

Article details: 2017-0029	
Title	A cross-sectional study on self-reported versus health administrative data: implications for assessing chronic illness burden in populations
Authors	Martin Fortin MD MSc, Jeannie Haggerty PhD, Steven Sanche MSc, José Almirall MD PhD
Reviewer 1	Mr. Michelle Rotermann
Institution	Health Information and Research Division, Statistics Canada, Ottawa, Ont.
General comments (author response in bold)	<p>Thanks very much for the opportunity to review this paper. I think the subject of multi-morbidity is of growing importance and contributions to the literature about data quality of administrative and/or survey data, especially less well-known data sources are always needed.</p> <p>This paper is well-written and concise but could be improved by adding more detail to the Methods and Discussion sections in particular. I have a number of questions/comments which I feel could help to make this paper stronger and more transparent.</p> <p>What age group does the study population include? Line 46, pg 4 it says 25 to 75; Table 1 says 25 to 75 but Abstract pg 3, Ln 13 says 45 or older? This was a mistake in the Abstract and it was corrected.</p> <p>Authors indicate that they are interested in estimating the prevalence of multi-morbidity in the general population. It is not clear from the information provided about the study design if results from the sample are generalizable to the Quebec population? E.g. 4 networks in one geographic area of Quebec using random digit dialing was used. Is this sufficient? We know from other research that response rates are falling and people are cord-cutting (disconnecting their landlines). Were cellular telephones included in the phone bank using for RDD? This is a very good point to clarify. Our intention was not to report on the prevalence of multimorbidity in the general population but to compare the prevalence obtained with two different methods. We acknowledge that it was confusing and we changed the wording in the Introduction. Estimation of the prevalence of multimorbidity would have required a random sample of the Quebec population which was not a requirement to compare the two methods like we did. As for the landline vs. portable telephones, we limited the study to landlines because we believed that it was a valid method for the age group targeted (25-75) in 2010. Inclusion of portable phones may have resulted in recruiting younger people that could have affected prevalence, but again, this was not the focus of this research.</p> <p>Because this is a less well-known dataset I think more explanation is needed about the research design, sample selection, response rate, potential sources of bias, whether respondents were reimbursed? Looking for a balance between the information provided and the limitation in the final word count, we estimated that information about the research design, sample selection, and response rate covered by the published research protocol (reference 7) and the current paper would be enough. However, we added more potential sources of bias in the Discussion.</p> <p>A link to the questionnaire would be useful as research on comparability of self-reported and administrative health data has regularly shown that question wording matters. As mentioned above, the questionnaire was included as an Appendix.</p> <p>Provide/confirm response rate? 48.9% 1178/2409 eligible respondents? Was there any difference between people who agreed to participate and those that refused? Also, we find that a proportion of people who are willing to participate in health surveys do not consent to linking their data to administrative records. Did respondents have to agree to both completing the questionnaire and linking to admin. Records to participate? As mentioned above, a total of 2,409 eligible respondents provided verbal consent to participate, but 1718 responded and completed the questionnaire. No data could be collected from the 691 subjects who provided consent but never responded.</p> <p>This work is part of a broader research study. Participants in the research study were free to consent or not to linking their data to administrative records. Data from the 1,178 participants who provided consent were included in the current work. The 540 subjects who did not provide consent were not overly dissimilar from those who provided consent. They were only different in regards to educational level.</p> <p>Data were collected in 2010. Are analyses based on these data still relevant? Prevalence of the individual conditions may have changed and we added this limitation to the Discussion. However, the main object of this study is the analysis of the agreement, and it is still relevant because little has changed in the system for the collection of administrative data.</p> <p>Research on the comparability of self-reported and administrative health data also shows that the look back period can affect case ascertainment. How was the 2 year look period selected? What are the advantages/disadvantage of choosing a relatively short look-back period? Would some conditions, particularly those more episodic in nature be expected to have worse agreement? As mentioned above, we considered that a two-year window would provide a good view of active chronic conditions. Unfortunately, we cannot expand this window now. We discussed, in the limitations of the study, the possibility that the use of additional years of administrative data could have increased the agreement between the two sources.</p> <p>We also addressed in the paper the probability of patients recalling the conditions for which they received medication and/or diet for treating it, whereas, within the constraints in the information that can be provided, physicians choose to report diagnoses with a higher level of disease burden. The episodic nature of some conditions does not seem to be an important element affecting agreement.</p> <p>Not familiar with Quebec health care system. In other provinces some physician's salaried, other fee-for-service and some both. Are all physician's fee-for-service in Quebec? It is not explicitly stated which types of physician consultations are captured by the fee-for-service payments? Office visits? ER? Day Surgery? In 2010, a great majority of physicians were remunerated by fee-for-service. Office visits, day surgery and ER visits were captured by the fee-for-services database.</p> <p>From reviewing other papers based on this dataset it seems that both prescription drug data were available and that information on more than 12 chronic conditions (CCs) was captured. - How were the 12 CCs included in this study selected?</p>

- What was the rationale for not including prescription data? The low level of agreement between some of your CCs and admin data would likely have been resolved by including prescription drugs e.g. statins for the treatment of high cholesterol. From the original questionnaire presented in the appendix, we limited our report to 12 diagnoses as some referred to symptoms or were judged too vague and difficult to match with HA data codes. We also combined under the term **"Any cardiac disease" two diagnoses referring to heart conditions. The list of original conditions and the amended list, along with the diagnostic codes, are provided in the appendix.** Prescription drug data in Quebec is only available for patients 65 years and older and those on welfare assistance. We could not use it for the whole sample. We agree that it could have helped identify some conditions from the HA Data. It may also have caused some confusion given that a medication may be indicated for several conditions.

Embedded in this study is a linkage between survey respondents and some administrative health data. Was this a deterministic linkage = exact match based on health card number? As one might expect (Table 2), not everyone who self-reported having a chronic condition had received treatment i.e. linked to at least one health record. Were missed links possible with this linkage project? **There was an exact match based on the health card number. We presumed that patients who declared having a chronic condition, likely had contact with the healthcare system within the two-year window to receive treatment for that condition.**

Other case ascertainment studies have required persons to be physically present and eligible for health coverage in the province during the entire study period. Was this a condition for study participation? **Although this was not explicitly considered or mentioned as a condition, we received valid data from the RAMQ showing that coverage was present for the whole study period.**

Results section. Suggest removing the good, very good adjective from your description of Kappas. The Kappa values are sufficient. Could include these descriptors in Discussion. **We removed the adjectives from the description of Kappas.**

More clarification needed. Pg 8 Ln 3-6. The author's state that participants more likely to be identified as multimorbid based on self-report than health administrative data. I wonder if it would be more accurate to specify the precise admin data as I suspect if drug, lab and other admin health sources were used the agreement would have been better. **To clarify that point, we expanded the discussion with a sentence addressing this comment.**

Agreement between admin and survey data for chronic conditions is known to vary as a function of type of data, # of years of data, # of contacts the patient had and patient characteristics, including age, sex, geography, rural/urban, SES. Wonder if it would be appropriate to explore agreement rates by different characteristics than overall. It is very possible that for some patients the agreements between self-report and admin data used might be better than the overall numbers. This would be important information to have. **We agree that this would be important information, but we believe that the results of such analysis could yield enough data for another paper. We take note of this suggestion for future work with a larger sample.**

Awkward sentence: pg 7 Line 6: "We could collect" ? Remove could. **It was removed.**

Unclear re: primary care. Pg 8. Ln 18. Authors use fee-for-service, inpatient hospital data and self-reports of chronic conditions and note **that agreement on cholesterol low because it is managed in primary care. Aren't these data included? Perhaps what the authors are trying to say is that high cholesterol is rarely used as the single diagnosis code for ambulatory care visits? Could the authors provide a list of the leading ambulatory care visit codes for people aged 25 or 45 to 75?** **Thanks for suggestion for rewording.** **As for the list of leading ambulatory care visit codes, we agree that it could have shed light on the results, but without changing them, we decided not to explore that.**

Pg 9. Reference #5 citation about accuracy of health administrative data. Add a Canadian reference. Perhaps one from the Canadian Institute for Health Information (CIHI); more relevant than one from Australia. **We could not find any Canadian study that could back the possibility of between-hospital variability on recording health conditions in routine hospital data as a possible factor in the accuracy of health administrative data. This element probably does not exist at all in our health system. We decided to remove the sentence, although we used the reference elsewhere about the agreement between both methods.**

Table 1. Age group counts add to 1118 but sample =1173. Error? **We checked the Table 1 and the counts add to 1,177 which is the final sample.**

Reviewer 2	Dr. Ruth Hall
Institution	Institute for Clinical Evaluative Sciences
General comments (author response in bold)	<p>The authors report on the use of administrative databases to identify multimorbidity (defined as 2 or more of 12 chronic conditions) and compare the prevalence of multimorbidity through self-report among a cohort of adults 45 years of age and older in from four regions in Quebec. They found low agreement between admin based vs self-report for 11/12 conditions. This is a well-written paper</p> <p>The study was nested within a larger research project (PRECISE) a brief summary would be helpful. Is there more literature comparing self-report to admin data to identify comorbid conditions? On page 4 line 25 there are only 2 references listed. We added a brief summary of the PRECISE program to the Methods section. Regarding the available literature, indeed, none of the studies comparing self-report data and administrative data intended to identify comorbid conditions in participants. We did not find any previous study on the subject. All publications that we found studied the agreement between both sources for individual diagnosis but not the occurrence of comorbid conditions in the same individual. The only study advancing the idea of using administrative data for the surveillance of multimorbidity was the one by Tonelli and colleagues (reference 3), but they did not compare results with self-reported data.</p> <p>A sensitivity analyses among the 65 and older age group may be helpful as this group is more likely to have chronic conditions and have more frequent MD/hospital visits. As mentioned above, we conducted a sensitivity analysis by age, and the Methods and the Results sections were</p>

expanded accordingly.

It would be helpful to provide the average number of MD visits and Hospitalization by age group.

We appreciate the suggestion. It could have shed some light on the results but we decided not to include that to keep the focus on our main message. This is something we consider investigating in the future.

An appendix that includes the survey questions would be informative to the reader.

We added an appendix with the survey questions.

The authors should provide reliability/validity information on the Quebec administrative databases. Eg. How many diagnostic codes are filled out per MD visit. What about the hospitalizations?

Thanks to the reviewer for raising this issue. It is a limitation of the Quebec administrative data not to have any information on reliability/validity as the dataset is mostly for billing (only one diagnostic code is required but the diagnosis may vary from one consultation to another) for the ambulatory part and for administrative use for hospitalizations. To our knowledge there is no reliability/validity study that could have informed our use of the data. This limitation was added to the Discussion.

Can the authors compare among the urban and rural participants?

It could be an interesting analysis and we take note of the suggestion, but we think that such analysis is beyond the scope of this study.

The data is 10 years old and therefore the findings may not be applicable with more recent data.

As mentioned above, prevalence of the individual conditions should have changed and we added this limitation to the Discussion. However, the main object of this study is the analysis of the agreement, and it is still relevant because little has changed in the system for the collection of administrative data. If we applied the same methods to more recent data, the agreement between the two sources would not be very different.

It would be nice to include in the discussion should include how their results compare to the CCHS for residents of Quebec.

We looked for data from the CCHS in the web site of Statistics Canada, but we could not find data about the same age groups that we used. Nevertheless, the data that we found for other age groups were about the prevalence of conditions and this could not be used to discuss the agreement.