.00	0.00	0.00
	UNINSUREDSERVICESCD	Uninsured Services Code	2.87	0.00	**97.13**

Appendix Table D.1 - Continued

Type	Variable Name	Variable Label	% Valid	% Invalid	% Missing
Date	ACQDT	Date record was acquired by MCHP	100.00	0.00	0.00
	BIRTHDT	Birth Date	100.00	0.00	0.00
	RECEIVEDDT	Received Date	100.00	0.00	0.00
	REPORTCOMMENTDT1	Report Comment Date 1st	5.50	0.00	**94.50**
	REPORTCOMMENTDT2	Report Comment Date 2nd	0.22	0.00	**99.78**
	REPORTCOMMENTDT3	Report Comment Date 3rd	0.02	0.00	**99.98**
	REQCOMMENTREPORTDT	Requisition Comment Report Date	3.43	0.00	**96.57**
	SPECIMENDT	Specimen Date	95.69	0.00	*4.31*
	STATDT	Status Date	100.00	0.00	0.00

Legend for Invalid and Missing Columns:
Regular font: None or Minimal; *Italics: Moderate* **; Bold: Significant**

Appendix Table D.2: Valid, Invalid, and Missing Data for Clinical Microbiology Results Data File, 1992/93–2009/10

Type	Variable Name	Variable Label	% Valid	% Invalid	% Missing
Identification	FILEPHIN	MH Scrambled PHIN	100.00	0.00	0.00
	SCRPHIN	MCHP Scrambled PHIN	100.00	0.00	0.00
Numeric	RECPOSN	Position of this record in requisition	100.00	0.00	0.00
Character	CMTESTTYPE	CM Test Type	100.00	0.00	0.00
	CPLPATIENTNUMBER	CPL Patient Number	100.00	0.00	0.00
	FILEPHINTYPE	Method to determine FILEPHIN	100.00	0.00	0.00
	POSNEG	Positive-Negative	70.36	0.00	**29.64**
	RECSEQUENCE	Record Sequence	100.00	0.00	0.00
	RECTYPE	Record Type	100.00	0.00	0.00
	REFERFACIL	Referring Facility	94.25	0.00	**5.75**
	REFEROUT	Referred Out	0.00	0.00	**100.00**
	REQUISITIONNUMBER	Requisition Number	100.00	0.00	0.00
	RESULTS1	CM type 1 Result 1 of 6	100.00	0.00	0.00
	RESULTS2	CM type 1 Result 2 of 6	39.59	0.00	**60.41**
	RESULTS3	CM type 1 Result 3 of 6	2.17	0.00	**97.83**
	RESULTS4	CM type 1 Result 4 of 6	0.48	0.00	**99.52**
	RESULTS5	CM type 1 Result 5 of 6	0.21	0.00	**99.79**
	RESULTS6	CM type 1 Result 6 of 6	0.06	0.00	**99.94**
	SCRPHINTYPE	Method to determine SCRPHIN	100.00	0.00	0.00
	SECTION	Section	100.00	0.00	0.00
	SPECIMENSOURCE	Specimen Source	100.00	0.00	0.00
	STATUS	Status	100.00	0.00	0.00
	TECHINIT	Technician Initials	99.97	0.00	0.03
	TESTSUBSECTION	Test Subsection	32.80	0.00	**67.20**
	VERIFIED	Verified	99.97	0.00	0.03
Date	ACQDT	Date record was acquired by MCHP	100.00	0.00	0.00
	RECEIVEDDT	Received Date	100.00	0.00	0.00
	REPORTDT	Report Date	100.00	0.00	0.00
	SPECIMENDT	Specimen Date	95.29	0.00	*4.71*

Legend for Invalid and Missing Columns:
Regular font: None or Minimal; *Italics: Moderate* **; Bold: Significant**

Appendix Table D.3: Valid, Invalid, and Missing Data for Parasitology and Serology Requisitions Data File, 1992/93–2009/10

Type	Variable Name	Variable Label	% Valid	% Invalid	% Missing
Identification	FILEPHIN	MH SCRAMBLED PHIN	100.00	0.00	0.00
	SCRPHIN	MCHP SCRAMBLED PHIN	100.00	0.00	0.00
Numeric	NPARASITOLOGY	N of Parasitology tests this requisition	100.00	0.00	0.00
	NSEROLOGY	N of Serology tests this requisition	100.00	0.00	0.00
	TESTCOUNT	Test Count	100.00	0.00	0.00
Character	ACUTECONVAL	Acute/Convalescent	2.22	0.00	**97.78**
	CHANGEDDEMODATACD	Changed Demo Data Code	14.49	0.00	**85.51**
	COMPLETEDIND	Completed Indicator	100.00	0.00	0.00
	CPLPATIENTNUMBER	CPL Patient Number	100.00	0.00	0.00
	FILEPHINTYPE	Method to determine FILEPHIN	100.00	0.00	0.00
	FREEFORMCOMMENT	Freeform Comment	8.21	0.00	**91.79**
	HISTORYCLINICALDIAG	History/Clinical Diagnosis	2.08	0.00	**97.92**
	HIVPREVHISTRESULT	HIV Previous History-Result	0.00	0.00	**100.00**
	HOSPCLINIC	Hospital/Clinic #	49.27	0.00	**50.73**
	INSUFFICIENTINFOCD	Insufficient Information Code	0.00	0.00	**100.00**
	LABNO	Laboratory Number	100.00	0.00	0.00
	LASTUSEDSEQ	Last Used Sequence	100.00	0.00	0.00
	MATCHREQNUMBER	Match Requisition Number	0.74	0.00	**99.26**
	MHREGION	MH Region Code	84.97	0.00	**15.03**
	MORETHAN18TESTS	More Than 18 Tests	0.00	0.00	**100.00**
	MUNCODE	Municipal Code	75.06	0.00	**24.94**
	PATIENTCATEGORY	Patient Category	100.00	0.00	0.00
	PHYSICIANNUMBER	Physician Number	92.41	0.00	**7.59**
	POSTAL	Patient Postal Code	75.05	0.01	**24.95**
	PREVHIST	Previous History	30.49	0.00	**69.51**
	QCREQUISITION	Quality Control Requisition	0.06	0.00	**99.94**
	RACKNUMBER	Rack Number	24.89	0.00	**75.11**
	RACKSPECIMENNUMBER	Rack Specimen Number	24.89	0.00	**75.11**
	RACKSUBSECTION	Rack Subsection	24.89	0.00	**75.11**
	RECSEQUENCE	Record Sequence	100.00	0.00	0.00
	REFERFACIL	Referring Facility	90.83	0.00	**9.17**
	REPORTCOMMENTCD1	Report Comment Code	12.23	0.00	**87.77**
	REPORTCOMMENTCD2	Report Comment Code	2.70	0.00	**97.30**
	REPORTCOMMENTCD3	Report Comment Code	0.30	0.00	**99.70**
	REPORTCOMMENTSTAT1	Report Comment Status	12.20	0.04	**87.77**
	REPORTCOMMENTSTAT2	Report Comment Status	2.69	0.01	**97.30**
	REPORTCOMMENTSTAT3	Report Comment Status	0.30	0.00	**99.70**
	REQCOMMENT	Requisition Comment	13.22	0.00	**86.78**
	REQCOMMENTSTAT	Requisition Comment Status	100.00	0.00	0.00
	REQUISITIONNUMBER	Requisition Number	100.00	0.00	0.00
	RHAMUNCODE	Regional Health Authority of MUN	94.66	0.00	**5.34**
	RHAPOSTAL	RHA of Postal Code	94.66	0.00	**5.34**
	RISKGROUPS1	Risk Groups 1	19.24	0.00	**80.76**
	RISKGROUPS2	Risk Groups 2	3.74	0.00	**96.26**
	RISKGROUPS3	Risk Groups 3	1.52	0.00	**98.48**
	SCRPHINTYPE	Method to determine SCRPHIN	100.00	0.00	0.00
	SECTION	Section	100.00	0.00	0.00
	SEX	Sex of Patient	100.00	0.00	0.00
	SPECIALSTUDYCD	Special Study Code	0.00	*3.54*	**96.46**
	SPECIMENSOURCE	Specimen Source on test	100.00	0.00	0.00

Appendix Table D.3 – Continued

Type	Variable Name	Variable Label	% Valid	% Invalid	% Missing
	STATREQUISITION	Stat Requisition	0.31	0.00	**99.69**
	SUBSECTION01	Subsection 01	100.00	0.00	0.00
	SUBSECTION02	Subsection 02	15.54	0.00	**84.46**
	SUBSECTION03	Subsection 03	4.06	0.00	**95.94**
	SUBSECTION04	Subsection 04	0.63	0.00	**99.37**
	SUBSECTION05	Subsection 05	0.09	0.00	**99.91**
	SUBSECTION06	Subsection 06	0.04	0.00	**99.96**
	SUBSECTION07	Subsection 07	0.00	0.00	**100.00**
	SUBSECTION08	Subsection 08	0.00	0.00	**100.00**
	SUBSECTION09	Subsection 09	0.00	0.00	**100.00**
	SUBSECTION10	Subsection 10	0.00	0.00	**100.00**
	SYMPTOMS1	Symptoms 1	28.13	0.00	**71.87**
	SYMPTOMS2	Symptoms 2	2.43	0.00	**97.57**
	SYMPTOMS3	Symptoms 3	0.45	0.00	**99.55**
	SYMPTOMS4	Symptoms 4	0.11	0.00	**99.89**
	SYMPTOMS5	Symptoms 5	0.03	0.00	**99.97**
	TESTSELECTION	Test Selection	99.98	0.00	0.02
	UNINSUREDSERVICESCD	Uninsured Services Code	3.79	0.00	**96.21**
	VIRALLOADANTIRETROVIRALMEDS1	Viral Load Antiretroviral Medications	0.63	0.00	**99.37**
	VIRALLOADANTIRETROVIRALMEDS2	Viral Load Antiretroviral Medications	0.61	0.00	**99.39**
	VIRALLOADANTIRETROVIRALMEDS3	Viral Load Antiretroviral Medications	0.55	0.00	**99.45**
	VIRALLOADANTIRETROVIRALMEDS4	Viral Load Antiretroviral Medications	0.15	0.00	**99.85**
	VIRALLOADREASON	Viral Load Reason	0.88	0.00	**99.12**
	VIRALLOADTYPE	Viral Load Type	0.90	0.00	**99.10**
	VLCD4COUNT	VL CD4 Count	0.87	0.00	**99.13**
Date	ACQDT	Date record was acquired at MCHP	100.00	0.00	0.00
	BIRTHDT	Birth Date	79.41	0.00	**20.59**
	HIVPREVHISTREPORTDT	HIV Previous History-Report Date	0.00	0.00	**100.00**
	RECEIVEDDT	Received Date	100.00	0.00	0.00
	REPORTCOMMENTDT1	Report Comment Date	12.23	0.00	**87.77**
	REPORTCOMMENTDT2	Report Comment Date	2.70	0.00	**97.30**
	REPORTCOMMENTDT3	Report Comment Date	0.30	0.00	**99.70**
	REQCOMMENTREPORTDT	Requisition Comment Report Date	12.23	0.00	**87.77**
	SPECIMENDT	Specimen Date	92.72	0.00	**7.28**
	STATDT	Status Date	100.00	0.00	0.00
	VLCD4DT	VL CD4 Date	0.79	0.00	**99.21**
	VLLASTTESTDT	VL Last Test Date	0.84	0.00	**99.16**

Legend for Invalid and Missing Columns:
Regular font: None or Minimal; *Italics: Moderate* **; Bold: Significant**

Appendix Table D.4: Valid, Invalid, and Missing Data for Parasitology Results Data File, 1992/93–2009/10

Type	Variable Name	Variable Label	% Valid	% Invalid	% Missing
Identification	FILEPHIN	MH SCRAMBLED PHIN	100.00	0.00	0.00
	SCRPHIN	MCHP SCRAMBLED PHIN	100.00	0.00	0.00
Numeric	RECPOSN	Test Subsection position this requisition	100.00	0.00	0.00
Character	CPLPATIENTNUMBER	CPL Patient Number	100.00	0.00	0.00
	FILEPHINTYPE	Method to determine FILEPHIN	100.00	0.00	0.00
	INTERPRETATION1	Interpretation	2.39	0.00	**97.61**
	INTERPRETATION2	Interpretation	0.73	0.00	**99.27**
	INTERPRETATION3	Interpretation	0.23	0.00	**99.77**
	INTERPRETATION4	Interpretation	0.08	0.00	**99.92**
	INTERPRETATION5	Interpretation	0.02	0.00	**99.98**
	PARASITOLOGYPARASITE1	Parasitology Parasite	2.73	0.00	**97.27**
	PARASITOLOGYPARASITE2	Parasitology Parasite	1.09	0.00	**98.91**
	PARASITOLOGYPARASITE3	Parasitology Parasite	0.31	0.00	**99.69**
	PARASITOLOGYPARASITE4	Parasitology Parasite	0.10	0.00	**99.90**
	PARASITOLOGYPARASITE5	Parasitology Parasite	0.03	0.00	**99.97**
	PARASITOLOGYRESULT1	Parasitology Result	100.00	0.00	0.00
	PARASITOLOGYRESULT2	Parasitology Result	2.20	0.00	**97.80**
	PARASITOLOGYRESULT3	Parasitology Result	0.64	0.00	**99.36**
	PARASITOLOGYRESULT4	Parasitology Result	0.20	0.00	**99.80**
	PARASITOLOGYRESULT5	Parasitology Result	0.06	0.00	**99.94**
	PARASITOLOGYTEST	Parasitology Test	100.00	0.00	0.00
	POSNEG	Positive-Negative	100.00	0.00	0.00
	RECSEQUENCE	Record Sequence	100.00	0.00	0.00
	REFERFACIL	Referring Facility	96.41	0.00	*3.59*
	REFEROUT	Referred Out	0.00	0.00	**100.00**
	REQUISITIONNUMBER	Requisition Number	100.00	0.00	0.00
	SCRPHINTYPE	Method to determine SCRPHIN	100.00	0.00	0.00
	SECTION	Section	100.00	0.00	0.00
	SPECIMENSOURCE	Specimen Source on test	100.00	0.00	0.00
	STATUS	Status	100.00	0.00	0.00
	TECHINIT	Technician Initials	100.00	0.00	0.00
	TESTSEQUENCE	Test Sequence	100.00	0.00	0.00
	TESTSUBSECTION	Test Subsection	100.00	0.00	0.00
	VERIFIED	Verified	100.00	0.00	0.00
Date	ACQDT	Date record was acquired at MCHP	100.00	0.00	0.00
	PARASITOLOGYREPORTDT	Parasitology Report Date	100.00	0.00	0.00
	RECEIVEDDT	Received Date	100.00	0.00	0.00
	SPECIMENDT	Specimen Date	89.06	0.00	**10.94**
	STATDT	Status Date	100.00	0.00	0.00

Legend for Invalid and Missing Columns:
Regular font: None or Minimal; *Italics: Moderate* **; Bold: Significant**

Appendix Table D.5: Valid, Invalid, and Missing Data for Serology Results Data File, 1992/93–2009/10

Type	Variable Name	Variable Label	% Valid	% Invalid	% Missing
Identification	FILEPHIN	MH SCRAMBLED PHIN	100.00	0.00	0.00
	SCRPHIN	MCHP SCRAMBLED PHIN	100.00	0.00	0.00
Numeric	RECPOSN	Test Subsection position this requisition	100.00	0.00	0.00
Character	CPLPATIENTNUMBER	CPL Patient Number	100.00	0.00	0.00
	FILEPHINTYPE	Method to determine FILEPHIN	100.00	0.00	0.00
	GROUPTESTCD	Group Test Code	1.26	0.00	**98.74**
	GROUPTESTIND	Group Test Indicator	97.68	0.00	*2.32*
	HOLDSHEETPRINTED	Hold Sheet Printed	2.13	0.00	**97.87**
	INTERPRETATION	Interpretation	6.30	0.00	**93.70**
	POSNEG	Positive-Negative	100.00	0.00	0.00
	QCINTERPRETATION	Quality Control Interpretation	0.00	0.00	**100.00**
	QCQUANTITY	Quality Control Quantity	0.00	0.00	**100.00**
	QCRESULT	Quality Control Result	0.00	0.00	**100.00**
	QUANTITY	Quantity	4.81	0.00	**95.19**
	RACKINDICATOR	Rack Indicator	100.00	0.00	0.00
	RECSEQUENCE	Record Sequence	100.00	0.00	0.00
	REFERFACIL	Referring Facility	90.38	0.00	**9.62**
	REFEROUT	Referred Out	0.00	0.00	**100.00**
	REQUISITIONNUMBER	Requisition Number	100.00	0.00	0.00
	SCRPHINTYPE	Method to determine SCRPHIN	100.00	0.00	0.00
	SECTION	Section	100.00	0.00	0.00
	SEROLOGYTEST	Serology Test	100.00	0.00	0.00
	SPECIMENSOURCE	Specimen Source on test	100.00	0.00	0.00
	SPRESULT	SP Result	100.00	0.00	0.00
	STATUS	Status	100.00	0.00	0.00
	STORAGE	Storage	2.95	0.00	**97.05**
	TECHINIT	Technician Initials	100.00	0.00	0.00
	TESTSEQUENCE	Test Sequence	100.00	0.00	0.00
	TESTSUBSECTION	Test Subsection	100.00	0.00	0.00
	VERIFIED	Verified	100.00	0.00	0.00
Date	ACQDT	Date record was acquired at MCHP	100.00	0.00	0.00
	RECEIVEDDT	Received Date	100.00	0.00	0.00
	REPORTDT	Report Date	100.00	0.00	0.00
	SPECIMENDT	Specimen Date	93.44	0.00	**6.56**
	STATDT	Status Date	100.00	0.00	0.00

Legend for Invalid and Missing Columns:
Regular font: None or Minimal; *Italics: Moderate* **; Bold: Significant**

Appendix Table D.6: Valid, Invalid, and Missing Data for Virus Detection Requisitions Data File, 1992/93–2009/10

Type	Variable Name	Variable Label	% Valid	% Invalid	% Missing
Identification	FILEPHIN	MH SCRAMBLED PHIN	100.00	0.00	0.00
	SCRPHIN	MCHP SCRAMBLED PHIN	100.00	0.00	0.00
Numeric	NVIRUSDETECTION	N of Virus Detection tests this requisition	100.00	0.00	0.00
Character	ACUTECONVAL	Acute/Convalescence	0.00	0.00	**100.00**
	CHANGEDDEMODATACD	Changed Demo Data Code	13.96	0.00	**86.04**
	COMPLETEDIND	Completed Indicator	100.00	0.00	0.00
	CPLPATIENTNUMBER	CPL Patient Number	100.00	0.00	0.00
	FILEPHINTYPE	Method to determine FILEPHIN	100.00	0.00	0.00
	FREEFORMCOMMENT	Freeform Comment	8.22	0.00	**91.78**
	HOSPCLINIC	Hospital/Clinic #	62.95	0.00	**37.05**
	INSUFFICIENTINFOCD	Insufficient Information Code	0.00	0.00	**100.00**
	LABNO	Laboratory Number	100.00	0.00	0.00
	LASTUSEDSEQ	Last Used Sequence	100.00	0.00	0.00
	MATCHREQNUMBER	Match Requisition Number	0.00	0.00	**100.00**
	MHREGION	MH Region Code	99.58	0.00	0.42
	MORETHAN18TESTS	More Than 18 Tests	0.00	0.00	**100.00**
	MUNCODE	Municipal Code	85.31	0.00	**14.69**
	PHYSICIANNUMBER	Physician Number	89.46	0.00	**10.54**
	POSTAL	Patient Postal Code	85.31	0.00	**14.69**
	QCREQUISITION	Quality Control Requisition	0.00	0.00	**100.00**
	RECSEQUENCE	Record Sequence	100.00	0.00	0.00
	REFERFACIL	Referring Facility	95.44	0.00	*4.56*
	REPORTCOMMENTCD1	Report Comment Code 1st	12.93	0.00	**87.07**
	REPORTCOMMENTCD2	Report Comment Code 2nd	3.88	0.00	**96.12**
	REPORTCOMMENTCD3	Report Comment Code 3rd	0.17	0.00	**99.83**
	REPORTCOMMENTSTAT1	Report Comment Status 1st	12.86	0.07	**87.07**
	REPORTCOMMENTSTAT2	Report Comment Status 2nd	3.87	0.01	**96.12**
	REPORTCOMMENTSTAT3	Report Comment Status 3rd	0.17	0.00	**99.83**
	REQCOMMENT	Requisition Comment	3.57	0.00	**96.43**
	REQCOMMENTSTAT	Requisition Comment Status	100.00	0.00	0.00
	REQUISITIONNUMBER	Requisition Number	100.00	0.00	0.00
	RHAMUNCODE	Regional Health Authority of MUNCODE	99.88	0.00	0.12
	RHAPOSTAL	RHA of Postal Code	99.88	0.00	0.12
	SCRPHINTYPE	Method to determine SCRPHIN	100.00	0.00	0.00
	SECTION	Section	100.00	0.00	0.00
	SEX	Sex of Patient	100.00	0.00	0.00
	SPECIALSTUDYCD	Special Study Code	0.00	*4.90*	**95.10**
	STATREQUISITION	Status of Requisition	0.05	0.00	**99.95**
	TESTCOUNT	Test Count	100.00	0.00	0.00
	UNINSUREDSERVICESCD	Uninsured Services Code	5.02	0.00	**94.98**
Date	ACQDT	Date record was acquired at MCHP	100.00	0.00	0.00
	BIRTHDT	Birth Date	100.00	0.00	0.00
	RECEIVEDDT	Received Date	100.00	0.00	0.00
	REPORTCOMMENTDT1	Report Comment Date 1st	12.93	0.00	**87.07**
	REPORTCOMMENTDT2	Report Comment Date 2nd	3.88	0.00	**96.12**
	REPORTCOMMENTDT3	Report Comment Date 3rd	0.17	0.00	**99.83**
	REQCOMMENTREPORTDT	Requisition Comment Report Date	2.87	0.00	**97.13**
	SPECIMENDT	Specimen Date	92.71	0.00	**7.29**
	STATDT	Status Date	100.00	0.00	0.00

Legend for Invalid and Missing Columns:
Regular font: None or Minimal; *Italics: Moderate* **; Bold: Significant**

Appendix Table D.7: Valid, Invalid, and Missing Data for Virus Detection Results Data File, 1992/93–2009/10

Type	Variable Name	Variable Label	% Valid	% Invalid	% Missing
Identification	FILEPHIN	MH SCRAMBLED PHIN	100.00	0.00	0.00
	SCRPHIN	MCHP SCRAMBLED PHIN	100.00	0.00	0.00
Numeric	RECPOSN	Position this Test in Virus requisition	100.00	0.00	0.00
Character	CPLPATIENTNUMBER	CPL Patient Number	100.00	0.00	0.00
	FILEPHINTYPE	Method to determine FILEPHIN	100.00	0.00	0.00
	POSNEG	Positive-Negative	100.00	0.00	0.00
	RECORDTYPE	Record Type	100.00	0.00	0.00
	RECSEQUENCE	Record Sequence	100.00	0.00	0.00
	REFERFACIL	Referring Facility	96.03	0.00	*3.97*
	REFEROUT	Referred Out	0.00	0.00	**100.00**
	REQUISITIONNUMBER	Requisition Number	100.00	0.00	0.00
	SCRPHINTYPE	Method to determine SCRPHIN	100.00	0.00	0.00
	SECTION	Section	100.00	0.00	0.00
	SEROTYPE	Serology Type	1.39	0.00	**98.61**
	SPECIMENSOURCE	Specimen Source	100.00	0.00	0.00
	STATUS	Status	100.00	0.00	0.00
	TECHINIT	Technician Initials	100.00	0.00	0.00
	TESTSEQUENCE	Test Sequence	100.00	0.00	0.00
	TESTSUBSECTION	Test Subsection	0.00	0.00	**100.00**
	VERIFIED	Verified	100.00	0.00	0.00
	VIRRESULT	Viral Result	90.51	0.00	**9.49**
	VIRTESTTYPE	Viral Test Type	100.00	0.00	0.00
	VIRUS	Virus	9.53	0.00	**90.47**
	VTREQUISITIONNUMBER	Requisition Number on virus test	100.00	0.00	0.00
Date	ACQDT	Date record was acquired at MCHP	100.00	0.00	0.00
	RECEIVEDDT	Received Date	100.00	0.00	0.00
	REPORTDT	Report Date	100.00	0.00	0.00
	SPECIMENDT	Specimen Date	92.05	0.00	**7.95**
	STATDT	Status Date	100.00	0.00	0.00

Legend for Invalid and Missing Columns:
Regular font: None or Minimal; *Italics: Moderate* **; Bold: Significant**

Appendix Table D.8: Description of CPL Clinical Microbiology Section Organisms, 1992–2010

Type	Variable Name	Variable Label	Minimum	Maximum	Mean	Standard Deviation	Outlier	Comment
Identification	FILEPHIN	MH Scrambled PHIN						
	SCRPHIN	MCHP Scrambled PHIN						
Numeric	RECPOSN	Position of this record in requisition	1	5	1.88	1.03	2.61	
Character	ANTIBIOTIC01	CM type 2 -Antibiotic 01	AM					
	ANTIBIOTIC02	CM type 2 -Antibiotic 02	CL					
	ANTIBIOTIC03	CM type 2 -Antibiotic 03	GM					
	ANTIBIOTIC04	CM type 2 -Antibiotic 04	TE					
	ANTIBIOTIC05	CM type 2 -Antibiotic 05	FD					
	ANTIBIOTIC06	CM type 2 -Antibiotic 06	SXT					
	ANTIBIOTIC07	CM type 2 -Antibiotic 07	C					
	ANTIBIOTIC08	CM type 2 -Antibiotic 08	E					
	ANTIBIOTIC09	CM type 2 -Antibiotic 09	CM					
	ANTIBIOTIC10	CM type 2 -Antibiotic 10	CLX, SO					Invalid Codes: SO
	ANTIBIOTIC11	CM type 2 -Antibiotic 11	P					
	ANTIBIOTIC12	CM type 2 -Antibiotic 12	NN					
	ANTIBIOTIC13	CM type 2 -Antibiotic 13	AN					
	ANTIBIOTIC14	CM type 2 -Antibiotic 14	FOX					
	ANTIBIOTIC15	CM type 2 -Antibiotic 15	CRO					
	ANTIBIOTIC16	CM type 2 -Antibiotic 16	VA					
	ANTIBIOTIC17	CM type 2 -Antibiotic 17	CZ					
	ANTIBIOTIC18	CM type 2 -Antibiotic 18	NOR, TIC					Invalid Codes: TIC
	ANTIBIOTIC19	CM type 2 -Antibiotic 19	CEF					
	ANTIBIOTIC20	CM type 2 -Antibiotic 20	CIP, NOR					
	ANTIBIOTIC21	CM type 2 -Antibiotic 21	PIP					
	ANTIBIOTIC22	CM type 2 -Antibiotic 22						
	ANTIBIOTIC23	CM type 2 -Antibiotic 23						
	ANTIBIOTIC24	CM type 2 -Antibiotic 24						
	ANTIBIOTIC25	CM type 2 -Antibiotic 25						
	CMTESTTYPE	CM Test Type	01, ... ,20					
	CPLPATIENTNUMBER	CPL Patient Number	*** SUPPRESSED ***					
	FILEPHINTYPE	Method to determine FILEPHIN	0, 4					
	ORGANISM	CM type 2 - Organism	ABDE, ... ,YERS					
	POSNEG	Positive-Negative	P					
	RECSEQUENCE	Record Sequence	1					
	RECTYPE	Record Type	2					
	REFERFACIL	Referring Facility	00560, ... ,15516					
	REFEROUT	Referred Out						
	REQUISITIONNUMBER	Requisition Number	*** SUPPRESSED ***					
	RESISTSUSCEPT01	CM type 2 -Resistant or Susceptible 01	R, S					
	RESISTSUSCEPT02	CM type 2 -Resistant or Susceptible 02	R, S					
	RESISTSUSCEPT03	CM type 2 -Resistant or Susceptible 03	R, S					
	RESISTSUSCEPT04	CM type 2 -Resistant or Susceptible 04	R, S					
	RESISTSUSCEPT05	CM type 2 -Resistant or Susceptible 05	R, S					
	RESISTSUSCEPT06	CM type 2 -Resistant or Susceptible 06	R, S					
	RESISTSUSCEPT07	CM type 2 -Resistant or Susceptible 07	R, S					

Appendix Table D.8 – Continued

Type	Variable Name	Variable Label	Minimum	Maximum	Mean	Standard Deviation	Outlier	Comment
	RESISTSUSCEPT08	CM type 2 -Resistant or Susceptible 08	R, S					
	RESISTSUSCEPT09	CM type 2 -Resistant or Susceptible 09	R, S					
	RESISTSUSCEPT10	CM type 2 -Resistant or Susceptible 10	R, S					
	RESISTSUSCEPT11	CM type 2 -Resistant or Susceptible 11	R, S					
	RESISTSUSCEPT12	CM type 2 -Resistant or Susceptible 12	R, S					
	RESISTSUSCEPT13	CM type 2 -Resistant or Susceptible 13	R, S					
	RESISTSUSCEPT14	CM type 2 -Resistant or Susceptible 14	R, S					
	RESISTSUSCEPT15	CM type 2 -Resistant or Susceptible 15	R, S					
	RESISTSUSCEPT16	CM type 2 -Resistant or Susceptible 16	R, S					
	RESISTSUSCEPT17	CM type 2 -Resistant or Susceptible 17	R, S					
	RESISTSUSCEPT18	CM type 2 -Resistant or Susceptible 18	R, S					
	RESISTSUSCEPT19	CM type 2 -Resistant or Susceptible 19	R, S					
	RESISTSUSCEPT20	CM type 2 -Resistant or Susceptible 20	R, S					
	RESISTSUSCEPT21	CM type 2 -Resistant or Susceptible 21	R, S					
	RESISTSUSCEPT22	CM type 2 -Resistant or Susceptible 22						
	RESISTSUSCEPT23	CM type 2 -Resistant or Susceptible 23						
	RESISTSUSCEPT24	CM type 2 -Resistant or Susceptible 24						
	RESISTSUSCEPT25	CM type 2 -Resistant or Susceptible 25						
	SCRPHINTYPE	Method to determine SCRPHIN	0, 4					
	SECTION	Section	5					
	SPECIMENSOURCE	Specimen Source	001, ... ,129					
	STATUS	Status	8, 9					
	TECHINIT	Technician Initials	A, ... ,Y					
	TESTSUBSECTION	Test Subsection	ST, TG					
	VERIFIED	Verified	N, Y					
Date	ACQDT	Date record was acquired by MCHP	2010-08-26	2010-08-26				
	RECEIVEDDT	Received Date	1990-10-24	2010-02-08				
	REPORTDT	Report Date	1990-10-26	2010-06-09				
	SPECIMENDT	Specimen Date	1888-05-21	2010-02-08				

Appendix Table D.9: Description of CPL Clinical Microbiology Section Requisitions, 1992–2010

Type	Variable Name	Variable Label	Minimum	Maximum	Mean	Standard Deviation	Outlier	Comment
Identification	FILEPHIN	MH Scrambled PHIN						
	SCRPHIN	MCHP Scrambled PHIN						
Numeric	NCMORGANISMS	N CM organisms sections this requisition	0	5	0.13	0.43	10.64	
	NCMRESULTS	N CM result sections this requisition	0	5	1.12	0.59	19.38	
Character	ACUTECONVAL	Acute/Convalescence						
	CHANGEDDEMODATACD	Changed Demo Data Code	N, Y					
	COMPLETEDIND	Completed Indicator	Y					
	CPLPATIENTNUMBER	CPL Patient Number	*** SUPPRESSED ***					
	FILEPHINTYPE	Method to determine FILEPHIN	0, 4					
	FREEFORMCOMMENT	Freeform Comment	" REPORT FROM NATIONAL MICROBIOLOGY, ..."					
	HOSPITALCLINIC	Hospital/Clinic	"0", ... ,{83-9099					
	INSUFFICIENTINFOCD	Insufficient Information Code	N					
	LABNO	Laboratory Number	0					
	LASTUSEDSEQ	Last Used Sequence	00, 01, 02, 03, 04, 05, 06, 07, 08, ...					
	MATCHREQNUMBER	Match Requisition Number						
	MHREGION	MH Region Code	A, B, C, D, E, F, G, H, I, J, X					
	MORETHAN18TESTS	More Than 18 Tests						
	MUNCODE	Municipal Code	001, ... ,A64					
	PHYSICIANNUMBER	Physician Number	00001, ... ,V0001					
	POSTAL	Patient Postal Code	012901, ... ,Y1A6K3					Invalid Codes: 012901, 024014, 025001, 032605, 033143, 033919, 043205, 055123, 055446, 056560, 056761, 057601, 058201, 058281, 058282, 058367, 058501, 058757, 067209, 067837, 068260, 070454, 077904, 078245, 080004, 080027, 085016, 089119, 090402, 090403
	QCREQUISITION	Quality Control Requisition	N					
	RECSEQUENCE	Record Sequence	1					
	REFERFACIL	Referring Facility	00560, ... ,15516					
	REPORTCOMMENTCD1	Report Comment Code 1st	01, ... ,L1					
	REPORTCOMMENTCD2	Report Comment Code 2nd	01, ... ,L2					
	REPORTCOMMENTCD3	Report Comment Code 3rd	03, ... ,L3					
	REPORTCOMMENTSTATUS1	Report Comment Status 1st	1, 8, 9					Invalid Codes: 1
	REPORTCOMMENTSTATUS2	Report Comment Status 2nd	1, 8, 9					Invalid Codes: 1
	REPORTCOMMENTSTATUS3	Report Comment Status 3rd	1, 8, 9					Invalid Codes: 1
	REQCOMMENT	Requisition Comment	01, ... ,99					
	REQCOMMENTSTATUS	Requisition Comment Status	0, 7, 9					
	REQUISITIONNUMBER	Requisition Number	*** SUPPRESSED ***					

Appendix Table D.9– Continued

Type	Variable Name	Variable Label	Minimum	Maximum	Mean	Standard Deviation	Outlier	Comment
	RHAMUNCODE	Regional Health Authority of MUNCODE	10, ... ,95					
	RHAPOSTAL	RHA of Postal Code	10, ... ,95					
	SCRPHINTYPE	Method to determine SCRPHIN	0, 4					
	SECTION	Section	5					
	SEX	Sex of Patient	1, 2					
	SPECIALSTUDYCD	Special Study Code	A1, ... ,ZZ					Invalid Codes: A1, A3, A7, A8, A9, AB, AD, AE, AG, AI, AJ, AM, AO, AQ, AR, AS, AU, AV, AX, AY, AZ, B1, B2, B3, B4, B5, B6, B8, B9, BA, BB, BC, BD, BE, BF, BG, BH, BI, BJ, BK, BL, BM, BN, BO, BQ, BT, BU, BW, BY, BZ, C, C2, C3, C4, C5, C6, C7, C8, C9, CA, C
	STATREQUISITION	Stat Requisition	N, Y					
	SUBSECTION01	Subsection 01	ST, TG					
	SUBSECTION02	Subsection 02						
	SUBSECTION03	Subsection 03						
	SUBSECTION04	Subsection 04						
	SUBSECTION05	Subsection 05						
	SUBSECTION06	Subsection 06						
	SUBSECTION07	Subsection 07						
	SUBSECTION08	Subsection 08						
	SUBSECTION09	Subsection 09						
	SUBSECTION10	Subsection 10						
	TESTCOUNT	Test Count	01, 02, 03, 04, 05					
	UNINSUREDSERVICESCD	Uninsured Services Code	1, 2, 3, 4, 5, 6					
Date	ACQDT	Date record was acquired by MCHP	2010-08-26	2010-08-26				
	BIRTHDT	Birth Date	1800-12-31	2010-01-15				
	RECEIVEDDT	Received Date	1990-10-24	2010-02-23				
	REPORTCOMMENTDT1	Report Comment Date 1st	1992-04-13	2010-04-29				
	REPORTCOMMENTDT2	Report Comment Date 2nd	1992-07-09	2010-04-29				
	REPORTCOMMENTDT3	Report Comment Date 3rd	1995-07-10	2010-02-06				
	REQCOMMENTREPORTDT	Requisition Comment Report Date	1992-04-14	2010-04-21				
	SPECIMENDT	Specimen Date	1888-05-21	2010-02-21				
	STATDT	Status Date	1992-05-21	2010-08-10				

Appendix Table D.10: Description of CPL Clinical Microbiology Section Results, 1992–2010

Type	Variable Name	Variable Label	Minimum	Maximum	Mean	Standard Deviation	Outlier	Comment
Identification	FILEPHIN	MH Scrambled PHIN						
	SCRPHIN	MCHP Scrambled PHIN						
Numeric	RECPOSN	Position of this record in requisition	1	5	1.22	0.55	16.13	
Character	CMTESTTYPE	CM Test Type	01, ... ,20					
	CPLPATIENTNUMBER	CPL Patient Number	*** SUPPRESSED ***					
	FILEPHINTYPE	Method to determine FILEPHIN	0, 4					
	POSNEG	Positive-Negative	N, P					
	RECSEQUENCE	Record Sequence	1					
	RECTYPE	Record Type	1					
	REFERFACIL	Referring Facility	00560, ... ,15516					
	REFEROUT	Referred Out						
	REQUISITIONNUMBER	Requisition Number	*** SUPPRESSED ***					
	RESULTS1	CM type 1 Result 1 of 6	ACNI, ... ,YNI					
	RESULTS2	CM type 1 Result 2 of 6	ACNI, ... ,YNI					
	RESULTS3	CM type 1 Result 3 of 6	ACNI, ... ,YNI					
	RESULTS4	CM type 1 Result 4 of 6	AFB, ... ,YNI					
	RESULTS5	CM type 1 Result 5 of 6	ANI, ... ,YNI					
	RESULTS6	CM type 1 Result 6 of 6	ATFC, ... ,VTTN					
	SCRPHINTYPE	Method to determine SCRPHIN	0, 4					
	SECTION	Section	5					
	SPECIMENSOURCE	Specimen Source	001, ... ,129					
	STATUS	Status	8, 9					
	TECHINIT	Technician Initials	A, ... ,ZS					
	TESTSUBSECTION	Test Subsection	ST, TG					
	VERIFIED	Verified	N, Y					
Date	ACQDT	Date record was acquired by MCHP	2010-08-26	2010-08-26				
	RECEIVEDDT	Received Date	1991-06-21	2010-02-23				
	REPORTDT	Report Date	1991-06-28	2010-06-17				
	SPECIMENDT	Specimen Date	1899-02-12	2010-02-21				

Appendix Table D.11: Description of CPL Parasitology and Serology Section Requisitions, 1992–2010

Type	Variable Name	Variable Label	Minimum	Maximum	Mean	Standard Deviation	Outlier	Comment
Identification	FILEPHIN	MH SCRAMBLED PHIN						
	SCRPHIN	MCHP SCRAMBLED PHIN						
Numeric	NPARASITOLOGY	N of Parasitology tests this requisition	0	6	0.25	0.83	9.69	
	NSEROLOGY	N of Serology tests this requisition	0	18	1.93	1.56	0.80	
	TESTCOUNT	Test Count	1	27	2.18	1.49	0.49	
Character	ACUTECONVAL	Acute/Convalescent	A, C					
	CHANGEDDEMODATACD	Changed Demo Data Code	N, Y					
	COMPLETEDIND	Completed Indicator	Y					
	CPLPATIENTNUMBER	CPL Patient Number	*** SUPPRESSED ***					
	FILEPHINTYPE	Method to determine FILEPHIN	0, 4					
	FREEFORMCOMMENT	Freeform Comment	"ANA PATTERN: FINE SPECKLED/ CHROMOS, ...					
	HISTORYCLINICALDIAG	History/Clinical Diagnosis	" HEPATITIS B CARRIER", ... , YOUR N, ...					
	HIVPREVHISTRESULT	HIV Previous History-Result	00					
	HOSPCLINIC	Hospital/Clinic #	#, ... ,¦					
	INSUFFICIENTINFOCD	Insufficient Information Code	N					
	LABNO	Laboratory Number	000000, ... ,999999					
	LASTUSEDSEQ	Last Used Sequence	01, ... ,29					
	MATCHREQNUMBER	Match Requisition Number	*** SUPPRESSED ***					
	MHREGION	MH Region Code	A, B, C, D, E, F, G, H, I, J, N, X					
	MORETHAN18TESTS	More Than 18 Tests	Y					
	MUNCODE	Municipal Code	001, ... ,A64					
	PATIENTCATEGORY	Patient Category	EA, ... ,VS					Invalid Codes: EA, EB, EC, EF, EO, ER, ES
	PHYSICIANNUMBER	Physician Number	*** SUPPRESSED ***					
	POSTAL	Patient Postal Code	000142, ... ,Z7H1S6					Invalid Codes: 000142, 0009 3, 001527, 001760, 002 00, 004005, 007419, 007626, 007873, 010006, 010024, 011211, 011322, 012901, 014850, 015101, 015208, 016508, 020008, 022042, 022815, 027613, 029526, 030030, 030350, 032507, 032578, 033014, 033102, 033143
	PREVHIST	Previous History	*					
	QCREQUISITION	Quality Control Requisition	N, Y					
	RACKNUMBER	Rack Number	01, ... ,99					
	RACKSPECIMENNUMBER	Rack Specimen Number	01, ... ,30					
	RACKSUBSECTION	Rack Subsection	HP, RO					
	RECSEQUENCE	Record Sequence	1, 2					
	REFERFACIL	Referring Facility	00560, ... ,15516					
	REPORTCOMMENTCD1	Report Comment Code	01, ... ,Q8					
	REPORTCOMMENTCD2	Report Comment Code	01, ... ,Q8					
	REPORTCOMMENTCD3	Report Comment Code	01, ... ,Q8					
	REPORTCOMMENTSTAT1	Report Comment Status	1, 3, 8, 9					Invalid Codes: 1, 3
	REPORTCOMMENTSTAT2	Report Comment Status	1, 8, 9					Invalid Codes: 1
	REPORTCOMMENTSTAT3	Report Comment Status	1, 8, 9					Invalid Codes: 1
	REQCOMMENT	Requisition Comment	01, ... ,Q8					
	REQCOMMENTSTAT	Requisition Comment Status	0, 7, 9					
	REQUISITIONNUMBER	Requisition Number	*** SUPPRESSED ***					
	RHAMUNCODE	Regional Health Authority of MUN	10, ... ,99					
	RHAPOSTAL	RHA of Postal Code	10, ... ,99					

Appendix Table D.11 – Continued

Type	Variable Name	Variable Label	Minimum	Maximum	Mean	Standard Deviation	Outlier	Comment
	RISKGROUPS1	Risk Groups 1	ACP, ... ,TRP					
	RISKGROUPS2	Risk Groups 2	ACP, ... ,TRP					
	RISKGROUPS3	Risk Groups 3	ACP, ... ,TRP					
	SCRPHINTYPE	Method to determine SCRPHIN	0, 4					
	SECTION	Section	1					
	SEX	Sex of Patient	1, 2					
	SPECIALSTUDYCD	Special Study Code	A7, ... ,YI					BM, BP, BS, C3, C6, C7, C8, C9, CA, CZ, D2, D5, D6, D7, D8, DM, DT, DW, DY, E8, EA, F1, F3, FC, FL, FM, FP, FX, FY, G7, GH, GI, GM, GY, GZ, H1, H2, H4, H8, HA, HB, HH, HJ, HR, HS, IH, J8, J9, JE, JL, JR,
	SPECIMENSOURCE	Specimen Source on test	001, ... ,130					
	STATREQUISITION	Stat Requisition	N, Y					
	SUBSECTION01	Subsection 01	BS, ... ,VS					
	SUBSECTION02	Subsection 02	BS, ... ,VS					
	SUBSECTION03	Subsection 03	BS, ... ,VS					
	SUBSECTION04	Subsection 04	BS, ... ,VS					
	SUBSECTION05	Subsection 05	BS, ... ,VS					
	SUBSECTION06	Subsection 06	HP, ... ,VS					
	SUBSECTION07	Subsection 07	HV, NA, PS, VS					
	SUBSECTION08	Subsection 08						
	SUBSECTION09	Subsection 09						
	SUBSECTION10	Subsection 10						
	SYMPTOMS1	Symptoms 1	01, ... ,91					
	SYMPTOMS2	Symptoms 2	01, ... ,91					
	SYMPTOMS3	Symptoms 3	01, ... ,91					
	SYMPTOMS4	Symptoms 4	01, ... ,91					
	SYMPTOMS5	Symptoms 5	01, ... ,90					
	TESTSELECTION	Test Selection	, 1, 2, 9					
	UNINSUREDSERVICESCD	Uninsured Services Code	1, 2, 3, 4, 5, 6					
	VIRALLOADANTIRETROVIRALMEDS1	Viral Load Antiretroviral Medications	094, ... ,ZID					Invalid Codes: 094, HAA, NIL, ON
	VIRALLOADANTIRETROVIRALMEDS2	Viral Load Antiretroviral Medications	ABA, ... ,ZID					Invalid Codes: ADA, DDI
	VIRALLOADANTIRETROVIRALMEDS3	Viral Load Antiretroviral Medications	ABA, ... ,ZID					Invalid Codes: ETA, SAW
	VIRALLOADANTIRETROVIRALMEDS4	Viral Load Antiretroviral Medications	ABA, ... ,ZID					Invalid Codes: HV, KET
	VIRALLOADREASON	Viral Load Reason	1, 2, 3, 4, 5, 6, 7					
	VIRALLOADTYPE	Viral Load Type	G, O, R, U					
	VLCD4COUNT	VL CD4 Count	00000, ... ,08892					
Date	ACQDT	Date record was acquired at MCHP	2010-08-26	2010-08-26				
	BIRTHDT	Birth Date	1852-02-08	2009-08-12				
	HIVPREVHISTREPORTDT	HIV Previous History-Report Date	1990-10-16	1990-10-16				
	RECEIVEDDT	Received Date	1922-08-25	2009-11-16				
	REPORTCOMMENTDT1	Report Comment Date	1992-04-14	2009-10-14				
	REPORTCOMMENTDT2	Report Comment Date	1992-04-14	2009-09-23				
	REPORTCOMMENTDT3	Report Comment Date	1992-05-07	2009-09-11				
	REQCOMMENTREPORTDT	Requisition Comment Report Date	1992-04-14	2009-10-07				
	SPECIMENDT	Specimen Date	1601-06-22	2009-11-14				
	STATDT	Status Date	1992-05-21	2010-06-10				
	VLCD4DT	VL CD4 Date	1989-05-14	2009-08-17				
	VLLASTTESTDT	VL Last Test Date	1989-03-31	2009-08-01				

Appendix Table D.12: Description of CPL Parasitology Section Results, 1992–2010

Type	Variable Name	Variable Label	Minimum	Maximum	Mean	Standard Deviation	Outlier	Comment
Identification	FILEPHIN	MH SCRAMBLED PHIN						
	SCRPHIN	MCHP SCRAMBLED PHIN						
Numeric	RECPOSN	Test Subsection position this requisition	1	6	2.03	1.07	0.00	
Character	CPLPATIENTNUMBER	CPL Patient Number	*** SUPPRESSED ***	N/A				
	FILEPHINTYPE	Method to determine FILEPHIN	0, 4	N/A				
	INTERPRETATION1	Interpretation	C, K, L, T	N/A				
	INTERPRETATION2	Interpretation	K, L, T	N/A				
	INTERPRETATION3	Interpretation	K, L	N/A				
	INTERPRETATION4	Interpretation	K, L, T	N/A				
	INTERPRETATION5	Interpretation	K, L	N/A				
	PARASITOLOGYPARASITE1	Parasitology Parasite	C, F, L, O, R, T, V, W, X	N/A				
	PARASITOLOGYPARASITE2	Parasitology Parasite	C, L, R, T, V, W, X	N/A				
	PARASITOLOGYPARASITE3	Parasitology Parasite	C, L, R, T, V	N/A				
	PARASITOLOGYPARASITE4	Parasitology Parasite	C, L, R, T, V, W	N/A				
	PARASITOLOGYPARASITE5	Parasitology Parasite	C, L, T, V	N/A				
	PARASITOLOGYRESULT1	Parasitology Result	00, ... ,54	N/A				
	PARASITOLOGYRESULT2	Parasitology Result	00, ... ,54	N/A				
	PARASITOLOGYRESULT3	Parasitology Result	03, ... ,54	N/A				
	PARASITOLOGYRESULT4	Parasitology Result	03, ... ,52	N/A				
	PARASITOLOGYRESULT5	Parasitology Result	03, ... ,54	N/A				
	PARASITOLOGYTEST	Parasitology Test	BS, ... ,XEL	N/A				
	POSNEG	Positive-Negative	N, P	N/A				
	RECSEQUENCE	Record Sequence	1	N/A				
	REFERFACIL	Referring Facility	00560, ... ,15516	N/A				
	REFEROUT	Referred Out		N/A				Blank or N does not apply
	REQUISITIONNUMBER	Requisition Number	*** SUPPRESSED ***	N/A				
	SCRPHINTYPE	Method to determine SCRPHIN	0, 4	N/A				
	SECTION	Section	1	N/A				Always 1 for Serology
	SPECIMENSOURCE	Specimen Source on test	001, ... ,114	N/A				
	STATUS	Status	8, 9	N/A				
	TECHINIT	Technician Initials	A, ... ,ZOS	N/A				Must Exist
	TESTSEQUENCE	Test Sequence	01, 02, 03, 04, 05, 06, 07, 08, 09	N/A				
	TESTSUBSECTION	Test Subsection	PA	N/A				Must Exist
	VERIFIED	Verified	N, Y	N/A				
Date	ACQDT	Date record was acquired at MCHP	2010-08-26	2010-08-26				
	PARASITOLOGYREPORTDT	Parasitology Report Date	1992-04-15	2009-10-19				
	RECEIVEDDT	Received Date	1990-09-07	2009-08-29				Must Exist
	SPECIMENDT	Specimen Date	1955-11-11	2009-08-29				
	STATDT	Status Date	1992-05-21	2010-01-09				

Appendix Table D.13: Description of CPL Serology Section Results, 1992–2010

Type	Variable Name	Variable Label	Minimum	Maximum	Mean	Standard Deviation	Outlier	Comment
Identification	FILEPHIN	MH SCRAMBLED PHIN						
	SCRPHIN	MCHP SCRAMBLED PHIN						
Numeric	RECPOSN	Test Subsection position this requisition	1	18	2.10	1.45	0.93	
Character	CPLPATIENTNUMBER	CPL Patient Number	*** SUPPRESSED ***					
	FILEPHINTYPE	Method to determine FILEPHIN	0, 4					
	GROUPTESTCD	Group Test Code	CHLAM, RESPB					
	GROUPTESTIND	Group Test Indicator	N, Y					
	HOLDSHEETPRINTED	Hold Sheet Printed	Y					
	INTERPRETATION	Interpretation	A, ... ,Z					
	POSNEG	Positive-Negative	N, P					
	QCINTERPRETATION	Quality Control Interpretation						
	QCQUANTITY	Quality Control Quantity						
	QCRESULT	Quality Control Result						
	QUANTITY	Quantity	A 1.36, ... ,U06840					
	RACKINDICATOR	Rack Indicator	N, Y					
	RECSEQUENCE	Record Sequence	1, 2					
	REFERFACIL	Referring Facility	00560, ... ,15516					
	REFEROUT	Referred Out						Blank or N does not apply
	REQUISITIONNUMBER	Requisition Number	*** SUPPRESSED ***					
	SCRPHINTYPE	Method to determine SCRPHIN	0, 4					
	SECTION	Section	1					
	SEROLOGYTEST	Serology Test	ACA, ... ,YFAB					Always 1 for Serology
	SPECIMENSOURCE	Specimen Source on test	001, ... ,130					
	SPRESULT	SP Result	00, ... ,45					
	STATUS	Status	8, 9					
	STORAGE	Storage	00, ... ,ZV 048					
	TECHINIT	Technician Initials	A, ... ,ZOZ					Must Exist
	TESTSEQUENCE	Test Sequence	01, ... ,29					
	TESTSUBSECTION	Test Subsection	BS, ... ,VS					
	VERIFIED	Verified	N, Y					
Date	ACQDT	Date record was acquired at MCHP	2010-08-26	2010-08-26				
	RECEIVEDDT	Received Date	1922-08-25	2009-11-16				Must Exist
	REPORTDT	Report Date	1992-04-13	2010-03-01				
	SPECIMENDT	Specimen Date	1601-06-22	2009-11-14				
	STATDT	Status Date	1992-05-21	2010-06-10				

Appendix Table D.14: Description of CPL Virus Detection Section Requisitions, 1992–2010

Type	Variable Name	Variable Label	Minimum	Maximum	Mean	Standard Deviation	Outlier	Comment
Identification	FILEPHIN	MH SCRAMBLED PHIN						
	SCRPHIN	MCHP SCRAMBLED PHIN						
Numeric	NVIRUSDETECTION	N of Virus Detection tests this requisition	1	10	1.57	0.95	2.53	
Character	ACUTECONVAL	Acute/Convalescence						
	CHANGEDDEMODATACD	Changed Demo Data Code	Y					
	COMPLETEDIND	Completed Indicator	Y					
	CPLPATIENTNUMBER	CPL Patient Number	*** SUPPRESSED ***					
	FILEPHINTYPE	Method to determine FILEPHIN	0, 4					
	FREEFORMCOMMENT	Freeform Comment	" EBV DETECTED 67 COPIES/ 1 UL BLOO, ...					
	HOSPCLINIC	Hospital/Clinic #	# 281, ... ,Z-828					
	INSUFFICIENTINFOCD	Insufficient Information Code						
	LABNO	Laboratory Number	000000					
	LASTUSEDSEQ	Last Used Sequence	00, 01, 02, 03, 04, 05, 06, 07, 08,, ...					
	MATCHREQNUMBER	Match Requisition Number						
	MHREGION	MH Region Code	A, B, C, D, E, F, G, H, I, J, X					
	MORETHAN18TESTS	More Than 18 Tests	Y					
	MUNCODE	Municipal Code	001, ... ,A64					
	PHYSICIANNUMBER	Physician Number	00001, ... ,P0013					
	POSTAL	Patient Postal Code	032605, ... ,X0E0R5					Invalid Codes: 032605, 055123, 068116, 078759, 090402
	QCREQUISITION	Quality Control Requisition						
	RECSEQUENCE	Record Sequence	1					
	REFERFACIL	Referring Facility	00560, ... ,15516					
	REPORTCOMMENTCD1	Report Comment Code 1st	01, ... ,9G					
	REPORTCOMMENTCD2	Report Comment Code 2nd	01, ... ,9G					
	REPORTCOMMENTCD3	Report Comment Code 3rd	01, ... ,9G					
	REPORTCOMMENTSTAT1	Report Comment Status 1st	1, 3, 8, 9					Invalid Codes: 1, 3
	REPORTCOMMENTSTAT2	Report Comment Status 2nd	1, 8, 9					Invalid Codes: 1
	REPORTCOMMENTSTAT3	Report Comment Status 3rd	1, 9					Invalid Codes: 1
	REQCOMMENT	Requisition Comment	01, ... ,9G					
	REQCOMMENTSTAT	Requisition Comment Status	0, 7, 9					
	REQUISITIONNUMBER	Requisition Number	*** SUPPRESSED ***					
	RHAMUNCODE	Regional Health Authority of MUNCODE	10, ... ,95					
	RHAPOSTAL	RHA of Postal Code	10, ... ,95					
	SCRPHINTYPE	Method to determine SCRPHIN	0, 4					
	SECTION	Section	4					
	SEX	Sex of Patient	1, 2					

Appendix Table D.14 – Continued

Type	Variable Name	Variable Label	Minimum	Maximum	Mean	Standard Deviation	Outlier	Comment
	SPECIALSTUDYCD	Special Study Code	A1, ... ,ZZ					Invalid Codes: A1, A2, A3, A5, A7, A9, AA, AB, AC, AD, AE, AF, AG, AH, AI, AJ, AK, AL, AM, AN, AO, AP, AQ, AR, AS, AT, AU, AV, AW, AX, AY, AZ, B, B2, B4, B5, B6, B8, B9, BA, BB, BC, BD, BE, BF, BG, BH, BI, BJ, BK, BL, BM, BO, BP, BQ, BR, BS, BT, BU, BV, B
	STATREQUISITION	Status of Requisition	N, Y					
	TESTCOUNT	Test Count	01, 02, 03, 04, 05, 06, 07, 08, 09,, ...					
	UNINSUREDSERVICESCD	Uninsured Services Code	1, 2, 4, 5, 6					
Date	ACQDT	Date record was acquired at MCHP	2010-08-26	2010-08-26				
	BIRTHDT	Birth Date	1889-10-12	2010-03-29				
	RECEIVEDDT	Received Date	1989-05-30	2010-04-05				
	REPORTCOMMENTDT1	Report Comment Date 1st	1992-05-15	2010-06-17				
	REPORTCOMMENTDT2	Report Comment Date 2nd	1994-09-21	2010-06-17				
	REPORTCOMMENTDT3	Report Comment Date 3rd	1995-08-23	2010-04-09				
	REQCOMMENTREPORTDT	Requisition Comment Report Date	1992-04-14	2010-03-23				
	SPECIMENDT	Specimen Date	1929-09-10	2010-04-05				
	STATDT	Status Date	1992-05-21	2010-07-10				

Appendix Table D.15: Description of CPL Virus Detection Section Results, 1992–2010

Type	Variable Name	Variable Label	Minimum	Maximum	Mean	Standard Deviation	Outlier	Comment
Identification	FILEPHIN	MH SCRAMBLED PHIN						
	SCRPHIN	MCHP SCRAMBLED PHIN						
Numeric	RECPOSN	Position this Test in Virus requisition	1	10	1.57	0.95	2.19	
Character	CPLPATIENTNUMBER	CPL Patient Number	*** SUPPRESSED ***					
	FILEPHINTYPE	Method to determine FILEPHIN	0, 4					
	POSNEG	Positive-Negative	N, P					
	RECORDTYPE	Record Type	1, 2					
	RECSEQUENCE	Record Sequence	1					Internal usage
	REFERFACIL	Referring Facility	00560, ... ,15516					
	REFEROUT	Referred Out						
	REQUISITIONNUMBER	Requisition Number	*** SUPPRESSED ***					
	SCRPHINTYPE	Method to determine SCRPHIN	0, 4					
	SECTION	Section	4					Always 4 for Virus
	SEROTYPE	Serology Type	00, ... ,91					
	SPECIMENSOURCE	Specimen Source	001, ... ,121					
	STATUS	Status	8, 9					
	TECHINIT	Technician Initials	A, ... ,YW					
	TESTSEQUENCE	Test Sequence	0, 1, 2, 3, 4, 5, 6, 7, 8, 9					
	TESTSUBSECTION	Test Subsection						Not used
	VERIFIED	Verified	N, Y					
	VIRRESULT	Viral Result	ADDIT, ... ,YSPAR					
	VIRTESTTYPE	Viral Test Type	01, ... ,37					
	VIRUS	Virus	AD, ... ,VZ					
	VTREQUISITIONNUMBER	Requisition Number on virus test	*** SUPPRESSED ***					
Date	ACQDT	Date record was acquired at MCHP	2010-08-26	2010-08-26				
	RECEIVEDDT	Received Date	1989-05-30	2010-04-05				Must Exist
	REPORTDT	Report Date	1989-06-06	2010-06-17				
	SPECIMENDT	Specimen Date	1929-09-10	2010-04-05				
	STATDT	Status Date	1992-05-21	2010-07-10				

Appendix Table D.16: Description of CPL Physician – Provider, 1992–2010

Type	Variable Name	Variable Label	Minimum	Maximum	Mean	Standard Deviation	Outlier	Comment
Character	HEALTHUNITOFFICE	Health Unit - Office	000, ... ,830					
	HEALTHUNITREGION	Health Unit- Region	00, ... ,95					
	MHREGION	MH Region	2, A, B, C, D, E, F, G, H, I, J, X					Invalid Codes: 2
	MUNCODE	Municipal Code	000, ... ,88					Invalid Codes: 88
	PHYSICIANNUMBER	Physician Number	00001, ... ,V0001					
	POSTAL	Postal Code	A1K1B3, ... ,V9V1C1					
	RHA	Regional Health Authority	02, ... ,95					Invalid Codes: 02
Date	ACQDT	Date record was acquired by MCHP	2010-06-21	2010-06-21				
	DELETEDT	Delete Date	2099-12-31	2099-12-31				
	STARTDT	Start Date	1992-04-10	2010-01-09				

Appendix Table D.17: Description of CPL Referring Facility Table, 1992–2010

Type	Variable Name	Variable Label	Minimum	Maximum	Mean	Standard Deviation	Outlier	Comment
Character	FACILADDR1	Facility Address #1	#1 - 705 BROADWAY AVE., ... ,Y3016 , ...					
	FACILADDR2	Facility Address #2	1 MORLEY ST, ... ,WPG MB					
	FACILADDR3	Facility Address #3	409 TACHE AVE, ... ,YORK LANDING					
	FACILADDR4	Facility Address #4	(COURIER), ... ,WPG MB					
	FACILNAME	Facility Name	RETURN TO , ... ,YOUVILLE CENTRE S,					
	FACILPOSTAL	Facility postal Code	G1K7P4, ... ,X1X1X1					
	HEALTHUNITOFFICE	Health Unit - Office	000, ... ,830					
	HEALTHUNITREGION	Health Unit- Region	11, ... ,95					
	MESSAGECD	Message Code	1					
	MHREGION	MH Region	A, B, C, D, E, F, G, H, I, J, X					
	MUNCODE	Municipal Code	000, ... ,321					
	REFERFACIL	Referring Facility #	00133, ... ,15516					
	RHA	Regional Health Authority	10, ... ,95					
Date	ACQDT		2010-06-21	2010-06-21				
	DELETEDT	Delete Date	1982-01-25	2099-12-31				
	STARTDT	Start Date	1992-04-10	2009-07-08				

RECENT MCHP PUBLICATIONS

2012

Perinatal Services and Outcomes in Manitoba by Maureen Heaman, Dawn Kingston, Michael Helewa, Marni Brownlee, Shelley Derksen, Bodgan Bogdanovic, Kari–Lynne McGowan, and Angela Bailly

How Are Manitoba's Children Doing? By Marni Brownlee, Mariette Chartier, Rob Santos, Oke Ekuma, Wendy Au, Joykrishna Sarkar, Leonard MacWilliam, Elaine Burland, Ina Koseva, and Wendy Guenette

Projecting Personal Care Home Bed Equivalent Needs in Manitoba Through 2036 by Dan Chateau, Malcolm Doupe, Randy Walld, Ruth–Ann Soodeen, Carole Ouelette, and Leanne Rajotte

Health and Healthcare Utilization of Francophones in Manitoba by Mariette Chartier, Gregory Finlayson, Heather Prior, Kari–Lynne McGowan, Hui Chen, Janelle de Rocquigny, Randy Walld, and Michael Gousseau

The Early Development Instrument (EDI) in Manitoba: Linking Socioeconomic Adversity and Biological Vulnerability at Birth to Children's Outcomes at Age 5 by Rob Santos, Marni Brownell, Oke Ekuma, Teresa Mayer, and Ruth–Ann Soodeen

The Epidemiology and Outcomes of Critical Illness in Manitoba by Alan Garland, Randy Fransoo, Kendiss Olafson, Clare Ramsey, Marina Yogendran, Dan Chateau, and Kari–Lynne McGowan

2011

Adult Obesity in Manitoba: Prevalence, Associations, and Outcomes by Randy Fransoo, Patricia Martens, Heather Prior, Dan Chateau, Chelsey McDougall, Jennifer Schultz, Kari–Lynne McGowan, and Angela Bailly

Manitoba Immunization Study by Tim Hilderman, Alan Katz, Shelley Derksen, Kari–Lynne McGowan, Dan Chateau, Carol Kurbis, Sandra Allison, Ruth–Ann Soodeen, and Jocelyn Nicole Reimer

Population Aging and the Continuum of Older Adult Care in Manitoba by Malcolm Doupe, Randy Fransoo, Dan Chateau, Natalia Dik, Charles Burchill, Ruth–Ann Soodeen, Songul Bozat–Emre, and Wendy Guenette

2010

Pharmaceutical Use in Manitoba: Opportunities to Optimize Use by Colette Raymond, Silvia Alessi–Severini, Colleen Metge, Matthew Dahl, Jennifer Schultz, and Wendy Guenette

Evaluation of the Healthy Baby Program by Marni Brownell, Mariette Chartier, Wendy Au, and Jennifer Schultz

Health Inequities in Manitoba: Is the Socioeconomic Gap in Health Widening or Narrowing Over Time? by Patricia Martens, Marni Brownell, Wendy Au, Leonard MacWilliam, Heather Prior, Jennifer Schultz, Wendy Guenette, Lawrence Elliott, Shelley Buchan, Marcia Anderson, Patricia Caetano, Colleen Metge, Rob Santos, and Karen Serwonka

Physician Integrated Network Baseline Evaluation: Linking Electronic Medical Records and Administrative Data by Alan Katz, Bogdan Bogdanovic, and Ruth–Ann Soodeen

Profile of Metis Health Status and Healthcare Utilization in Manitoba: A Population–Based Study by Patricia Martens, Judith Bartlett, Elaine Burland, Heather Prior, Charles Burchill, Shamima Huq, Linda Romphf, Julianne Sanguins, Sheila Carter, and Angela Bailly

The Additional Cost of Chronic Disease in Manitoba by Gregory Finlayson, Okechukwu Ekuma, Marina Yogendran, Elaine Burland, and Evelyn Forget

2009

Effects of Manitoba Pharmacare Formulary Policy on Utilization of Prescription Medications by Anita Kozyrskyj, Colette Raymond, Matt Dahl, Oke Ekuma, Jenn Schultz, Mariana Sklepowich, and Ruth Bond

Manitoba RHA Indicators Atlas 2009 by Randy Fransoo, Patricia Martens, Elaine Burland, *The Need to Know* Team, Heather Prior, and Charles Burchill

Composite Measures/Indices of Health and Health System Performance by Colleen Metge, Dan Chateau, Heather Prior, Ruth–Ann Soodeen, Carolyn De Coster, and Louis Barre

The Direct Cost of Hospitalizations in Manitoba, 2005/06 by Greg Finlayson, Julene Reimer, Matthew Stargardter, and Kari–Lynne McGowan

Physician Resource Projection Models by Alan Katz, Bogdan Bogdanovic, Oke Ekuma, Ruth–Ann Soodeen, Dan Chateau, and Chris Burnett

2008

Manitoba Child Health Atlas Update by Marni Brownell, Carolyn De Coster, Robert Penfold, Shelley Derksen, Wendy Au, Jennifer Schultz, and Matthew Dahl

An Initial Analysis of Emergency Departments and Urgent Care in Winnipeg, by Malcolm Doupe, Anita Kozyrskyj, Ruth–Ann Soodeen, Shelley Derksen, Charles Burchill, and Shamima Huq

What Works? A First Look at Evaluating Manitoba's Regional Health Programs and Policies at the Population Level, by Patricia Martens, Randy Fransoo, *The Need to Know* Team, Elaine Burland, Heather Prior, Charles Burchill, Linda Romphf, Dan Chateau, Angela Bailly, and Carole Ouelette

Manitoba Centre for Health Policy
University of Manitoba, Faculty of Medicine
Department of Community Health Sciences
408-727 McDermot Avenue
Winnipeg, Manitoba R3E 3P5
Tel: (204) 789-3819
Fax: (204) 789-3910
Email: reports@cpe.umanitoba.ca
Web: umanitoba.ca/medicine/units/mchp