Student Name: Sari Yakubovich **Date:** Aug. 8, 2014

Project Title: Data Quality in Primary Care Electronic Medical Records in Manitoba

Primary Supervisor Name: Dr. Alexander Singer

Department: Family Medicine

SUMMARY: (no more than 250 words single spaced)

Background: Evaluation of primary care EMR data quality is crucial since data must be of high quality in order to maximize patient care and use databases for secondary purposes including improved chronic disease management. Completeness evaluates data for gaps that may limit it's ability to represent what it should. This study aims to evaluate the baseline problem list completeness for Manitoba primary care EMRs.
Methods: We conducted a retrospective analysis of the QHR Accuro® EMR database within 9 salaried Winnipeg Regional Health Authority (WRHA) and 3 fee for service primary care clinics in Manitoba. Queries were designed in the Accuro® EMR query builder. Aggregates were used to calculate sensitivity as a measure of completeness. The seven chronic diseases evaluated include, hypertension, diabetes, hypothyroidism, asthma, COPD, CHF, and CAD. Only searchable, structured data with ICD-9 coding was assessed. The 12 clinic types were divided into four categories: teaching, access centres, community, and fee for service and mean completeness was calculated for each. One way ANOVA and post hoc contrast analyses were conducted to identify differences between salaried and fee for service clinics.
Results: Fee for service clinics exhibited significantly lower problem list completeness rates than salaried clinics for hypothyroidism, asthma, COPD, and CAD. Sensitivities calculated for each disease were significantly worse than those reported from previous UK research.
Conclusion: This study demonstrates the need for better understanding of data quality in Canada and improvements so that primary care data can be reliably used for secondary purposes.

[redacted]

[redacted]

ACKNOWLEDGEMENTS:
I gratefully acknowledge the support by one or more of the following sponsors;
H.T. Thorlakson Foundation
Dean, Faculty of Medicine
Manitoba Health Research Council
Manitoba Institute of Child Health
Kidney Foundation of Manitoba
Leukemia and Lymphoma Society of Canada
CancerCare Manitoba
Manitoba Medical Service Foundation
Associate Dean (Research), Faculty of Medicine
Heart and Stroke Foundation
Health Sciences Centre Research Foundation
Other:

Introduction and Background

The implementation and utilization of information technology in health care settings to improve patient care continues to be a goal within the health sector (Lau et al., 2010). Electronic medical records (EMRs) have the ability to fulfill this goal in primary care and greatly benefit both patients and clinicians. As recently outlined by Canada Health Infoway (2012), EMRs can be beneficial by allowing faster access to patient information, collaboration between care providers, providing decision support at the point of care, improved efficiency by avoiding duplications and improving information management. However, in order for these benefits to be realized it is important that clinicians are using EMRs appropriately. There remains a long way to go in order for clinicians to fully implement and use health information technology in their practices (Canada Health Infoway, 2012b).

It has been previously reported that only 14% of physicians in Canada are using their EMRs in a clinically significant way (Dermer & Morgan, 2010; Price et al., 2011), however EMR use in Canada has increased from 37% in 2009 to 56% in 2012 (Schoen et al., 2012). The eHealth Observatory at the University of Victoria has developed a framework to evaluate EMR adoption and use (Price et al., 2011). It utilizes ten different functional categories such as health information, lab management, referrals, and decision support. The adoption framework has been used to assess practices in Manitoba and the clinics involved had EMR adoption scores ranging from 2.3-3.0 out of a possible 5.0 (Price, et al., 2013). However measuring adoption does not provide information about what is actually recorded in EMR databases.

Assessment of data quality is a more sophisticated way of measuring how effectively clinicians are actually managing their patients using an EMR (Bowen & Lau, 2012), as it evaluates the data being recorded. As a concept, data quality can be very ambiguous so it is important to clearly define it, however the term does not have one generally accepted definition (Bowen & Lau, 2012). For the purposes of this study we have adopted the data quality definition “fitness for use” (Juran & Godfrey 1999; Wang & Strong, 1996). For example if we want to assess the fitness of an EMR problem list for use in identifying patients with hypothyroidism, we can design a query to find all patients with a hypothyroidism diagnosis in their problem list. We can then compare that to a reference population, such as all patients who have been prescribed levothyroxine (a medication specifically used to treat hypothyroidism) (Bowen, 2012). Ideally, the numbers should match – if not, then the data quality would be considered below expectation and the EMR problem list would not be fit to identify patients with hypothyroidism.

Data quality is crucial because with the transition from paper to electronic records there is an expectation that patient care will not only be improved, but that EMR databases will have the capacity to be used for secondary purposes. Some of these include, research, public health surveillance, chronic disease management, and quality improvement (Byrd et al., 2013; Birtwhistle et al., 2009; Griever et al., 2011; Shephard et al., 2011; Tate et al., 2014). If data is of poor quality then its use for all these different purposes will lead to unreliable results.

Studies that have been conducted to assess data quality have found a multitude of problems with the quality of data entered into EMRs, such as inconsistent diagnostic coding, misspelled words, inconsistent word strings, free text instead of structured text, and a lack of structured data in multiple areas of EMRs (Birtwhistle et al., 2009; Griever et al., 2011). Therefore, it is essential to evaluate EMR data quality especially in the the Canadian context, since there have been very few studies dedicated to this within Canada (Bowen & Lau, 2012). In a systematic review of studies assessing the quality of EMR data, only one of the 37 articles included Canadian data. (Thiru et al., 2003). In another more recent review of data quality research, none

of the 35 studies were based in Canada (Chan et al., 2010). With the development of initiatives such as the Canadian Primary Care Sentinal Surveillance Network (CPCSSN) database, which is a primary care database for chronic disease surveillance (Kadhim-Saleh et al., 2013), there is a great urgency for data quality to be evaluated in primary care EMRs across Canada.

Currently the Winnipeg Regional Health Authority (WRHA) has implemented the QHR Accuro® EMR within it's funded community care network. To date the WRHA Community EMR has been implemented at 37 sites, and holds approximately 100,000 distinct, active patient records. Implementation of the EMR began after an October 2010 announcement by the Manitoba Minister of Health stating that EMR funding would be increased (Canada Health Infoway, 2010). Other fee for service primary care clinics throughout Manitoba have also implemented the QHR Accuro® EMR at various points in time. However, since these clinics are not part of the WRHA they each have their own EMR database. To date, no coordinated assessment of data quality has been done within primary care clinics in Manitoba using the QHR Accuro® EMR.

This study focusses on one data quality parameter, completeness. Completeness evaluates the data to determine if there are any gaps limiting it from representing what it is supposed to (Bowen & Lau, 2012). The goal of this study is to evaluate the baseline problem list completeness of primary care EMR databases in Manitoba. Completeness is a very common parameter assessed in the literature. One data quality review reports that 54% of 37 studies assessed completeness (Thiru et al., 2003), another review reports that 57% of 35 studies evaluate it (Chan et al., 2010), and in a third review 64% of 95 articles researched completeness (Weiskopf & Weng, 2013).

More specifically, this study evaluates EMR problem list completeness of seven different chronic diseases. Chronic disease management and surveillance as mentioned previously, are main secondary uses of EMR databases, so as the first step it is important to ensure that these disease diagnoses are being properly recorded in EMR problem lists. In the context of this study we define properly as, data searchable via queries and structured data coded with ICD-9 (International Classification of Disease, version 9) codes.

The chronic diseases evaluated in this study are, hypertension, diabetes, hypothyroidism, asthma, chronic obstructive pulmonary disease (COPD), congestive heart failure (CHF), and coronary artery disease (CAD). Hypertension, diabetes, asthma, CHF, and CAD were chosen because the management of these diseases is part of Primary Care Quality Indicators (PCQIs) in Manitoba (Government of Manitoba, 2014). These indicators are a way for primary care providers to track management and screening involved in caring for patients with these chronic diseases and essentially serve as reminders to providers. These diseases were chosen for evaluation because without proper coding in the EMR, the functionality that assists in management and screening would not function. COPD and hypothyroidism were evaluated as well, since their management is not part of PCQIs. In order to fulfill the study goal of evaluating the baseline problem list completeness of primary care databases in Manitoba, it is important to assess a range of diseases.

The understanding of primary care EMR data quality is in its infancy in Canada, therefore we hypothesize that problem list completeness within the QHR Accuro® EMR will be poor and improvements will be necessary before the databases can be reliably and easily used for secondary purposes.

Materials and Methods

Study design
This study was conducted as a retrospective analysis of the QHR Accuro® EMR database within Winnipeg Regional Health Authority (WRHA) primary care clinics and fee for service primary care clinics in urban and rural Manitoba. Queries were created with the QHR Accuro® EMR query builder. The query design was based on the University of Victoria's eHealth Observatory Data Quality Evaluation Guide (Bowen, 2012). When creating queries, rules are added in order to restrict the data captured. In this study query rules included, the clinic office, office provider (i.e., the physician or nurse practitioner), bill date (see below for date restrictions), bill diagnosis code, and problem list diagnosis code. Only structured data that used the ICD-9 coding system was captured for both billing and problem list diagnoses. A log of the queries created can be viewed in Appendix A.

Manitoba eHealth created a separate research database for this study which mirrored the QHR Accuro® database used throughout the WRHA. Since the WRHA version of the Accuro® EMR is a shared database, we were able to run queries from any computer workstation connected to the network. For clinics that were not part of the WRHA, data was collected through remote access to their EMR or by travelling to the site to collect data directly.

Queries were restricted to patients that had received care from providers within an 18 month time period. This is the timeframe used to determine a provider's truly active patient panel (Murray, et al., 2007), meaning that the patient does not only have an "active" status in the EMR, but has actually received care within an 18 month period. The 18 month time period used was dependent on the time of data collection. Data was collected from May to August 2013 and May to July 2014 and the database was frozen in time at two different points: June 18, 2013 for data collected in summer 2013 and June 3, 2014 for collection in the summer of 2014. In the WRHA database, queries captured active patients seen in the 18 months preceding the time of the creation of the mirror (December 18, 2011 to June 18, 2013 and December 3, 2012 to June 3, 2014). For the fee for service clinics the 18 month period used was December 3, 2012 to June 3, 2014.

We received ethics approval for this study from the WRHA and the University of Manitoba Research Ethics Board.

Participants
Primary care clinics in Manitoba were recruited based on their use of the Accuro® EMR for at least 18 months. Clinics were first recruited based on being part of the WRHA, but recruitment was then expanded to include fee for service primary care clinics not part of the WRHA. We visited each clinic and obtained consent from primary care providers, including physicians and nurse practitioners. A total of 12 clinics and 96 practices with approximately 48,184 active patients participated in the study.

Outcome measure
Two queries were created; one that determined the patients billed for a certain chronic disease visit, and a second that identified those patients who had had that disease recorded in their problem list in addition to being billed for a disease visit. We have assumed that if a patient is billed for a chronic disease visit then they actually have the disease.

Analysis
For each disease, the data from every provider's practice within a clinic was added up to determine the clinic totals. The collected aggregate data was then used to determine EMR problem list completeness by calculating sensitivity (i.e., the capacity of the test to identify the disease when it is present (Zhou, et al., 2011)) (Bowen and Lau, 2012). The number of patients billed for a disease visit was used as the denominator, while the number of patients billed who also had the disease coded in their problem list was used as the numerator. We are therefore evaluating the probability that an EMR problem list will indicate a patient has a disease among those patients who actually have the disease (i.e., the sensitivity of the EMR problem list). This was calculated for each of the seven diseases within each of the 12 clinics. Mean sensitivities among all the clinics were calculated for each disease.

Clinics were also grouped into four categories based on their type: academic teaching clinics, access centre clinics, community clinics, and fee for service clinics. Academic teaching clinics are those which primarily serve as training sites for family medicine residents. Access centres are primary care facilities operated by the WRHA that integrate health and social services within certain higher need Winnipeg communities (Winnipeg Regional Health Authority, 2014). Community Clinics are primary care clinics within Winnipeg which host inter-professional teams that are funded by the WRHA but have independent administration. Clinicians working in Academic Teaching sites, Access Centres and Community Clinics are salaried or block funded, however providers at these clinics still record billing data in the EMR. Fee for service clinics are funded directly through physician billing for visits by Manitoba's single payer insurance system administered by Manitoba's Ministry of Health. Mean sensitivities for all seven diseases were calculated for each clinic type.

One way analysis of variance (ANOVA) was conducted for each disease comparing the mean sensitivities among the four clinic types. A post-hoc contrast analysis was performed for each disease to determine if there is a significant difference between completeness at salaried WRHA clinics (teaching, access, and community) and non-WRHA clinics (fee for service). Statistical significance was set at $p \leq 0.05$. All statistical analysis was performed in Microsoft® Excel® 2008 for Mac.

Disease prevalence rates at every clinic were calculated to examine the relationship, if any, between problem list completeness and prevalence. Since we have assumed that if a patient is billed for a disease visit it means they have the disease, the number of patients billed for a visit within the appropriate 18 month period was used as the numerator in the calculation. The denominator was the total number of patients that had been seen for an appointment at the clinic within the same 18 month period. Therefore, this calculation represents the disease prevalence rate within a clinic's active patient population. The mean prevalence rates were calculated for all 12 clinics, as well as among clinic types.

Results

Individual Clinics
Table 1 presents problem list completeness of the seven chronic diseases evaluated in this study for individual clinics, as well as the mean (standard deviation, SD) sensitivity for every disease. Percentages in bolded italics indicate the clinic with the lowest sensitivity for that particular disease, while those bolded indicate the highest. Clinic 7 has the highest problem list completeness most frequently, while clinic 11 has the lowest most frequently. Figure 1 shows the mean sensitivities of each disease; hypertension and diabetes are the highest at 74.47

(7.74)% and 75.20 (14.60)%, while COPD and CHF are the lowest at 47.93 (20.03)% and 46.50 (19.21)%, respectively.

Table 2 shows the disease prevalence rates at every clinic, as well as the mean (SD) prevalence rates. Hypertension and diabetes consistently have the highest prevalence rates at all clinics, while COPD, CHF, and CAD generally are the lowest. Figure 2 presents mean disease prevalence rates for all 12 clinics.

Table 5 presents the Accuro® EMR "go-live" dates for each of the 12 clinics. These are the dates that the providers at a particular clinic began using the Accuro® EMR to record patient information. The dates are variable ranging from August 2009 to December 2012, with most occurring in 2011. This table also shows the total number of active patients at each clinic, i.e. the total number of patients who have had an appointment at the clinic within the appropriate 18 month time period.

Clinic Types

Figure 3 presents mean problem list completeness by clinic type for each disease. As the figure demonstrates, the fee for service clinics generally have lower completeness than the other three clinic types, while for the most part the community clinics have higher completeness. The exceptions to this include hypertension, for which the teaching clinics have the lowest completeness and diabetes for which the access centres have the highest.

Figure 4 presents mean disease prevalence rates among the four clinic types. Diabetes and hypertension have the highest prevalence rates for each clinic type, and COPD and CHF have the lowest. However, CHF and CAD, not COPD have the lowest prevalence rates at the access centres.

Table 3 presents results from the one way ANOVA analysis, which revealed that the differences in mean sensitivities among clinic types were significant for, hypertension, $F(3,8)= 4.05$, $p= 0.050$, hypothyroidism, $F(3,8)= 36.8$, $p= 0.0000498$, asthma, $F(3,8)= 5.49$, $p= 0.0241$, and CAD, $F(3.8)= 6.48$, $p= 0.0156$.

Table 4 presents results from post-hoc contrast analyses conducted comparing salaried WRHA clinics (teaching, access, and community) to non-WRHA clinics (fee for service). Results demonstrated that problem list completeness is significantly lower at non-WRHA clinics for, hypothyroidism, $F(1,8)= 108.9$, $p= 0.00000616$, asthma, $F(1,8)= 7.77$, $p= 0.0236$, COPD, $F(1,8)= 7.54$, $p= 0.0252$, and CAD, $F(1,8)= 13.9$, $p= 0.00576$.

Discussion

To the best of our knowledge this is the first large scale study reporting EMR problem list completeness for this range of diseases within the Canadian context. This study has established the baseline problem list completeness for certain chronic disease diagnoses within primary care databases in Manitoba. As we hypothesized quality is poor and requires improvements before data can be used reliably for secondary purposes.

Few studies in Canada have reported sensitivities for the diseases investigated in the current study. However, the mean sensitivities in this study were much lower than those found in previous UK research. Hassey et al. (2001) reported sensitivities for hypertension, diabetes, hypothyroidism, asthma, and CAD at 97.8%, 98.3%, 82.1%, 87.3%, and 95.5% respectively.

Faulconer & de Lusignan (2004), report COPD sensitivity as 79%. This further highlights the urgent need for EMR data quality improvement in Canada.

Completeness vs. Prevalence
Problem list completeness was compared to disease prevalence among the 12 individual clinics as well as the four clinic types. There is a minor association between EMR problem list completeness and disease prevalence rates for certain diseases among the 12 clinics. The data suggests that diseases with higher prevalence rates have higher problem list completeness, while those with lower prevalence rates have lower completeness. This holds true for hypertension, diabetes, COPD, and CHF. However, when evaluating hypothyroidism, asthma, and CAD that association does not apply. Therefore, the association between disease prevalence rate and problem list completeness may be disease specific.

On assessment of completeness and disease prevalence rates among the four clinic types, the pattern was similar but with more exceptions. Diabetes and hypertension have the highest prevalence rates at the teaching clinics, but diabetes and hypothyroidism have the highest completeness. Similarly at the access centres, COPD and CHF have the lowest completeness, but CHF and CAD have the lowest prevalence rates. Also at the fee for service clinics COPD and CHF have the lowest prevalence rates, but hypothyroidism has the lowest completeness.

Therefore at this point in time, no definitive conclusions can be drawn regarding the association between disease prevalence rates and problem list completeness. Hassey et al. (2001) reported completeness (sensitivity) as well as disease prevalence rates, and although they did not comment on it, there does not appear to be any association between the two in their study.

Completeness vs. EMR Go-Live Dates
Problem list completeness was also compared to EMR Go-live dates for the 12 clinics which can be seen in Table 5. The EMR Go-live date is a measure of the length of time a clinic has been using the Accuro® EMR for recording patient information. Clinic 9 was the last to implement the EMR and has therefore been using it for the shortest period of time and never had the lowest problem list completeness. However, clinic 11 has been using the EMR for the second shortest length of time, and had the lowest problem list completeness most frequently. Clinic 10 has been using the EMR for the longest period of time, but never had the highest completeness rate. Clinic 7 implemented in 2011, the most common year of implementation, and had the highest problem list completeness most frequently. It is out of scope of this study to infer why that clinic has better completeness than the other clinics which have been using the EMR for similar lengths of time, but research on this should be done in the future. These results suggest that there is not a definite association between EMR Go-Live dates (i.e. length of time using the EMR) and problem list completeness.

Primary Care Quality Indicators
In general, diabetes and hypertension consistently have the highest problem list completeness within individual clinics as well as among the different clinic types. CAD generally has the next highest completeness but not quite as consistently among the different clinic types (hypothyroidism is higher at both the teaching clinics and access centres). However, the other diseases that are part of the PCQI initiative including asthma and CHF, exhibit variable completeness rates. CHF for the most part has one of the lower completeness rates across individual clinics as well as clinic types. It is promising that three out of five diseases part of the PCQI initiative tend to exhibit higher problem list completeness, however it is still not at the level it should be. Therefore, in order for these indicators to function to their full potential in helping

providers track management and screening of chronic diseases, there needs to be improvements in data quality.

Clinic Types

The data from this study suggests that EMR problem list completeness is worse at non-WRHA (fee for service) clinics compared to the salaried WRHA clinics (teaching, access, and community), for hypothyroidism, asthma, COPD and CAD. One potential contributory element is Chronic Disease Management Tariffs. These allow primary care providers to bill for the annual management of certain chronic diseases, including hypertension, diabetes, asthma, CHF, and CAD (Government of Manitoba, 2014a). This applies to providers at fee for service clinics as they bill directly for visits, while providers at the other clinic types are salary or block funded. These tariffs not only provide incentive for providers to spend more time with patients that have these diseases, but also potentially more time with their EMR records ensuring that information is properly recorded.

COPD and hypothyroidism are the only two diseases in the current study that do not have chronic disease tariff codes, and they have significantly lower problem list completeness at fee for service clinics than the other clinic types. As well, they have the two lowest mean problem list completeness rates at fee for service clinics. This could mean that providers at these clinics are not as focussed on properly recording data related to these diseases, which could potentially affect problem list completeness. However, it is important to note that COPD completeness was among the lowest for every clinic type. As well, asthma and CAD completeness were also significantly lower for fee for service than the WRHA clinics and they do have tariff codes, so other factors most likely play a role. Further research delving into the reasons for problem list completeness variability among different diseases as well as different clinic types will be necessary in the future.

Limitations

Limitations for this study include an issue with using billing data from EMRs to calculate sensitivities and disease prevalence rates. This is due to the fact that billing data may contribute to both under and over reporting of patients billed for disease visits. For example, if a patient has both hypertension and diabetes and they come in for an appointment regarding both, a provider can only bill for one, which would cause under reporting of the other disease. If a patient comes in for a diabetes screening visit, the provider will bill the appointment as a diabetes visit. However, because the the patient does not yet have a diagnosis, this may lead to over reporting. Nevertheless, since both under and over reporting can potentially occur, it will most likely balance out having little effect on the study results.

In addition to problems with billing data, the WRHA EMR database was frozen in time at June 18, 2014 and was therefore not up to date. We were unable to collect data entered into the database after that date as it was not feasible to continually have Manitoba eHealth refresh the database.

Future Research

As previously mentioned, future research will need to be conducted to determine what clinic factors contribute to better or worse problem list completeness and EMR data quality in general. However, this will be a challenge as there are certain elements that will be difficult to develop solutions for. For example, at teaching clinics there are many residents trained on a constant basis, sometimes for only a couple of months. It could potentially be difficult to properly ensure that they are reporting data of high quality when they are there for a short time and may be unfamiliar with the EMR.

As well, interventions must be established to improve EMR data quality once the baseline level has been determined. This is beginning to be done within Canada. Greiver et al. (2011) implemented an intervention utilizing a data clerk to re-enter and clean up data of poor quality in the EMR. Data quality was assessed before and after the intervention, and they found that quality improved.

Conclusion

This study has identified the baseline problem list completeness of seven chronic diseases within primary care QHR Accuro® EMR databases throughout Manitoba. We found that data quality is below expectation and must be improved before the databases can be used reliably and easily for secondary purposes and EMR benefits can be fully realized. As well, there may be some differences in data quality between various clinic types, and this must be researched further in order to establish meaningful interventions that will lead to improvements. This study contributes to EMR research within the Canadian context and demonstrates the need for more work to be done to improve EMR data quality.

References

Bowen, M. (2012). EMR data quality evaluation guide, version 1.0. Retrieved from University of Victoria eHealth Observatory website: http://ehealth.uvic.ca/resources/tools/EMRsystem/EMRsystem.php

Bowen, M. & Lau, F. (2012). Defining and evaluating electronic medical record data quality within the Canadian context. *Electronic Healthcare, 11*(1), e5-e13.

Birtwhistle, R., Keshavjee, K., Lambert-Lanning A., Godwin, M., Greiver, M., Manca, D., & Lagace C. (2009). Builidng a pan-Canadian primary care sentinel surveillance network: Initial development and moving forward. *Journal of the American Board of Family Medicine, 22*(4), 412-422.

Byrd, J.B., Vigen, R., Plomondon, M.E., Rumsfeld, J.S., Box, T.L., Fihn, S.D., & Maddox, T.M. (2013). Data quality of an electronic health record tool to support VA cardiac catheterization laboratory quality improvement: the VA clinical assessment, reporting, and tracking system for cath labs (CART) program. *American Health Journal, 165*(3), 434-440.

Canada Health Infoway. (2012). Benefits to Clinicians. Retrieved from https://www.infoway-inforoute.ca/index.php/progress-in-canada/knowing-is-better/knowing-is-better-for-clinicians/benefits-to-clinicians

Canada Health Infoway. (2012b). *Canada Health Infoway Annual Report* 2011-2012. Retrieved from: https://www.infoway-inforoute.ca/index.php/resources/infoway-corporate/annual-reports?view=docman

Canada Health Infoway. (2010). Manitoba launches electronic medical record program. Retrieved from https://www.infoway-inforoute.ca/index.php/news-media/2010-news-releases/manitoba-launches-electronic-medical-record-program

Chan, K.S., Fowles, J.B., & Weiner, J.P. (2010). Electronic health records and reliability and

validity of quality measures: a review of the literature. *Medical Care Research and Review.*

Dermer, M., & Morgan, M. (2010). Certification of primary care electronic medical records: lessons learned from Canada. *Journal of healthcare information management: JHIM*, *24*(3), 49-55.

Faulconer, E.R., & de Lusignan, S. (2004). An eight-step method for assessing diagnostic data quality in practice: chronic obstructive pulmonary disease as an exemplar. *Informatics in Primary Care, 12*(4)*,* 243-254.

Government of Manitoba. (2014a). Chronic Disease Management (CDM) Tariffs. Retrieved from http://www.gov.mb.ca/health/primarycare/providers/chronicdisease/cdmtariffs.html

Government of Manitoba. (2014). Physician Integrated Network (PIN): Quality measurement. Retrieved from http://www.gov.mb.ca/health/primarycare/public/pin/qm.html

Greiver, M., Barnsley, J., Aliazadeh, B., Krueger, P., Moineddin, R., Butt, D.A., Dolabchian, E., Jaakkimainen, L., Keshavjee, K., White, D., & Kaplan, D. (2011). Using a data entry clerk to improve data quality in primary care electronic medical records: a pilot study. *Informatics in Primary Care, 19,* 241-250.

Hassey, A., Gerrett, D., & Wilson, A. (2001). A survey of validity and utility of electronic patient records in a general practice. *Bmj, 322*(7299), 1401-1405.

Juran, J.M., & Godfrey, A.B. (1999). Quality handbook. *Republished McGraw-Hill.*

Kadhim-Saleh, A., Green, M., Williamson, T., Hunter, D., & Birtwhistle, R. (2013). Validation of the diagnostic algorithms for 5 chronic conditions in the Canadian Primary Care Sentinel Surveillance Network (CPCSSN): a Kingston Practice-based Research Network (PBRN) report. *The Journal of the American Board of Family Medicine, 26*(2), 159-167.

Lau, F., Kuziemsky, C., Price, M., & Gardner, J. (2010). A review on systematic reviews of health information system studies. *Journal of the American Medical Informatics Association, 17,* 637-645.

Price, M., Lau, F., & Lai, J. (2011). Measuring EMR adoption: A framework and case study. *Electronic Healthcare, 10*(1), e23-e28.

Price, M., Singer, A., & Kim, J. (2013). Adopting electronic medical records Are they just electronic paper records?. *Canadian Family Physician, 57*(7), e322-e329.

Murray, M., Davies, M., & Boushon, B. (2007). Panel size: how many patients can one doctor manage? *Family Practice Management*, *14*(4), 44-51.

Schoen, C., Osborn, R., Squires, D., Doty, M., Rasmussen, P., Pierson, R., & Applebaum, S. (2012). A survey of primary care doctors in ten countries shows progress in use of health information technology, less in other areas. *Health Affairs*, *31*(12), 2805-2816.

Shephard, E., Stapley, S., & Hamiltion, W. (2011). The use of electronic databases in primary care research. *Family Practice, 28*(4), 352-354.

Tate, A.R., Beloff, N., Al-Radwan, B., Wickson, J., Puri, S., Williams, T., Van Staa, T., & Bleach, A. (2014). Exploiting the potential of large databases of electronic health records for research using rapid search algorithms and intuitive query interface. *Journal of American Medical Informatics Association, 21*(2), 292-298.

Thiru, K., Hassey, A., & Sullivan F. (2003). Systematic review of scope and quality of electronic patient record data in primary care, *Bmj, 326*(7398), 1070.

Wang, R.Y. & Strong, D.M. (1996). Beyond accuracy: what data quality means to data consumers. *Journal of Management Information Systems*, *12*(4), 5-34.

Weiskopf, N.G., & Weng, C. (2013). Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of the American Medical Informatics Association, 20*(1), 144-151.

Winnipeg Regional Health Authority. (2014). Winnipeg Integrated Services. Retrieved from http://www.wrha.mb.ca/community/wis/index.php.

Zhou, X. H., Obuchowski, N. A., & McClish, D. K. (2011). *Statistical methods in diagnostic medicine* (Vol. 712). John Wiley & Sons.

Appendix A: Queries created in the QHR Accuro® EMR query builder based on the University of Victoria's EMR Data Quality Guide (Bowen, 2012).

To illustrate the queries created, the time frame here is the 18 month period used for data collection in summer 2014. All queries were calculated for each individual provider within a clinic and then added up to determine the clinic totals. All queries utilized ICD-9 coding to represent billing and problem list diagnoses.

i. Total patients who have had an appointment within the appropriate 18 month period (used as the denominator to calculate prevalence rates):
1. Office contains 'Office Name'
2. Office Provider= 'Last name, First name'
3. Patient Status= 'Active'
4. Appointment Date Between 2012-Dec-3 and 2014-Jun-3 (Patient records only)

ii. Hypertension

Total patients billed for hypertension within in the appropriate 18 month period (denominator to calculate sensitivity):
1. Office Contains 'Office Name'
2. Office Provider= 'Last name, First name'
3. Patient Status= 'Active'
4. Bill Date Between 2012-Dec-3 and 2014-June 3 AND Bill Diagnosis Code Starts With '401' (Patient Records Only)*

Total patients billed for hypertension within the appropriate 18 month period and have hypertension recorded in their problem list (numerator to calculate sensitivity and disease prevalence):
1. Office Contains 'Office Name'

2. Office Provider= ‘Last name, First name’
3. Patient Status= ‘Active’
4. Bill Date Between 2012-Dec-3 and 2014-June 3 AND Bill Diagnosis Code Starts With ‘401’ (Patient Records Only)*
5. Diagnosis Diagnosis Code Starts With ‘401’ (Patient Records Only)*

For the remaining diseases rules 1-3 are the same as the hypertension queries so only rules 4 and 5 will be presented to show the ICD-9 codes used for each disease.

iii. Diabetes

4. Bill Date Between 2012-Dec-3 and 2014-June 3 AND Bill Diagnosis Code Starts With ‘250’ (Patient Records Only)
5. Diagnosis Diagnosis Code Starts With ‘250’ (Patient Records Only)

iv. Hypothyroidism

4. Bill Date Between 2012-Dec-3 and 2014-June 3 AND Bill Diagnosis Code Starts With ‘243’ OR Starts With ‘244’ (Patient Records Only)
5. Diagnosis Diagnosis Code Starts With ‘243’ OR Starts With ‘244’ (Patient Records Only)

v. Asthma

4. Bill Date Between 2012-Dec-3 and 2014-June 3 AND Bill Diagnosis Code Starts With ‘493’ (Patient Records Only)
5. Diagnosis Diagnosis Code Starts With ‘493’ (Patient Records Only)

vi. COPD

4. Bill Date Between 2012-Dec-3 and 2014-June 3 AND Bill Diagnosis Code Starts With ‘491’ OR= ‘492’ OR’ ‘494’ OR= ‘496’ (Patient Records Only)
5. Diagnosis Diagnosis Code Starts With ‘491’ OR= ‘492’ OR’ ‘494’ OR= ‘496’ (Patient Records Only)

vii. CHF

4. Bill Date Between 2012-Dec-3 and 2014-June 3 AND Bill Diagnosis Code Starts With ‘428’ (Patient Records Only)
5. Diagnosis Diagnosis Code Starts With ‘428’ (Patient Records Only)

viii. CAD

4. Bill Date Between 2012-Dec-3 and 2014-June 3 AND Bill Diagnosis Code Starts With ‘410’ OR Starts With ‘411’ OR Starts With ‘412’ OR Starts With ‘413’ OR Starts With ‘414’ (Patient Records Only)
5. Diagnosis Diagnosis Code Starts With ‘410’ OR Starts With ‘411’ OR Starts With ‘412’ OR Starts With ‘413’ OR Starts With ‘414’ (Patient Records Only)

*For clinic 10 the code ‘405’ was also added as that was the ICD-9 code most providers used for hypertension at that clinic.

Table 1. EMR problem list completeness (%) for seven chronic diseases at 12 primary care clinics in Manitoba. *Standard deviation.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Mean (SD*)**
Hypertension	76.67	63.73	***63.19***	69.62	74.53	78.76	**87.29**	83.36	80.12	66.19	71.54	78.58	74.47 (7.74)
Diabetes	83.98	74.45	64.54	84.08	70.20	89.43	61.94	87.31	**89.95**	***42.94***	65.33	88.20	75.20 (14.60)
Hypothyroidism	73.91	62.05	77.78	65.77	75.00	**80.37**	76.40	79.80	78.46	28.45	***15.89***	32.18	62.17 (23.06)
Asthma	67.56	50.26	55.88	50.33	70.33	53.05	75.00	**80.95**	75.71	45.16	***36.50***	59.69	60.04 (13.82)
COPD	62.30	40.48	41.67	49.21	60.38	37.23	**83.33**	58.73	61.36	***8.33***	21.21	50.94	47.93 (20.03)
CHF	55.77	50.00	46.15	50.00	53.85	41.18	**75.00**	64.52	38.00	21.05	***3.45***	58.97	46.50 (19.21)
CAD	74.07	58.51	71.43	66.67	68.00	63.92	**87.50**	85.39	66.32	51.39	***47.50***	57.68	66.53 (12.13)

Table 2. Disease prevalence rate (%) for seven chronic diseases at 12 primary care clinics in Manitoba. *Standard deviation.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Mean (SD*)**
Hypertension	16.02	21.92	8.38	13.64	12.23	16.93	12.67	15.19	16.04	9.37	18.63	21.69	15.23 (4.28)
Diabetes	4.91	9.37	8.87	7.44	5.01	14.45	8.32	7.21	6.87	4.38	5.52	6.41	7.40 (2.74)
Hypothyroidism	3.97	5.02	1.70	3.53	2.79	3.41	4.78	2.66	4.27	1.41	3.94	4.62	3.51 (1.17)
Asthma	4.31	4.97	4.28	3.58	3.02	5.22	3.87	2.26	4.60	1.51	5.04	5.07	3.98 (1.19)
COPD	1.17	1.08	0.76	1.49	1.76	2.99	0.97	1.69	1.45	0.87	1.22	1.41	1.40 (0.59)
CHF	1.00	0.62	0.82	0.99	0.43	1.08	0.43	0.83	1.64	0.46	1.07	1.55	0.91 (0.40)
CAD	2.07	2.42	0.88	1.21	0.83	3.09	0.43	0.43	3.12	2.39	1.47	4.24	1.92 (1.17)

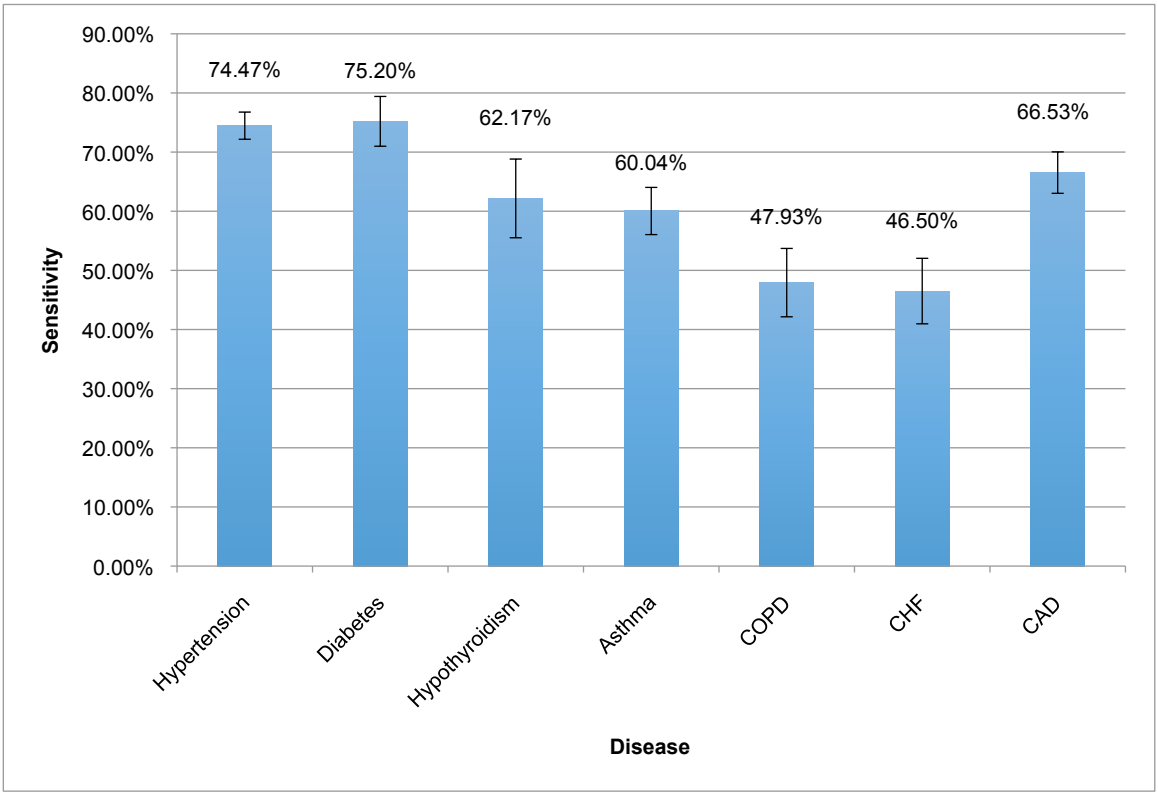


Figure 1. Mean EMR problem list completeness of seven chronic diseases from 12 Manitoba primary care clinics. Error bars represent standard error of the mean.

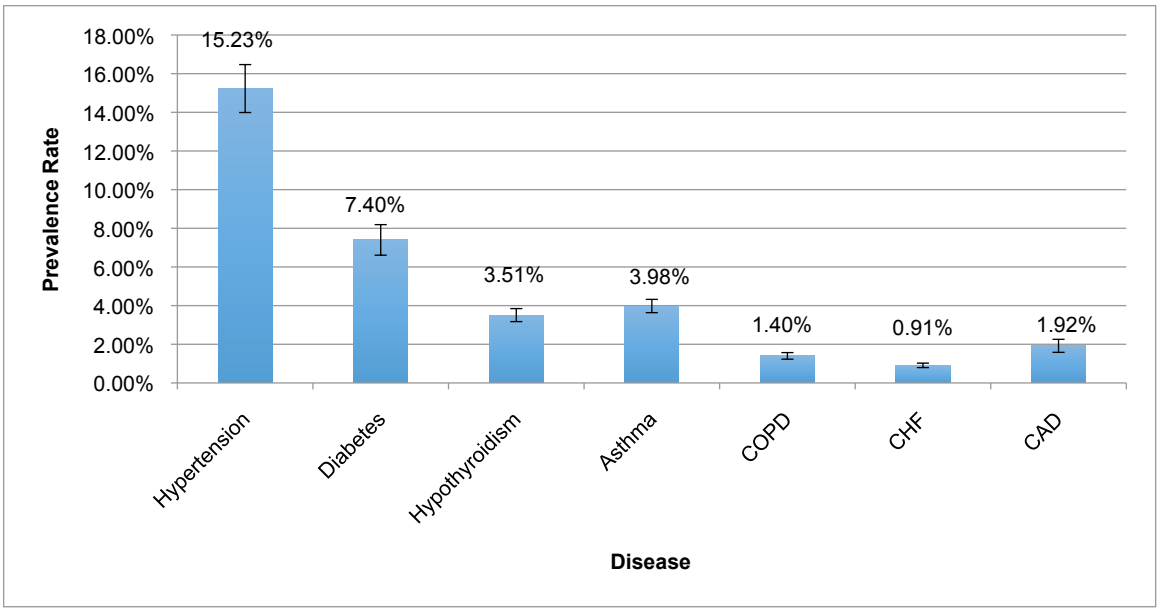
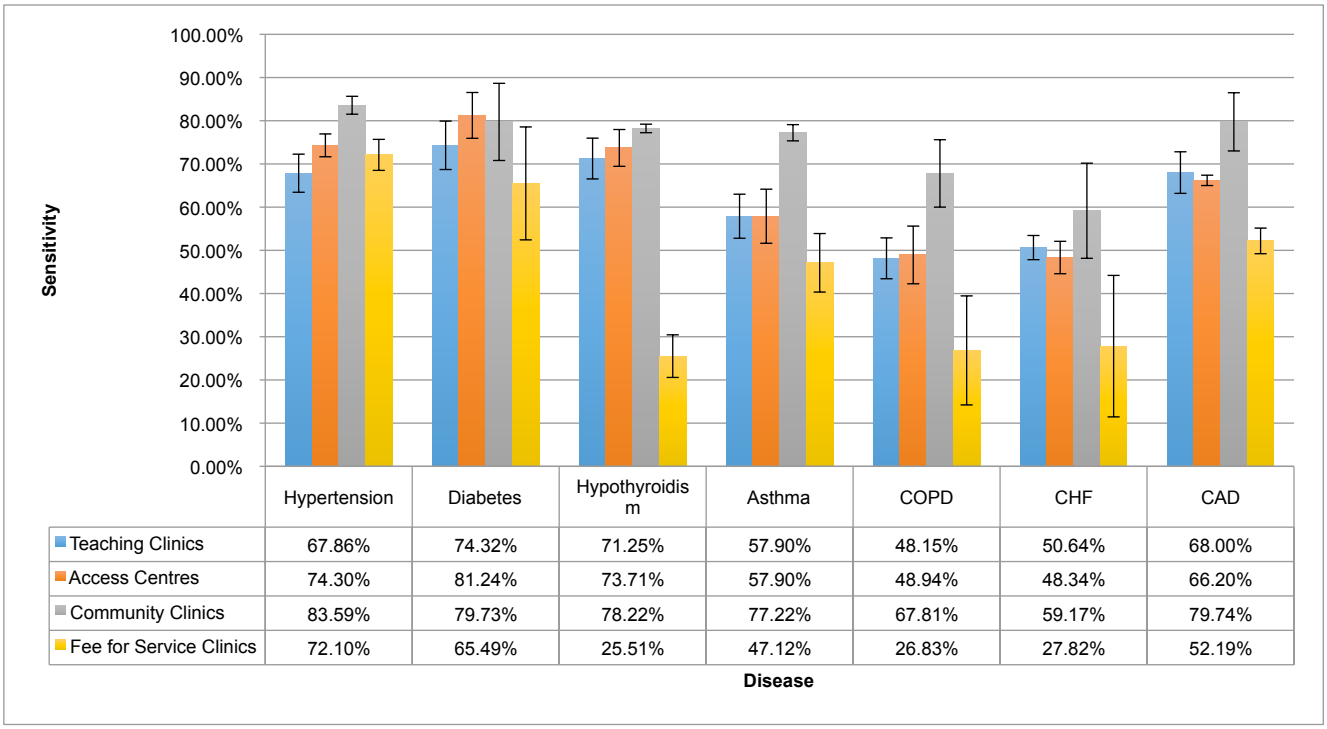
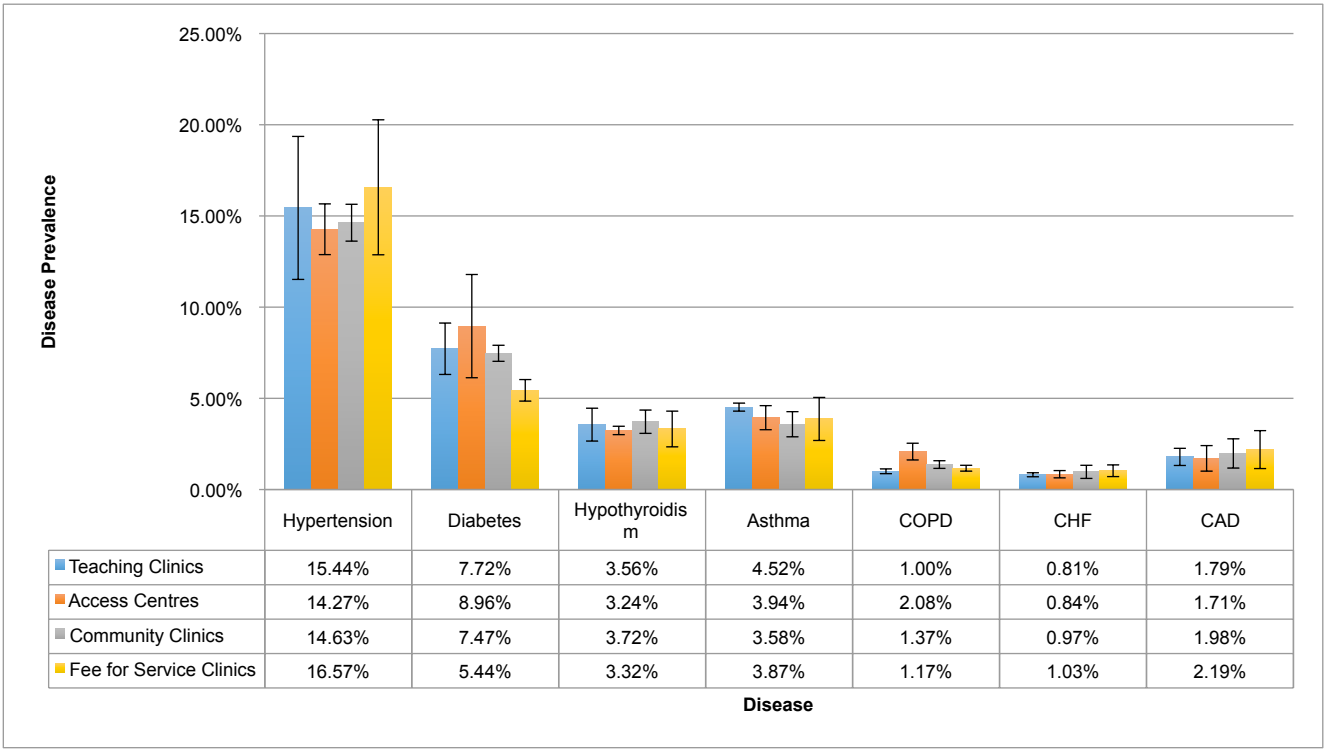


Figure 2. Mean disease prevalence rates of seven chronic diseases from 12 Manitoba primary care clinics. Error bars represent standard error of the mean.



	Hypertension	Diabetes	Hypothyroidism	Asthma	COPD	CHF	CAD
Teaching Clinics	67.86%	74.32%	71.25%	57.90%	48.15%	50.64%	68.00%
Access Centres	74.30%	81.24%	73.71%	57.90%	48.94%	48.34%	66.20%
Community Clinics	83.59%	79.73%	78.22%	77.22%	67.81%	59.17%	79.74%
Fee for Service Clinics	72.10%	65.49%	25.51%	47.12%	26.83%	27.82%	52.19%

Figure 3. Mean EMR problem list completeness of seven chronic diseases from 12 Manitoba primary care clinics. Error bars represent standard error of the mean.



	Hypertension	Diabetes	Hypothyroidism	Asthma	COPD	CHF	CAD
Teaching Clinics	15.44%	7.72%	3.56%	4.52%	1.00%	0.81%	1.79%
Access Centres	14.27%	8.96%	3.24%	3.94%	2.08%	0.84%	1.71%
Community Clinics	14.63%	7.47%	3.72%	3.58%	1.37%	0.97%	1.98%
Fee for Service Clinics	16.57%	5.44%	3.32%	3.87%	1.17%	1.03%	2.19%

Figure 4. Mean disease prevalence by primary care clinic type in Manitoba. Error bars represent standard error of the mean.

Table 3. One way ANOVA results comparing differences in mean problem list completeness of seven chronic diseases between four primary care clinic types in Manitoba.

	df_{clinic}, df_{error}	*F*	*p*-value
Hypertension	3, 8	4.05	0.050*
Diabetes	3, 8	0.644	0.608
Hypothyroidism	3, 8	36.8	0.0000498**
Asthma	3, 8	5.49	0.0241*
COPD	3, 8	3.56	0.0669
CHF	3, 8	1.72	0.240
CAD	3, 8	6.48	0.0156*
Note: *p≤0.05, **p≤0.001 *df*= degrees of freedom			

Table 4. Contrast analysis results comparing mean problem list completeness of seven chronic diseases between WRHA vs. non-WRHA primary care clinics in Manitoba.

	df_{clinic}, df_{error}	*F*	*p*-value
Hypertension	1, 8	0.683	0.432
Diabetes	1, 8	1.60	0.242
Hypothyroidism	1, 8	108.9	0.00000616**
Asthma	1, 8	7.77	0.0236*
COPD	1, 8	7.54	0.0252*
CHF	1, 8	4.52	0.0662
CAD	1, 8	13.9	0.00576*
Note: *p≤0.05, **p≤0.001 *df*= degrees of freedom			

Table 5. Accuro® EMR go-live dates for 12 primary care clinics in Manitoba and the total number of active patients at each clinic.

Clinic	**Accuro® EMR Go-live date**	**Total Active Patients**
1	December 2011	5218
2	November 2011	3886
3	March 2011	1589
4	May 2011	4223
5	June 2011	3016
6	Fall 2010	3142
7	November 2011	1862
8	December 2011	3719
9	December 2012	3043
10	August 2009	8238
11	March 2012	2716
12	August 2010	7532