Comparing the Validity of Different Sources of Information on Emergency Department Visits

A Latent Class Analysis

Nandini Dendukuri, PhD,‡ Jane McCusker, MD, DrPH,†‡ François Bellavance, PhD,§ Sylvie Cardin, PhD,¶ Josée Verdon, MD, MSc,‖ Igor Karp, MD, MPH,‡ and Éric Belzile, MSc†*

Background: Emergency department (ED) use in Québec may be measured from varied sources, eg, patient's self-reports, hospital medical charts, and provincial health insurance claims databases. Determining the relative validity of each source is complicated because none is a gold standard.

Objective: We sought to compare the validity of different measures of ED use without arbitrarily assuming one is perfect.

Subjects: Data were obtained from a nursing liaison intervention study for frail seniors visiting EDs at 4 university-affiliated hospitals in Montreal.

Measures: The number of ED visits during 2 consecutive follow-up periods of 1 and 4 months after baseline was obtained from patient interviews, from medical charts of participating hospitals, and from the provincial health insurance claims database.

Methods: Latent class analysis was used to estimate the validity of each source. The impact of the following covariates on validity was evaluated: hospital visited, patient's demographic/clinical characteristics, risk of functional decline, nursing liaison intervention, duration of recall, previous ED use, and previous hospitalization.

Results: The patient's self-report was found to be the least accurate (sensitivity: 70%, specificity: 88%). Claims databases had the greatest validity, especially after defining claims made on consecutive days as part of the same ED visit (sensitivity: 98%, specificity: 98%). The validity of the medical chart was intermediate. Lower sensitivity (or under-reporting) on the self-report appeared to be associated with higher age, low comorbidity and shorter length of recall.

Conclusion: The claims database is the most valid method of measuring ED use among seniors in Quebec compared with hospital medical charts and patient-reported use.

Key Words: emergency visits, latent class analysis, validation, self-report, seniors

(*Med Care* 2005;43: 266–275)

Valid measurement of emergency department (ED) use is important for routine monitoring/surveillance purposes as well as for research on determinants of ED use. In Québec, Canada, information on a patient's ED use may be obtained directly from a patient (or an informant), from the physicians' billing database maintained by the provincial health insurance organization (Régie de l'Assurance Maladie du Québec [RAMQ]), or from hospital medical charts. The ease of collecting information directly from a patient is balanced by concerns about validity, particularly among seniors. Several authors comparing patient-reported health services use to data from medical charts[1–3] or insurance claims databases[4–8] have reported that the degree of validity of the self-report is related to the type of service. Visits to an outpatient clinic or physician's office, which are routine in nature, tend to be recalled with the least validity,[1–8] while hospitalizations, which are rare, are recalled with near perfect validity.[3,5,7,8] ED visits appear to be recalled with intermediate validity.[6–8] Frequent users of physician visits have been consistently found to under-report them.[1–3,5,7,8] On the contrary, one study found that patients who reported frequent ED use also had a tendency to over-report use compared with a computerized utilization record.[8]

So far, studies of validity of self-reports have tended to treat medical chart or claims databases as the perfect reference standard. However, many studies list drawbacks of these assumed "gold standards,"[1,4,7,8] similar to problems with the

From the *Technology Assessment Unit, McGill University Health Center; the †Department of Clinical Epidemiology and Community Studies, St. Mary's Hospital; the ‡Department of Epidemiology and Biostatistics, McGill University; the §Department of Management Sciences, HEC Montréal; the ¶Centre de recherche du Centre Hospitalier de l'Université de Montréal; and the ‖McGill University Health Center, Montreal, Canada.

Supported by the grant titled "Groupe interuniversitaire de recherche sur les urgences" (#25300-2700) funded by Fonds de la recherche en Santé du Québec (FRSQ). The first author receives salary support from a Chercheur Boursier Junior 1 award from the FRSQ.

Reprints: Nandini Dendukuri, Technology Assessment Unit, Royal Victoria Hospital, R4.14 Ross Pavilion, 687 Pine Avenue West, Montreal, PQ H3A 1A1, Canada. E-mail: nandini.dendukuri@mcgill.ca.

ISSN: 0025-7079/05/4303-0266

RAMQ claims database and medical charts considered in the current study. In Québec, over 90% of residents were covered by health insurance from RAMQ during the period of this study.[9] The RAMQ claims database records ED claims from all hospitals in Québec making it the most comprehensive of the 3 data sources we consider. A drawback of using the database is that it is not possible to identify separate visits to the same ED if these visits were made on the same day or on consecutive days. This is because multiple claims made for the same patient cannot be linked to a specific visit. Also visits to facilities outside Québec will be missed. Finally, data entry errors could lead to inaccuracies. The third source we evaluate, hospital medical charts, may be the most accurate source of information for measuring ED use at a given hospital. However, these data have to be extracted separately at each hospital that a patient visited. This can be a problem when there are several EDs in the same area. Because chart data are often not computerized, this could be an expensive and cumbersome process. Further, problems with filing could cause some visits to be missed.

Assuming that the reference standard (ie, the medical chart, or insurance claims database) is perfect could seriously bias estimates of the self-report's validity.[10] To avoid this problem we propose using a statistical method called latent class analysis. This approach allows us to simultaneously estimate the validity of multiple imperfect measurements in the absence of a "gold standard" or perfect measurement. It has found widespread use in the analysis of medical diagnostic tests,[11,12] and in other areas, such as psychology and sociology,[25] where "gold standard" measures are usually not available. Our objective is to develop a latent class model to estimate the validity of 3 imperfect sources of ED use (self-report, medical chart and claims database), without arbitrarily considering one source to be a gold standard.

Figure 1 provides a schematic diagram of a simple latent class model. The unknown parameters in the model are represented by ovals and the observed data by rectangles. The arrows indicate which parameters influence each observed variable. Here, ED use is measured using a dichotomous variable taking values "any ED use" or "no ED use" during a certain period. Under this model all 3 observed variables are influenced by the probability of the "latent" variable, namely the true probability of any ED use by a patient. Each observed variable is also influenced by its own sensitivity and specificity, parameters that determine its validity. The sensitivity is defined as the probability of observing ED use given that the patient truly visited an ED during the period of interest. The specificity is defined as the probability of not observing any ED use given that the patient truly did not make any ED visits during the period of interest.

The latent class model in Figure 1 can be extended to compare more than 3 imperfect measures. It can also be extended to deal with situations where the observed variables or the latent variable are ordinal or continuous (ex. ED use measured in number of visits during the period of interest). However, the complexity of the model must be balanced by the number of degrees of freedom available. In Figure 1, where we assume that only dichotomous results are available from all 3 variables, we can observe 8 possible combinations of the data from the 3 sources: ED use according to all 3 sources; ED use according to self-report only; ED use according to medical chart only; ED use according to RAMQ only; ED use according to self-report and medical chart; ED use according to self-report and RAMQ; ED use according medical chart and RAMQ; ED use not reported by any of the 3 sources. Thus, we have 7 degrees of freedom. This means we can estimate a maximum of 7 unknown parameters. In our case we have exactly 7 parameters—sensitivity and specificity of each observed variable and the true probability of any ED use. If we had only 2 observed variables (say, the self-report and the medical chart) then it would not have been possible to estimate all 5 unknown parameters involved in this case as only 3 degrees of freedom would be available.

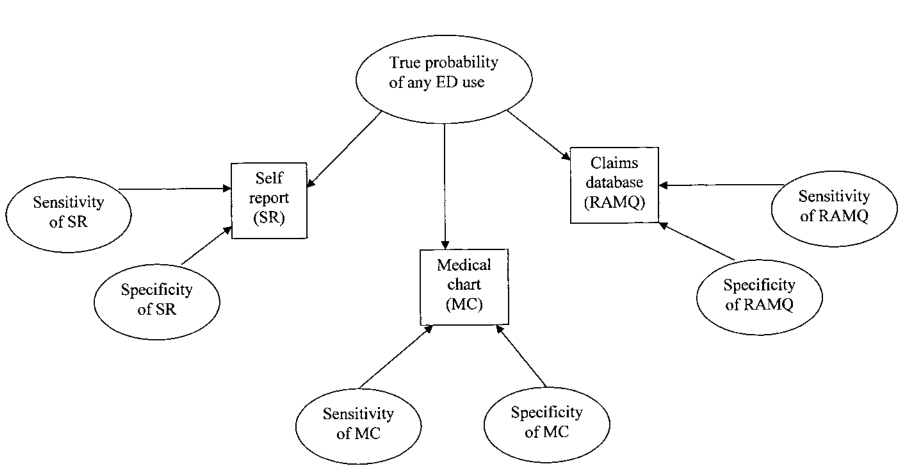

FIGURE 1. Schematic diagram of a latent class model for a binary measure of ED use.

For such situations some prior information would need to be available about some of the unknown parameters.[11,12]

The classic latent class model that we use in the current article makes the important assumption that the 3 sources of information on ED use are independent of each other conditional on knowing the value of the latent variable alone, ie, conditional on whether or not a patient truly had an ED visit. This means if we knew that a patient had truly made an ED visit, the fact that the visit was correctly self-reported does not influence observations made in the medical chart or RAMQ database and vice versa. An example of when conditional independence would be violated is when the following 2 statements are true: (1) if we know that when a person correctly self-reports, they are less likely to be a frequent user (as suggested by the literature[8]); and (2) if being a less frequent user has an effect on administrative or medical chart recording given the person's true ED status.

METHODS

Data for this project were previously collected for a randomized controlled trial of a nursing liaison intervention aimed at reducing functional decline following an ED visit.[13] Patients were recruited from EDs of 4 university-affiliated hospitals in the Montreal area between September 1998 and April 1999. In brief, inclusion criteria were age 65 years or older; stable (noncritical) medical status and orientation to time and place or availability of an informant; English or French speaking; resident in Montreal; not resident in a nursing home; not transferred from another hospital; and not admitted to hospital at the ED visit. Patients were recruited based on their performance on the Identification of Seniors at Risk (ISAR) screening tool for functional decline.[14] A score of 2–6 on this 6-item questionnaire predicts increased risk of functional decline as well as more ED visits during follow-up[15,16] as measured by the RAMQ database. All patients who scored 2–6 on the ISAR were randomized by day of visit to receive the intervention or usual care. A random sample (8%) of those who scored 0–1 was also recruited. Patient consent was obtained to extract data on health services use from RAMQ databases and medical charts during the course of the study. The different sources were linked together using a unique identification number created by us for each patient. Ethics committees at the 4 participating hospitals approved the study. For the current analyses we further excluded those patients who were missing information from any of the 3 sources on the number of ED visits during follow-up or the hospital at which these EDs were located. Further, we only included patient visits made to the 4 participating hospitals.

Patient's Self-Report

During telephone interviews with the patient 1 and 4 months after baseline information on the number of repeat ED visits in the interim was obtained. Patients were asked "Since your visit to the emergency department on (date of visit), have you visited any hospital emergency (room) or been admitted to hospital?" We considered ED visits during which the patient was either admitted to hospital or discharged home. If the patient reported having visited an emergency department they were asked the name of the hospital and whether they were admitted to the hospital at that visit. When the patient was impaired an informant was questioned. In the event that the patient was contacted at the 4- but not the 1-month follow-up the total number of ED visits between baseline and the 4-month interview was recorded. 75% of reported visits were to the 4 participating hospitals.

RAMQ Claims Database

Data were extracted from this source for each patient between the dates of the patient's baseline and four-month interviews. Only visits made to the 4 hospitals participating in the study were considered. Individual ED visits were defined in 2 ways using this database: (1) RAMQ-SEQ: a sequence of claims made on consecutive days constituted a single visit; and (2) RAMQ-DAY: each day on which a claim was made was considered to be a separate visit.

Hospital Medical Chart

Details of each ED visit between the dates of the patient's baseline and 4-month interviews were extracted from medical charts at the 4 participating hospitals.

Covariates

Patient's demographic characteristics (age, sex) and clinical characteristics (cognitive impairment, comorbidity, depression) were measured at recruitment. Cognitive impairment was defined as a score of 7 or more on the Blessed Orientation, Memory and Concentration (BOMC) scale.[17,18] High comorbidity was defined as a score of 12 or more on a self-report comorbidity questionnaire that was validated against the chart-based Charlson comorbidity index.[19] Depression was defined as a score of 6 or more on the 15-item Geriatric Depression Scale (GDS).[20,21] Previous ED users were identified as those who had made at least one visit to any ED during the month prior to the index ED visit. Previous hospitalization was defined as any hospital discharge in the year prior to the index ED visit. This information was obtained from the Med-Echo database on hospitalizations maintained by the provincial Ministry of Health and Social Services.

Statistical Analysis

Two separate latent class models were used to estimate the validity of the different sources in measuring: 1) any ED use (Model 1), and 2) the number of ED visits (Model 2). The models are described in greater detail in the Appendix. The number of ED visits was treated as an ordinal variable having values 0, 1, and > 1. From Model 1 we estimated the validity of each source in terms of sensitivity (probability that the observed number of ED visits > 0 when the true number of ED visits> 0) and specificity (probability that observed number of ED visits was 0 when the true number of ED visits was 0). From Model 2 we were able to further separate the probability of over- and underestimation by the volume of the true number of visits. For example, we could separate the probability of overestimation when the true number of visits was 0 from the probability of over-estimation when the true number of visits was 1. Both models were first fit within the 2 samples (ISAR 0–1 and ISAR 2+). Because results were very similar in these 2 groups, the remainder of the analysis was performed in the combined group. In secondary analyses, we fit Model 1 within subsets defined by the following covariates to evaluate their effect on the sensitivity and specificity of each measure: age (< 80 years, 80+ years), sex (male, female), comorbidity (low, high), cognitive impairment (yes, no), depression (yes, no), prior ED use (none, any), prior hospitalization (none, any) and length of recall (1 month, 3 months). We evaluated the effect of the nursing liaison intervention among patients with an ISAR score of 2+. Impact of high ED use during follow-up was not examined as too few patients had multiple visits according to any source. We also did not evaluate the impact of using an informant of the validity of the self-report because of the small percentage of self-reports that were completed by an informant. Secondary analyses of Model 2 were not possible due to small number of subjects with more than 1 visit. A Bayesian approach was used for parameter estimation and inference of the latent class models.[22] WinBUGS[23] programs for implementing these methods can be obtained from the corresponding author. The suitability of the conditional independence assumption was verified using methods suggested by Berkof et al.[24] For each latent class, this method involves calculating the probability that the observed number of visits is greater than its expected value under the hypothesized conditional independence model. Very low or high probabilities would suggest that the observed data are very different from the expected data under the conditional independence model, and therefore the model is not appropriate.

RESULTS

The inclusion criteria were met for 520 patients. Of these, 132 had an ISAR score of 0–1 and 388 had an ISAR score of 2–6. For the purpose of the current analyses, we excluded a further 84 patients with missing information on ED visits during follow-up (81 were missing information on number of visits from either self-report, RAMQ database or medical chart, 3 self-reports were missing information on the hospital where the ED was located). Information from a common time frame was available from all 3 sources for 436 patients.

Table 1 summarizes the distribution of patient characteristics at baseline. The majority of patients were aged between 65 and 79 years, were women and were able to self-complete the questionnaire at baseline. Almost a third of the sample reported visiting the ED in the month prior to the index visit and 41.3% had been hospitalized in the year prior to the ED visit. Patients who were eliminated from the analysis were more likely to be 80+ years of age, have an informant complete the self-report, be at greater risk of functional decline and be cognitively impaired.

TABLE 1. Distribution of Patient Characteristics at Baseline

	Excluded From Study (Total = 84*)		Included in Study (Total = 436)	
	No. Patients	% of Total	No. Patients	% of Total
Age (80+ years)	35	41.7	149	34.2
Female	47	56.0	265	60.8
Self-report completed by informant	14	16.7	42	9.6
ISAR 0–1	17	20.2	115	26.4
ISAR 2–6 intervention†	30	35.7	148	33.9
ISAR 2–6 control	37	44.1	173	39.7
ED use in month before index date (missing observations)	29 (3)	35.8	132	30.3
Hospitalization in year before index date	38	45.2	180	41.3
High comorbidity	40	47.6	211	48.4
Cognitive impairment (missing observations)	41 (12)	56.9	133 (38)	33.4
Depression (missing observations)	18 (15)	26.1	95 (50)	24.6

*Patients without information from any of the 3 sources (self-report, medical chart, and RAMQ claims database) were excluded.
†Intervention was randomized in the ISAR 2–6.
ISAR indicates Identification of Seniors at Risk screening tool; ED, Emergency Department.

For the 436 patients during the 4-month follow-up period there were a total of 190 visits according to patients' self-reports, 267 visits according the medical chart, 290 visits according to the RAMQ-SEQ definition and 496 visits according to the RAMQ-DAY definition. The maximum length of a sequence under the RAMQ-SEQ definition was 8 days, though only 1.5% of the 290 sequences were greater than 3 days. To check whether sequences of greater than 1 day corresponded to multiple visits, we matched the RAMQ database with the hospital medical chart data by date. We found that there were no such instances in our data.

After examining the cross-tabulation of results from the 3 data sources (not shown) during the 4-month follow-up period we decided to pool patients who had more than 2 visits into one category (> 1 visit). Percentage agreement between the 3 sources was 67% when using the RAMQ-SEQ definition, and 58% when using the RAMQ-DAY definition. While a large percentage of the disagreement is due to greater number of visits recorded by the RAMQ compared with the other 2 sources, there were 5–7% of patients for whom the number of self-reported visits was greater compared with RAMQ and about 0.1–3% of patients for whom the medical chart reported more visits than the RAMQ.

Table 2 summarizes the results of Model 1 when using the RAMQ-SEQ or RAMQ-DAY definitions. We can see that there is little effect on the sensitivity and specificity of the self-report and the medical chart whether they are analyzed together with RAMQ-SEQ or RAMQ-DAY definition. The estimates of the sensitivity and specificity of the self-report and medical chart were similar in both cases. The patient's self-report had the lowest sensitivity and specificity. The probability of under-reporting (approximately 36-38%) on the self-report was greater than the probability of over-reporting (about 8%). The specificity of the medical chart was very high, but there was approximately 9-10% probability of under-reporting ED visits. The RAMQ-SEQ definition performed the best with near perfect sensitivity as well as specificity. The RAMQ-DAY definition had a lower specificity (approximately 11% over-reporting), but near perfect sensitivity. This is to be expected since the RAMQ-DAY definition tends to overestimate the true number of visits by treating claims made on consecutive days during the same visit, to be made during different visits. The probability that the observed data exceeds its expected value under the conditional independence model was estimated to be 0.49 among those who truly did not visit the ED and 0.50 among those who did visit the ED, suggesting that the conditional independence assumption was appropriate.

Figure 2 summarizes the impact of covariates on the sensitivity and specificity of the self-report. These results were obtained from a model that included the RAMQ-SEQ definition. Similar results for the validity of the self-report were obtained when using the RAMQ-DAY definition. Sensitivity appeared to be lower among patients aged 80+, with low comorbidity and when the length of recall was shorter. Specificity appeared to be lower among those with high comorbidity. However, none of these associations were statistically significant, ie, the 95% credible intervals in different subsets overlapped substantially. It is of interest to note that the point estimates of sensitivity and specificity were similar within subgroups of patients with

TABLE 2. Results of Latent Class Analysis of Model 1 (Any versus No Emergency Department Visits)

Results When Using the RAMQ-SEQ Definition

	Sensitivity (%) Median* (95% CI)*	Specificity (%) Median (95% CI)	Under-Reporting (%) Median (95% CI)	Over-Reporting (%) Median (95% CI)
Self-report	64 (57, 71)	92 (88, 95)	36 (29, 43)	8 (5, 12)
Medical chart	92 (86, 96)	97 (94, 99)	8 (4, 14)	3 (1, 6)
RAMQ-SEQ	98 (94, 100)	98 (95, 100)	2 (0, 6)	2 (0, 5)

Results When Using the RAMQ-DAY Definition

	Sensitivity (%) Median (95% CI)	Specificity (%) Median (95% CI)	Under-Reporting (%) Median (95% CI)	Over-Reporting (%) Median (95% CI)
Self-report	62 (55, 70)	92 (88, 95)	38 (30, 45)	8 (5, 12)
Medical chart	90 (83, 96)	99 (98, 100)	10 (4, 17)	1 (0, 2)
RAMQ-DAY	99 (97, 100)	89 (83, 94)	1 (0, 3)	11 (6, 17)

*Median and 95% credible intervals were obtained from the posterior distributions. The 95% credible interval can be interpreted as an interval within which there is 95% probability of finding the true value of the parameter.

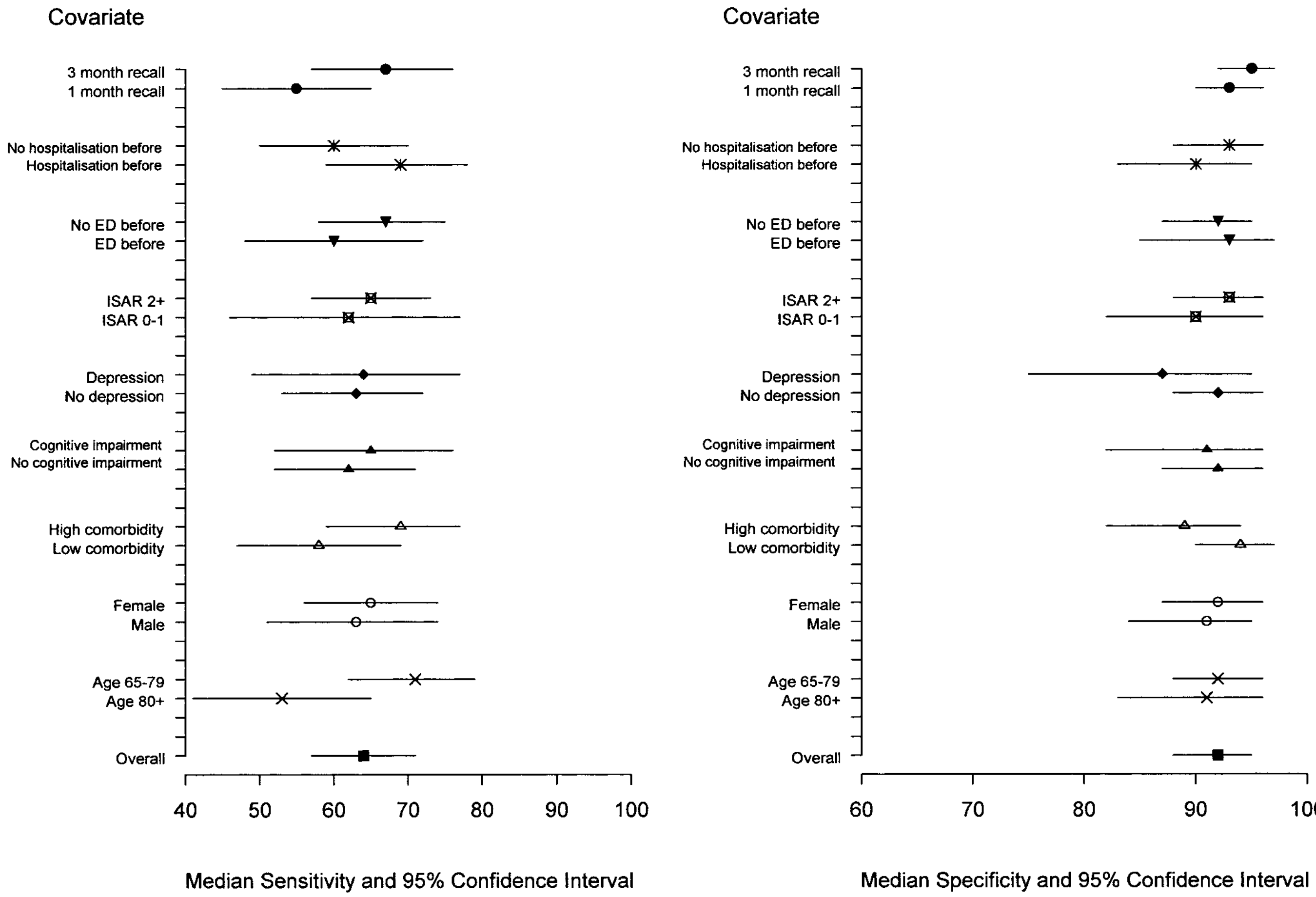

FIGURE 2. Impact of patient characteristics on sensitivity and specificity of self-report from Model 1.

and without cognitive impairment. Thus, the relatively large percentage of patients with cognitive impairment (33%) did not affect our overall estimates of the validity of the patient's self-report.

The intervention appeared to be associated with a tendency to over-report ED use (intervention group: n = 145, specificity = 0.88, 95% confidence interval 0.80–0.95; control group: n = 173, specificity = 0.95 (95% confidence interavl 0.89–0.98). There was no apparent effect of either patient characteristics or admitting hospital on the RAMQ-SEQ, RAMQ-DAY or MC variables (results not shown).

Figure 3 summarizes the results of Model 2. From the left-hand panel, we see that the self-report has the lowest probability of estimating the true number of ED visits correctly while the RAMQ-SEQ definition demonstrates the best overall performance. The ability of the medical chart to estimate correctly decreased with increasing numbers of visits. The sensitivity of the RAMQ-DAY definition dropped when only one visit was made, due to its tendency to overestimate the number of visits in the event of any visits. For the same reason it had a higher sensitivity when the true number of visits was > 1. Probability of overestimation was greatest for the RAMQ-DAY visit, followed by the self-report when the true number of visits was 1. The probability of underestimation was greatest on the self-report in the event that the true number of visits was 1. The probability that the observed data exceeds its expected value under the conditional independence model was estimated to be 0.43, 0.47, and 0.52, respectively, among patients whose true number of ED visits were 0, 1, and > 1, suggesting that the conditional independence assumption was appropriate.

DISCUSSION

This study describes an alternative approach to estimating the relative validity of 3 different sources of information on ED utilization, without arbitrarily considering any one of them a "gold standard." Although medical charts or claims databases are probably more sensitive than self-reports, assuming that they detect visits 100% of the time results in exaggeration of the tendency of the self-report to overesti-

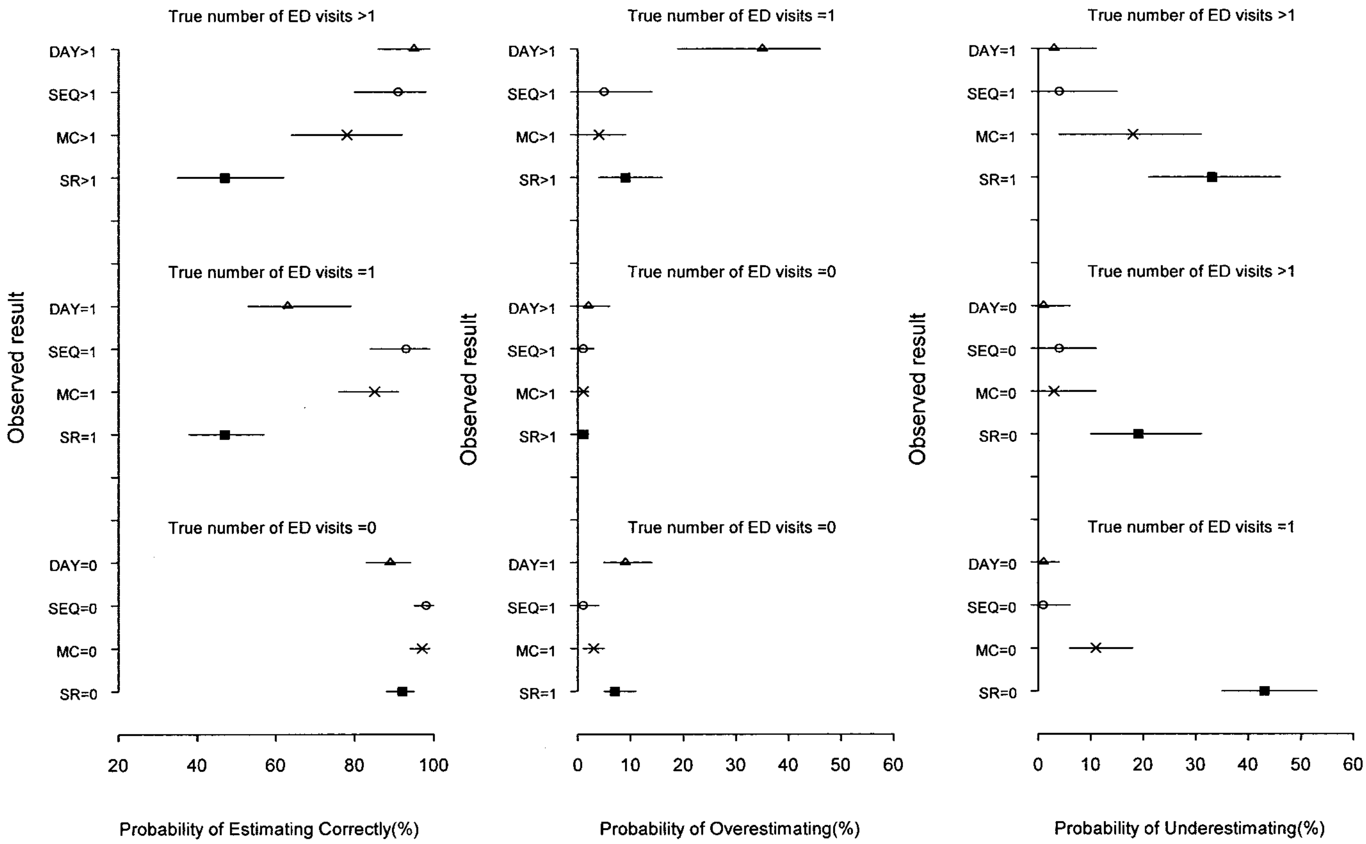

FIGURE 3. Probabilities of estimating correctly, overestimating, and underestimating (along with 95% credible intervals) obtained from Model 2. The probability of the "observed result" on the y-axis is conditional on the "true number of visits." For example, the probability of observing SR $>$ 1 given "True number of ED visits is $>$ 1" is 47% [35%; 62%].

mate visits. Similarly, ignoring the overestimation by the RAMQ-DAY definition would have lead us to conclude that the problem of underestimating by the self-report is greater than it truly is. While the drawbacks of medical charts and claims databases have been highlighted earlier, the lack of perfect validity of these sources has not been taken into account when estimating the validity of a self-report. The disagreement between hospital medical charts and the RAMQ database highlights the difficulty in selecting one of these as a "gold standard." While latent class analyses have been used widely for analysis of diagnostic test results, they have not been used for comparing imperfect sources of information on health services use.

Like earlier studies, we found that concordance between pairs of measures of ED use was between 70% and 90%.[6–8] In our study, the insurance claims database maintained by the RAMQ was the most accurate source in terms of both sensitivity and specificity, particularly when treating claims on consecutive days as part of the same ED visit. Patient's self-report had the least sensitivity and specificity, suggesting that as much as 30% of ED use would be under-reported while 10% of reported visits, in fact, did not take place. The medical chart was found to be intermediate between the self-report and the RAMQ-SEQ definitions with very high specificity but a lower sensitivity than the RAMQ-SEQ. Most of the covariates we examined did not appear to impact the validity of the self-report. However, age greater than 80 years, lower comorbidity and shorter length of recall appeared to be associated with worse sensitivity, while higher comorbidity was associated with worse specificity. When the observed number of visits was divided into 3 groups (0, 1 and $>$ 1 visits), we found that the RAMQ-SEQ definition had the greatest validity followed by the medical chart. The RAMQ-DAY definition tended to have a high probability of overestimation when the true number of visits was 1, while the self-report tended to have a high probability of underestimation in this category.

The validity of the latent class analysis is dependent on whether the assumption of conditional independence is satisfied. Based on our model checking method, this does appear to be the case. A drawback of the latent class model is that a minimum of 3 data sources is needed. To be used when only 2 sources are available, prior information on the validity of at least one source is required.[12] Extensions of the latent class model to adjust for conditional dependence have been proposed but we did not use these as we felt our models were adequate.[26]

The results of this study are generalizable only to the population studied and to the specific claims database and medical charts studied here. Further, elimination of patients without complete information on all 3 sources may have resulted in a selection bias. The small sample size may have obscured an important difference between the ISAR 0–1 and ISAR 2+ groups. It may have also contributed to the nonstatistically significant effect of the covariates on validity of the self-report. Because we have a nonrepresentative distribution of the ISAR score in this study, which in turn is associated with ED visits measured by the RAMQ database, this may have slightly overestimated the self-report's sensitivity and under-estimated its' specificity. In conclusion, use of the RAMQ database with the RAMQ-SEQ definition as described in this paper, is a more valid method of measuring ED use among seniors in Quebec, compared with patients' self-reports and hospital medical charts.

REFERENCES

1. Cleary PD, Jette AM. The validity of self-reported physician utilization measures. *Med Care*. 1984;22:796–803.
2. Bellón JÁ, Lardelli P, de Dios Luna J, et al. Validity of self reported utilisation of primary health care services in an urban population in Spain. *J Epidemiol Community Health*. 2000;54:544–551.
3. Roberts RO, Bergstralh EJ, Schmidt L, et al. Comparison of self-reported and medical record health care utilization measures. *J Clin Epidemiol*. 1996;49:989–995.
4. Carsjö K, Thorslund M, Wärneryd B. The validity of survey data on utilization of health and social services among the very old. *J Gerontol*. 1994;49:S156–S164.
5. Raina P, Torrance-Rynard V, Wong M, et al. Agreement between self-reported and routinely collected health-care utilization data among seniors. *Health Serv Res*. 2001;37:751–774.
6. Ungar WJ, Coyte PC, Pharmacy Medication Monitoring Program Advisory Board. Health services utilisation reporting in respiratory patients. *J Clin Epidemiol*. 1998;51:1335–1342.
7. Wallihan DB, Stump TE, Callahan CM. Validity of self-reported health services use and patterns of care among urban older adults. *Med Care*. 1999;37:662–670.
8. Ritter PL, Stewart AL, Kaymaz H, et al. Self-reports of health care utilization compared to provider records. *J Clin Epidemiol*. 2001;54:136–141.
9. Régie de l'Assurance Maladie duq Uébec. Tableau 1. 01: Nombre de personnes inscrites et admissibles au régime d'assurance maladie duq Uébec selon le sexe, le groupe d'āge et la région sociosanitaire, Québec 1999. [Web Page] 2000; Available at: http://www.ramq.gouv.qc.ca/crc/etudes/pdf/1999/tab101.pdf. Accessed December 16, 2004.
10. Valenstein P. Evaluating diagnostic tests with imperfect standards. *Am J Clin Path*. 1990;93:252–258.
11. Walter S, Irwig L. Estimation of test error rates, disease prevalence and relative risk from misclassifed data: a review. *J Clin Epidemiol*. 1988;41:923–937.
12. Joseph L, Gyorkos T, Coupal L. Bayesian Estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *Am J Epidemiol*. 1995;141:263–272.
13. McCusker J, Verdon J, Tousignant P, et al. Rapid emergency department intervention for elders reduces risk of functional decline: results of a multi-centre randomized trial. *J Am Geriatr Soc*. 2001;49:1272–1281.
14. McCusker J, Bellavance F, Cardin S, et al. Detection of older people at increased risk of adverse health outcomes after an emergency visit: the ISAR screening tool. *J Am Geriatr Soc*. 1999;47:1229–1237.
15. McCusker J, Cardin S, Bellavance F, et al. Return to the emergency department among elders: patterns and predictors. *Acad Emerg Med*. 2000;7:249–259.
16. Dendukuri N, McCusker J, Belzile E. The identification of seniors at risk (ISAR) screening tool: further evidence on concurrent and predictive validity. *J Am Geriatr Soc*. 2004;52:290–296.
17. Katzman R, Brown T, Fuld P, et al. Validation of a short orientation-memory-concentration test of cognitive impairment. *Am J Psychiatry*. 1983;140:734–739.
18. Davous P, Lamour Y, Debrand E, et al. A comparative evaluation of the short orientation memory concentration test of cognitive impairment. *J Neurol Neurosurg Psychiatry*. 1987;50:1312–1317.
19. Katz J, Chang L, Sangha O, et al. Can comorbidity be measured by questionnaire rather than medical record review. *Medical Care*. 1996;34:73–84.
20. Burke W, Roccaforte W, Wengel S. The short form of the Geriatric Depression Scale: a comparison with the 30-item form. *J Geriatr Psychiatry Neurol*. 1991;4:173–178.
21. Lesher E, Berryhill J. Validation of the Geriatric Depression Scale—short form among inpatients. *J Clin Psychol*. 1994;50:256–260.
22. Gelman A, Carlin JB, Stern HS, et al. *Bayesian Data Analysis.* 1st ed. New York: Chapman & Hall; 1997.
23. Spiegelhalter DJ, Thomas A, Best NG. *WinBUGS Version 1.2 User Manual*. Cambridge: MRC Biostatistics Unit; 1999.
24. Berkhof J, van Mechelen I, Gelman A. A Bayesian approach to the selection and testing of mixture models. *Statistica Sinica*. 2003;13:423–442.
25. McCutcheon AL. *Latent Class Analysis*. Sage University Paper Series on Quantitative Applications in the Social Sciences. Beverly Hills, CA: Sage Publications; 1987.
26. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics*. 2001;57:158–167.

APPENDIX

For both models described here the observed data from the RAMQ could follow either the RAMQ-SEQ or RAMQ-DAY definitions. WinBUGS programs for implementing these algorithms may be obtained from the first author.

Model 1: Latent Class Model for Binary Measure of Emergency Department Use (0 Visits, 1 or More Visits)

Define SR = 0 when the observed number of visits on the self-report is 0, and SR = 1 when the observed number of visits on the self-report is greater than 0. Similarly, define MC for the medical chart, and RAMQ for the RAMQ database. Let T denote the true number of visits. Then T = 1 when there was truly at least one ED visit and T = 0 otherwise.

The observed data takes the form of a $2 \times 2 \times 2$ contingency table with 8 cells. Let P_{ijk} denote the probability that a patient will be observed with the combination SR = i, MC = j, RAMQ = k (i, j, k = 0 or 1), and n_{ijk} denote the number of patients observed to have this combination. The probability of observing each cell of the contingency table can be split into 2 groups—the probability of observing that cell among those who truly used an ED and those that truly did not use the ED.

Using the probability notation we have:

$$P_{ijk} = P(SR = i, MC = j, RAMQ = k) = P(SR = i, MC = j, RAMQ = k|T = 1)\ P(T = 1) + P(SR = i, MC = j, RAMQ = k|T = 0)\ P(T = 0) = P(S_{SR} = i|T = 1)\ P(MC = j|T = 1)\ P(RAMQ = k|T = 1)\ P(T = 1) + P(S_{SR} = i|T = 0)\ P(MC = j|T = 0)\ P(RAMQ = k|T = 0)\ P(T = 0)$$

(due to the assumption of conditional independence).

Further, these probabilities can be written in terms of the sensitivity and specificity of each source. For example, the probability of observing 1 or more visits on all 3 sources is given by:

$$P_{111} = P(SR = 1, MC = 1, RAMQ = 1) = P(SR = 1, MC = 1, RAMQ = 1|T = 1)\ P(T = 1) + P(SR = 1, MC = 1, RAMQ = 1|T = 0)\ P(T = 0) = P(S_{SR} = 1|T = 1)\ P(MC = 1|T = 1)\ P(RAMQ = 1|T = 1)\ P(T = 1) + P(S_{SR} = 1|T = 0)\ P(MC = 1|T = 0)\ P(RAMQ = 1|T = 0)\ P(T = 0) = S_{SR}\ S_{MC}\ S_{RAMQ}\pi + (1 - C_{SR})(1 - C_{MC})(1 - C_{RAMQ})(1 - \pi),$$

where S_{SR}, S_{MC}, S_{RAMQ} denote the sensitivities of the 3 sources, C_{SR}, C_{MC}, C_{RAMQ} denote the specificities, and π denotes the true probability of any ED use by a patient. We estimated these 7 unknown parameters using a Bayesian approach. The observed data follows a multinomial distribution with 8 possible outcomes. The likelihood is thus proportional to the product of the probability of each cell in the table raised to the power of the number of patients in that cell:

Likelihood

$$L \propto \prod_{i,j,k=0}^{1} P_{ijk}^{n_{ijk}}$$

Each P_{ijk} was further expanded in terms of the sensitivities and specificities of the 3 sources and the true probability of an ED visit, as explained above.

Prior Distributions

Noninformative Uniform (0,1) prior distributions were used for all 7 unknown parameters.

Posterior Distributions

As it was not possible to obtain an analytical solution for the marginal posterior distribution of each parameter we used a Gibbs sampler to obtain samples from these distributions instead. The Gibbs sampler was implemented by providing the above likelihood and prior distributions to the WinBUGS software package.

Model 2: Latent Class Model for Ordinal Measure of Emergency Department Use (0 Visits, 1 Visit, 2 or More Visits)

In Model 2, SR, MC and RAMQ take values 0, 1 and 2 respectively when a patient makes no ED visits, 1 ED visit, and 2 or more ED visits. Further, we assume that patients with each observed pattern are a mixture of 3 groups for whom T = 0, T = 1, or T = 2. Using a similar notation as for Model 1, the probability of observing 1 visit on all 3 sources is now expressed as follows:

$$P_{111} = P(SR = 1, MC = 1, RAMQ = 1) = P(SR = 1, MC = 1, RAMQ = 1|T = 0)\ P(T = 0) + P(SR = 1, MC = 1, RAMQ = 1|T = 1)\ P(T = 1) + P(SR = 1, MC = 1, RAMQ = 1|T = 2)\ P(T = 2) = P(SR = 1|T = 0)\ P(MC = 1|T = 0)\ P(RAMQ = 1|T = 0)\ P(T = 0) + P(SR = 1|T = 1)\ P(MC = 1|T = 1)P(RAMQ = 1|T = 1)\ P(T = 1) + P(SR = 1|T = 2)\ P(MC = 1|T = 2)\ P(RAMQ = 1|T = 2)\ P(T = 2).$$

Once again this probability can be expressed in terms of parameters measuring the validity of each source. There are 9 such parameters for each source. For example, in the case of the self-report, there are 3 probabilities associated with "estimating correctly": $P(SR = 0|T = 0)$, $P(SR = 1|T = 1)$, $P(SR = 2|T = 2)$, 3 probabilities associated with "overestimating": $P(SR = 1|T = 0)$, $P(SR = 2|T = 0)$, $P(SR = 2|T = 1)$ and 3 probabilities associated with underestimating $P(SR = 0|T = 1)$, $P(SR = 0|T = 2)$, $P(SR = 1|T = 2)$. Similar probabilities can be defined for the medical chart and RAMQ database. Note, however, that $P(SR = 2|T = 1) = 1 - P(SR = 0|T = 1) - P(SR = 1|T = 1)$, where $1 = 0,1,2$. This leaves us with 6 independent, unknown parameters associated with the self-report. Similar relationships exist between the conditional probabilities for the medical chart and RAMQ. Thus, the total number of unknown parameters is 20 (18 validity parameters + 2 parameters to describe the distribution of T). The observed data takes the form of a $3 \times 3 \times 3$ contingency table with 27 cells giving us 26 degrees of freedom, which is sufficient to estimate all parameters. We estimated the 20 unknown parameters using a Bayesian approach.

Likelihood

The observed data follows a multinomial distribution with 27 possible outcomes. The likelihood is thus proportional to the product of the probability of each cell in the table raised to the power of the number of patients in that cell:

$$L \propto \prod_{i,j,k=0}^{2} P_{ijk}^{n_{ijk}}$$

Each P_{ijk} was further expanded in terms of the validity parameters of the 3 sources and the true probability of 0, 1, or 2 ED visits, as explained above.

Prior Distributions

Noninformative Dirichlet (1,1,1) prior distributions were used over the validity parameters for each data source for a given value of T.

Posterior Distributions

As it was not possible to obtain an analytical solution for the marginal posterior distribution of each parameter, we used a Gibbs sampler to obtain samples from these distributions instead. The Gibbs sampler was implemented by providing the above likelihood and prior distributions to the WinBUGS software package.

