REPORTING OF INFLUENZA-RELATED EVENTS

COMPARISON OF SELF-REPORT AND MEDICAL RECORD DOCUMENTATION OF INFLUENZA RELATED EVENTS

By

ANGELA M. BARBARA, B.Sc., M.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Doctor of Philosophy

in Health Research Methodology

McMaster University

DOCTOR OF PHILOSOPHY (2011)

Health Research Methodology

McMaster University

Hamilton, Ontario, Canada

TITLE:	Comparison of Self-Report and Medical Record Documentation of Influenza-Related Events
AUTHOR:	Angela M. Barbara B.Sc. (University of Toronto), M.Sc. (Wake Forest University)
SUPERVISOR:	Professor Mark B. Loeb, MD, M.Sc., FRCPC
NUMBER OF PAGES:	xii, 145

ABSTRACT

Our objective was to evaluate the comparability of influenza-related events self-reported by research participants and their outpatient medical records using data collected from the Hutterite Influenza Prevention Trial. We also explored the implications of using data on influenza symptoms from both data sources, independently and in combination, as predictors of laboratory-confirmed influenza.

Self (and maternal) report of ten common influenza symptoms, physician-diagnosed otitis media and antibiotics prescribed at outpatient consultations was collected from participants in the Hutterite Influenza Prevention Study. Similar data were also collected by fax requests for medical record information to the medical facilities. We calculated prevalence of each event (by each data source); sensitivity, specificity, predictive values and likelihood ratios of self-reported otitis media and prescription antibiotics; and agreement indices between sources for each symptom. We also calculated how frequently influenza-like illness (ILI) surveillance definitions correlated to laboratory-confirmed influenza and the predictive value of surveillance definitions.

We found lower rates of self-reported prevalence for fever, sore throat, earache and otitis media and higher rates of antibiotic prescriptions compared to the medical records. Total agreements between self-report and medical report of symptoms varied between 61% and 88%. Negative agreement was considerably higher than positive agreement for each symptom, except cough. Self report of otitis media was a very specific measure (93%), but had lower sensitivity (47%). Positive predictive value was moderate at 64% but negative predictive value was good at 86%. Self-reported antibiotic

prescription was a highly sensitive measure (98%), but had low specificity (50%). Positive predictive value was high at 91% but negative predictive value was modest at 65%.

Fever (on its own) and combined with cough and/or sore throat were highly correlated with laboratory-confirmed influenza for all data sources. The ILI surveillance definition of fever and sore throat, based on combined symptoms by both medical records and self report, was the best predictor laboratory confirmed influenza.

ACKNOWLEDGEMENTS

This work was accomplished with the help and support of several individuals. I thank my thesis supervisor Mark Loeb who provided invaluable guidance, perspective and support of my research interests. I also thank my committee members, Lisa Dolovich and Kevin Brazil, for their helpful insights and direction, and the fourth author on the manuscripts, Margaret Russell, for her feedback and diligent editing. I must also acknowledge the work of my colleagues who are responsible for the successful undertaking of the Hutterite Influenza Prevention Trial, especially Lorraine Moss, Cassandra Howse, Pardeep Singh and the research nurses working in the prairie provinces. I also acknowledge the supportive environment provided by everyone in the Mark Loeb Research Group, particularly my fellow graduate students. Above all, I am grateful for the love and support of my family throughout this endeavor.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

This PhD thesis in the Health Research Methodology program has been written as a series of inter-related papers. It consists of an introductory chapter, five chapters written as manuscripts, and a concluding chapter. I am first author on all five manuscripts which have been submitted for publication. The submission or publication status of the manuscripts is provided at the start of each chapter.

All of the manuscripts included in this these have used data from first year of the CIHR and NIH funded randomized controlled trial on influenza prevention in Hutterite communities, designed and led by Dr. Mark Loeb. I have been the research coordinator for this study, independently of my thesis work. For each manuscript, I developed the research question and analysis plan, performed the analyses, and wrote the paper, with guidance from the supervisory committee.

The five papers, in order of presentation in the thesis, are entitled:

1) *"Agreement between self-report and medical records on influenza-related symptoms"*
2) *"Self and parental report of physician-identified otitis media in a rural sample"*
3) *"Accuracy of self-reported antibiotic prescription in a Hutterite sample"*
4) *"A comparison of self-report and medical record data to assess surveillance definitions of influenza-like illness in outpatients"*
5) *"Measuring agreement for binary data between patient self-report and medical records"*

LIST OF ABBREVIATIONS

CDC	Centre for Disease Control and Prevention
CI	Confidence interval
CIHR	Canadian Institute of Health Research
GRRAS	Guidelines for Reporting Reliability and Agreement Studies
ILI	Influenza-like illness
NIH	National Institute of Health
OR	Odds ratio
P	*P*-value
PCR	Polymerase Chain Reaction
RCT	Randomized controlled trial
RT-PCR	Real Time Polymerase Chain Reaction
Se	Sensitivity
SD	Standard deviation
Sp	Specificity
TA	Total agreement

CHAPTER 1

INTRODUCTION

This introduction chapter sets the context for the work presented in this thesis. This chapter presents the objectives and goals of the thesis, background information, methodology, study population, and brief description of each manuscript.

Objectives and Goals

The overall objective of this thesis was to evaluate the comparability of influenza-related events self-reported by Hutterite research participants and their outpatient medical records. Using data collected from the Hutterite Influenza Prevention Trial, we set the following goals of the thesis:

1. To assess the agreement of the presence or absence of influenza-related symptoms (fever, cough, runny nose, sore throat, headache, sinus problems, muscle aches, fatigue, earache, and chills) between outpatient research participants' concurrent report and health care providers' documentation in the medical records.
2. To examine how well research participant self-report (and maternal report) of otitis media corresponds with physician identification in the medical records.
3. To determine the validity of research participant's self-report of prescribing of antibiotics for respiratory-related illness compared to the medical charts.
4. To assess the utility of two data sources (self-report and medical records) in determining the surveillance definition of influenza-like illness (ILI).

5. To discuss the measurement of agreement of binary data between two raters in family medicine research.

Background and Motivation

Epidemiological and family medicine research must ensure the quality of the raw data generated for analysis. Investigators must bear in mind the importance of ensuring accurate and reproducible data on which to base sound recommendations from research findings. Self-report is an important source of data in epidemiological and family medicine research. The presence of symptoms is typically assessed by patient-reportable information. However, little is known about the quality of self-reported data on influenza symptoms, outpatient diagnosis of otitis media and prescription antibiotics. If possible, research participant responses should be compared with individual patient records to ensure accuracy; yet due to difficulties in obtaining records from physicians and hospitals, it is not always possible to confirm details of an outpatient visit. It is uncertain whether self-report responses alone can be reliably used in Hutterite outpatients seeking treatment for respiratory-related illness. Our objective in this thesis was to examine the quality and reproducibility of self-report responses. This thesis includes a specific population: Hutterites living in the Canadian prairie provinces. However, other clinical trials and epidemiological surveys should consider reproducibility and agreement between data sources as part of their quality control.

No medical record is perfectly accurate because it includes subjective assessments by individual physicians and also relies on patient reports. Nevertheless, the medical

record is an important source of patient information for prospective randomized trials and retrospective observational clinical studies. Validating self- and parent report and medical record documentation is essential for research and clinical care. Carefully conducted research may depict erroneous conclusions if relying on flawed data sources. To answer the question of whether self-report is an adequate tool for monitoring influenza events, we investigated the comparability of medical records and self- and parental report. Although literature exists about the accuracy of self-reported health information, the results are inconsistent and fragmentary. Agreement between self-report and the medical record varies, depending upon the specific condition that is being evaluated.[1] Remarkably few studies have evaluated the comparability of self-report of influenza-related symptoms, respiratory-related diagnoses or prescriptions for respiratory illness with medical records.

Accurate diagnosis of influenza is predicated on the history, physical examination, and results of laboratory testing. A key part of the history involves obtaining a comprehensive description of the symptoms that motivated the patient to seek care. The documentation of symptoms provides clinicians and researchers with important information about patients' experiences and a thorough assessment may lead to better symptom management. Accurate and complete medical records should be available as a reference to ensure quality patient care. Some studies have assessed the agreement between patient self-report and medical record data, and found that agreement differs depending on the medical issue.[4] We are not aware of studies that address the quality of self-report in Hutterite patients regarding influenza-related events.

Study Population

The study population consisted of residents of Hutterite colonies from the provinces of Alberta, Saskatchewan, and Manitoba. Hutterites make up the largest rural group in these provinces.[5] Hutterites are Anabaptists, who live in communal, self-governing, mostly thriving, technologically advanced, farming colonies of about ten to 25 families with anywhere from 60 to 200 people on one colony. There are approximately 347 colonies in Canada: 179 in Alberta, 61 in Saskatchewan, 105 in Manitoba, and two colonies in British Columbia. Their practice of communal living and sharing of material goods differentiates the Hutterites from other Anabaptist groups, such as the Amish and the Mennonites.[6]

Hutterite individuals are culturally integrated within their group. Behaviours that are known to affect health (such as food intake and exercise) follow cultural demands and expectations. For example, although each family lives in their own house, meals are prepared in the common kitchen and eaten in the communal (adult or children) dining halls; men eat with men, women eat with women, children eat together. Hard work and pacifism are basic tenets of their religious and community life. Hutterite individuals do not have wealth or income; colonies do. Work, food, clothing, shelter, and care after childbirth and old age are guaranteed for every member.[7]

Compared to other communities, Hutterite colonies have a relatively higher rate of within colony social interactions and lower rate of social interactions external to the colonies. Thus, they were an ideal setting for participation in the Hutterite Influenza

Prevention Trial testing the concept of herd immunity; that is, whether immunization of healthy children and adolescents with inactivated influenza vaccine would reduce laboratory-confirmed influenza in other community residents.

The Hutterites as health care users

General health conditions and mortality rates of Hutterites have been reported as slightly above those of the general population in the Unites States.[8] Alienation and social collapse are virtually absent; mental illness is rare. [9] Compared to other farming populations, Hutterites are less likely to be exposed to homes with signs of dampness or heated with natural gas. Hutterites do not have pets, suppress smoking and are not exposed to cigarette smoking in the home. Hutterite children also have a lower prevalence of asthma and allergies compared to non-Hutterite farm children.[10] Although colonies are independent and geographically detached from the larger society, Hutterites make good use of available health care services. Individuals who are ill go to physicians in nearby towns. Hutterites are willing to use their own resources for special treatment of chronically ill colony members, i.e. sending members to clinics in large cities following physician recommendation.[7]

Good health is important to the Hutterites. Poor health is disruptive when it prevents a person from performing his or colony work. Loss of the ability to work is a loss of status. As much as possible, the Hutterites try to maintain continuity of health care.[8] However, the best physicians are usually in large urban centres and so unavailable to some Hutterites because of the great distances.[7] Therefore, Hutterites may be relying

on emergency hospital services (and walk-in clinics) for primary and non-urgent medical care.

Documentation in the research literature regarding *current* health resource utilization by Hutterites is largely absent. Boycott and colleagues (2008) claim that "many members of the Hutterite population advocate for state-of-the art health care service delivery for their community" based on surveyed Hutterites family's attitudes and feelings towards genetic testing for cystic fibrosis.[11] Independent systematic surveys indicate that adult Hutterites seek medical care more often that non-Hutterites. One population-based study of Manitoba Hutterites found that the Hutterites used medical services 300% more than non-Hutterite controls after the age of 30.[12, 13] Local physicians in the prairies have noted that Hutterites tend to seek medical help early in an illness rather than late.[12] Physicians who see Hutterites have also reported that headaches, constipation and a range of neurotic problems make up the main medical complaints.[7] To our knowledge, research has not been undertaken to assess why these issues are prevalent among the Hutterites.

Hutterites as research subjects

Epidemiological studies of unique communities are important for assessing disease prevalence, health services utilization, health care needs and health economic analyses. The Hutterian concept of health is largely understood in spiritual terms and closely aligned with the shared ideal of living a life based on selflessness. Good physical

health is a gift from God. Illness is not considered a punishment, but rather, a burden that one must bear, partly as a test. Rather than praying for good health, they pray for the wisdom to know how to live a healthy life or bear their suffering without complaint.[15] Evidence of this phenomenon can be seen in pregnancy. There is no recognition of pregnancy on the colony, no modification to the pregnant woman's work schedule and practically no information given regarding labour and delivery. Women and girls have been taught to ignore pain or discomfort and to accept without complaint whatever be their lot.[7] This attitude toward health and their regular use of health services may impact their self-report of influenza-related events.

Although our sample is not representative of the general outpatient population, information on the accuracy of self-reports is equally important in non-representative samples, especially since non-representative samples are often entered into intervention studies.

Factors that may influence participant-physician agreement

Baseline data were collected on participants following enrolment into the Hutterite Influenza Prevention Trial that could potentially influence participant-provider agreement, such as at sex, age, and participant's risk status for seasonal influenza. Other studies have looked ethno-cultural background, income, and socio-economic status as moderators of agreement. In our sample, all three variables were uniform. As mentioned above, Hutterite individuals do not have a personal income; all earnings are held in common and the funds for essentials are distributed according to need.[7]

In Chapters 2, 3 and 4, agreement and validity statistics were calculated in four strata defined by sex, age group, level of risk for influenza complications, and number of sick days. Logistic regression modelling was used predict the odds of the research participant and the medical record being in agreement. The dependent variable was agreement, coded as 1 for agreement (if the participant and medical report both reported the presence of the symptom or both reported the absence of the symptom) or 0 for disagreement. Sex, age groups, risk level for influenza complications, and number of sick days were categorical variables. Results were expressed as odds ratios and their 95% confidence intervals, which estimate how each independent variable affects the probability of symptom agreement.

The age groups used for the analyses were specifically chosen to reflect specific demographic characteristics. Children less than seven years of age are too young to provide assent; it assumed that the mother or other family member provided self-reported information regarding the child's symptoms and medical visits. Education and marital status are captured by age and, so, excluded as stratification variables. The education levels of individuals have little variation within colonies; traditionally, it is compulsory that children complete grade eight. Individuals up to 15 years of age are schoolchildren. While the local school boards provide a teacher and the standard provincial curriculum is followed all children must attend "German school" (which teaches German and the Hutterite way of life) before and after "regular" school classes. At the age of 16 years, adolescents finish school and become apprentices in the farming and household duties of the colony. Work is assigned along the lines of traditional gender roles: males are

assigned farming and agricultural jobs; women are assigned gardening, cleaning and kitchen duties.[6]

Individual are baptized in their early twenties as adult colony members. Most Hutterites are married by the age of 23, following baptism. Few Hutterites fail to get married and divorce is forbidden. Only 1.9 percent of Hutterite men and 5.4 percent of Hutterite women over the age of 30 have never been married, and only one divorce and four desertions have been reported since 1875. Since 1980, the average age of marriage has been 24.9 for women and 26.0 for men.[6] Therefore, the age group 16 to 22 years includes unmarried young adults and the age group 23 to 49 years includes baptized, married and working Hutterites. The age group 50 to 64 years includes "retired" adults who have been relieved of their colony jobs, having earned their rest.[7] At the age of 65 years, individuals are considered at high risk for influenza complications.

The high risk group included subjects with chronic medical conditions, person 65 years or older, children 23 months and younger, and pregnant women. Sick days were calculated based on self-reported data from the study diaries and nurse interviews, and categorized as: (1) one to three sick days at the time of the medical visit, representing the acute period of potential infection; and (2) four or more days of experiencing symptoms at the time seeking medical attention.

The influence of specific factors on agreement as an outcome have been studied either by stratification or regression models. Stratification requires fewer assumptions than regression. For example, we do not need to make a formal assumption about the relationship between age and participant-provider agreement on the presence of

symptoms using stratification. In contrast, in regression we formally assume that the presence of influenza like illness will predict the likelihood that a patient has influenza.[16]

Agreement and reliability

This thesis assesses reproducibility, or the degree to which medical records and self-report provide similar results regarding the presence or absence of influenza-related events. To study the methods of data collection is necessary to evaluate whether data are reproducible for research purposes and whether the prevalence of influenza-related events can be estimated dependably. Agreement estimates assess the degree of congruency for repeated measurements by estimating measurement error and are used for evaluative purposes. Agreement can be distinguished from reliability, which also deals with reproducibility. Reliability estimates assess how well study objects can be distinguished from each other, despite measurement errors, and are used for discriminative purposes.[17]

For continuous measurements, such as blood pressure, agreement may be assessed by calculating the correlation statistic, intraclass correlation, regression, line fitting, absolute differences, and other summary analyses. Because influenza-related events were binary measurements (e.g. yes/no, present/absent, positive/negative), the following estimates were calculated in Chapters 2 and 6: total agreement, kappa, positive agreement and negative agreement. Chapters 3 and 4 are validity studies; we analyzed the validity of patient reports of physician-diagnosed otitis media and antibiotic prescribing compared with medical record documentation. Chapter 5 builds on the findings in Chapter 2 and examines the issue of combining symptom data from different

sources, in addition to comparing that data as predictors of laboratory-confirmed influenza.

Thesis Overview

The first paper "*Agreement between self-report and medical records on influenza-related symptoms*" looks at the reporting of ten influenza-related signs/symptoms (fever, cough, runny nose, sore throat, headache, sinus problems, muscle aches, fatigue, earache, and chills) by first calculating prevalence of each symptom by self-report and medical recording documentation, and then calculating total agreement, kappa, positive agreement and negative agreement between the two data sources.

The second paper in the thesis "*Self and parental report of physician-identified otitis media in a rural sample*" assesses the validity of otitis media reporting by study participants and their mothers (for young children) compared to the medical records. The third paper "*Accuracy of self-reported antibiotic prescription in a Hutterite sample*" also assesses the validity of self-report compared to medical record documentation focusing on antibiotic prescription receipt for influenza-related illness.

The fourth paper "*A comparison of self-report and medical record data to assess surveillance definitions of influenza-like illness in outpatients"* extends the findings in the second paper by looking at not only the comparison of symptoms by data sources, but also combining the available data. In the second paper, the symptoms were the outcomes; here, symptoms and symptoms complexes within ILI surveillance definitions are treated as predictors of laboratory confirmed influenza.

The fifth paper *"Measuring agreement for binary data between patient self-report and medical records"* is a concise methodological paper intended for a family physician audience elaborating and summarizing the assessment indicators that were used in the second paper.

There is some overlap in the methods sections of the five papers that was necessary for each paper to be able to stand alone as a publication. However, each introduction focuses on the issues specific to the paper's objectives, and each methods section gives only enough detail about the overall study design to allow the reader to understand the study, while focusing on the methods specific to each paper's main objectives.

References

1. Englert H, Muller-Nordhorn J, Seewald S, et al. Is patient self-report an adequate tool for monitoring cardiovascular conditions in patients with hypercholesterolemia? *J Public Health (Oxf)*. 2010.

2. Monto AS, Gravenstein S, Elliott M, Colopy M, Schweinle J. Clinical signs and symptoms predicting influenza infection. *Arch Intern Med*. 2000;160(21):3243-3247.

3. Horwitz RI. Comparison of epidemiologic data from multiple sources. *J Chronic Dis*. 1986;39(11):889-896.

4. Nagurney JT, Brown DF, Sane S, Weiner JB, Wang AC, Chang Y. The accuracy and completeness of data collected by prospective and retrospective methods. *Acad Emerg Med*. 2005;12(9):884-895.

5. Loeb M, Russell ML, Moss L, et al. Effect of influenza vaccination of children on infection rates in Hutterite communities: A randomized trial. *JAMA*. 2010;303(10):943-950.

6. Hostetler JA, Huntington GE. *The Hutterites in North America.* Mason, Ohio: Cengage Learning; 2002.

7. Hostetler JA. *Hutterite Society.* Baltimore: Johns Hopkins University Press; 1997.

8. Spencer V. Study of the Hutterites and a specific health program. 1963.

9. Ingoldsby B, Smith R. Public school teacher's perspectives on the contemporary Hutterite family. *J Comp Fam Studies.* 2005;36:249-265.

10. Rennie DC, Dosman J, Senthilselvan A. Respiratory symptoms and asthma in two farming populations: A comparison of Hutterite and non-Hutterite children. *Can Respir J*. 2002;9(5):313-318.

11. Boycott KM, Parboosingh JS, Chodirker BN, et al. Clinical genetics and the Hutterite population: A review of Mendelian disorders. *Am J Med Genet A*. 2008;146A(8):1088-1098.

12. Ross RT, Cheang M. Common infectious diseases in a population with low multiple sclerosis and varicella occurrence. *J Clin Epidemiol*. 1997;50(3):337-339.

13. Ross RT, Nicolle LE, Cheang M. Varicella zoster virus and multiple sclerosis in a Hutterite population. *J Clin Epidemiol*. 1995;48(11):1319-1324.

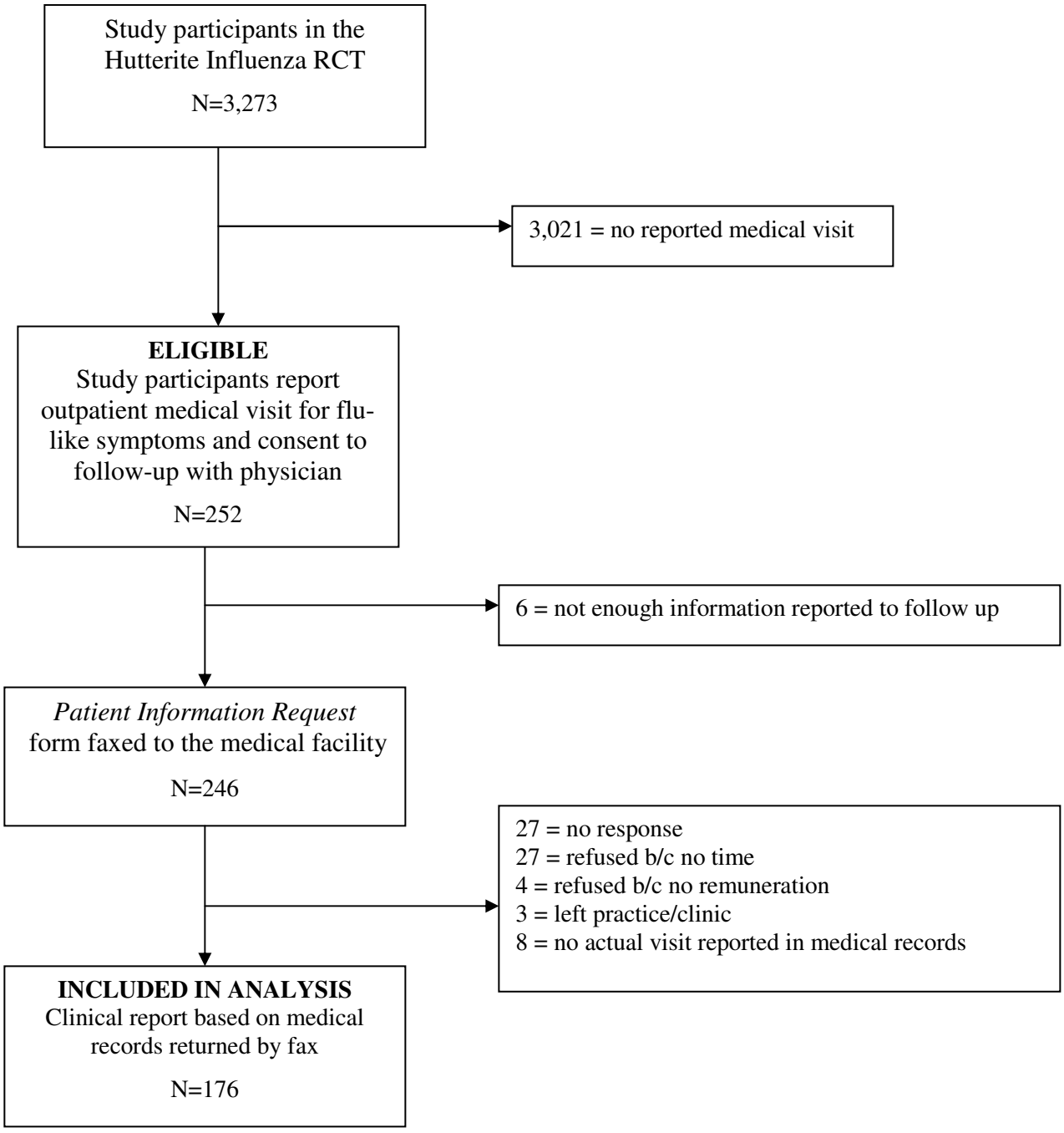
14. Nimgaonkar VL, Fujiwara TM, Dutta M, et al. Low prevalence of psychoses among the Hutterites, an isolated religious community. *Am J Psychiatry*. 2000;157(7):1065-1070.

15. Brunt JH, Lindsey E, Hopkinson J. Health promotion in the Hutterite community and the ethnocentricity of empowerment. *Can J Nurs Res*. 1997;29(1):17-28.

16. McNamee R. Regression modelling and other methods to control confounding. *Occup Environ Med*. 2005;62(7):500-6, 472.

17. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol*. 2006;59(10):1033-1039.

Figure 1: Flow diagram of participants included in the thesis



CHAPTER 2

REPORTING OF INFLUENZA SYMPTOMS

This manuscript has been submitted for publication to *Canadian Family Physician:*

> **Barbara AM,** Loeb M, Dolovich L, Brazil K, Russell MK. Agreement between self-report and medical records on influenza-related symptoms.

Agreement between self-report and medical records on influenza-related symptoms

Angela M. Barbara[1], Mark Loeb[2], Lisa Dolovich[3], Kevin Brazil[4], Margaret L. Russell[5]

[Quantitative Research]

[1]Department of Pathology and Molecular Medicine, McMaster University

[2]Department of Pathology and Molecular Medicine, Department of Medicine, Department of Clinical Epidemiology and Biostatistics, McMaster University

[3]Department of Family Medicine, Center for Evaluation of Medicines, McMaster University

[4]St. Joseph's Health System Research Network; Department of Clinical Epidemiology and Biostatistics, McMaster University

[5] Department of Community Health Sciences, University of Calgary

Running head: Agreement for influenza symptoms

Word Count: 2,481

Tables: 2

Abstract

Objective To assess the agreement between self-report and documentation in the medical records of influenza-related signs/symptoms (fever, cough, runny nose, sore throat, headache, sinus problems, muscle aches, fatigue, earache, and chills).

Design Cross sectional study.

Setting Research participants in the Hutterite Influenza Prevention Study who made outpatient visits for influenza-related signs/symptoms during the 2008-2009 influenza season.

Participants The sample included individuals with information about the presence or absence of influenza signs/symptoms from both self-report and medical records.

Main outcomes measures We calculated prevalence of each sign/symptom by source of data. We measured agreement between self report and medicals records using total agreement, kappa, positive agreement and negative agreement.

Results In comparison to the medical record, we found lower rates of self-reported prevalence for fever, sore throat and earache. Total agreements between self-report and medical report of symptoms varied between 61% (for sore throat) and 88% (for muscle aches and earache). Negative agreement was considerably higher (68% for sore throat to 93% for muscle aches and earache) than positive agreement (13 % for chills to 58% for earache) for each symptom; except cough, where positive agreement (77%) was higher than negative agreement (64%).

Conclusion Agreements were variable depending on the specific symptom. Contrary to research in other patient populations that suggests clinicians report fewer symptoms that

patients, we found that the medical record captured more symptoms compared to self-report. The choice of data source for symptom evaluation should depend on the specific population, outcome of interest, and whether the results will be used for clinical decision making, research, or surveillance.

Keywords: agreement, symptom, self-report, medical records

Abstract Word Count: 253 (without titles)

Agreement on patient symptoms across data sources is relevant to both clinical practice and research. Agreement between clinicians and patients regarding the assessment of presenting clinical symptoms is important for patient satisfaction[1-5] and symptom resolution[6-8] In research, differences in estimations of the prevalence or incidence of symptoms that are dependent on data source (i.e., patient self report versus documentation in medical record) can lead to substantial differences in estimated disease parameters.[9] An understanding of the relative agreement between the occurrence of symptoms by self report compared to medical records can be useful in the interpretation of the clinical and research literature.

Some studies have assessed the agreement between patient self-report and medical record data, and found that agreement differs depending on the medical issue.[10-15] Symptom research has looked at patient-clinician agreement of symptoms related to angina pectoris,[16] myocardial infarction;[17, 18] psychological and somatic disorders,[1, 19-22] HIV infection,[23] and cancer.[24, 25] One study found fair to substantial agreement between adult self-assessment and clinician assessment on the symptoms related to "strep throat".[26] However, there is a relative paucity of reports for infectious diseases, particularly for influenza.

A large clinical trial on influenza among Hutterite community members used both self-report and medical records to collect data on influenza-related symptoms and allowed for the assessment of agreement between sources. The specific objective of the current study was to quantify the agreement between research participants' concurrent

report of the presence or absence of ten influenza-relevant signs/symptoms and documentation of these in the medical records.

Methods

Study design and population

The present study is a cross-sectional analysis of data collected for a cluster randomized trial of vaccinating children in Hutterite communities against influenza. Details of the Hutterite randomized controlled trial are described elsewhere.[27]

Patient reports of influenza-related symptoms

Self report data were collected by study diaries (completed by a family representative) and in-person interviews by trained research nurses from December 28, 2008 to June 23, 2009. During this period, trial participants completed family diaries of daily checklists for signs and symptoms of influenza: fever, cough, runny nose, sore throat, headache, sinus problems, muscle ache, fatigue, ear ache, and chills. Fever was defined as a temperature >=38 degrees Celsius. Participating families were given similar thermometers to take oral temperatures for this purpose.

Trained research nurses visited the Hutterite colonies twice per week to check diary entries and interviewed individual participants (or parents, in the case of infants; typically the mother) to confirm the reported symptoms and assess other symptoms. The focus was on "new" or episodic symptoms that are predictive of influenza, rather than chronic or ongoing symptoms. The research nurses also collected information regarding outpatient health care visits made for flu-like symptoms, including physician name or health care facility, location, and date of each medical visit.

Requests for information from medical records

Medical information was gathered from the health care clinics and hospitals. The Canadian Medical Directory (2009 edition) and online physician registries were used to obtain contact information of physicians for whom participants had provided incomplete addresses. For each reported medical visit, a one-page "Patient Information Request" form was faxed to the medical facility asking for individual patient record data regarding presenting symptoms, with an equivalent list of symptoms as on the study diaries (Table 1). Clinicians were blind to the patient's self reported symptoms. The institutional review boards at McMaster University, University of Calgary, University of Saskatchewan, and the University of Manitoba approved the study. Faxes were sent to the physician offices or medical facilities between March and September 2009. The primary analysis was restricted to an individual's first confirmed medical visit to maintain independence of observations.

Statistical analyses

Prevalence and individual two-by-two contingency tables were calculated for each of the ten symptoms. To test for differences in mean number of reports per source, we used the paired Student *t* test. Significance levels were set at $p < 0.05$. For symptom agreement, neither of the two data sources was assigned as the criterion or "gold" standard index. The analyses focused on the concurrence of the presence or absence of each symptom by the two data sources: medical record and self-report. Total agreement (number of concordant pairs / total sample) and kappa coefficient (and standard deviation) were computed. Kappa measures the strength of agreement beyond that

expected solely by chance ([observed agreement – chance agreement] / [1 – chance agreement]), where 0 = chance agreement and 1 = perfect agreement.[28] Due to the challenges associated with interpreting kappa values,[29, 30] we also calculated positive agreement (concordance in positive responses by both sources) and negative agreement (concordance in negative responses by both sources). [31, 32] Crosstabulations and kappa values were computed using SPSS 16.0 (SPSS Inc., Chicago, IL).

Results

Availability of data for both medical records and self-reports

Of the 3,273 trial participants, 252 individuals (8%) on 37 of the 46 enrolled colonies (80%) reported at least one outpatient medical visit during the study influenza season. We included only the first medical visit reported for the present analyses. Of the 252, 246 provided sufficient information to contact the care source utilized. Despite multiple attempts to contact care source, replies were received for only 184 of the 246 participants; and the occurrence of a medical visit was confirmed by care source for 176 of the 184. These 176 individuals were included in the sample for analyses.

Sample characteristics

The mean age was 24 years; more than a third of the sample (36%) was under the age of seven years. Just over a quarter (26%) were between the ages of 23 and 49 years and 15% were over the age of 50 years. Sixty three percent were female. Thirty-nine percent of individual were considered to be at high risk for influenza complications, because of chronic medical conditions, age (children under 24 months and adults 65 years and older), or pregnancy. Medical visits were made between January and June 2009.

Because we used data from participants' first confirmed medical visits reported during the influenza season, 141 (80%) were made prior to the introduction of the novel H1N1 pandemic influenza in Canada on April 23, 2009.[33] At least one of the ten symptoms was self reported by 142 (81%) persons. Of the 142, 48% were symptomatic for less than four days at the time of the medical visit; the average number of sick days was 3.7 (SD = 4.5). According to the medical records, 162 (92%) individuals were diagnosed with a respiratory illness; including otitis media (24%), upper respiratory tract infection (17%), sinusitis (12%), pharyngitis (12%), bronchitis (12%), pneumonia (4%) and influenza (3%). Most received care from a family physician office (80%), while 17% visited a hospital emergency department.

Prevalence of symptoms by data source

Table 2 shows the prevalence of the ten symptoms as estimated from the medical record and from self-report. There were three symptoms/signs for which there was a statistically significant difference in prevalence by data source: fever, sore throat, and earache; in all cases the self reports were underestimations compared to medical records. The differences in prevalence were 19% for fever, 9% for sore throat and 5% for earache.

The medical records indicated that the sample experienced a higher number of the ten symptoms compared to self-report (paired t = 2.2, $p = 0.03$). Patients self-reported an average of 2.1 of the ten symptoms (SD = 1.8, minimum to maximum = 0 – 8); the medical records indicated a mean of 2.5 symptoms (SD = 1.4, minimum to maximum = 0 – 8) per subject.

Symptom agreement between patient self-report and medical report

The proportion of total agreement between self-report and medical record report varied between 61% (for sore throat) and 88% (for muscle aches and earache) (Table 1). The highest kappa values were for earache (0.51) and cough (0.41). Other Kappa values varied from 0.38 (for sinus problems) to 0.05 (chills). The *p* values (<0.05) for fever, cough, runny nose, sore throat, headache, sinus problems, muscle aches and earache indicated that agreements for these symptoms were not due to chance. Negative agreement was considerably higher (ranging from 68% for sore throat to 93% for muscle aches and earache) than positive agreement (ranging from 13 % for chills to 58% for earache) for each symptom; except cough, where positive agreement (77%) was higher than negative agreement (64%).

Discussion

Influenza research and clinical care rely on the monitoring of respiratory symptoms. This information is often obtained directly from research participants by self-administered surveys or interview,[34] which can be relatively cost efficient and organizationally straightforward to implement.[35, 36] However, the limitations of self-report relate to accuracy, recall, interviewer skills, and willingness to report.[37] Another common method for assessing symptoms is the review of written medical records, which can be costly, labour-intensive and time-consuming,[38] especially in settings like large province-wide or nation-wide studies where study participants access different medical services across diverse geographic areas. Medical record abstraction is further limited by illegibility, varying levels of completeness, and inaccuracies resulting from delayed documentation by busy physicians.[39, 40]

In this study, we compared data collected from self reports and from medical records. Because of the active surveillance conducted by research nurses, the self-reported data were collected in near 'real time' and should be considered prospective and unlikely to be biased by memory recall. The data in the medical records are most appropriately considered to be retrospective. The two concurrent data sources in this study provided insights into the congruence between methods.

Agreements between medical record data and self-report were symptom-dependent. Rates of agreement were good for fever, cough, headache, sinus problems, muscle aches, fatigue, earache and chills (minimum to maximum = 72% to 88%), but less adequate for sore throat (61%) and runny nose (64%). The moderate overall agreement for fever (74%) may have been affected by differences in measurement. Fever was explicitly and objectively defined for the randomized trial as a temperature >=38 degrees Celsius, and measured using consistent methods for each participant experiencing symptoms. Physicians may not have been so consistent in methods used to measure fever (i.e., may have been documented on the basis of patient complaint without measurement). The occurrence of fever at the time of the medical visit might also have been influenced by phase of infection or use of antipyretics.[41]

Rates for positive agreement and negative agreement suggest that any corrective action to improve agreement should be concentrated on reporting the *presence* of influenza-related symptoms. For nine symptoms, there were lower positive agreements (13% to 77%) and higher negative agreements (64% to 93%). This indicates poorer agreement regarding the presence of symptoms with an imbalance weighted towards the

absence of symptoms. The exception to this pattern was cough; positive agreement was 77% and negative agreement was 64%. Agreement regarding the absence of cough and runny nose can also be improved.

The grouping of fever and cough has been established as predictive of clinical diagnosis of influenza during a seasonal epidemic.[42-44] It is possible that by focusing on these obvious symptoms of influenza, such as fever and cough, physicians did not prompt for or document other symptoms. This may partly explain why both fever and cough had higher prevalence according to medical records compared to self-report. It has also been shown that physicians reliably record data about their patients' main complaints or classic symptoms, but not the less typical symptoms.[45]

Symptom disagreement may have stemmed from the individual and combined influences of the following: the symptoms themselves, the perspectives of the observer, measurement error, and the context in which the symptoms were observed and recorded.[46] Other studies have found that agreement is better with regards to concrete, objective clinical signs that require less interpretation from others.[13, 47-50] Fatigue has been referred to as a "subjective" symptom.[18, 51] Prevalence of fatigue in the medical records was extremely low (8%), suggesting the possibility that clinicians are less likely to recognize or document this symptom. The low prevalence also contributed to the low kappa value (0.13) despite fairly good agreement (84%). The difference between directly experiencing a symptom and externally observing a behaviour that is indicative of a symptom should lead to different evaluations. The situational or contextual basis of the

judgement also differs.[52] This suggests that both perspectives are worthwhile and have unique contributions.

The observed variability in prevalence by data source in our study may have resulted from limitations or errors at each source. Self-reported information can be imprecise for various reasons, including better understanding of some symptoms than others, underreporting, lack of motivation to report accurately, and poor compliance. Medical records can also be problematic. Several studies have found non-reporting and misreporting in medical records.[53, 54] Busier physicians may record less in the medical record or delay recording, leading to errors in recall.[40] The process of abstracting information from the medical chart itself is also subject to imprecision.[16, 55] Furthermore, medical records accessed for this analysis were not written or kept for the purposes of this study and were guided by institutional policy, provider training and provider preference.[55, 56]

Discordances may also be attributed to the differences between the research setting and the clinical setting; such as the nature of patient-physician (or participant-researcher) interaction, differential elicitation of symptoms, variation in reporting styles (specifics of the documentation system[16] versus the research protocol for data collection), and environment (community or medical facility). For example, in the clinical setting, symptom information is often collected passively during the patient visit and then documented in the medical chart.[24] For the clinical trial, participant checklists were used, followed by face-to-face interviews. Checklists have been shown to capture more symptom complaints compared to open ended and passive reports.[57]

Our findings may also be explained by very different motivations for reporting symptoms in each context. As part of the study protocol, the research nurses obtained nasopharyngeal specimens (or nasal swabs) if a participant reported two or more symptoms. We conjecture that patients may report more specific symptoms to their physicians to get a prescription for antibiotics, but the same individuals as research participants will under-report symptoms to avoid the discomfort of a nasopharyngeal swab.

By utilizing a homogeneous population and focusing on a specific set of symptoms, generalizability is limited. We found that that, overall, medical record captured more symptoms compared to self-report. This is contrary to symptom research in other populations that suggests clinicians report fewer symptoms than patients.[24, 26, 58, 59] We suspect that participants may have underestimated their symptoms; especially fever, sore throat and earache; and not that medical records overrated their symptoms. The Hutterites are known as being "stoic" and bearing pain and physical ailments without complaint.[60] Participants may have hesitated to report symptoms to the research nurse to avoid being perceived as complainers, or as previously mentioned, to avoid the discomfort of nasopharyngeal swabbing.

The findings of the study indicate that deciding which data source to use for symptom evaluation depends on the population and outcome of interest and whether the results will be used for clinical decision making, research, or surveillance.[40] Information from patient medical records might be a valuable supplement to self-report; thus, enhancing the probability that symptoms are fully captured by research investigators.

References

1. Zastrow A, Faude V, Seyboth F, Niehoff D, Herzog W, Lowe B. Risk factors of symptom underestimation by physicians. *J Psychosom Res*. 2008;64(5):543-551.

2. Staiger TO, Jarvik JG, Deyo RA, Martin B, Braddock CH,3rd. BRIEF REPORT: Patient-physician agreement as a predictor of outcomes in patients with back pain. *J Gen Intern Med*. 2005;20(10):935-937.

3. Cremeans-Smith JK, Stephens MA, Franks MM, Martire LM, Druley JA, Wojno WC. Spouses' and physicians' perceptions of pain severity in older women with osteoarthritis: Dyadic agreement and patients' well-being. *Pain*. 2003;106(1-2):27-34.

4. Vermeire E, Hearnshaw H, Van Royen P, Denekens J. Patient adherence to treatment: Three decades of research. A comprehensive review. *J Clin Pharm Ther*. 2001;26(5):331-342.

5. Richards HL, Fortune DG, Weidmann A, Sweeney SK, Griffiths CE. Detection of psychological distress in patients with psoriasis: Low consensus between dermatologist and patient. *Br J Dermatol*. 2004;151(6):1227-1233.

6. Silverman J, Draper J. Identifying the agenda in the consultation. *Br J Gen Pract*. 1995;45(390):52-53.

7. Starfield B, Wray C, Hess K, Gross R, Birk PS, D'Lugoff BC. The influence of patient-practitioner agreement on outcome of care. *Am J Public Health*. 1981;71(2):127-131.

8. Stewart MA, McWhinney IR, Buck CW. The doctor/patient relationship and its effect upon outcome. *J R Coll Gen Pract*. 1979;29(199):77-81.

9. Zhu K, McKnight B, Stergachis A, Daling JR, Levine RS. Comparison of self-report data and medical records data: Results from a case-control study on prostate cancer. *Int J Epidemiol*. 1999;28(3):409-417.

10. Tisnado DM, Adams JL, Liu H, et al. Does concordance between data sources vary by medical organization type? *Am J Manag Care*. 2007;13(6 Part 1):289-296.

11. Stange KC, Zyzanski SJ, Smith TF, et al. How valid are medical records and patient questionnaires for physician profiling and health services research? A comparison with direct observation of patients visits. *Med Care*. 1998;36(6):851-867.

12. Fowles JB, Fowler EJ, Craft C. Validation of claims diagnoses and self-reported conditions compared with medical records for selected chronic diseases. *J Ambul Care Manage*. 1998;21(1):24-34.

13. Sneeuw KC, Sprangers MA, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease. *J Clin Epidemiol*. 2002;55(11):1130-1143.

14. Sprangers MA, Sneeuw KC. Are healthcare providers adequate raters of patients' quality of life--perhaps more than we think? *Acta Oncol*. 2000;39(1):5-8.

15. Velikova G, Wright P, Smith AB, et al. Self-reported quality of life of individual cancer patients: Concordance of results with disease course and medical records. *J Clin Oncol*. 2001;19(7):2064-2073.

16. Pakhomov SV, Jacobsen SJ, Chute CG, Roger VL. Agreement between patient-reported symptoms and their documentation in the medical record. *Am J Manag Care*. 2008;14(8):530-539.

17. Fukuoka Y, Dracup K, Ohno M, Kobayashi F, Hirayama H. Symptom severity as a predictor of reported differences of prehospital delay between medical records and structured interviews among patients with AMI. *Eur J Cardiovasc Nurs*. 2005;4(2):171-176.

18. DeVon HA, Ryan CJ, Zerwic JJ. Is the medical record an accurate reflection of patients' symptoms during acute myocardial infarction? *West J Nurs Res*. 2004;26(5):547-560.

19. Greer J, Halgin R. Predictors of physician-patient agreement on symptom etiology in primary care. *Psychosom Med*. 2006;68(2):277-282.

20. Comino EJ, Silove D, Manicavasagar V, Harris E, Harris MF. Agreement in symptoms of anxiety and depression between patients and GPs: The influence of ethnicity. *Fam Pract*. 2001;18(1):71-77.

21. Rost KM, Dickinson WP, Dickinson LM, Smith RC. Multisomatoform disorder: Agreement between patient and physician report of criterion symptom explanation. *CNS Spectr*. 2006;11(5):383-388.

22. Joiner TE,Jr, Rudd MD, Rajab MH. Agreement between self- and clinician-rated suicidal symptoms in a clinical sample of young adults: Explaining discrepancies. *J Consult Clin Psychol*. 1999;67(2):171-176.

23. Fontaine A, Larue F, Lassauniere JM. Physicians' recognition of the symptoms experienced by HIV patients: How reliable? *J Pain Symptom Manage*. 1999;18(4):263-270.

24. Basch E, Iasonos A, McDonough T, et al. Patient versus clinician symptom reporting using the national cancer institute common terminology criteria for adverse events: Results of a questionnaire-based study. *Lancet Oncol*. 2006;7(11):903-909.

25. Cirillo M, Venturini M, Ciccarelli L, Coati F, Bortolami O, Verlato G. Clinician versus nurse symptom reporting using the national cancer institute-common terminology criteria for adverse events during chemotherapy: Results of a comparison based on patient's self-reported questionnaire. *Ann Oncol*. 2009;20(12):1929-1935.

26. Xu J, Schwartz K, Monsur J, Northrup J, Neale AV. Patient-clinician agreement on signs and symptoms of 'strep throat': A MetroNet study. *Fam Pract*. 2004;21(6):599-604.

27. Loeb M, Russell ML, Moss L, et al. Effect of influenza vaccination of children on infection rates in Hutterite communities: A randomized trial. *JAMA*. 2010;303(10):943-950.

28. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20:37-46.

29. Sim J, Wright CC. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Phys Ther*. 2005;85(3):257-268.

30. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. the problems of two paradoxes. *J Clin Epidemiol*. 1990;43(6):543-549.

31. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. resolving the paradoxes. *J Clin Epidemiol*. 1990;43(6):551-558.

32. Chen G, Faris P, Hemmelgarn B, Walker RL, Quan H. Measuring agreement of administrative data with chart data using prevalence unadjusted and adjusted kappa. *BMC Med Res Methodol*. 2009;9:5.

33. Cutler J, Schleihauf E, Hatchette TF, et al. Investigation of the first cases of human-to-human infection with the new swine-origin influenza A (H1N1) virus in Canada. *CMAJ*. 2009;181(3-4):159-163.

34. Okura Y, Urban LH, Mahoney DW, Jacobsen SJ, Rodeheffer RJ. Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. *J Clin Epidemiol*. 2004;57(10):1096-1103.

35. Englert H, Muller-Nordhorn J, Seewald S, et al. Is patient self-report an adequate tool for monitoring cardiovascular conditions in patients with hypercholesterolemia? *J Public Health (Oxf)*. 2010.

36. Newell SA, Girgis A, Sanson-Fisher RW, Savolainen NJ. The accuracy of self-reported health behaviors and risk factors relating to cancer and cardiovascular disease in the general population: A critical review. *Am J Prev Med*. 1999;17(3):211-229.

37. Goldman N, Lin IF, Weinstein M, Lin YH. Evaluating the quality of self-reports of hypertension and diabetes. *J Clin Epidemiol*. 2003;56(2):148-154.

38. Phillips KA, Milne RL, Buys S, et al. Agreement between self-reported breast cancer treatment and medical records in a population-based breast cancer family registry. *J Clin Oncol*. 2005;23(21):4679-4686.

39. Skinner KM, Miller DR, Lincoln E, Lee A, Kazis LE. Concordance between respondent self-reports and medical records for chronic conditions: Experience from the Veterans health study. *J Ambul Care Manage*. 2005;28(2):102-110.

40. Ferrante JM, Ohman-Strickland P, Hahn KA, et al. Self-report versus medical records for assessing cancer-preventive services delivery. *Cancer Epidemiol Biomarkers Prev*. 2008;17(11):2987-2994.

41. Thursky K, Cordova SP, Smith D, Kelly H. Working towards a simple case definition for influenza surveillance. *J Clin Virol*. 2003;27(2):170-179.

42. Boivin G, Hardy I, Tellier G, Maziade J. Predicting influenza infections during epidemics with use of a clinical case definition. *Clin Infect Dis*. 2000;31(5):1166-1169.

43. Monto AS, Gravenstein S, Elliott M, Colopy M, Schweinle J. Clinical signs and symptoms predicting influenza infection. *Arch Intern Med*. 2000;160(21):3243-3247.

44. Ohmit SE, Monto AS. Symptomatic predictors of influenza virus positivity in children during the influenza season. *Clin Infect Dis*. 2006;43(5):564-568.

45. Romm FJ, Putnam SM. The validity of the medical record. *Med Care*. 1981;19(3):310-315.

46. Kraemer HC, Measelle JR, Ablow JC, Essex MJ, Boyce WT, Kupfer DJ. A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *Am J Psychiatry*. 2003;160(9):1566-1577.

47. Wennman-Larsen A, Tishelman C, Wengstrom Y, Gustavsson P. Factors influencing agreement in symptom ratings by lung cancer patients and their significant others. *J Pain Symptom Manage*. 2007;33(2):146-155.

48. Lobchuk MM, Degner LF. Patients with cancer and next-of-kin response comparability on physical and psychological symptom well-being: Trends and measurement issues. *Cancer Nurs*. 2002;25(5):358-374.

49. Tang ST, McCorkle R. Use of family proxies in quality of life research for cancer patients at the end of life: A literature review. *Cancer Invest*. 2002;20(7-8):1086-1104.

50. von Essen L. Proxy ratings of patient quality of life--factors related to patient-proxy agreement. *Acta Oncol*. 2004;43(3):229-234.

51. Canaris GJ, Tape TG, Smith LM, Nickol DR, Wigton RS. Gender differences in patient-provider symptom agreement in reporting respiratory complaints on a questionnaire. *Gend Med*. 2008;5(2):186-193.

52. Achenbach TM, McConaughy SH, Howell CT. Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychol Bull*. 1987;101(2):213-232.

53. Marrie TJ, Durant H, Sealy E. Pneumonia--the quality of medical records data. *Med Care*. 1987;25(1):20-24.

54. Bush TL, Miller SR, Golden AL, Hale WE. Self-report and medical record report agreement of selected medical conditions in the elderly. *Am J Public Health*. 1989;79(11):1554-1556.

55. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. *Am J Health Syst Pharm*. 2008;65(23):2276-2284.

56. Fathelrahman AI. Agreement between questionnaire and medical records on some health and socioeconomic problems among poisoning cases. *BMC Res Notes*. 2009;2:183.

57. Bent S, Padula A, Avins AL. Brief communication: Better ways to question patients about adverse medical events: A randomized, controlled trial. *Ann Intern Med*. 2006;144(4):257-261.

58. Fromme EK, Eilers KM, Mori M, Hsieh YC, Beer TM. How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the quality-of-life questionnaire C30. *J Clin Oncol*. 2004;22(17):3485-3490.

59. Vogelzang NJ, Breitbart W, Cella D, et al. Patient, caregiver, and oncologist perceptions of cancer-related fatigue: Results of a tripart assessment survey. The fatigue coalition. *Semin Hematol*. 1997;34(3 Suppl 2):4-12.

60. Brunt JH, Lindsey E, Hopkinson J. Health promotion in the Hutterite community and the ethnocentricity of empowerment. *Can J Nurs Res*. 1997;29(1):17-28.

Table 1: Content of *Patient Information Request Form* faxed to medical offices

PATIENT INFORMATION REQUEST

The Hutterite Influenza Study is being conducted by researchers from McMaster University to better understand whether immunizing school-age children against influenza can protect high-risk members of their community.

Your patient, identified on the attached consent form, has agreed to participate in this study and has given us consent to contact you about his/her recent visit on _____ to you for treatment of respiratory infection symptoms.

Please answer the following questions:

<table>
<tr><td colspan="4">1. What was <u>actual date</u> of the patient's visit?</td></tr>
<tr><td>2. What were the patient's symptoms? Check all that apply.</td><td colspan="3">○ Fever (≥38° C)
○ Cough
○ Runny nose
○ Sore throat
○ Headache
○ Sinus problems
○ Muscle aches
○ Fatigue
○ Ear ache
○ Chills
○ Other, specify:
_____</td></tr>
<tr><td>3. What was the diagnosis?</td><td colspan="3">○ Pneumonia
○ Otitis media
○ Other, specify: _____</td></tr>
<tr><td>4. Was an x-ray ordered?</td><td>○ No
○ Yes</td><td>If yes, was there opacity on chest x-ray compatible with pneumonia?</td><td>○ No ○ Yes</td></tr>
</table>

Table 2: Prevalence of each symptom according to each data source and total agreement, kappa statistic, positive agreement and negative agreement of symptoms between medical record and self-report

Symptom	Medical report *n* (%)	Self-report *n* (%)	p-value*	Total Agreement	Kappa (SD)	Positive Agreement	Negative Agreement
Fever (>=38 degrees Celsius)	58 (33)	24 (14)	<0.001	0.74	0.31 (0.07) †	0.44	0.83
Cough	112 (64)	102 (58)	0.16	0.72	0.41 (0.07) †	0.77	0.64
Runny nose	52 (30)	56 (32)	0.62	0.64	0.15 (0.08) †	0.41	0.74
Sore throat	78 (44)	61 (35)	0.04	0.61	0.19 (0.07) †	0.50	0.68
Headache	24 (14)	25 (14)	0.86	0.81	0.21 (0.10) †	0.32	0.89
Sinus problems	35 (20)	27 (15)	0.16	0.82	0.38 (0.09) †	0.48	0.89
Muscle aches	13 (7)	16 (9)	0.51	0.88	0.21 (0.12) †	0.28	0.93
Fatigue	14 (8)	23 (13)	0.10	0.84	0.13 (0.10)	0.22	0.91
Ear ache	33 (19)	19 (11)	0.003	0.88	0.51 (0.09) †	0.58	0.93
Chills	13 (7)	18 (10)	0.34	0.85	0.05 (0.08)	0.13	0.92

*Paired-sampled t-test for proportion
† $p<0.05$

CHAPTER 3

REPORTING OF OTITIS MEDIA

This manuscript has been submitted for publication to *The Canadian Journal of Speech-Language Pathology and Audiology:*

> **Barbara AM,** Loeb M, Dolovich L, Brazil K, Russell MK. Agreement between self-report and medical records on influenza-related symptoms.

Self and parental report of physician-identified otitis media in a rural sample

Angela M. Barbara PhD(C)[1], Mark Loeb MD[2], Lisa Dolovich PharmD[3], Kevin Brazil PhD[4], Margaret L. Russell MD, PhD[5]

[1]Department of Pathology and Molecular Medicine, McMaster University

[2]Department of Pathology and Molecular Medicine, Department of Medicine, Department of Clinical Epidemiology and Biostatistics, McMaster University

[3]Department of Family Medicine, Center for Evaluation of Medicines, McMaster University

[4]St. Joseph's Health System Research Network; Department of Clinical Epidemiology and Biostatistics, McMaster University

[5] Department of Community Health Sciences, University of Calgary

Abstract

Otitis media is a leading cause of family medicine consultations. Rates of otitis media are traditionally determined by review of medical charts, which can be costly and time consuming. This information can also be obtained directly from patients (or parents) by self-administered surveys or personal interview. To ensure the quality of self-reported otitis media as a proxy for physician-recorded diagnosis, we assessed its accuracy compared to medical report documentation. Self (and maternal) report of otitis media at outpatient consultations was collected prospectively by interview from participants in the Hutterite Influenza Prevention Study. Similar data were also collected by fax requests for medical record information to the medical facilities. We calculated prevalence (by each data source), sensitivity, specificity, predictive values and likelihood ratios. Compared to the medical records, the prevalence of otitis media was underestimated by participants (22% versus 16%), but this difference did not reach significance ($p = 0.07$). Self report of otitis media was a very specific measure (93%), but had lower sensitivity (47%). Positive predictive value was moderate at 64% but negative predictive value was good at 86%. The positive likelihood ratio was 6.71, while the negative positive likelihood ratio was 0.57. Self-report of otitis media in our sample had high specificity and good negative predictive value. However, reliance on self-report without verification by medical record may result in a number of false negatives, which may affect enrollment eligibility or outcome analyses in medical research.

Key words: otitis media, self report, medical records

Otitis media (or middle ear infection) is a frequent complication of influenza virus infection and a leading cause of family physician visits (Charles, Pan, & Britt, 2004; Heikkinen & Chonmaitree, 2003; Heikkinen et al., 2004; Vergison et al., 2010). Medical records and parent report (for children) or self reports are common data sources for epidemiological studies of otitis media.

Medical chart review, commonly used for assessing medical events, can be costly, labour-intensive and time-consuming (Phillips et al., 2005). For province-wide or nation-wide studies where study participants access different medical services across large geographic areas, multiple personnel must obtain the data. However, the advantage of medical record review is that it removes the burden of data collection from research participants to the research team (Fukuoka et al., 2005).

When it is not possible to perform clinical tests or consultations, individuals' self-reports are used to measure disease status (Strauss et al., 2001). Information is often obtained directly from research participants by self-administered surveys or personal interview (Okura et al., 2004). Self-report has disadvantages; it can be inaccurate because participants may not be aware of their diagnoses, may misunderstand their diagnoses, may not recall their diagnoses, or may simply not be willing to report (Goldman et al., 2003). However, self reports can be relatively cost efficient and organizationally straightforward to implement, especially in large community samples (Englert et al., 2010; Newell et al., 1999).

Errors in self-reports of disease status can lead to errors in epidemiological estimates, such as prevalence and relative risks related to exposures that are being

studied, flawed research conclusions and inadequate health care planning (Paganini-Hill & Chao, 1993). There are reports on the predictive value of parental reports of otitis media in infants under the age of 27 months using otoscopy, tympanometry, and audiometry as the "gold standard". These studies focused on parental recognition of otitis media before screening and evaluation, rather than validation of physician identification of otitis media cases (Anteunis et al., 1999; J. Engel et al., 2000; J. A. Engel et al, 1999). The validity of retrospectively reported otitis media, or childhood history of otitis media, has also been addressed in the research literature (Anteunis et al., 1999; J. Engel et al., 2000; J. A. Engel et al., 1999). We set out to investigate how well self- (or parental-) reports of otitis media correspond with physician identification in the medical records. To our knowledge, a prospective study evaluating self-reported (or parent-reported) otitis media as a proxy for physician-recorded diagnosis has not been reported in the literature.

Methods

Study design and population

The present study is a cross-sectional analysis of data collected for a cluster randomized trial of vaccinating children in Hutterite communities against influenza. Hutterites are a communal religious group who live in self-governing, mostly thriving, technologically advanced, farming colonies and seek to actively detach themselves from the impact of the outside world. Participants from 46 Hutterite colonies participated in the trial; 22 in Alberta, 22 in Saskatchewan, and two colonies in Manitoba. Children, between the ages of 36 months and 15 years, were randomly assigned, according to

colony and in a blinded manner, to receive either a standard dosing of inactivated trivalent influenza vaccine or hepatitis A vaccine. All colony members were then monitored during the influenza season for signs of respiratory-related illness. Details of the Hutterite randomized controlled trial (RCT) are described elsewhere (Loeb et al., 2010).

Self reports of otitis media

Self report data were collected by study diaries (completed by a family representative) and in-person interviews by trained research nurses from December 28, 2008 to June 23, 2009. During this period, RCT participants recorded their influenza-related symptoms using family diaries to record influenza-related signs and symptoms on a daily basis. Participants and mothers (of infants) were instructed that ear infection was established by consultation with a physician and to be distinguished from the subjective symptom of earache (also on the family diary); that is, ear infection should be reported on the day that it was diagnosed by a health care provider at a medical consultation. Research nurses visited the Hutterite colonies twice per week to check diary entries and interviewed individual participants regarding outpatient health care visits made for flu-like symptoms, including physician name or health care facility, location (town or city, and address, if possible), and date of medical visits. This surveillance approach ensured a limited time period between medical visits and verification of self-report data, e.g. one to three days, on average; and up to seven days if a participant was away from the colony at the time of the nurse visit and data were obtained at the next visit.

Physician requests for information

Written permission was obtained from study participants and parents to request influenza-related medical record information from health care providers visited during study surveillance. The Canadian Medical Directory (2009 edition) and online physician registries were used to obtain contact information of physicians for whom subjects had provided incomplete addresses. For each reported medical visit, a one-page "Patient Information Request" form was faxed to the medical facility asking for individual patient record data regarding diagnosis (influenza, otitis media or other respiratory illness (Table 1). The procedure was approved by the institutional review boards at McMaster University, the University of Calgary, the University of Saskatchewan, and the University of Manitoba.

Faxed requests for information were sent to medical facilities. Reminders were faxed after one month if a response had not been received. A response indicating that there was "no visit" was followed up by (at least one) fax to an alternative medical facility, based on feedback from the original responder or geography. Participation by health care provider or medical institution was voluntary. Physicians were blind to patient's self reported data. In cases where a copy of the patient record, rather than the completed form, was faxed back, one investigator (AB) transferred the medical record information to the study form. All faxes were sent between March and September 2009.

Statistical analyses

Descriptive statistics were used to examine the study sample demographics. Prevalence and individual two-by-two contingency tables were calculated for otitis media cases. For the primary analysis the medical report was considered the criterion standard,

since participants were asked to report "physician-identified" ear infection. Validity of self-report in comparison to medical record documentation was assessed by calculating the following estimates: sensitivity (correctly reported positive participant reports / all positive medical records); specificity (correctly reported negative participant reports / all negative medical records); positive predictive value (correctly reported positive participant reports / all positive self-reports); negative predictive value (correctly reported negative participant reports / all negative self-reports); likelihood ratio for a positive test (sensitivity / [1 – specificity]); and likelihood ratio for a negative test ([1 – sensitivity] / specificity). Higher specificity and fewer false positive reports can lead to a higher likelihood ratio for a positive test and lower likelihood ratio for a negative test, both of which indicate better precision of reporting (Haynes et al., 2006).

Total agreement (number of concordant pairs / total sample) and kappa coefficient (and standard deviation) were computed. Kappa measures the strength of agreement beyond that expected solely by chance ([observed agreement – chance agreement] / [1 – chance agreement]), where 0 = chance agreement and 1 = perfect agreement (Cohen, 1960). To test for differences in mean number of reports per source, we used the paired Student *t* test.

Statistics were also calculated in four strata defined by sex, age group, level of risk for influenza complications, and number of sick days. The association between the stratification variables and agreement was further investigated using logistic regression analysis. The dependent variable was agreement, coded as 1 for agreement (if the participant and medical report both reported the presence of otitis media or both reported

the absence of otitis media) or 0 for disagreement. All analyses were conducted using the Statistical Package of the Social Sciences (SPSS) version 16.0 for Windows (SPSS Inc., Chicago, IL). Significance levels were set at $p < 0.05$.

Results

Characteristics of the participant sample

Of the 3,273 participants in the trial, 252 (8%) reported at least one outpatient medical visit during the study influenza season. Six participants were unable to provide sufficient identifying information for the doctor or medical facility to be contacted. The first medical visit reported by participants and confirmed by medical record information Therefore, 176 unique medical visits (70%) were included in the sample.

The mean age of the 176 participants was 24 years; more than a third of the sample (36%) was under the age of seven years. Just over a quarter of the sample (26%) were between the ages of 23 and 49 years. Over half of the sample (56%) resided in Saskatchewan, 36% in Alberta and 8% in Manitoba. Sixty three percent were female and 39% were at high risk for influenza. Medical visits were made between January and June 2009.

Characteristics of the medical facilities

Three hundred and six initial faxes were sent to the physician offices or medical facilities; 131 fax reminders and 34 additional follow-up requests were also sent. A small number (6%), particularly from hospital emergency departments, opted to fax back a copy of the patient record for the specified visit.

The 176 participants primarily visited family practice offices (80%), while almost a fifth of the sample (17%) accessed a hospital emergency department. Eighty nine individual physicians were visited at 42 medical centres, 13 hospitals and two walk-in clinics. Almost half of the sample (48%) visited a medical facility in an urban centre, defined as an area that has more than 400 people per square kilometre and more than 1,000 people residing there (Statistics Canada 2002, 2005). Fifty-six percent sought medical care in their home province of Saskatchewan and only 8% lived and accessed medical care in Manitoba. In total, participants visited 32 towns, cities or villages across the three provinces for medical care.

Prevalence of otitis by data source

The prevalence of physician-identified otitis media (22%) was underestimated by participant self-report (16%) by 5%. This difference did not reach significance (95% CI = 4.0 to 11.8, $p = 0.068$ (Table 2).

Otitis media is a common illness in young children (Vergison et al., 2010). Of the 38 cases of otitis media documented in the medical record, 28 (74%) were six years old or younger. Of the 63 (36%) children under the age of seven in the sample, 28 (44%) had a classification of otitis media in the medical record; 16 (25%) were 24 months or younger.

Otitis media is also a common infection for which antibiotics are prescribed, especially in children (Autret-Leca et al., 2002; Nyquist et al., 1998). Physicians reported that 90% (34 out of 38) of otitis media cases were prescribed antibiotics. According to the research participants, 96% (27 out of 28) of cases were given a prescription.

Assessment of self-reported otitis

Self report of otitis media was a very specific measure (93%), but had lower sensitivity (47%) (Table 2). The high specificity indicates the participant's very good ability to accurately report not having otitis media. However, the sensitivity means that participant self report failed to identify more than half of otitis media cases. The probability of otitis media in a participant who reported otitis media, or positive predictive value, was moderate with an estimate of 64%. That is, the medical records confirmed 64% of the self reports of otitis media. The probability of not having otitis media in a participant who did not report the diagnosis, or negative predictive value, was good at 86%. The likelihood ratio of having otitis media was 6.71 and the likelihood ratio of not having the otitis media was 0.57, indicating moderate exactness with the medical record. The kappa value was 0.44, which according to Landis and Koch (1977) indicates moderate agreement (Landis & Koch, 1977). Prerequisites for high kappa are good agreement and a fairly even distribution between positive versus negative responses. That is, the kappa coefficient is sensitive to both prevalence and bias (Feinstein & Cicchetti, 1990; Sim & Wright, 2005).

Indices for self-reported otitis media were not affected by age, sex, number of sick days or influenza risk status. The results of logistic regression analyses showed that none of the variables examined were significantly associated with agreement.

Discussion

Research studies commonly have access to only one source of data and may not verify self-reports. However, the two concurrent data sources in this study provided

insights into the congruence between methods. Data collected from research participants were prospective. Unlike research using retrospective questionnaires, participants reported this information "in real time," thereby, avoiding recall bias. In many studies, participants are asked if they have a medical problem, but not whether it has been identified by a health care professional (Okura et al., 2004). In this study, participants were asked about physician identification of otitis media immediately following the medical visit.

We found estimates indicating that participants were quite good at identifying they did *not* have otitis media, but poor at identifying the actual diagnosis. Specificity of self-report remained high across all stratified variables (>=89%). The overall sensitivity was modest (47%). In the absence of confirmation from clinical records, it is important to know the positive predictive value, which was a moderate 64%.

Low sensitivity is a bigger threat to study validity than low specificity (Strom, 2000). The low sensitivity found in this study may have resulted from limitations at each source. Medical records are not the perfect criterion standard for the presence of otitis media. Self-reported information can be imprecise for various reasons, including underreporting, lack of motivation to report accurately, and poor compliance. It is possible self-report data may be systematically biased. For example, it may be that the Hutterites, because they have limited formal education and low health care literacy were more likely to make errors in reporting the details of their outpatient medical visits. Another explanation is that clinicians provided insufficient information about otitis media or communicated ineffectively so that patients or their parents misunderstood or quickly

forgot the diagnosis (Westbrook et al., 1998). However, it is unknown exactly what physicians communicated to patients.

Clinicians were not asked to provide information about how the diagnosis was made. According to the American Academy of Pediatrics, otitis media is confirmed if all of the following three criteria are present: 1) recent or abrupt onset of symptoms, 2) the presence of middle ear effusion (defined by one of the following: bulging of the tympanic membrane, limited or absent mobility of the tympanic membrane, air fluid level behind the tympanic membrane, otorrhea), 3) evidence of middle ear inflammation (either distinct erythema of the tympanic membrane or distinct otalgia) (American Academy of Pediatrics, Subcommittee on Management of Acute Otitis Media, 2004). We cannot determine whether physicians followed these or other criteria to diagnose otitis media. Given this possible limitation, documentation in the medical record was used as a proximate measure of the gold standard. Severity and recurrence may also have impacted agreement between sources, but this was not assessed. Therefore, reproducibility, rather than accuracy, of self-report, was examined.

To compare data sources, we limited the analyses to participants whose physician or hospital had provided medical record information, i.e. participants who had data from both sources. Thirty percent of the fax requests were not completed; responding physicians or medical facilities may differ from non-responders, resulting in self-selection bias. Two medical offices in Saskatchewan declined to participate in the study; 26 (10%) participants visited one of two offices. It is possible that agreement between sources would be different for these participants with missing medical record data.

Research on health services utilization by Hutterite colony members is lacking in the medical literature. This study contributes to the knowledge of this understudied group. Despite the inclusion of medical records abstracted from diverse medical facilities, generalizability is limited by utilizing a homogeneous cohort. Agreement with the medical record may vary for other outpatient populations.

These findings suggest that the prevalence of otitis media based on reports from study participants may not be entirely accurate and may result in a number of false negatives without supplemental data collected from medical records. Reliance on self-report without verification by medical record may lead to errors in determination of otitis media rates and may affect enrollment eligibility or outcome analyses in research studies. The decision regarding which source of data to use will depend on the outcome of interest; whether finding are used for clinical decision making, population surveillance, outcome studies or other research purposes; availability of resources; and whether a false positive or false negative is of more concern (Ferrante et al., 2008).

References

American Academy of Pediatrics, Subcommittee on Management of Acute Otitis Media. (2004). Diagnosis and management of acute otitis media. *Pediatrics, 113*(5), 1451–1465.

Anteunis, L. J., Engel, J. A., Hendriks, J. J., & Manni, J. J. (1999). A longitudinal study of the validity of parental reporting in the detection of otitis media and related hearing impairment in infancy. *Audiology: Official Organ of the International Society of Audiology, 38*(2), 75-82.

Autret-Leca, E., Giraudeau, B., Ployet, M. J., & Jonville-Bera, A. P. (2002). Amoxicillin/clavulanic acid is ineffective at preventing otitis media in children with presumed viral upper respiratory infection: A randomized, double-blind equivalence, placebo-controlled trial. *British Journal of Clinical Pharmacology, 54*(6), 652-656.

Charles, J., Pan, Y., & Britt, H. (2004). Trends in childhood illness and treatment in Australian general practice, 1971-2001. *The Medical Journal of Australia, 180*(5), 216-219.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational Psychological Measures, 20*, 37-46.

Engel, J., Anteunis, L., Volovics, A., Hendriks, J., & Marres, E. (2000). Predictive value of parent-reported symptoms in the assessment of otitis media with effusion during infancy. *Scandinavian Journal of Primary Health Care, 18*(1), 25-29.

Engel, J. A., Anteunis, L. J., Volovics, A., Hendriks, J. J., & Manni, J. J. (1999). Chronic otitis media with effusion during infancy, have parent-reported symptoms prognostic

value? A prospective longitudinal study from 0 to 2 years of age. *Clinical Otolaryngology and Allied Sciences, 24*(5), 417-423.

Englert, H., Muller-Nordhorn, J., Seewald, S., Sonntag, F., Voller, H., Meyer-Sabellek, W., Willich, S. N. (2010). Is patient self-report an adequate tool for monitoring cardiovascular conditions in patients with hypercholesterolemia? *Journal of Public Health (Oxford, England),* doi:10.1093/pubmed/fdq013

Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low kappa: I. the problems of two paradoxes. *Journal of Clinical Epidemiology, 43*(6), 543-549.

Ferrante, J. M., Ohman-Strickland, P., Hahn, K. A., Hudson, S. V., Shaw, E. K., Crosson, J. C., & Crabtree, B. F. (2008). Self-report versus medical records for assessing cancer-preventive services delivery. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology, 17*(11), 2987-2994.

Fukuoka, Y., Dracup, K., Ohno, M., Kobayashi, F., & Hirayama, H. (2005). Symptom severity as a predictor of reported differences of prehospital delay between medical records and structured interviews among patients with AMI. *European Journal of Cardiovascular Nursing: Journal of the Working Group on Cardiovascular Nursing of the European Society of Cardiology, 4*(2), 171-176.

Goldman, N., Lin, I. F., Weinstein, M., & Lin, Y. H. (2003). Evaluating the quality of self-reports of hypertension and diabetes. *Journal of Clinical Epidemiology, 56*(2), 148-154.

Haynes, R. B., Sackett, D. L., Guyatt, G. H., Tugwell, P., & Ovid Technologies, I. (2006). *Clinical epidemiology: How to do clinical practice research* (3rd ed.). Philadelphia: Lippincott Williams & Wilkins.

Heikkinen, T., & Chonmaitree, T. (2003). Importance of respiratory viruses in acute otitis media. *Clinical Microbiology Reviews, 16*(2), 230-241.

Heikkinen, T., Silvennoinen, H., Peltola, V., Ziegler, T., Vainionpaa, R., Vuorinen, T., Juven, T. (2004). Burden of influenza in children in the community. *The Journal of Infectious Diseases, 190*(8), 1369-1373.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174.

Loeb, M., Russell, M. L., Moss, L., Fonseca, K., Fox, J., Earn, D. J., Walter, S. D. (2010). Effect of influenza vaccination of children on infection rates in Hutterite communities: A randomized trial. *JAMA: The Journal of the American Medical Association, 303*(10), 943-950.

Monto, A. S., Gravenstein, S., Elliott, M., Colopy, M., & Schweinle, J. (2000). Clinical signs and symptoms predicting influenza infection. *Archives of Internal Medicine, 160*(21), 3243-3247.

Newell, S. A., Girgis, A., Sanson-Fisher, R. W., & Savolainen, N. J. (1999). The accuracy of self-reported health behaviors and risk factors relating to cancer and cardiovascular disease in the general population: A critical review. *American Journal of Preventive Medicine, 17*(3), 211-229.

Nyquist, A. C., Gonzales, R., Steiner, J. F., & Sande, M. A. (1998). Antibiotic prescribing for children with colds, upper respiratory tract infections, and bronchitis. *JAMA: The Journal of the American Medical Association, 279*(11), 875-877.

Okura, Y., Urban, L. H., Mahoney, D. W., Jacobsen, S. J., & Rodeheffer, R. J. (2004). Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. *Journal of Clinical Epidemiology, 57*(10), 1096-1103.

Paganini-Hill, A., & Chao, A. (1993). Accuracy of recall of hip fracture, heart attack, and cancer: A comparison of postal survey data and medical records. *American Journal of Epidemiology, 138*(2), 101-106.

Phillips, K. A., Milne, R. L., Buys, S., Friedlander, M. L., Ward, J. H., McCredie, M. R., Hopper, J. L. (2005). Agreement between self-reported breast cancer treatment and medical records in a population-based breast cancer family registry. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology, 23*(21), 4679-4686.

Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy, 85*(3), 257-268.

Statistics Canada 2002. (2005). *2001 community profiles.* Catalogue no. 93F0053XIE. Ottawa: Statistics Canada.

Strauss, S. M., Rindskopf, D. M., Deren, S., & Falkin, G. P. (2001). Concurrence of drug users' self-report of current HIV status and serotest results. *Journal of Acquired Immune Deficiency Syndromes (1999), 27*(3), 301-307.

Strom, B. L. (2000). *Pharmacoepidemiology* (3rd ed.). Chichester; New York: Wiley.

Vergison, A., Dagan, R., Arguedas, A., Bonhoeffer, J., Cohen, R., Hooge, I., Pelton, S. I. (2010). Otitis media and its consequences: Beyond the earache. *The Lancet Infectious Diseases, 10*(3), 195-203.

Westbrook, J. I., McIntosh, J. H., Rushworth, R. L., Berry, G., & Duggan, J. M. (1998). Agreement between medical record data and patients' accounts of their medical history and treatment for dyspepsia. *Journal of Clinical Epidemiology, 51*(3), 237-244.

Table 1: Request for Medical Record Information Form

PATIENT INFORMATION REQUEST

The Hutterite Influenza Study is being conducted by researchers from McMaster University to better understand whether immunizing school-age children against influenza can protect high-risk members of their community.

Your patient, identified on the attached consent form, has agreed to participate in this study and has given us consent to contact you about his/her recent visit on _____ to you for treatment of respiratory infection symptoms.

1. What was actual date of the patient's visit?		
2. What were the patient's symptoms? *Check all that apply.*	○ Fever (≥38° C) ○ Cough ○ Runny nose ○ Sore throat ○ Headache	○ Sinus problems ○ Muscle aches ○ Fatigue ○ Ear ache ○ Chills ○ Other, specify: _____
3. What was the diagnosis?	○ Pneumonia ○ Otitis media ○ Other, specify: _____	

Table 2: Prevalence by source and validity estimates of self-reported otitis media

	Value
Prevalence – Medical record, n (%)	38 (22)
Prevalence – Participant report, n (%)	28 (16)
Total Agreement	0.83
Sensitivity	0.47
Specificity	0.93
Positive Predictive Value	0.64
Negative Predictive Value	0.86
Positive Likelihood Ratio	6.71
Negative Likelihood Ratio	0.57
Kappa (SD)	0.44 (0.08)

*SD = standard deviation

CHAPTER 4

REPORTING OF ANTIBIOTIC PRESCRIPTIONS

This manuscript has been submitted for publication to *Family Practice*:

> **Barbara AM,** Loeb M, Dolovich L, Brazil K, Russell MK. Accuracy of self-reported antibiotic prescription in a Hutterite sample.

Accuracy of self-reported antibiotic prescription in a Hutterite sample

Angela M. Barbara PhD(C)[1], Mark Loeb MD[2], Lisa Dolovich PharmD[3], Kevin Brazil PhD[4], Margaret L. Russell MD, PhD[5]

[1]Department of Pathology and Molecular Medicine, McMaster University

[2]Department of Pathology and Molecular Medicine, Department of Medicine, Department of Clinical Epidemiology and Biostatistics, McMaster University

[3]Department of Family Medicine, Center for Evaluation of Medicines, McMaster University

[4]St. Joseph's Health System Research Network; Department of Clinical Epidemiology and Biostatistics, McMaster University

[5] Department of Community Health Sciences, University of Calgary

Word count: 2,031

Tables: 4

References: 32

Abstract

Background. Antibiotics are extensively prescribed for respiratory infections. In family practice research, rates of prescribing are traditionally measured by physician survey or review of medical charts. These types of data collection are costly and labour intensive. In contrast, self-report is relatively cost efficient and organizationally straightforward to implement.

Objective. To compare antibiotic prescription based upon medical records to self reports of having been prescribed an antibiotic.

Methods. Data regarding antibiotic prescriptions for influenza or respiratory illness at outpatient consultations were collected prospectively from participants in the Hutterite Influenza Prevention Study using nurse-administered questionnaires. Similar data were also collected by fax requests for medical record information to the clinicians. We calculated prevalence (by each data source), sensitivity, specificity, predictive values, likelihood ratios and kappa.

Results. Compared to the medical records, the prevalence of antibiotic prescriptions was overestimated by self-report (73% versus 86%). The most frequently reported antibiotics by both self-report and medical records were amoxicillin, azithromycin, and clarithromycin. Compared to the medical records, self-reported antibiotic prescription was a highly sensitive measure (98%), but had low specificity (50%). Positive predictive value was high at 91% but negative predictive value was modest at 65%. The positive likelihood ratio was 1.96, while the negative likelihood ratio was 0.04.

Conclusions. Self-report of antibiotic prescription had high specificity and good positive predictive value. Assessment of the quality and reliability of self-reported antibiotic prescription in other patient populations is recommended.

Keywords. Antibiotics, self-report, medical records, prescribing, respiratory illness

Introduction

Antibiotics are prescribed extensively for respiratory infections. In family medicine research, rates of prescribing are traditionally measured by physician survey, dispensary records, or review of medical charts.[1-3] These types of data collection can be costly, labour-intensive and time-consuming,[4] especially in large multi-center studies, where research participants access a variety of medical services across large geographic areas. Information on antibiotic prescriptions can also be obtained directly from research participants by self-administered surveys or interview.[5] While self-report has potential limitations related to accuracy, recall and willingness to report,[6] it can be relatively cost efficient and organizationally straightforward to implement.[7, 8]

If concordance between data sources is not high, research findings may differ substantially depending on the method of data collection used.[9] The accuracy of patients' self-reports of prescription medication use and adherence has been reported in the research literature.[10-15] Past studies have focused on medications for chronic conditions, rather than short–term courses of antibiotics for episodes of respiratory infection.[16] One cross-sectional telephone survey of United States veterans found that self-reported antibiotic exposure within the previous six months had a 53% sensitivity and 88% specificity.[17] The present study is concerned with the validity of self-reported prescription receipt for antibiotics concurrently, rather than actual use or medication history. The objective of the current study was to determine the validity of research participant's self-report of receiving a prescription for antibiotics for respiratory-related illness.

Methods

Study design and population

This is a cross-sectional analysis of data collected for a cluster randomized trial of vaccinating children in Hutterite communities against influenza. The Hutterites are a communal religious group, who live in self-governing, technologically advanced farming colonies and seek to actively detach themselves from the impact of the outside world. Participants from 46 Hutterite communities in the Canadian Prairie Provinces participated in the trial; 22 in Alberta, 22 in Saskatchewan, and two colonies in Manitoba. The design, methods and results of trial have been described elsewhere.[18]

Self reports of antibiotic prescription

Study surveillance took place from December 28, 2008 to June 23, 2009. During this period, trained research nurses visited the Hutterite colonies twice per week. Self-report data were obtained by nurse-administered questionnaires regarding outpatient health care visits made for flu-like symptoms, including names of physician or health care facility, location (town or city, and address, if possible), and date of medical visits. Participants were also asked if antibiotics were prescribed at the medical visits. They were asked, in advance, to have their medication containers available for the interview.

Physician requests for information

The Canadian Medical Directory (2009 edition) and online physician registries were used to obtain contact information of physicians for whom subjects had provided incomplete addresses. For each reported medical visit, a one-page "Patient Information Request" form was faxed to the medical facility asking for individual patient record data

regarding prescribed antibiotics, including name of antibiotic, dose, frequency and duration (Table 1). The institutional review boards at McMaster University, University of Calgary, University of Saskatchewan, and the University of Manitoba approved the study.

Faxes were sent to the physician offices or medical facilities between March and September 2009. The primary analysis was restricted to an individual's first confirmed medical visit to maintain independence of observations. Physicians were blind to patient's self reports of receiving a prescription.

Statistical analyses

Along with descriptive statistics, prevalence and individual two-by-two contingency tables were calculated for antibiotic prescription. The validity of self-report was assessed by calculating the following estimates: sensitivity (correctly reported positive self-reports / all positive medical records); specificity (correctly reported negative self-reports / all negative medical records); positive predictive value (correctly reported positive participant reports / all positive self-reports); negative predictive value (correctly reported negative self-reports / all negative self-reports); likelihood ratio for a positive test (sensitivity / [1 – specificity]); and likelihood ratio for a negative test ([1 – sensitivity] / specificity). Total agreement (number of concordant pairs / total sample) and kappa coefficient (and standard deviation) were also computed.

Statistics were also calculated in three strata defined by sex, age group, and level of risk for influenza complications. The association between the stratification variables and agreement was further investigated using univariable logistic regression analysis.

The dependent variable was agreement, coded as 1 for agreement (if the self- and medical report both reported the presence of antibiotic prescription or both reported the absence of antibiotic prescription) or 0 for disagreement. All analyses were conducted using SPSS 16.0 (SPSS Inc., Chicago, IL). The criterion for statistical significance was set at alpha = 0.05.

Results

Characteristics of the participant sample

Of the 3,273 participants in the trial, 252 (8%) reported at least one outpatient medical visit during the study influenza season. Sufficient information to contact physicians or medical facilities was available for 246 persons. The first medical visit reported by participants was included in the primary analysis. A total of 176 initial medical visits (70%) were confirmed by medical record information; this is the sample include in the analyses.

The mean age of the sample was 24 years; more than a third (36%) was under the age of seven years. Just over a quarter of the sample (26%) were between the ages of 23 and 49 years. Over half of the sample (56%) resided in Saskatchewan, 36% in Alberta and 8% in Manitoba. Sixty three percent were female and 39% were at "high risk" for influenza complications. Medical visits were made between January and June 2009. The majority of the sample (92%) had a diagnosis of respiratory-related illness in the medical charts. The most common diagnoses were otitis media (24%); upper respiratory tract infection (17%); sinusitis, pharyngitis and bronchitis (12% each), tonsillitis (7%), strep throat and pneumonia (4% each).

Characteristics of sources of care

The sample primarily visited family practice offices (80%), but almost a fifth of the sample (17%) accessed a hospital emergency department; 3% went to a walk-in clinic. Eighty nine individual physicians were visited at 42 medical practices, 13 hospitals and two walk-in clinics in 32 towns, cities or villages across the three provinces. While most respondents filled in the faxed form, a small number (6%), particularly from emergency departments opted to fax back a copy of the medical record for the specified visit. In these cases, the information on the medical record was transferred to the study form by one author (AB).

Prevalence of antibiotic prescription by data source

Compared to medical records (73%), the prevalence of antibiotic prescriptions was significantly overestimated by participants (86%) by 13% (95% CI 7.8 to 18.3, $p < 0.0001$) (Table 2). Of the 162 participants diagnosed with respiratory illness, 90% self-reported receiving an antibiotic prescription, but this was documented in medical records for only 77%.

Accuracy of self-reported antibiotic prescription

Self-report of antibiotic prescription was a highly sensitive measure (98%), but had a high rate of false positives and low specificity (50%) (Table 2). Positive predictive value was high at 91% but negative predictive value was modest at 65%. However, the positive likelihood ratio of 1.96 indicates low accuracy of self-report, while the negative positive likelihood ratio of 0.04 indicates higher accuracy (Table 2). Of the 151 participants who self-reported having been prescribed antibiotics, 142 (94%) provided the

name of the medication and 96 (64%) showed the research nurse the medication container. Moreover, 138 participants (91%) correctly named the prescribed medication and 100 (66%) provided the correct prescription details (dose, frequency and duration). Of the 96 participants with a medication bottle, 83 (86%) had medical documentation of the prescription.

The most frequently reported antibiotic by both sources were amoxicillin, azithromycin, and clarithromycin (Table 3). Sensitivity and specificity for amoxicillin were both high (87%, 98%). Azithromycin had perfect sensitivity (100%) and very high specificity (97%). Clarithromycin had good sensitivity (83%) and very high specificity (99%).

Characteristics associated with accuracy of self-report

Stratified indices of validity for self-reported medication prescriptions are presented in Table 4. Medication sensitivity was high for all strata. Slightly higher medication specificity was associated with age. The results of logistic regression analyses showed that none of the variables examined were significant predictors of agreement.

Discussion

One strength of the study was the prospective data collection. Unlike most retrospective questionnaire research, participants were asked about prescribed medications immediately following the medical visit in near real time, thereby, avoiding memory recall bias.

Research studies commonly have access to only one source of data and are unable to verify self-report responses. The interpretation of such studies is aided by having insight into the potential validity of self-reports, as provided by the present research. The sensitivity of prescription receipt was high, regardless of the demographic characteristics. High sensitivity will result in the identification of more cases of antibiotic prescription, but at the expense of increasing the numbers of false "positives". As pointed out by Zimmerman and colleagues, high sensitivity is not surprising from a behavioural perspective. It is easier to report an action that did happen (getting a prescription) rather than something that did not happen (not getting a prescription).[19] The positive predictive value was good at 82%. According to the stratification and logistic regression analyses, the demographic and variables did not appreciably impact self-report validity.

In many pharmacy studies, prescription records at the pharmacy level are typically seen as the gold standard due to their high accuracy are often used to evaluate the completeness of medical charts.[20, 21] However, the use of pharmacy records has the limitation of an underlying assumption that people actually always fill their prescriptions. Because the focus of our study was prescription receipt, we used the medical records as the gold standard.

In our study, the observed variability of prevalence by data source may have resulted from specific factors related to each source. Self-reported information can be imprecise for various reasons. Many patients expect an antibiotic when they visit their physician for respiratory infection[22, 23] and this expectation may impact self-reporting. Likewise, medical records are not necessarily accurate sources of information. Several

studies have found non-reporting and misreporting in medical records.[24, 25] Busier physicians may document less in the medical record or delay recording, leading to errors in recall and recording.[26] This may explain why 14% of the participants who were able to verify their antibiotic prescription with the actual medication bottle did not have the appropriate documentation in their medical records. The process of abstracting information from the medical chart itself is also subject to imprecision.[27] Furthermore, medical records accessed for this analysis were not written or kept for the purposes of this study; thus, it is difficult to assess the level to which medical records were complete.[28] In contrast, we used a comprehensive approach to data collection, including open-ended questions and condition-specific prompts, which has been shown to increase the ability of subjects to self-report antibiotic exposures.[17] research nurses were trained to collect and document information consistent manner from all

The study population included only Hutterite outpatients participating in the randomized trial; therefore generalizability of the findings might be difficult. The antibiotic prescription rate for Hutterite patients with respiratory infection was 77% (according to the medical records). This is comparable to findings in other populations, with varying rates for specific condition: for example, 46% to 79% for lower respiratory tract infection, 68% for upper respiratory tract infections, 69% to 80% for sinusitis, 75% to 78% for bronchitis, 81% for tonsillitis, 56% to 80% for otitis media, and 54% for pharyngitis.[2, 3, 29-32]

Another limitation of the present research is that thirty percent of the fax requests for information were not completed; responding physicians or medical facilities may

differ from non-responders in their data recording or antibiotic prescription behaviours. Two medical offices in Saskatchewan declined to participate in the study. Twenty-six (10%) participants visited one of the two offices. Thus, potential bias must be considered when interpreting the findings. A third limitation is the relatively low sample size which may be the reason for not having found any factors associated with better reporting, in particular in the regression analysis.

In summary, we found good agreement between self-report by the Hutterite sample and their medical records. Self-report of antibiotic prescriptions was a highly sensitive measure and may be a useful source of data for family practice research. However, we strongly recommend further research in other patient populations and using a larger sample size before applying these findings into practice.

References

1. Gonzales R, Steiner JF, Sande MA. Antibiotic prescribing for adults with colds, upper respiratory tract infections, and bronchitis by ambulatory care physicians. *JAMA*. 1997; **278**: 901-4.

2. Nyquist AC, Gonzales R, Steiner JF, Sande MA. Antibiotic prescribing for children with colds, upper respiratory tract infections, and bronchitis. *JAMA*. 1998; **279**: 875-7.

3. Steinman MA, Landefeld CS, Gonzales R. Predictors of broad-spectrum antibiotic prescribing for acute respiratory tract infections in adult primary care. *JAMA*. 2003;**289**: 719-25.

4. Phillips KA, Milne RL, Buys S, et al. Agreement between self-reported breast cancer treatment and medical records in a population-based breast cancer family registry. *J Clin Oncol*. 2005;**23**: 4679-86.

5. Okura Y, Urban LH, Mahoney DW, Jacobsen SJ, Rodeheffer RJ. Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. *J Clin Epidemiol*. 2004; **57**: 1096-103.

6. Goldman N, Lin IF, Weinstein M, Lin YH. Evaluating the quality of self-reports of hypertension and diabetes. *J Clin Epidemiol*. 2003;**56**: 148-54.

7. Englert H, Muller-Nordhorn J, Seewald S, et al. Is patient self-report an adequate tool for monitoring cardiovascular conditions in patients with hypercholesterolemia? *J Public Health*. 2010; **32**: 387-94.

8. Newell SA, Girgis A, Sanson-Fisher RW, Savolainen NJ. The accuracy of self-reported health behaviors and risk factors relating to cancer and cardiovascular disease in the general population: A critical review. *Am J Prev Med*. 1999; **17**:211-29.

9. Zhu K, McKnight B, Stergachis A, Daling JR, Levine RS. Comparison of self-report data and medical records data: Results from a case-control study on prostate cancer. *Int J Epidemiol*. 1999; **28**: 409-17.

10. Abdolrasulnia M, Weichold N, Shewchuk R, et al. Agreement between medical record documentation and patient-reported use of nonsteroidal antiinflammatory drugs. *Am J Health Syst Pharm*. 2006; **63**: 744-7.

11. Solomon DH, Stedman M, Licari A, Weinblatt ME, Maher N, Shadick N. Agreement between patient report and medical record review for medications used for rheumatoid arthritis: The accuracy of self-reported medication information in patient registries. *Arthritis Rheum*. 2007; **57**: 234-9.

12. Garber MC, Nau DP, Erickson SR, Aikens JE, Lawrence JB. The concordance of self-report with other measures of medication adherence: A summary of the literature. *Med Care*. 2004; **42**: 649-52.

13. Cotterchio M, Kreiger N, Darlington G, Steingart A. Comparison of self-reported and physician-reported antidepressant medication use. *Ann Epidemiol*. 1999; **9**: 283-9.

14. Sjahid SI, van der Linden PD, Stricker BH. Agreement between the pharmacy medication history and patient interview for cardiovascular drugs: The Rotterdam elderly study. *Br J Clin Pharmacol*. 1998; **45**: 591-5.

15. Lamiae GB, Michel R, Elodie A, et al. Agreement between patients' self-report and physicians' prescriptions on cardiovascular drug exposure: The PGRx database experience. *Pharmacoepidemiol Drug Saf*. 2010; **19**: 591-5.

16. Strom BL. *Pharmacoepidemiology.* 3rd ed. Chichester ; New York: Wiley; 2000.

17. Metlay JP, Hardy C, Strom BL. Agreement between patient self-report and a Veterans affairs national pharmacy database for identifying recent exposures to antibiotics. *Pharmacoepidemiol Drug Saf*. 2003; **12**: 9-15.

18. Loeb M, Russell ML, Moss L, et al. Effect of influenza vaccination of children on infection rates in Hutterite communities: A randomized trial. *JAMA*. 2010; **303**: 943-50.

19. Zimmerman RK, Raymund M, Janosky JE, Nowalk MP, Fine MJ. Sensitivity and specificity of patient self-report of influenza and pneumococcal polysaccharide vaccinations among elderly outpatients in diverse patient care strata. *Vaccine*. 2003; **21**: 1486-91.

20. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol*. 2005; **58**: 323-37.

21. West SL, Strom BL, Freundlich B, Normand E, Koch G, Savitz DA. Completeness of prescription recording in outpatient medical records from a health maintenance organization. *J Clin Epidemiol*. 1994; **47**: 165-71.

22. Butler CC, Rollnick S, Pill R, Maggs-Rapport F, Stott N. Understanding the culture of prescribing: Qualitative study of general practitioners' and patients' perceptions of antibiotics for sore throats. *BMJ*. 1998; **317**: 637-42.

23. Dosh SA, Hickner JM, Mainous AG,3rd, Ebell MH. Predictors of antibiotic prescribing for nonspecific upper respiratory infections, acute bronchitis, and acute sinusitis. *J Fam Pract*. 2000; **49**: 407-14.

24. Marrie TJ, Durant H, Sealy E. Pneumonia--the quality of medical records data. *Med Care*. 1987; **25**: 20-4.

25. Bush TL, Miller SR, Golden AL, Hale WE. Self-report and medical record report agreement of selected medical conditions in the elderly. *Am J Public Health*. 1989; **79**: 1554-6.

26. Ferrante JM, Ohman-Strickland P, Hahn KA, et al. Self-report versus medical records for assessing cancer-preventive services delivery. *Cancer Epidemiol Biomarkers Prev*. 2008; **17**: 2987-94.

27. Pakhomov SV, Jacobsen SJ, Chute CG, Roger VL. Agreement between patient-reported symptoms and their documentation in the medical record. *Am J Manag Care*. 2008; **14**: 530-9.

28. Skinner KM, Miller DR, Lincoln E, Lee A, Kazis LE. Concordance between respondent self-reports and medical records for chronic conditions: Experience from the Veterans health study. *J Ambul Care Manage*. 2005; **28**: 102-10.

29. Hopstaken RM, Butler CC, Muris JW, et al. Do clinical findings in lower respiratory tract infection help general practitioners prescribe antibiotics appropriately? An observational cohort study in general practice. *Fam Pract*. 2006; **23**: 180-7.

30. Akkerman AE, Kuyvenhoven MM, van der Wouden JC, Verheij TJ. Determinants of antibiotic overprescribing in respiratory tract infections in general practice. *J Antimicrob Chemother*. 2005; **56**: 930-6.

31. Teng CL, Achike FI, Phua KL, et al. General and URTI-specific antibiotic prescription rates in a Malaysian primary care setting. *Int J Antimicrob Agents*. 2004; **24**: 496-501.

32. Grijalva CG, Nuorti JP, Griffin MR. Antibiotic prescription rates for acute respiratory tract infections in US ambulatory settings. *JAMA*. 2009; **302**: 758-66.

Table 1: Patient Information Request

The Hutterite Influenza Study is being conducted by researchers from McMaster University to better understand whether immunizing school-age children against influenza can protect high-risk members of their community.

Your patient, identified on the attached consent form, has agreed to participate in this study and has given us consent to contact you about his/her recent visit on _____ to you for treatment of respiratory infection symptoms.

What was actual date of the patient's visit?		
What was the diagnosis?	○ Pneumonia ○ Otitis media ○ Other, specify: _____	
Was an x-ray ordered?	○ No ○ Yes	*If yes,* was there opacity on chest x-ray compatible with pneumonia? ○ No ○ Yes
Were antibiotics prescribed?	○ No ○ Yes	*If yes*, record Name: Dose: Frequency: Duration:

Table 2: Prevalence according to medical record and self-report, total agreement, sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, negative likelihood ratio, and kappa statistic for antibiotic prescription receipt

	Value
Prevalence – Medical record, n (%)	128 (73)
Prevalence – Participant report, n (%)	151 (86)
Total Agreement	0.85
Sensitivity	0.98
Specificity	0.50
Positive Predictive Value	0.91
Negative Predictive Value	0.61
Positive Likelihood Ratio	1.96
Negative Likelihood Ratio	0.04
Kappa (SD)*	0.57 (0.07)

*standard deviation

Table 3: Prevalence of antibiotic prescriptions by class, generic name and brand name according to medical record and self report

Antibiotic – Class	Generic Name	Brand Name	Medical record	Self report
All, n (% total sample)			128 (73)	151 (86)
Did not report antibiotic name			1	9
Cephalosporins				
	Cefalexin	Keflex	7 (5.5)*	6 (4.0)**
	Cefprozil	Cefzil	4 (3.1)	4 (2.6)
	Cefuroxime	Ceftin	3 (2.3)	4 (2.6)
	Ceftriaxone	Rocephin	1 (0.8)	1 (0.7)
Macrolides				
	Azithromycin	Zithromax	10 (7.8)	15 (10.0)
	Clarithromycin	Biaxin	13 (10.2)	14 (9.3)
	Erythromycin	Erythrocin	2 (1.6)	1 (0.7)
Penicillins				
	Amoxicillin	Amoxil	62 (48.4)	62 (41)
	Amoxicillin/ clavunate	Clavulin	7 (5.5)	6 (4.0)
	Cloxacillin	Tegopen	1 (0.8)	1 (0.7
	Penicillin V	Pen-Vee-K	4 (3.1)	6 (4.0)
Quinolones				
	Ciprofloxacin	Cipro	2 (1.6)	1 (0.7
	Levofloxacin	Levaquin	2 (1.6)	2 (1.3)
	Moxifloxacin	Avelox	1 (0.8)	1 (0.7
Sulfonamides				
	Co-trimoxazole	Bactrim	1 (0.8)	3 (2.0)
Tetracyclines				
	Doxycycline		4 (3.1)	4 (2.6)
Antiemetics				
	Metoclopramide	Maxeran	1 (0.8)	1 (0.7)
Other prescription			1 (0.8)[a]	7 (4.6)[b]
Over the counter product				3 (2.0)[c]

***** Percentage of reports in the medical record

** Percentage of self-reports

[a] Avapro

[b] Dexamethasone, Flovent, Hydrocod Phenyltolox, Nasonex, Salbutamol (n=2)

[c] Advil, Aerius (n=2)

Table 4: Prevalence by data source, Sensitivity (Se), Specificity (Sp), Total Agreement (TA), Kappa Statistic for medication prescription stratified by participant characteristics and univariable logistic regression analysis of predictors of agreement

Participant Characteristic	Data source		Se	Sp	TA	Kappa	Odds Ratio (CI)	*p*-value
	Medical record (n)	Self report (n)						
All (n= 176)	128	151	0.98	0.50	0.85	0.57		
Female (n= 110)	79	95	0.99	0.45	0.84	0.52	0.61 (0.24-1.54)	0.29
Male (n = 66)*	49	56	0.98	0.59	0.88	0.64		
Age group								
0-6 years (n = 63)	45	54	1.00	0.50	0.86	0.58	1.04 (0.29-3.73)	0.95
7-22 years (n = 40)	28	32	0.93	0.58	0.83	0.55	1.08 (0.26-4.47)	0.92
23-49 years (n = 46)	33	39	1.00	0.54	0.87	0.63	1.07 (0.28-4.03)	0.92
50 + years (n = 27)*	22	26	1.00	0.20	0.85	0.29		
High risk** (n = 68)	50	60	1.00	0.44	0.85	0.54	0.94 (0.39-2.22)	0.88
Not high risk (n =108)*	78	91	0.97	0.53	0.85	0.58		

*reference group for logistic regression
**high risk = high risk for influenza complications

CHAPTER 5

DATA SOURCES FOR DETERMINING INFLUENZA-LIKE ILLNESS

This manuscript has been submitted for publication to *The Canadian Journal of Public Health:*

> **Barbara AM,** Loeb M, Dolovich L, Brazil K, Russell MK. A comparison of self-report and medical record data to assess surveillance definitions of influenza-like illness in outpatients.

A Comparison of Self-Report and Medical Record Data to Assess Surveillance Definitions of Influenza-Like Illness in Outpatients

Type of Submission: Quantitative Research

Running Title: Data Sources for Influenza-Like Illness

Word Count: 2,334

References: 24

Tables: 5

Objective: Several surveillance definitions of influenza-like illness (ILI) have been proposed, based on the presence of symptoms. Symptom data can be obtained from patients, medical records, or both. Past research has found that agreements between health record data and self-report are variable depending on the specific symptom. Therefore, we aimed to explore the implications of using data on influenza symptoms extracted from medical records, similar data collected prospectively from outpatients, and the combined data from both sources as predictors of laboratory-confirmed influenza.

Methods: Using data from Hutterite Influenza Prevention Study, we calculated: (1) the sensitivity, specificity and predictive values of individual symptoms within surveillance definitions; (2) how frequently surveillance definitions correlated to laboratory-confirmed influenza; and (3) the predictive value of surveillance definitions.

Results: Of the 176 participants with reports from participants and medical records, 142 (81%) individuals were test for influenza and 37 (26%) were influenza virus positive. Fever (on its own) and combined with cough and/or sore throat were highly correlated with laboratory-confirmed influenza for all data sources. The ILI surveillance definition of fever and sore throat, based on combined symptoms by both medical records and self report, was the best predictor laboratory confirmed influenza with the odds ratio of 9.53 (p=0.0001; 95% CI, 4.01-22.63) and positive predictive value of 61%.

Discussion: The choice of data source to determine ILI will depend on the patient population, outcome of interest, and use for clinical decision making, research, or surveillance.

Key words: Influenza, influenza-like illness, surveillance definition, data source

Word Count: 239

As part of most influenza surveillance systems, patients who meet specific symptom criteria will have culture samples taken for laboratory testing.[1, 2] Several surveillance definitions of influenza-like illness (ILI) have been proposed.[2-6] The Centre for Disease Control and Prevention (CDC) US Sentinel Providers Surveillance Network defines ILI as fever (temperature of 38 degrees Celsius or greater) and a cough and/or sore throat in the absences of a *known* cause other than influenza (www.cdc.gov/flu/weekly/fluactivity.htm). Health Canada's Flu Watch uses a variant of the CDC definition of ILI: fever and cough and with one or more of the following - sore throat, arthralgia, myalgia, or prostration (www.phac-aspc.gc.ca/fluwatch). Several studies have found that the grouping of high fever and cough is the best predictor of influenza.[7-10] What these ILI definitions have in common is the presence of fever plus one or more symptoms of respiratory illness.

Data about influenza symptoms can be obtained from multiple sources. For example, symptoms can be reported by multiple informants, such as self-reports and health care providers; or by multiple methods, such as symptom checklists and medical record data. In prior work, we have found agreements between health record data and self-report varied by symptom [unpublished manuscript]. Therefore, factors that might influence the sensitivity, specificity and predictive values of the ILI include the actual surveillance definitions, but also the data source from which the symptoms data contained in the ILI definition is taken. The impact of these factors will be relevant to both public health researchers and clinicians in determining choice of ILI definitions.

The goal of the current study was assess the utility of two sources of data in determining the surveillance definitions for ILI and their association with laboratory-confirmed influenza. Using data from the Hutterite Influenza Prevention Study,[11] we compared data collected retrospectively from medical record extraction, similar data collected prospectively from research participants, and the combined or "pooled" data from both sources.

Methods

Study design and population

Residents of 46 Hutterite communities in the Canadian prairie provinces participated in a cluster randomized controlled trial to determine if the vaccination of healthy children and adolescents with inactivated influenza vaccine would reduce laboratory-confirmed influenza in other residents of these colonies. Details of the trial are described elsewhere.[11]

Participant reports of influenza-related symptoms

Study surveillance for influenza took place from December 28, 2008 to June 23, 2009. Participants recorded their influenza-related symptoms (fever, cough, runny nose, sore throat, headache, sinus problems, muscle ache, fatigue, ear ache, and chills) using daily diaries. Fever was defined as a temperature $\geq$38 degrees Celsius; each participating family was given a thermometer for this purpose. Trained research nurses visited the Hutterite colonies twice per week to check diary entries and interviewed individual participants (or parents, in the case of infants) to confirm the reported symptoms, assess other symptoms and collect information regarding outpatient visits made to medical

offices and hospital emergency departments for flu-like symptoms, including physician name, health care facility, and location.

Health care provider reports of influenza-related symptoms

For each reported medical visit, a one-page "Patient Information Request" form was faxed to the medical facility asking for individual patient record data regarding presenting symptoms, using the list of symptoms as on the participant study diaries. The standardized form was designed to easily extract information from the patient medical record. Clinicians were blinded to the patient's self reported symptoms. The institutional review boards at McMaster University, University of Calgary, University of Saskatchewan, and the University of Manitoba approved the study. All participants gave written consent to allow us to obtain health record information if they visited a doctor or hospital with flu-related symptoms during the 2008-2009 influenza season.

Faxes were sent to the physician offices or medical facilities between March 2009 and September 2009. A response indicating that there was "no visit" was followed up by (at least one) fax to an alternative medical facility, based on feedback from the original responder or geography. Data from the first medical visit reported by participants and confirmed by the health care provider were included in the analysis.

Laboratory confirmation of influenza

During the colony visits, research nurses took nasopharyngeal swab samples from study participants who reported two or more symptoms, or a physician-diagnosed ear infection. Specimens were submitted to the public health laboratories in the respective provinces to be tested for influenza by Polymerase Chain Reaction (PCR). Influenza was

confirmed by the detection of viral Ribonucleic Acid on the basis of reverse transcriptase Real Time Polymerase Chain Reaction (RT-PCR) targeting matrix gene for influenza A and non-structural gene for influenza B.[11] PCR has been demonstrated to be more sensitive to viral culture alone; compared to direct immunofluorescence and cell culture assay, fluorogenic RT-PCR was 95% sensitive and 100% specific for detecting influenza. It is, therefore, considered the "gold standard" for detecting influenza.[12-15]

Statistical analyses

The frequency of occurrence of individual symptoms was calculated using three strategies: (a) those identified only from medical records; (b) those identified only from research participant reports; (c) symptoms identified from both medical record and participant report. For each data strategy, we tested for differences in mean number of symptom reports between participants with and without PCR-confirmed influenza using the Student *t* test. We then calculated the sensitivity, specificity, positive predictive value and negative predictive value for each of the ten symptoms by individual data source and the combined strategy, using laboratory results as the gold standard for diagnosis of influenza. Univariate logistic regression analysis was used to evaluate the association of each symptom with laboratory confirmed influenza. Odds ratios were calculated to determine the strength of association between symptom and PCR-confirmed influenza; 95% confidence intervals (CI) were calculated to estimate the precision of each odds ratio. Individual symptoms were included in the ILI definitions to be further analysed if the individual symptom was associated with laboratory confirmed influenza (where alpha = 0.05).

We classified each ILI definition according to four strategies: (a) Medical record reports were analysed separately. (b) Research participant reports were analysed separately. (c) Data were combined based on ILI classifications. A participant was considered to have ILI if *either* the medical record *or* research participant reported the combinations of symptoms; e.g. fever or cough ware reported by either source equalled ILI. (d) Data were combined based on joint identification of individual symptoms. A symptom was considered present if it was reported by *either* the research participant *or* recorded in the medical record; e.g., fever reported by either source plus cough reported by either source equalled ILI. A combination of symptoms was only included in the analysis if there was a minimum of ten confirmed influenza cases for each data strategy. [16]

We used the Student *t* test to test for differences in mean number of cases for each ILI definition between participants with and without influenza, for each data strategy. The sensitivity, specificity, predictive values, and odds ratios were calculated for ILI definitions using the four data strategies; laboratory confirmed influenza was considered the gold standard.

All analyses were conducted using SPSS 16.0 (SPSS Inc., Chicago, IL).

Results

Of the 176 participants with both self-report and physician-recorded data, 142 (81%) individuals were tested for influenza and 37 (26%) were influenza virus positive. A description of the sample is displayed in Table 1. Laboratory confirmed cases of influenza were younger (mean = 17 years versus mean = 24 years, p = 0.07). The

influenza A virus was found in 19 of the 37 influenza virus positive (51%) participants and influenza B was found in 18 (49%) participants. The mean time from onset of symptoms to providing a swab sample was 3.5 days for confirmed cases of influenza and 4.9 days for subjects for subjects without influenza (t=2.19, p=0.03). Cases with influenza self-reported being symptomatic for 2.86 days at the time of the medical visit compared to 4.0 sick days for participants without influenza (t=2.07, p=0.04). Confirmed cases of influenza also experienced a higher number of the ten influenza related symptoms, but the difference was only significant according to participant report, not medical record data.

Table 2 compares the prevalence of symptoms by data source, independently and combined, between participants who tested positive and those who tested negative for influenza. Compared to uninfected participants, influenza cases were significantly more likely to have fever (regardless of the data source), participant-reported cough, and participant-reported muscle aches. They were less likely to have earache, although participant report did not reach significance. Influenza cases also had significantly more sore throat, but only when combining the data from both sources. Table 3 presents the sensitivity and logistic regression analyses for individual symptoms. Cough had the highest sensitivity for each data source (76% - 86%). Physician-recorded fever had the highest positive predictive value (56%) and odds ratio (8.9; p=0.0001; 95% CI, 3.81-20.58). Based on these findings, we further analyzed three surveillance definitions for ILI: fever and cough, fever and sore throat, fever and cough or sore throat (CDC definition). Because of the low prevalence among influenza cases, we did not analyze the

symptom combinations of fever and fatigue (physician, n=4; participant, n= 6; combined ILI, n=10; combined symptoms, n=11) or fever and muscle aches (physician, n=3; participant, n= 4; combined ILI, n=7; combined symptoms, n=10).

Table 4 compares the prevalence of each surveillance definition, according to each data strategy, between participants who tested positive and those who tested negative for influenza. Individuals with influenza had significantly more ILI according to each surveillance definition, regardless of data source. Table 5 presents the sensitivity and logistic regression analyses for the three surveillance definitions. Overall, the symptom complex of fever and sore throat, based on combined symptoms by both medical records and self report, was the best predictor laboratory confirmed influenza with the odds ratio of 9.53 (p=0.0001; 95% CI, 4.01-22.63) and positive predictive value of 61%. For each ILI definition, the positive predictive value was higher when based on medical record data. Physician-recorded fever and sore throat had the highest positive predictive value overall (62%). Physician-reported fever plus cough or sore throat, based on combined symptoms, had the highest sensitivity (70%).

Discussion

Most studies evaluating the surveillance definitions of influenza have relied on physician or clinical record data[3, 7, 17, 18] Some also included a patient survey following entry into the study and physician examination or review of medical records.[12, 13] Nicholson and colleagues (1997) had weekly phone surveillance for symptoms and then home visits for systematic patients.[19] Vaccine effectiveness studies have also used clinical data, as well as self-report from research participants.[11, 20, 21]

In this study, we explored the implications of using two different data sources independently and jointly as predictor variables to evaluate surveillance definitions of ILI. Cough alone had the highest sensitivity, regardless of data strategy, (76% - 86%) similar to other studies.[3, 8] We found that positive predictive value for ILI based on physician records was higher than ILI based on participant data for each surveillance definition. This is consistent with previous findings by Govaert and colleagues (1998) that predictive values are higher in subpopulations that consult a general practitioner for influenza symptoms. They found a positive predictive of 30% for fever, cough and acute onset based on questionnaire data compared to 40% for the sample symptom complex according to physician records. Family physicians, having clinical experience with patient consultations for influenza, may be well placed to infer the significance of symptom combinations.[13] Indeed, physicians have been found to correctly diagnose influenza infection in more than 60% to 70% of patients on the basic of clinical symptoms alone.[22]

Laboratory-confirmation of influenza may have been influenced by other factors. For example, the influenza cases visited the doctor and were swabbed sooner after symptom onset compared to the non-influenza participants. In contrast to other studies, [7, 8, 10] we found that the combination of fever and sore throat, rather than fever and cough, had the highest positive predictive value for each data strategy (57% - 62%). Unlike self-reported cough and sore throat, physician-recorded cough and sore throat were *not* more prevalent in influenza-infected subjects compared to non-influenza subjects. Cough and sore throat are non-specific symptoms. Participants were prompted by our research

nurses not to report these symptoms if they were unambiguously unrelated to respiratory illness. Physicians, who were blind to participant responses and interested in participant's overall health, including symptoms unrelated to influenza, would likely record cough and sore throat regardless of etiology.

A recognized limitation of this study is that it was conducted within a specific cultural and religious population of outpatients during a single influenza season. The results may not been generalizable in all patient populations during other influenza seasons. The Hutterite perception that good physical health is a gift from God and ill health is a burden one must bear [23] may lead to less awareness and reluctance to report or complain about bodily symptoms, which may explain the lower prevalence of self-reported ILI and individual symptoms.

Another limitation is that the sample size was modest, resulting in odds ratios with wide confidence intervals. It is probable that a larger sample size would yield similar estimates with narrower confidence intervals. However, because we limited the analyses to participants who had data from both sources, it is also possible that responding physicians or medical facilities had different rates of symptom recording from non-responders.

Researchers and public health clinicians should consider the issue of measurement error and reporting variations when designing studies. Different data sources should correspond with the study question or objective. For example, the data source used will have implications for studies evaluating the effectiveness of influenza vaccination or other interventions. Our findings indicate that using medical record data to determine

ILI, due to its higher positive predictive values, will maximize the effectiveness of an intervention. For overall disease burden and use of health services, pooled data from both health records and participant reports may be more appropriate because of their higher sensitivities.[5, 24] To identify all potential cases of influenza or ILI, either of the combined data strategies will result in a higher rate of detection.

REFERENCES

1. Public Health Agency of Canada. Flu watch. . Accessed 06-16, 2010.

2. Thursky K, Cordova SP, Smith D, Kelly H. Working towards a simple case definition for influenza surveillance. *J Clin Virol*. 2003;27(2):170-179.

3. Babcock HM, Merz LR, Dubberke ER, Fraser VJ. Case-control study of clinical features of influenza in hospitalized patients. *Infect Control Hosp Epidemiol*. 2008;29(10):921-926.

4. Aguilera JF, Paget WJ, Mosnier A, et al. Heterogeneous case definitions used for the surveillance of influenza in Europe. *Eur J Epidemiol*. 2003;18(8):751-754.

5. Nichol KL, Mendelman P. Influence of clinical case definitions with differing levels of sensitivity and specificity on estimates of the relative and absolute health benefits of influenza vaccination among healthy working adults and implications for economic analyses. *Virus Res*. 2004;103(1-2):3-8.

6. Taylor JL, Dwyer DM, Coffman T, Groves C, Patel J, Israel E. Nursing home outbreak of influenza A (H3N2): Evaluation of vaccine efficacy and influenza case definitions. *Infect Control Hosp Epidemiol*. 1992;13(2):93-97.

7. Boivin G, Hardy I, Tellier G, Maziade J. Predicting influenza infections during epidemics with use of a clinical case definition. *Clin Infect Dis*. 2000;31(5):1166-1169.

8. Monto AS, Gravenstein S, Elliott M, Colopy M, Schweinle J. Clinical signs and symptoms predicting influenza infection. *Arch Intern Med*. 2000;160(21):3243-3247.

9. Ohmit SE, Monto AS. Symptomatic predictors of influenza virus positivity in children during the influenza season. *Clin Infect Dis*. 2006;43(5):564-568.

10. Call SA, Vollenweider MA, Hornung CA, Simel DL, McKinney WP. Does this patient have influenza? *JAMA*. 2005;293(8):987-997.

11. Loeb M, Russell ML, Moss L, et al. Effect of influenza vaccination of children on infection rates in Hutterite communities: A randomized trial. *JAMA*. 2010;303(10):943-950.

12. van Elden LJ, van Essen GA, Boucher CA, et al. Clinical diagnosis of influenza virus infection: Evaluation of diagnostic tools in general practice. *Br J Gen Pract*. 2001;51(469):630-634.

13. Govaert TM, Dinant GJ, Aretz K, Knottnerus JA. The predictive value of influenza symptomatology in elderly people. *Fam Pract*. 1998;15(1):16-22.

14. Habib-Bein NF, Beckwith WH,3rd, Mayo D, Landry ML. Comparison of SmartCycler real-time reverse transcription-PCR assay in a public health laboratory with direct immunofluorescence and cell culture assays in a medical center for detection of influenza A virus. *J Clin Microbiol*. 2003;41(8):3597-3601.

15. Schweiger B, Zadow I, Heckler R, Timm H, Pauli G. Application of a fluorogenic PCR assay for typing and subtyping of influenza viruses in respiratory samples. *J Clin Microbiol*. 2000;38(4):1552-1558.

16. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol*. 1996;49(12):1373-1379.

17. Carrat F, Tachet A, Rouzioux C, Housset B, Valleron AJ. Evaluation of clinical case definitions of influenza: Detailed investigation of patients during the 1995-1996 epidemic in France. *Clin Infect Dis*. 1999;28(2):283-290.

18. Puzelli S, Valdarchi C, Ciotti M, et al. Viral causes of influenza-like illness: Insight from a study during the winters 2004-2007. *J Med Virol*. 2009;81(12):2066-2071.

19. Hayden FG, Osterhaus AD, Treanor JJ, et al. Efficacy and safety of the neuraminidase inhibitor zanamivir in the treatment of influenza virus infections. GG167 influenza study group. *N Engl J Med*. 1997;337(13):874-880.

20. Quach C, Newby D, Daoust G, Rubin E, McDonald J. QuickVue influenza test for rapid detection of influenza A and B viruses in a pediatric population. *Clin Diagn Lab Immunol*. 2002;9(4):925-926.

21. Rodriguez WJ, Schwartz RH, Thorne MM. Evaluation of diagnostic tests for influenza in a pediatric practice. *Pediatr Infect Dis J*. 2002;21(3):193-196.

22. Mandell GL. *Essential Atlas of Infectious Diseases.* 2d ed. Phila.: Current Medicine; 2001.

23. Brunt JH, Lindsey E, Hopkinson J. Health promotion in the Hutterite community and the ethnocentricity of empowerment. *Can J Nurs Res*. 1997;29(1):17-28.

24. Nichol KL. Heterogeneity of influenza case definitions and implications for interpreting and comparing study results. *Vaccine*. 2006;24(44-46):6726-6728.

Table 1: Descriptive characteristics of all participants who provided swab specimens, PCR positive influenza cases and PCR negative influenza participants

Characteristic	All *n* (%)*	PCR positive for influenza		*p*-value**
		Yes *n* (%)	No *n* (%)	
Total	142	37	105	
Age groups, years				0.003
Less than 7	52 (36.6)	12 (32.4)	40 (38.1)	
7 -15	22 (15.5)	13 (35.1)	9 (8.6)	
16 - 22	11 (7.7)	3 (8.1)	8 (7.6)	
23 - 49	38 (26.8)	7 (18.9)	31 (29.5)	
50 - 64	11 (7.7)	0	11 (10.5)	
65 and older	8 (5.6)	2 (5.4)	6 (5.7)	
Female	86 (60.6)	22 (59.5)	64 (61)	0.87
Vaccination status				
Study vaccine (influenza or hepatitis A)	41 (28.9)	17 (46)	24 (22.8)	0.03
Influenza vaccine	21 (14.8)	2 (5.4)	19 (18.9)	0.06
				p-value***
Mean age, years (SD)	22.1 (21.5)	17.03 (18.2)	23.9 (22.3)	0.07
Mean time from symptom onset to swabbing, days (SD)	4.6 (4.9)	3.5 (2.2)	4.9 (5.5)	0.03
Mean number of participant-reported sick days at medical visit (SD)	3.7 (4.1)	2.9 (1.9)	4 (4.6)	0.04
Mean number of symptoms, medical records	2.4 (1.3)	2.8 (1.1)	2.3 (1.4)	0.08
Mean number of symptoms, participant report	2.3 (1.8)	2.9 (2)	2.1 (1.6)	0.03

*Percentage of total per row

** *p*-value for Pearson chi square test for confirmed cases of influenza compared with participants without influenza

*** *p*-value for independent samples t-test between confirmed cases of influenza and participants without influenza

Table 2: Symptoms experienced by PCR positive influenza cases and PCR negative influenza participants, according to each data strategy

Symptoms, data source	**All** *n* (%*)	**PCR Positive** *n* (%)	**PCR Negative** *n* (%)	***p*-value****
All	142	37	105	
Fever				
Physician	45 (31.7)	25 (67.6)	20 (19.5)	<0.0001
Participant	22 (15.5)	10 (27)	12 (11.4)	0.02
Combined	51 (35.9)	26 (70.3)	25 (23.8)	<0.0001
Cough				
Physician	92 (64.8)	28 (75.7)	64 (60.9)	0.10
Participant	89 (62.7)	29 (78.4)	60 (57.1)	0.02
Combined	111 (78.2)	32 (86.5)	79 (75.2)	0.16
Sore throat				
Physician	65 (45.8)	19 (51.4)	46 (43.8)	0.43
Participant	55 (38.7)	19 (51.4)	36 (34.3)	0.07
Combined	89 (62.7)	30 (81.1)	59 (56.2)	0.01
Runny nose				
Physician	44 (31)	9 (24.3)	35 (33.3)	0.30
Participant	51 (35.9)	15 (40.5)	36 (34.3)	0.50
Combined	75 (52.8)	19 (51.4)	56 (53.3)	0.84
Headache				
Physician	16 (11.3)	4 (10.8)	12 (11.5)	0.92
Participant	23 (16.2)	6 (16.2)	17 (16.2)	1.00
Combined	32 (22.5)	8 (21.6)	24 (22.8)	0.88
Sinus problems				
Physician	26 (18.3)	5 (13.5)	21 (20)	0.35
Participant	25 (17.6)	5 (13.5)	20 (19.1)	0.45
Combined	38 (26.8)	7 (18.9)	31 (29.5)	0.18
Muscle aches				
Physician	9 (6.3)	3 (8.1)	6 (5.7)	0.61
Participant	15 (10.6)	8 (21.6)	7 (6.7)	0.05
Combined	20 (14.1)	9 (24.3)	11 (10.5)	0.08
Fatigue				
Physician	12 (8.5)	5 (13.5)	7 (6.7)	0.28
Participant	20 (14.1)	9 (24.3)	11 (10.5)	0.08
Combined	29 (20.4)	12 (32)	17 (16)	0.06
Earache				
Physician	26 (18.3)	2 (5.4)	24 (22.9)	0.002
Participant	17 (12)	2 (5.4)	15 (14.3)	0.09
Combined	30 (21)	3 (8.1)	27 (25.7)	0.006
Chills				

Physician	9 (6.3)	2 (5.4)	7 (6.7)	0.79
Participant	15 (10.6)	4 (10.8)	11 (10.5)	0.96
Combined	22 (15.5)	6 (16.2)	16 (15.2)	0.89

*Percentage of total per row

** *p*-value for independent samples t-test between participants with influenza and participants without influenza.

Table 3: Symptoms, as reported by data source, predicting influenza

Symptoms, data source						Logistic Regression		
	n	**Sensitivity**	**Specificity**	**Positive Predictive Value**	**Negative Predictive Value**	**Odds Ratio**	**95% Confidence Intervals**	**p-value**
Fever								
Physician	45	0.68	0.81	0.56	0.88	8.90	3.81 – 20.58	<0.0001
Participant	22	0.27	0.89	0.45	0.78	2.87	1.12 – 7.37	0.03
Combined	51	0.70	0.76	0.51	0.88	7.56	3.28 – 17.45	<0.0001
Cough								
Physician	92	0.76	0.39	0.30	0.82	1.99	0.85 – 4.65	0.11
Participant	89	0.78	0.43	0.33	0.85	2.72	1.13 – 6.51	0.03
Combined	111	0.86	0.29	0.25	0.84	2.11	0.74 – 5.97	0.16
Sore throat								
Physician	65	0.51	0.56	0.29	0.77	1.35	0.64 – 2.87	0.43
Participant	55	0.51	0.66	0.35	0.79	2.02	0.95 – 4.33	0.07
Combined	89	0.81	0.44	0.34	0.87	3.34	1.35 – 8.29	0.01
Runny nose								
Physician	44	0.24	0.67	0.21	0.71	0.64	0.27 – 1.51	0.31
Participant	51	0.41	0.66	0.29	0.76	1.31	0.61 – 2.82	0.50
Combined	75	0.51	0.47	0.25	0.73	0.92	0.44 – 1.96	0.84
Headache								
Physician	16	0.11	0.89	0.25	0.74	0.94	0.28 – 3.12	0.92
Participant	23	0.16	0.84	0.26	0.74	1.00	0.36 – 2.77	1.00
Combined	32	0.22	0.77	0.25	0.74	0.93	0.38 – 2.30	0.88
Sinus problems								
Physician	26	0.14	0.80	0.19	0.72	0.63	0.22 – 1.80	0.38
Participant	25	0.14	0.81	0.20	0.73	0.66	0.23 – 1.92	0.45
Combined	38	0.19	0.71	0.18	0.71	0.56	0.22 – 1.40	0.21

Table 3 continued: Symptoms, as reported by data source, predicting influenza

Symptoms, data source						Logistic Regression		
	n	**Sensitivity**	**Specificity**	**Positive Predictive Value**	**Negative Predictive Value**	**Odds Ratio**	**95% Confidence Intervals**	**p-value**
Muscle aches								
Physician	9	0.08	0.94	0.33	0.74	1.46	0.35 – 6.14	0.61
Participant	15	0.22	0.93	0.53	0.77	3.86	1.29 – 11.55	0.02
Combined	20	0.24	0.90	0.45	0.77	2.75	1.03 – 7.30	0.04
Fatigue								
Physician	12	0.14	0.93	0.42	0.75	2.19	0.65 – 7.37	0.21
Participant	20	0.24	0.90	0.45	0.77	2.75	1.03 – 7.30	0.04
Combined	29	032	0.84	0.42	0.78	2.49	1.05 – 5.89	0.04
Earache								
Physician	26	0.05	0.77	0.08	0.70	0.19	0.04 – 0.86	0.03
Participant	17	0.05	0.86	0.12	0.72	0.34	0.08 – 1.58	0.17
Combined	30	0.08	0.74	0.10	0.70	0.26	0.07 – 0.90	0.03
Chills								
Physician	9	0.05	0.93	0.22	0.74	0.80	0.16 – 4.03	0.79
Participant	15	0.11	0.90	0.27	0.74	1.04	0.31 – 3.48	0.96
Combined	22	0.19	0.85	0.27	0.74	1.08	0.39 – 3.00	0.89

Table 4: Symptom combinations of participants that provided swab specimens, positive influenza cases and negative influenza participants, according to each data strategy

Symptoms, data source	**All** *n* (%*)	**Influenza Positive** *n* (%)	**Influenza Negative** *n* (%)	***p*-value****
All	142	37	105	
Fever and cough				
Physician	32 (22.5)	18 (48.6)	14 (13.3)	<0.0001
Participant	20 (14.1)	10 (27)	10 (9.5)	0.01
Combined ILI	43 (30.3)	22 (59.5)	21 (20)	<0.0001
Combined symptoms	45 (31.7)	23 (62.2)	22 (21)	<0.0001
Fever and sore throat				
Physician	21 (14.8)	13 (35.1)	8 (7.6)	<0.0001
Participant	14 (9.9)	8 (21.6)	6 (5.7)	0.01
Combined ILI	28 (19.7)	17 (45.9)	11 (10.5)	<0.0001
Combined symptoms	36 (25.3)	22 (59.5)	14 (13.3)	<0.0001
Fever and (cough or sore throat)				
Physician	38 (26.8)	22 (59.5)	16 (15.2)	<0.0001
Participant	22 (15.5)	10 (27)	12 (11.4)	0.02
Combined ILI	48 (33.8)	25 (67.6)	23 (21.9)	<0.0001
Combined symptoms	51 (35.9)	26 (70.3)	25 (23.8)	<0.0001

*Percentage of total per row

***p*-value for independent samples t-test between participants with influenza and participants without influenza.

Table 5: Surveillance definitions, as reported by data source, predicting influenza

Symptoms, data source						Logistic regression		
	n	**Sensitivity**	**Specificity**	**Positive Predictive Value**	**Negative Predictive Value**	**Odds Ratio**	**95% Confidence Intervals**	**p-value**
Fever and cough								
Physician	32	0.49	0.87	0.56	0.83	6.16	2.61 – 14.49	<0.0001
Participant	20	0.27	0.90	0.50	0.78	3.52	1.33 – 9.33	0.01
Combined ILI	43	0.59	0.80	0.51	0.85	5.87	2.61 – 13.21	<0.0001
Combined sym	45	0.62	0.79	0.51	0.86	6.20	2.75 – 13.99	<0.0001
Fever and sore throat								
Physician	21	0.35	0.94	0.62	0.80	6.57	2.45 – 17.63	<0.0001
Participant	14	0.22	0.95	0.57	0.77	4.55	1.46 – 14.18	0.01
Combined ILI	28	0.46	0.90	0.61	0.82	7.26	2.96 – 17.85	<0.0001
Combined sym	36	0.59	0.87	0.61	0.86	9.53	4.01 – 22.63	<0.0001
Fever and (cough or sore throat)								
Physician	38	0.59	0.85	0.58	0.86	8.16	3.51 – 18.99	<0.0001
Participant	22	0.27	0.89	0.45	0.78	2.87	1.12 – 7.37	0.03
Combined ILI	48	0.68	0.78	0.52	0.87	7.43	3.24 – 17.02	<0.0001
Combined sym	51	0.70	0.76	0.51	0.88	7.56	3.28 – 17.45	<0.0001

CHAPTER 6

MEASURING AGREEMENT FOR BINARY DATA

This manuscript has been accepted for publication in the Hypothesis section of the June 2011 issue of *Canadian Family Physician.* This manuscript is reprinted here with copyright permission from the College of Family Physicians of Canada for *Canadian Family Physician*.

Barbara AM, Loeb M, Dolovich L, Brazil K, Russell MK. Patient self report and medical records: Measuring agreement for binary data. *Canadian Family Physician, 2011;* 57: 737-738.

Measuring agreement for binary data between patient self report and medical records

Angela M. Barbara[1]*, Mark Loeb[2], Lisa Dolovich[3], Kevin Brazil[4], Margaret L. Russell[5]

[1]Department of Pathology and Molecular Medicine, McMaster University

[2]Department of Pathology and Molecular Medicine, Department of Medicine, Department of Clinical Epidemiology and Biostatistics, McMaster University

[3]Department of Family Medicine, Center for Evaluation of Medicines, McMaster University

[4]St. Joseph's Health System Research Network; Department of Clinical Epidemiology and Biostatistics, McMaster University

[5] Department of Community Health Sciences, University of Calgary

*Correspondence to:

Infectious Diseases Research Unit

Pathology & Molecular Medicine

1280 Main Street West, MDCL 3200

Hamilton, ON, L8N 4K1

barbara@mcmaster.ca

Phone: (905) 525-9140 ext.21478

Fax: (905) 389-5822

Research data in family medicine often comes from two sources: self-report and medical record review. Frequently, the quality of these data sources is assumed, but measuring the reproducibility of these data is essential to evaluate the quality of the information collected. In ideal circumstances, data obtained from either data source would be equivalent. However, no source of data is without error. If you have low agreement between data sources, research findings will differ depending on the method of data collection used[1] and you will not know which estimate is correct. Comparisons of data from different sources can provide family medicine researchers with insight into the most appropriate data source to use to answer a specific research question or where efforts should be made to improve the collection and recording of health data.[2]

Imagine that we are interested in the prevalence of fever or cough in outpatients over the past influenza season. Neither the medical record nor patient self-report is considered the true criterion or "gold standard" for symptoms. Therefore, we are not assessing the accuracy of one data source compared to another; rather, we are examining agreement between the sources of data. The presence or absence of patient symptoms is considered a binary variable; a categorical variable in which there are two possible conditions (e.g. yes/no, positive/negative). This paper describes indicators for determining agreement between binary variables: total agreement, kappa, positive agreement and negative agreement.

The Table below displays data from the Hutterite Influenza Prevention Study in 2x2 contingency tables.[3] Symptoms reported by Hutterite community members were compared to documentation in the medical records. Total agreement is the number of

concordant pairs divided by the total sample. In A, total agreement is 74%, which is the number of concordant yes's for fever (18) plus the concordant no's (112) divided by 176 participants. This simple measure, however, does not take into account that a certain amount of agreement between medical charts and self-report is expected by chance alone.[4] Kappa, on the other hand, measures the strength of agreement beyond what we expect solely by chance. The calculation for kappa is: total agreement - chance agreement / 1 - chance agreement. The answer is on a scale of –1 to 1, where 0 = chance agreement and 1 = perfect agreement. Landis and Koch (1977) proposed the following guidelines for understanding kappa values: < 0 = no agreement; 0.01 - 0.20 = slight; 0.21 - 0.40 = fair; 0.41 - 0.60 = moderate; 0.61 - 0.80 = substantial; and 0.81 - 1.0 = almost perfect agreement.[5] These guidelines are widely used and cited. However, the cut-offs are not universally accepted and have been criticized for being arbitrary divisions based on personal opinion rather than evidence.[6, 7]

Kappa is not simple to interpret because it is influenced by the prevalence of the variable being measured.[8] In the Table, A and C have similar total agreements, but kappas differ according to distributions. Kappa represents the proportion of total variance which is not attributable to chance or random error. Because total variance is minimal in a uniform (homogeneous) population where there is a relatively high (or low) prevalence, kappa will be low even though total agreement may be high (D). Because chance agreement is smallest in a mixed (heterogeneous) population, kappa will be higher when prevalence is closer to 50% (B, C). This makes it difficult to compare kappa values between patient symptoms or other variables with different prevalences.[9]

Kappa is also influenced by "bias" or the disagreement in the proportion of positive or negative cases (number of discordant responses).[6] That is, there is a mismatch of positive or negative cases or when the disagreements are not random and go in one direction rather than another,[8, 11] which tends to happen when the prevalence of a symptom is high or low. This may result in a low kappa even though agreement is substantial (Table, D). Kappa is higher when there is a large bias and lowest when bias is absent.[10]

Kappa does not distinguish between various types and sources of agreement and disagreement.[6, 8, 13-14] The aim of measuring agreement is to discover the bases of differences and reduce them if possible, rather than only quantifying the degree of disagreement per se.[9] It may be that no single agreement statistic can adequately capture agreement.[10] Positive agreement is calculated as 2 x concordant positives / (positive pair + positive pair; and negative agreement is 2 x concordant negatives / negative pair + negative-. Both have been recommended to help interpret kappa.[10,12] Using these indices also provides insight into the agreement and imbalance in the proportion of positive or negative responses. This information is useful in determining where actions to improve data quality should focus on depending on what is most important and would be missed by calculating only kappa and total agreement.[2, 10, 12] Low positive agreement indicates there is poor concordance between both sources in reporting the presence of the symptom (Table, D). Whereas, high negative agreement means that there is good concordance between both sources in identifying that the symptom was not experienced[12] (Table, A, C, D).

Family medicine practitioners should consider the above concepts when evaluating various aspects of clinical care, e.g. data collection for a new practice quality assurance process. Although total agreement and kappa are commonly reported in agreement studies, we recommend the additional use of positive agreement and negative agreement.

References

1. Zhu K, McKnight B, Stergachis A, Daling JR, Levine RS. Comparison of self-report data and medical records data: Results from a case-control study on prostate cancer. *Int J Epidemiol*. 1999;28(3):409-17.

2. Westbrook JI, McIntosh JH, Rushworth RL, Berry G, Duggan JM. Agreement between medical record data and patients' accounts of their medical history and treatment for dyspepsia. *J Clin Epidemiol*. 1998;51(3):237-44.

3. Loeb M, Russell ML, Moss L, et al. Effect of influenza vaccination of children on infection rates in Hutterite communities: A randomized trial. *JAMA*. 2010;303(10):943-50.

4. Banerjee M, Capzzoli M, McSweeney L, Sinha D. Beyond kappa: A review of interrater agreement measures. *Canadian Journal of Statistics*. 1999;27(1):3-23.

5. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-74.

6. Sim J, Wright CC. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Phys Ther*. 2005;85(3):257-8.

7. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. *BMJ*. 1992;304(6840):1491-4.

8. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. the problems of two paradoxes. *J Clin Epidemiol*. 1990;43(6):543-9.

9. Sargeant JM, Martin SW. The dependence of kappa on attribute prevalence when assessing the repeatability of questionnaire data. *Prev Vet Med*. 1998;34(2-3):115-23.

10. Chen G, Faris P, Hemmelgarn B, Walker RL, Quan H. Measuring agreement of administrative data with chart data using prevalence unadjusted and adjusted kappa. *BMC Med Res Methodol*. 2009;9:5.

11. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol*. 1993;46(5):423-9.

12. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. resolving the paradoxes. *J Clin Epidemiol*. 1990;43(6):551-8.

13. Thompson WD, Walter SD. A reappraisal of the kappa coefficient. *J Clin Epidemiol*. 1988;41(10):949-58.

14. Lantz CA, Nebenzahl E. Behavior and interpretation of the kappa statistic: Resolution of the two paradoxes. *J Clin Epidemiol*. 1996;49(4):431-4.

Table:

A. **Fever**

	Medical record		
		Yes	No
Self-report	Yes	18	6
	No	40	112

Total Agreement = 0.74
Kappa = 0.31
Positive Agreement = 0.44
Negative Agreement = 0.83

B. **Earache**

	Medical record		
		Yes	No
Self-report	Yes	15	4
	No	18	139

Total Agreement = 0.88
Kappa = 0.51
Positive Agreement = 0.58
Negative Agreement = 0.93

C. **Cough**

	Medical record		
		Yes	No
Self-report	Yes	82	20
	No	30	44

Total Agreement = 0.72
Kappa = 0.41
Positive Agreement = 0.77
Negative Agreement = 0.64

D. **Chills**

	Medical record		
		Yes	No
Self-report	Yes	2	16
	No	11	147

Total Agreement = 0.85
Kappa = 0.05
Positive Agreement = 0.13
Negative Agreement = 0.92

CHAPTER SEVEN

CONCLUSION

The chapters comprising this thesis yield several interesting findings both from epidemiological and health services research methodology perspectives. This concluding chapter discusses the main findings, additional findings that could not easily fit into versions of the manuscripts to be submitted for publication, overall methodological issues and limitations, and suggests potential implications and areas for future research.

Major and Additional Findings

Summary of Findings by Symptoms

Agreement between self-report and medical records for fever was moderate (74%). Fever is often an early systemic symptom of influenza, usually with short duration,[1] as reflected in the sample where average duration was 2.5 days. Fever was explicitly and objectively defined for the RCT as a temperature >=38 degrees Celsius as measured by the same type of thermometer and same route (oral) for each participant experiencing symptoms. It is possible that physicians were not as consistent in measuring and recording fever for patients; fever may have been assessed subjectively (as per patient complaint) and may have been influenced by phase of infection and use of antipyretics.[2] Agreement for fever improved for males and as participant age increased.

For nine symptoms, there were lower positive agreements (13% to 58%) and higher negative agreements (74% to 93%). This finding indicated poorer agreement

regarding the presence of symptoms with an imbalance weighted towards the absence of symptoms. The exception to this pattern was cough; positive agreement was 77% and negative agreement was 64%. Positive agreement may be more suitable than total agreement when there are many negative reports, i.e. low reporting of symptoms. The value of positive agreement has been considered comparable to sensitivity.[3]

Cough is a common symptom of respiratory illness[1] and accounts for the largest single cause of primary care visits in the United Kingdom.[4] In this thesis, it was the most frequently reported symptom by both the participants (58%) and medical records (62%).

The grouping of fever and cough has been established as predictive of clinical diagnosis of influenza during a seasonal epidemic.[5-7] It is possible that by focusing on these two obvious symptoms of influenza, physicians did not prompt for or document other symptoms. This may partly explain why both fever and cough had higher prevalence according to medical records compared to self-report. It has also been shown that physicians reliably record data about their patients' main complaints or classic symptoms, but not their less typical symptoms.[8]

The local symptom of rhinorrhoea, or runny nose, had an overall agreement of 64%. Sore throat is often the first symptom of a respiratory illness.[1] Agreement for this symptom improved for adults aged 23 to 49 years. Total agreement for headache, another common early systemic symptom related to a respiratory illness[1] was fairly good (81%) reflecting a high negative agreement (89%), but low positive agreement (32%). Participant characteristics improved that estimate, including younger age, high influenza risk status, being sick longer, and being female. The total agreement of 81% for sinus

problems improved for males. Agreement also had an inverse relationship with age; as age group increased, agreement decreased.

Muscle aches, or myalgia, are a common symptom of respiratory infections and experienced by as many as 50% of patients with common cold.[1] This symptom showed good agreement between sources (88%). However, this reflects the low prevalence of muscle aches and high agreement regarding the absence of the symptom. The low prevalence limits the assessment of the subgroup comparisons.

Fatigue has been referred to as a "subjective" symptom.[9, 10] Studies that compared the evaluation of symptoms between patient and both significant others and health care professionals found that agreement was better with regards to concrete, objective and observable symptoms that required less interpretation from others.[11-15] Prevalence of fatigue in the medical records was extremely low (8%), suggesting the possibility that clinicians are less likely to recognize or document this symptom. The low prevalence also contributed to the low kappa estimate (0.13) despite fairly good agreement (84%). Similar to muscle aches, agreement for fatigue increased for the youngest and oldest age groups and those at high risk for influenza complications.

Earache is a subjective complaint added to study diary checklist to distinguish it from the clinical diagnosis of otitis media. Total agreement for earache (88%) was good. Agreement was better for females and those with four or more sick days.

Chills are an early symptom of common cold and sometimes described as the first stage of fever.[1] Agreement was also good for this symptom (85%), but the prevalence was too low for adequate assessment by stratification.

Marginal homogeneity (i.e. differences between categories) was evaluated for each binomial distribution using McNemar's chi-square test. There were statistically significant differences ($p < 0.05$) in grading for fever, sore throat and chills. This indicates *systematic difference*s between the proportions of "yes" responses for these three symptoms from the two data sources.

Impact of specific factors on agreement

Compared to the medical records, self-reported prevalence of earache was underestimated by 8% in the sample. This difference was significant in males (14%, 30%), but not females (9%, 12%). Chills were recorded significantly less often in the medical record compared to self-report, but only for participants who were symptomatic for three days or less (4%, 22%). Men and women reported the same mean number of symptoms (2.15, 2.08, p=0.80). Men self-reported fever more often than women (21%, 9%). While the same proportion (58%) of both men and women reported to have cough, the medical records indicated more men (62%) compared to women (45%) had cough. Women reported slightly less chills compared to the medical record (8%, 10%), while men reported more (14%, 3%).

Compared to females, males had somewhat higher overall agreement for fever (77%, 72%) and muscle aches (91%, 86%). Males were also better at agreeing on presence of runny nose (47%, 37%). Women were better at agreeing with the medical records on the presence of headache (38%, 23%). While women were better at agreeing on the presence of sinus problems (53%, 31%), men were better at agreeing on the absence of sinus problems (92%, 87%).

Canaris and colleagues[9] administered similar surveys to adult patients (19 years and older) and health care providers at the time of medical consultation. Among patients, women reported 17 (of 26) respiratory symptoms more often than men. However, among the providers, men reported 17 symptoms more frequently than women. Conversely, we found that the adult males in our study self-reported five (of the ten) symptoms more often than women and that women had seven symptoms reported more often than men in the medical records. Compared to this survey study; our Hutterite sample of adults had lower total agreements for fever (81%, 86%), cough (71%, 93%), runny nose (64%, 83%), sore throat (64%, 89%), headache (71%, 87%), sinus problems (75%, 81%) and chills (78%, 85%); and comparable agreements for muscle aches (82%, 85%), and fatigue (80%, 84%). Because the participants in this study were surveyed immediately before seeing the clinician, it is unsurprising to find higher rates of agreement. Interestingly, our adult sample had better agreement with the medical records for earache (92%, 85%).

Total agreement for headache (95%) was much higher for children under seven years of age (compared to 81% for overall sample). The results of the logistic regression analyses also reflected this. However, agreements for fever (68%) and earache (77%) were lowest for the youngest age group (compared to 74% and 88% for overall sample). Prevalence of sinus problems, muscle aches, fatigue and chills were really low in this age group (varying from n = 0 to n=4) and negative agreements were very high (varying from 95% to 98%). The results of the logistic regression analyses showed that agreements for sinus problems and chills were significantly better for children under the age of seven years. Almost one-thirds of the sample (n = 63) consisted of children up to the ages of

six years of age. The research nurse would interview the mother (in rare cases, the father or other guardian) about the child's symptoms and medical visits. Because some symptoms are subjective, self-reporting is preferred, but when the individual is too young, it is common to obtain parent proxy reports.[16, 17] However, the reliability of parental reports of acute respiratory symptoms of their children has been questioned.[18]

Adults aged 23 to 49 years of ages had the worst agreement for fatigue (72%), but best for sore throat (72%) (compared to 84% and 61% for overall sample). Older people (ages 50 years and older) had better agreements for fever (81%), runny nose (74%) and muscle aches (96%), but significantly worse agreement for chills (74%) (compared to 74%, 64%, 88% and 85% respectively for overall sample).

Agreement for muscle aches (96%) and fatigue (94%) was higher for high-risk participants (compared to 83% and 77% for non-high risk sample). The high total agreements and low positive agreements for headache (87%, 18%) and fatigue (96%, 0%) reflected the low prevalence of the symptoms in this group; i.e. according to the medical records, 3 (4%) high risk participants had muscle aches and 1 (1%) had fatigue; and 4 (6%) self-reported muscle aches and 3 (4%) self-reported fatigue. The results of the logistic regression analyses showed individuals at high risk for influenza agreed with their physicians more often about muscle aches and fatigue.

Individuals over 65 years of age are considered high risk for developing complications of influenza and ILI. However, this age group was underrepresented in our sample (n = 12, 7%). Other studies have found that older persons experienced a lower frequency of signs of upper respiratory tract dysfunction,[19] reported less influenza

symptoms and ILI.[20, 21] Also, the Hutterite population is younger than the overall Canadian population: 5.1% aged 65 years or older in contrast to 13.0% of the Canadian population. Older people make up a small proportion of any Hutterite colony population because of the high fertility rates and large number of children found in all colonies.[22]

The stratification by number of sick days impacted agreement for four symptoms. Total agreement for runny nose was better, although still low, for participants who had been sick for less than four days (66%, 55%). Total agreement for headache (86%, 71%), earache (89%, 82%) and chills (91%, 76%) was better for participants who had more than three sick days. The results of the logistic regression analyses also showed that agreements for headache and chills were improved if the participants reported these symptoms for four or more days.

The results of the logistic regression analyses showed that agreements for cough and for runny nose were not significantly affected by the four variables analyzed.

The prevalence of symptoms among the 162 participants diagnosed with respiratory illness was similar to overall sample. Prevalence of cough increased slightly in this group (63% to 68% according to medical records; 58% to 61% according to self-report).

Some studies have looked at whether the agreement between patients and clinicians are influenced by clinical socio-demographic characteristics. The findings have been inconsistent regarding the effects of patient age, sex, education, income level, and health status on agreement.[23] The research is also difficult to evaluate due to methodological differences.

Data sources for determining influenza-like illness

The observed discrepancies in symptom prevalence between self-report and medical records in Chapter 2 led to the comparison of data collected retrospectively from medical record extraction, similar data collected prospectively from research participants, and the combined or "pooled" data from both sources to assess the sensitivity, specificity and predictive values of the influenza-like illness (ILI) categories in Chapter 5. This analysis would have been superfluous had we found high to perfect agreement between sources in regards to the presence of fever, cough and sore throat in Chapter 2.

Table 1 provides information on the agreement between self- (and parent) reports and medical records for the three influenza like illness (ILI) surveillance definitions in the 142 individuals with medical visit information that provided specimens, the influenza cases and participants who tested negative for influenza. Negative agreements were considerably higher than positive agreements, except for cough and sore throat in the influenza positive group, where positive agreement was higher than negative agreement. Total agreements for ILI were higher for the non-infected group (83% to 92%) compared to the infected group (51% to 57%).

In the overall trial population, there were 254 (8%) participants with laboratory confirmed influenza; 70 (30%) were among the 252 that reported at least one medical visits during influenza season; 37 (21%) were among the 176 participants whose medical visit information was confirmed by clinical report. Of the 176 participants with both self-report and physician-recorded data, 142 (81%) individuals were tested for influenza; 131 provided nasopharyngeal swabs, ten nasal swabs and one throat swab. We obtained a

nasopharyngeal aspirate from over 90% of participants who self-reported two or more symptoms, as per our study design. However, because medical visits were made independently of the clinical trial, a lower proportion of eligible participants with physician-record symptoms (78 – 80%) were tested for influenza. Seventeen (46.0%) influenza cases and 24 (22.8%) non-infected subjects were healthy children and adolescents who received the study vaccine: either inactivated seasonal influenza vaccine or hepatitis A vaccine. Two (5%) influenza cases and 19 (18%) of non-influenza subjects were immunized with the seasonal influenza vaccine from health care providers (independent of trial procedures).

Because we used data from participants' first confirmed medical visits reported during the influenza season, most swab samples (n=117, 82%) were collected prior to the introduction of the novel H1N1 pandemic influenza in Canada on April 23, 2009.[24] Therefore, only 4 (11%) of the 37 positive cases of influenza included in the analyses were made during the H1N1 pandemic. All 37 participants with influenza were diagnosed with a respiratory related illness according to the medical record; the most common were upper respiratory tract infection (n=8), bronchitis (n=5), sinusitis (n=5), pneumonia (n=4), pharyngitis (n=4), and otitis media (n=3).

The criteria for ILI were more often found in children and adolescents under the age of 16 years. Self-reported (or maternal-reported) ILI was found in 16 (22%) participants under the age of 16 years compared to 6 (9%) participants 16 years of age and older. Physician-recorded ILI was found in 25 (34%) individuals less than 16 years of age compared to 16 (23%) individuals 16 years of age and over. The combined data

showed the same results; 32 (43%) versus16 (23%) for combined ILI, and 34 (46%) versus 17 (24%) for combined symptoms.

Simple information-combining strategies may be as good as or better than more complex ones (such as choosing the optimal informant or differential weighting of symptoms) at approximating clinical diagnosis.[25] An advantage of the two pooling schemes using the "or" algorithms, whereby a symptom is considered present if reported by either informant and the classification of ILI is considered present if the symptom criteria is reported by either informant, is that all the available data is included in the analyses. Combining data in this way increases sample size and completes potentially missing data in one source.[26] In contrast, had we used the "and" algorithm, whereby a symptom is considered present if reported by both informants, discordant data would be excluded from the analyses.[27] Analyses would also be limited given the low prevalence for each ILI classification (n<7). We also recommend combining data in a systematic way, rather than the more practical nonsystematic manner.

Reporting of antibiotic name

Of the 151 participants who self-reported having been prescribed antibiotics in Chapter 4, 142 (94%) provided the name of the medication. Of the 128 participants with documentation of an antibiotic in their chart, clinicians provided us with the name of the antibiotic on 127 (99%) information request forms. The most frequently reported antibiotic by both sources were amoxicillin, azithromycin, and clarithromycin. For 122 participants, both sources provided the name of the antibiotic. Using this data, we calculated sensitivity estimates for the top three antibiotics. Sensitivity, specificity,

positive predictive values and negative predictive value for amoxicillin were all high (87%, 98%, 98%, 88%); positive predictive value was good Azithromycin had perfect sensitivity and negative positive predictive value (100%, 100%), very high specificity (97%) and moderate positive predictive value (77%). Clarithromycin had good sensitivity (83%), very high specificity (99%), high positive predictive value (91%) and very high negative predictive value (98%).

Of the 151 participants who self-reported having been prescribed antibiotics, 96 (64%) showed their medication container to the research nurse. Forty-one bottles contained amoxicillin; of these, 37 (90%) had documentation of amoxicillin in the medical chart. Eleven bottles contain azithromycin and 8 (72%) were notes in the medical record. Ten participants had clarithromycin bottles; 7 (70%) of these were documented in the medical record.

Methodological Issues

Chapters 3, 4 and 5 in this thesis focused on the validity of influenza-related event reporting. In Chapter 5, we assessed the validity of influenza-like illness (ILI) surveillance definitions based on symptom reports from the different data sources, using laboratory-confirmed influenza as the criterion or "gold" standard index.

In Chapter 3, we used the medical record as the gold standard since participants were specifically asked to report "physician-identified" ear infection. Reports by medical personnel are often considered the gold standard. In chapter 5, we did find that medical record data on symptom complexes did have greater predictive utility laboratory-

confirmed influenza compared to self-report. However, we do acknowledge the possible limitations of the medical record as a perfect reference. Several studies have found non-reporting and misreporting in medical records.[28, 29] Busier physicians may record less in the medical record or delay recording, leading to errors in recall.[30] The process of abstracting information from the medical chart itself is also subject to imprecision.[31] Furthermore, medical records accessed for this analysis were not written or kept for the purposes of this thesis[26] and that documentation may have been guided by institutional policy, physician training and physician preference. [32] Furthermore, in Chapter 3, we did not ask physicians to provide information about how the diagnosis of otitis media was made.

In many pharmacy studies, prescription records at the pharmacy level are typically seen as the good standard due to their high accuracy are often used to evaluate the completeness of medical charts. .[33, 34] However, the use of pharmacy records has the limitation of an underlying assumption that people actually always fill their prescriptions. Because the focus of our study was prescription receipt, we used the medical records as the gold standard. Nevertheless, one study found <5% discrepancy between pharmacy billing records and medical charts regarding antibiotics administered for influenza.[35]

The medical record was in actuality used as a proximate measure of the gold standard to assess accuracy of self-reported otitis media and antibiotic prescribing. Therefore, convergent (or concurrent) validity was examined in Chapters 3 and 4 since neither data source could unequivocally be considered the gold standard. Criterion validity was assessed only in Chapter 5.

Whereas it is difficult to assess the level to which medical records were complete,[36] the research nurses were trained to collect and document information consistent manner from all participants. They were also trained to follow-up with participants to minimize any missing data. We used a comprehensive approach to data collection, including open-ended questions and condition-specific prompts, which has been shown to increase the ability of subjects to self-report antibiotic exposures.[37] Therefore, one could argue that it is also appropriate to evaluate the performance of the medical record for documentation of antibiotic prescriptions using self-report as the gold standard (Table 2). The medical record has good sensitivity (84%), high specificity (96%) and almost perfect positive predictive values (99%) compared to self-report. Disagreements occurred most when the patient reported a prescription for antibiotics, but the medical record did not have this documented.

For health-related parameters, such as patient symptoms, one could make an argument for either self-report or medical record as the gold standard. In the absence of a clear, undisputed criterion, neither of the two data sources was assigned as the gold standard for the presence or absence of influenza symptom. Therefore, the analyses in Chapter 2 focused on agreement. While high validity implies high agreement, high agreement does *not* necessarily imply high validity.

While we evaluated the potential effects of a number of demographic variables on agreement, the stratification and logistic regression analyses showed no clear discernable pattern between agreement rates and demographic variables. Rather, it may be methodological factors that influence the accuracy and reliability of report, including: the

importance of the event to the individual, the way the trait of interest is defined, the time frame in which the event occurs, the way a question is asked (i.e., a list, open-ended, or a closed question), and the individual's knowledge about the exposure.[38]

Overall Discussion

The main question of this thesis was to determine which data source provides the most dependable information on influenza-related vents. The answer depends on the type of event. Our data supports the use of medical records for influenza-related symptoms and otitis media and self-reported prescription receipt for antibiotics, especially if we are interested in obtaining information on all possible influenza-related events, as was the aim of the Hutterite Influenza Prevention RCT. In general, the decision regarding which source of data to use will depend on the outcome of interest; whether finding are used for clinical decision making, population surveillance, outcome studies or other research purposes; availability of resources; and whether a false positive or false negative is of more concern.[30]

Validation studies are usually conducted in one location or a limited number of medical facilities; they rarely include a sample from the wide geographic area from which we obtained medical record information. However, this thesis was an observational study with no information on how health providers elicited patient information.

In this thesis, both the data reported by the medical records and by research participants are self-reports. Unlike survey methods that result in “unfiltered” self-reports, both medical record data and our research data were filtered through the

additional questioning of the physician or research nurse.[39] The adequate capture of self-report data relies on both the participant's reporting and the research nurse's recording of the information. Likewise, the adequate capture of clinical data requires the patient's reporting and the health care provider's documentation of the information.[40] However, the contexts and perspectives of the self-reports are different.[41] Differences in physician elicitation of symptoms can be compared to the clinical trial's focus on consistent and systematic collection of data, whereby the research nurses were trained to spend additional time probing for the presence of symptoms. Whereas it is difficult to assess the level to which medical records were complete,[36] the research nurses were trained to collect information regarding the signs and symptoms of influenza in a consistent manner from all participants.

Our findings may also be explained by very different motivations for reporting symptoms in each context. As part of the study protocol, the research nurses obtained nasopharyngeal specimens (or nasal swabs) if a participant reported two or more symptoms. We conjecture that patients may have reported more specific symptoms to their physicians to get a prescription for antibiotics, but the same individuals as research participants may have under-reported symptoms to avoid the discomfort of a nasopharyngeal swab. This is in accordance to one Dutch study that found that general practitioners overestimated symptoms when prescribing antibiotic therapy for respiratory tract infections.[42]

While some symptoms, otitis media and prescription receipt did have high agreement, self-report cannot be viewed as an equivalent substitute for the medical

record. The difference between directly experiencing a symptom and externally observing a behaviour that is indicative of a symptom may lead to different evaluations. This may be especially pertinent for "subjective" symptoms. Studies that compared the evaluation of symptoms between patient and both significant others and professionals found that agreement was better with regards to concrete, objective and observable symptoms that required less interpretation from others.[11-15] Fatigue has been referred to as a "subjective" symptom.[9, 10] Prevalence of fatigue in the medical records was extremely low (8%), suggesting the possibility that clinicians are less likely to recognize or document this symptom. The situational or contextual basis of the judgement also differs,[43] suggesting that both perspectives are worthwhile and have unique contributions. Although the symptom list on the study diaries and medical record request forms used simplified terms that were meant to be clear and unambiguous, there may have been discrepancies related to diverse definitions of particular symptoms.

Limitations

First, the study was not initially designed to analyze the reliability of data sources. However, influenza-related events were similarly recorded on family diaries and during participant interviews and in the medical record questionnaires. Hence, it was feasible to analyze reliability without serious methodological problems.

Second, the study population included only Hutterite participants; therefore generalizability of the findings is unclear. Compared to the medical records, participants under-reported influenza symptoms; especially fever, sore throat and earache. The

Hutterites are known as being "stoic" and bearing pain and physical ailments without complaint.[44] Participants may have hesitated to report symptoms to the research nurses to avoid being perceived as complainers. Hutterites may share their attitudes towards health problems with other rural populations. One study in The Netherlands found that rural populations, compared to urban inhabitants, reported less acute complaints and better health status.[45]

On the one hand, by utilizing a homogeneous population, generalizability is limited. On the other hand, epidemiological studies of unique communities are important for assessing disease prevalence, health services utilization, and health care needs. Such research in the Hutterite population is lacking in the medical literature even though independent systematic surveys indicate that adult Hutterites seek medical care more often that non-Hutterites in the Canadian prairie provinces.[46] This study contributes to the knowledge of this understudied group. The inclusion of a large group of diverse medical facilities and practitioners may contribute to the overall generalizability.

Third, missing data about outpatient medical services for influenza were not considered in the analyses. That is, we could only follow up (with fax requests for medical record information) on medical visits reported by participants or family members. Of the 252 individuals reporting at least one outpatient visit, 73 (29%) were under the age of seven years and 77 (31%) were under between the ages of 23 and 49 years. To our knowledge, health care service utilization of Hutterites living in Canada has not been quantitatively studied or reported in the medical literature. Therefore, there

is no data to indicate how accurate self-reported health service utilization is in this population.

To compare data sources, we limited the analyses to participants whose physician or hospital had provided medical record information, i.e. participants who had data from both sources. It is possible that agreement between sources would be different for participants with missing medical record data. There were no differences as regards to participant sex, level of risk for complications due to influenza, and timing of the visit between the 176 participants with a confirmed medical visit and the 76 participants without a confirmed visit, who were excluded from the analyses. Participation differed by certain demographic characteristics: age, province of residence and location of medical facility. The participants included in the analyses were younger in age (mean=23.7 years, SD=22.6 vs. mean=32.0 years, SD=21.4; p=0.003). The 76 unconfirmed visits were mostly for participants who resided in Saskatchewan (80%) and visited a medical facility in an urban centre (78%). This also limits the generalizability of our results to the general Hutterite population.

Fourth, the analysis looked at the presence or absence of symptoms, but did not evaluate symptom severity. Perception of severity may have influenced recording for both participants and physicians. Also, the study included outpatients only; therefore, severe or disabling symptoms that require hospitalization were not captured.

A fifth limitation was that we relied on clinic or hospital personnel to abstract the medical record data; we cannot assume that data were abstracted in a methodologically consistent manner. No reply was received for 11% of requests, despite follow-up

attempts; another 11% refused to fill out the study form because of the time involved. In some cases, a participant's entire medical record was received, despite clearly requesting only records information that pertained to respiratory illness. These observations suggested that many physician offices or hospital information records departments are too busy to respond to research requests from studies such as this one.

A sixth limitation is that the sample size was modest, resulting in odds ratios with wide confidence intervals. Wide confidence intervals indicate less precision, which could be due to an inadequate sample size. To interpret the results, we looked at both ends of the confidence intervals. For self-report, the range of confidence intervals mean we could not reach a strong conclusion; there was an association but we require more data to know whether it was scientifically trivial or important. For physician and combined data, the low ends of the confidence intervals represented an association large enough to be considered important; therefore, we can conclude that the association is strong enough to be scientifically relevant. It is probable that a larger sample size would yield similar estimates with narrower confidence intervals. However, because we limited the analyses to participants who had data from both sources, it is also possible that responding physicians or medical facilities had different rates of symptom recording from non-responders.

Potential Implications

This thesis adds to the understanding of the reproducibility and concurrent validity of outpatient-reported influenza events. We found that supplementation of self-

report with medical record data was optimal for influenza symptoms and diagnosis of otitis media, while self-report alone for receipt of antibiotics prescription was accurate enough for study purposes.

To obtain accurate data, researcher must choose which main data source to use. Research studies commonly have access to only one source of data and are unable to verify responses from their main data source. The qualities of data sources must be deliberated, especially under budgetary restrictions. In this thesis, there were two sources of data available to calculate prevalence of influenza-related events. We found a discrepancy between the two sources, with the medical records giving a statistically higher prevalence of fever, sore throat and earache and statistically lower prevalence of antibiotic prescriptions. These discrepancies in results between data sources could alter study conclusions depending on which data source (if only one) was used.

By studying the validity of self-reported otitis media and antibiotic prescriptions compared to medical records, our findings suggest that the choice of data source could also impact the inclusion of participants into intervention studies. For example, of the study participants with documentation of otitis media in their medical charts, only 18 / 38 (47%) reported this to the research nurse. If we were conducting a study in which individuals with otitis media were assigned to the treatment group, 20 individuals would be inappropriately assigned to the control group based on self- and parental reports. Of those who reported otitis media to the research nurse, 18 / 28 (64%) had the diagnosis in their medical records. Therefore, ten individuals would be inappropriately assigned to the control group if we relied on self- and parental report for the inclusion criteria. Of

course, the statistical analysis for this fictitious study would lead to imprecise (and potentially biased) results.

The clinical implications include the requisite for clearer communication between influenza patients and physicians. Physicians should be clear with patients and parents about diagnosis and type of medications to be taken (i.e. prescription antibiotics or over-the-counter medications).

The results of this thesis show that epidemiologists may encounter challenges in the successful investigation of influenza-related events. Methodological studies that examine issues of data reproducibility and validity in other outpatient populations are required to ensure that conclusions about influenza symptoms and outcomes are accurate. In cases where gold standards are not clearly discernable and available, researchers must think carefully about the impact of choice of data source on their findings.

There are different methods for obtaining self-reports of influenza-related events. In this thesis, we used were obtained by a two-pronged structured approach: a checklist-type diary followed by face-to-face interview. Telephone interviews, self-administered questionnaires, open-ended methods and electronic interfaces are other modes of obtaining self-report. Future research should assess the relative accuracy and reproducibility of different self-report methods.

Reliability and agreement issues in quality assurance, classification and the conduct of clinical studies have been recognized as important areas of investigation. This is evident in the recent publication of "Guidelines for Reporting Reliability and Agreement Studies (GRRAS)" in the *Journal of Clinical Epidemiology*.[47] We

recommend that these guidelines be used for similar work in other patient populations and communities.

References

1. Eccles R. Understanding the symptoms of the common cold and influenza. *Lancet Infect Dis*. 2005;5(11):718-725.

2. Thursky K, Cordova SP, Smith D, Kelly H. Working towards a simple case definition for influenza surveillance. *J Clin Virol*. 2003;27(2):170-179.

3. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol*. 1990;43(6):551-558.

4. Morice AH. Epidemiology of cough. *Pulm Pharmacol Ther*. 2002;15(3):253-259.

5. Boivin G, Hardy I, Tellier G, Maziade J. Predicting influenza infections during epidemics with use of a clinical case definition. *Clin Infect Dis*. 2000;31(5):1166-1169.

6. Monto AS, Gravenstein S, Elliott M, Colopy M, Schweinle J. Clinical signs and symptoms predicting influenza infection. *Arch Intern Med*. 2000;160(21):3243-3247.

7. Ohmit SE, Monto AS. Symptomatic predictors of influenza virus positivity in children during the influenza season. *Clin Infect Dis*. 2006;43(5):564-568.

8. Romm FJ, Putnam SM. The validity of the medical record. *Med Care*. 1981;19(3):310-315.

9. Canaris GJ, Tape TG, Smith LM, Nickol DR, Wigton RS. Gender differences in patient-provider symptom agreement in reporting respiratory complaints on a questionnaire. *Gend Med*. 2008;5(2):186-193.

10. DeVon HA, Ryan CJ, Zerwic JJ. Is the medical record an accurate reflection of patients' symptoms during acute myocardial infarction? *West J Nurs Res*. 2004;26(5):547-560.

11. Wennman-Larsen A, Tishelman C, Wengstrom Y, Gustavsson P. Factors influencing agreement in symptom ratings by lung cancer patients and their significant others. *J Pain Symptom Manage*. 2007;33(2):146-155.

12. Lobchuk MM, Degner LF. Patients with cancer and next-of-kin response comparability on physical and psychological symptom well-being: Trends and measurement issues. *Cancer Nurs*. 2002;25(5):358-374.

13. Sneeuw KC, Sprangers MA, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease. *J Clin Epidemiol*. 2002;55(11):1130-1143.

14. Tang ST, McCorkle R. Use of family proxies in quality of life research for cancer patients at the end of life: A literature review. *Cancer Invest*. 2002;20(7-8):1086-1104.

15. von Essen L. Proxy ratings of patient quality of life--factors related to patient-proxy agreement. *Acta Oncol*. 2004;43(3):229-234.

16. Jokovic A, Locker D, Guyatt G. How well do parents know their children? implications for proxy reporting of child health-related quality of life. *Qual Life Res*. 2004;13(7):1297-1307.

17. Sherifali D, Pinelli J. Parent as proxy reporting: Implications and recommendations for quality of life research. *J Fam Nurs*. 2007;13(1):83-98.

18. Hoek G, Wypij D, Brunekreef B. Self-reporting versus parental reporting of acute respiratory symptoms of children and their relation to pulmonary function and air pollution. *Int J Epidemiol*. 1999;28(2):293-299.

19. Govaert TM, Dinant GJ, Aretz K, Knottnerus JA. The predictive value of influenza symptomatology in elderly people. *Fam Pract*. 1998;15(1):16-22.

20. Smolderen KG, Vingerhoets AJ, Croon MA, Denollet J. Personality, psychological stress, and self-reported influenza symptomatology. *BMC Public Health*. 2007;7:339.

21. Thumin F, Wims E. The perception of the common cold, and other ailments and discomforts, as related to age. *Int J Aging Hum Dev*. 1975;6(1):43-49.

22. Statistics Canada. *Religion (95A), Age Groups (7A) and Sex (3) for Population, for Canada, Provinces, Territories, Census Metropolitan Areas1 and Census Agglomerations, 1991 and 2001 Censuses - 20% Sample Data.* 2003rd ed. Ottawa: Statistics Canada; 2001.

23. Tisnado DM, Adams JL, Liu H, et al. Does concordance between data sources vary by medical organization type? *Am J Manag Care*. 2007;13(6 Part 1):289-296.

24. Cutler J, Schleihauf E, Hatchette TF, et al. Investigation of the first cases of human-to-human infection with the new swine-origin influenza A (H1N1) virus in Canada. *CMAJ*. 2009;181(3-4):159-163.

25. Offord DR, Boyle MH, Racine Y, et al. Integrating assessment data from multiple informants. *J Am Acad Child Adolesc Psychiatry*. 1996;35(8):1078-1085.

26. Fathelrahman AI. Agreement between questionnaire and medical records on some health and socioeconomic problems among poisoning cases. *BMC Res Notes*. 2009;2:183.

27. Horton NJ, Fitzmaurice GM. Regression analysis of multiple source and multiple informant data from complex survey samples. *Stat Med*. 2004;23(18):2911-2933.

28. Marrie TJ, Durant H, Sealy E. Pneumonia--the quality of medical records data. *Med Care*. 1987;25(1):20-24.

29. Bush TL, Miller SR, Golden AL, Hale WE. Self-report and medical record report agreement of selected medical conditions in the elderly. *Am J Public Health*. 1989;79(11):1554-1556.

30. Ferrante JM, Ohman-Strickland P, Hahn KA, et al. Self-report versus medical records for assessing cancer-preventive services delivery. *Cancer Epidemiol Biomarkers Prev*. 2008;17(11):2987-2994.

31. Pakhomov SV, Jacobsen SJ, Chute CG, Roger VL. Agreement between patient-reported symptoms and their documentation in the medical record. *Am J Manag Care*. 2008;14(8):530-539.

32. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. *Am J Health Syst Pharm*. 2008;65(23):2276-2284.

33. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol*. 2005;58(4):323-337.

34. West SL, Strom BL, Freundlich B, Normand E, Koch G, Savitz DA. Completeness of prescription recording in outpatient medical records from a health maintenance organization. *J Clin Epidemiol*. 1994;47(2):165-171.

35. Wilkes JJ, Leckerman KH, Coffin SE, et al. Use of antibiotics in children hospitalized with community-acquired, laboratory-confirmed influenza. *J Pediatr*. 2009;154(3):447-449.

36. Skinner KM, Miller DR, Lincoln E, Lee A, Kazis LE. Concordance between respondent self-reports and medical records for chronic conditions: Experience from the Veterans health study. *J Ambul Care Manage*. 2005;28(2):102-110.

37. Metlay JP, Hardy C, Strom BL. Agreement between patient self-report and a Veterans affairs national pharmacy database for identifying recent exposures to antibiotics. *Pharmacoepidemiol Drug Saf*. 2003;12(1):9-15.

38. Olson JE, Shu XO, Ross JA, Pendergrass T, Robison LL. Medical record validation of maternally reported birth characteristics and pregnancy-related events: A report from the children's cancer group. *Am J Epidemiol*. 1997;145(1):58-67.

39. Corser W, Sikorskii A, Olomu A, Stommel M, Proden C, Holmes-Rovner M. Concordance between comorbidity data from patient self-report interviews and medical record documentation. *BMC Health Serv Res*. 2008;8:85.

40. Babcock HM, Merz LR, Dubberke ER, Fraser VJ. Case-control study of clinical features of influenza in hospitalized patients. *Infect Control Hosp Epidemiol*. 2008;29(10):921-926.

41. Kraemer HC, Measelle JR, Ablow JC, Essex MJ, Boyce WT, Kupfer DJ. A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *Am J Psychiatry*. 2003;160(9):1566-1577.

42. Akkerman AE, Kuyvenhoven MM, van der Wouden JC, Verheij TJ. Determinants of antibiotic overprescribing in respiratory tract infections in general practice. *J Antimicrob Chemother*. 2005;56(5):930-936.

43. Achenbach TM, McConaughy SH, Howell CT. Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychol Bull*. 1987;101(2):213-232.

44. Brunt JH, Lindsey E, Hopkinson J. Health promotion in the Hutterite community and the ethnocentricity of empowerment. *Can J Nurs Res*. 1997;29(1):17-28.

45. Kroneman M, Verheij R, Tacken M, van der Zee J. Urban-rural health differences: Primary care data and self reported data render different results. *Health Place*. 2010;16(5):893-902.

46. Ross RT, Nicolle LE, Cheang M. Varicella zoster virus and multiple sclerosis in a Hutterite population. *J Clin Epidemiol*. 1995;48(11):1319-1324.

47. Kottner J, Audige L, Brorson S, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol*. 2011;64(1):96-106.

Table 1: Measures of agreement for surveillance definitions between the medical record and self-report for outpatient participants that provided swab specimens, positive influenza cases and negative influenza participants

Symptom combination, agreement	All	Influenza Positive	Influenza Negative
Fever and cough			
Total agreement	0.76	0.57	0.83
Positive agreement	0.35	0.43	0.25
Negative agreement	0.85	0.65	0.90
Fever and sore throat			
Total agreement	0.85	0.64	0.92
Positive agreement	0.40	0.38	0.43
Negative agreement	0.92	0.75	0.96
Fever and cough or sore throat			
Total agreement	0.75	0.51	0.83
Positive agreement	0.40	0.44	0.36
Negative agreement	0.83	0.57	0.85

Table 2: Relative validity estimates for antibiotic prescription reporting of each data source using the other data source as the gold standard

Estimate	Value of *self report*; *medical record* as gold standard	Value of *medical record; self-report* as gold standard
Sensitivity	0.98	0.84
Specificity	0.50	0.96
Positive Predictive Value	0.91	0.99
Negative Predictive Value	0.65	0.50
Positive Likelihood Ratio	1.96	0.80
Negative Likelihood Ratio	0.04	0.13
Total Agreement	0.85	
Kappa (SD)	0.57 (0.07)	