Agreement between Self-reported and Routinely Collected Health-care Utilization Data among Seniors

Parminder Raina, Vicki Torrance-Rynard, Micheline Wong, and Christel Woodward

Objective. To examine the agreement between self-reported and routinely collected administrative health-care utilization data, and the factors associated with agreement between these two data sources.

Data Sources/Study Setting. A representative sample of seniors living in an Ontario county within Canada was identified using the Ontario Ministry of Health's Registered Persons Data Base in 1992. Health professional billing information and hospitalization data were obtained from the Ontario Ministry of Health and Long-Term Care (OMH) and the Ontario Health Insurance Plan (OHIP).

Study Design. A cross-sectional survey was carried out to assess *any* contact and *frequency* of contacts with health professionals and hospital admissions. Similar information was obtained from routinely collected administrative data. The level of agreement was assessed using the proportion of absolute agreement, Cohen's kappa statistic (κ), and the intraclass correlation coefficient (ICC). Logistic and linear regressions were used to identify factors that were associated with the magnitude and direction of disagreement respectively.

Data Collection/Extraction Methods. Telephone interviews were conducted on 1,054 seniors, and complete data were available for 1,038 seniors. Each respondent's personal health number was used to electronically link survey data with health professional billing and hospitalization databases.

Principal Findings. Substantial to almost perfect agreement was found for the contact utilization measures, while agreement on volume utilization measures varied from poor to almost perfect. In surveys, seniors overreported contact with general practitioners and physiotherapists or chiropractors, and underreported contact with other medical specialists. Seniors also underreported the number of contacts with general practitioners and other medical specialists. The odds of agreement decreased if respondents were male, aged 75 years and older, had incomes of less than $25,000, had poor/fair/good self-assessed health status, or had two or more chronic conditions.

Conclusion. The findings of this study indicate that there are substantial discrepancies between self-reported and administrative data among older adults. Researchers seeking to examine health-care use among older adults need to consider these

discrepancies in the interpretation of their results. Failure to recognize these discrepancies between survey and administrative data among older adults may lead to the establishment of inappropriate health-care policies.

Key Words. Health services utilization, seniors, self-reports, agreement, billing data

---

Epidemiological and health services research has increasingly focused attention on health-care utilization among the aging population (Ory and Bond 1989). Many of these studies have consistently shown strong associations between health problems, aging, and the use of health-care services. However, it remains to be seen whether improved health practices and changes to health-care delivery will translate into better individual health for seniors in the future. The increasing numbers of seniors in many industrialized countries has already placed a growing demand on the health-care system (Mossey, Havens, and Wolinsky 1989). In recent years, health services research has closely examined the development of accurate measures of health-care utilization by seniors, as well as factors affecting the use of health-care services (Glandon, Counte, and Tancredi 1992; Wan 1989; Muller 1986). In many of these studies, self-reported questionnaires have been the most common method of collecting information on respondents' health status and their use of health-care services. Roos and Havens (1991) found that self-reported health was a strong predictor of ''successful aging'' or remaining independent, and that those who aged successfully made markedly fewer demands on the health-care system. Other studies in the literature have also found a positive association between self-reported health and the use of health-care services among the elderly

---

Address correspondence to Parminder Raina, Ph.D., Associate Professor, Department of Clinical Epidemiology and Biostatistics, Courthouse T-27 Building, 3rd Floor, 1280 Main Street West, Hamilton, ON L 8S 4L8. Dr. Raina is an Associate Professor in the Department of Clinical Epidemiology and Biostatistics. McMaster University, Hamilton, Ontario, Canada. He is also the Director of the McMaster Evidence-based Practice Center. Dr. Raina is the recipient of Health Canada's National Health Scholar Award. Vicki Torrance-Rynard, M.Sc., is a Research Coordinator in the Department of Clinical Epidemiology and Biostatistics, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada. Micheline Wong, M.H.Sc., is a Research Assistant at the Department of Health Care and Epidemiology, Canada. British Columbia, Canada. Christel Woodward, Ph.D., is a Professor in the Department of Clinical Epidemiology and Biostatistics, Faculty of Health Sciences, McMaster University, and the Centre for Health Economics and Policy Analysis, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada.

population (Prigerson, Maciejewski, and Rosenheck 1999; Raina 1995; Branch, Jette, and Evashwick 1981).

Self-reported questionnaires, however, rely on respondents' ability to recall past events. It is important to understand the accuracy of survey responses among seniors, especially since seniors have been known to both over and underreport their use of health-care services more often than younger individuals (Cummings and Nevitt 1988; National Center for Health Statistics 1987). The National Center for Health Statistics in the United States examined the accuracy of self-reported utilization and found significant underreporting of hospitalization days by respondents aged 55 years and older compared with those under the age of 55 (National Center for Health Statistics 1987). Studies specifically on seniors have found substantial over- and underreporting of health-care services (Cummings, Nevitt, and Kidd 1988; Bush, Miller, Golden et al. 1989; Mackenbach, Looman, and Van der Meer 1996; Wallihan, Stump, and Callahan 1999). Wallihan and colleagues (1999), for example, studied a group of 4,506 patients aged 60 years and older. They found that approximately one-quarter of older adults failed to report a hospitalization in the past 12 months and almost one-half underestimated the number of hospitalizations by at least one episode.

The ability or willingness to accurately report use of health-care services may decline with physical and cognitive function (National Center for Health Statistics 1987; Wallihan, Stump, and Callahan 1999). The National Center for Health Statistics found that more acute and chronic health conditions, and longer length of stay in the hospital were related to less underreporting (National Center for Health Statistics 1987). The possibility that discrepancies exist has important implications for future delivery and accessibility of health-care services, particularly if survey findings play an important role in policy-making decisions. If the extent of over- and underreporting of health-care utilization do not balance with each other, the average utilization of services in the elderly population may be much different than is currently believed (Glandon, Counte, and Tancredi 1992). Moreover, the impact that inaccurate findings have on more complicated statistical analyses relating to health utilization is likely to be even more dramatic. For instance, if older and sicker individuals tend to overreport the use of services systematically, then the empirical findings from studies will tend to estimate a greater effect of age and health status on health utilization than actually exists. These reasons emphasize the importance of studies to assess the factors that affect reported health-care utilization. The differences observed between self-reported and routinely collected health utilization data in community-based seniors population are

not well documented in the literature (Glandon, Counte, and Tancredi 1992). Few studies, if any, have attempted to identify factors that are associated with the level of agreement observed.

This study used a representative sample of community-based noninstitutionalized seniors to study the level of agreement between self-reported and routinely collected health-care utilization data. In addition, this study explored salient factors that may be associated with agreement or disagreement between the two data sources.

METHODS

Study Sample

The target population for this study was community-dwelling adults aged 65 years and older. A sample ($n = 1{,}500$) of older people stratified by age and sex was identified using the Ontario Ministry of Health's (MOH) Registered Persons Data Base (RPDB). The RPDB is an ideal sampling frame for selecting a random sample of subjects because of Canada's universal health-care system. The RPDB contains demographic information such as names, date of birth, sex, and address for all residents registered for healthcare coverage in Ontario.

From the sample of 1,500 seniors, 1,296 consented to participate in the study. However, when contacted for the telephone interview, only 1,054 agreed to participate. Complete data for the study were available for 1,038 seniors. Comparison between respondents and nonrespondents showed that both groups were similar for factors such as gender, marital status, education, physical activity, and perceived health status. However, respondents were significantly more likely to be younger in age and have a higher household income than nonrespondents (Raina, Waltner-Toews, Bonnett et al. 1999). In November 1992, trained interviewers conducted telephone interviews approximately 30 minutes in length. Further description of the study and survey design is available elsewhere (Raina, Waltner-Toews, Bonnett et al. 1999).

Questionnaire

The questionnaire used in the study assessed two types of health-care utilization, contacts with health professionals and hospital admissions. For contacts with health professionals, respondents were asked, ''During the past 12 months, have you had contact with any of the following about your physical or mental health: (a) general practitioner or family physician? (b) medical

specialist (such as a cardiologist, urologist, or psychiatrist)? (c) physiotherapist or chiropractor?'' If a respondent answered ''yes,'' they were then asked about the number of times they saw that particular health professional. For hospital admissions, respondents were asked, ''Were you in a hospital overnight during the past 12 months?'' If a respondent answered ''yes'' they were then asked for the number of nights they spent in the hospital. These questions were adopted from the Statistics Canada's 1994 National Population Health Survey (NPHS) (Statistics Canada 1994). A 12-month recall period for each question on health care was used to compare the results with national surveys. The test–retest reliabilities for these questions were found to be moderate to high; interclass correlation coefficient (ICC) ranged from 0.69 to 0.91 (Raina, Bonnett, Waltners-Toews et al. 1999).

The questionnaire also asked about a variety of variables that have been shown in the literature to be associated with health-care utilization (Wan 1989; Roos and Havens 1999; Chappell, Strain, and Blandford 1986). The following sociodemographic variables were included: gender (female or male); age (i.e., 65–69 years old, 70–74 years old, 75–79 years old, or 80 years and older); gross household income level (under $11,999, $12,000–$24,999, $25,000–$49,999, or $50,000 and greater); education (less than high school graduation, or high school and above); current marital status (married, widowed, or other, which included never married, divorced, and separated); living arrangement (alone or living with someone); and type of dwelling (house, apartment, or other, which included trailer, townhouse, and other types of dwellings). In addition, current status of smoking cigarettes (current smokers, past smokers, or never smokers) and alcohol drinking (never drinkers, sometimes drinkers, or regular) were also assessed. For the category of current status of alcohol drinking, ''never drinkers'' were respondents who did not drink alcohol at all, ''sometime drinkers'' drank one to three times a month or less than once a month, and ''regular drinkers'' drank once a week or more (Raina 1995).

Health status variables in the study included self-perceived health status, self-perceived emotional health status, and number of chronic health conditions. Self-perceived health status was assessed by asking the question, ''In general, compared to other persons your age, would you say your health is...excellent, very good, good, fair, or poor?'' Self-perceived emotional health status was based on the question, ''How would you rate your mental or emotional health at the present time...excellent, very good, good, fair, poor, or very poor?'' The number of chronic conditions experienced by respondents were assessed by answering ''yes'' or ''no'' to whether a doctor had ever diagnosed the respondent with any of the following health problems: arthritis,

asthma, back pain, chronic bronchitis, diabetes, epilepsy, heart trouble, cancer, high blood pressure, stroke, cataract, glaucoma, or stomach/intestinal ulcers. Respondents were then classified as having zero, one, two, or three or more chronic illnesses. These questions were also adopted from the national surveys and were shown to have moderate to high test–retest reliabilities. The interclass correlations (ICC) ranged from 0.69 to 0.96 (Raina, Bonnett, Waltner-Toews, et al. 1999).

Ministry of Health and Long-Term Care Utilization Data

This study used two databases of routinely collected administrative data from the Ontario Ministry of Health and Long-Term Care (OMH) and the Ontario Health Insurance Plan (OHIP): health professional billing information and hospitalizations. Respondents provided written permission to access their health-care data from the OMH. As part of the public-funded health-care system in Canada, heath-care utilization data on both outpatient and inpatient services are routinely maintained by each province. Residents who are covered by the health-care system are issued a unique personal health number (PHN) that entitles them to health-care services paid for by the province's health insurance plan. OHIP maintains a health professional billing database on the specific services paid to health professionals when they submit a claim for services rendered. This database includes information on the following: the patient's personal health number, provider identification code, type of specialty, diagnostic and procedure codes, referring physician, date of service, fee schedule code, number of services, and fee(s) paid.

Similarly, the hospitalization database maintained by the OMH records information on all inpatient admissions to any Ontario hospital, including personal health number, date of admission, date of discharge, diagnostic and procedure codes, and length of stay. In this study, each respondent's PHN was used to electronically link his/her survey data with the OHIP billing and hospitalization database. Information from billing and hospital databases was extracted for the year immediately preceding the survey data (November 1, 1991 to October 31, 1992) to ensure that the time frame associated with all sources of data coincided. For the purposes of the study, we identified physicians by their area of specialty rather than the specific services that were billed to OHIP. To verify the physician's area of specialty, an OHIP demographic database containing information on the original specialty declared by the health professional and other related data, such as age, gender, place of graduation, year of graduation, type of specialty, and provider

identification code, was used in the study. Discrepancies between the billing database and the demographic database, however, occurred in less than 2 percent of listed clinicians.

Analysis

Descriptive statistics were calculated for all independent variables available in the study. The level of agreement for any stay in a hospital and any contact with a health professional, as well as for the number of nights stayed in hospital and the number of contacts with a health professional were examined. Ten records from the health professional utilization database and two records from the hospitalization database were dropped due to missing data, leaving a final sample size of 1,028 respondents in the health professional analyses and 1,036 in the hospital analyses. For the purposes of this study, the measures of any stay in hospital (yes or no) and any contact with a health professional (yes or no) were termed ''contact utilization measures.'' The measures related to the number of nights stayed in hospital and the number of contacts with a health professional were termed ''volume utilization measures.''

The level of agreement for dichotomous contact utilization was assessed using the proportion of absolute agreement and beyond chance agreement using Cohen's kappa statistic (κ). Agreement for continuous volume utilization measures was assessed using the random effects intraclass correlation coefficient (ICC) (Strout and Fleiss 1979; Fleiss 1986). The 95 percent confidence intervals around the and ICC were calculated (Strout and Fleiss 1979; Fleiss 1986). The benchmark for determining the closeness of the comparison for both κ and ICC was based on Landis and Koch (1977). In their scale, the strength of agreement was as follows: 0.00–0.20 (poor), 0.21–0.40 (fair), 0.41–0.60 (moderate), 0.61–0.80 (substantial), and 0.81–1.00 (almost perfect).

Agreement for the contact utilization measures included ''yes/yes'' and ''no/no'' responses in the self-reported and administrative data respectively. Disagreement, therefore, included ''yes/no'' (overreporting in survey) and ''no/yes'' (underreporting in survey) responses. The volume utilization measures were also analyzed in a similar manner.

Polytomous logistic regression for categorical outcomes would have been the most appropriate to assess the predictors of under- or overreporting for the contact utilization measures. However, due to small sample size, each dependent variable for the contact utilization measures was dichotomized. Each dependent variable was coded as ''1'' indicating agreement between the

two data sources or "0" indicating disagreement (including over- and underreporting). Each independent variable known to be associated with health-care utilization was selected and used in the logistic regression to identify predictors of agreement/disagreement for contact utilization measures. Separate logistic regressions were run for the following four dependent variables: any stay in hospital, any contact with GP (general practitioner), any contact with physiotherapist (PT) or chiropractor, and any contact with other medical specialties. Each variable used in the study was examined for its effect on each model. Variables were omitted from the analyses if they were not significantly associated with agreement in health-care utilization measures ($p > 0.05$) and did not contribute to the overall model. The variables of age and income level were collapsed into smaller categories due to small sample sizes. Age was regrouped into two categories, 65–74 years old and 75 years and older. Household income was regrouped according to whether individuals earned $25,000 or less, or more than $25,000.

For volume utilization measures, linear regression was used to assess the predictors of under- or overreporting. The dependent variable was the difference between self-reported and MOH data. Positive values imply that self-reported measures of utilization exceed the measures of utilization in the MOH data (e.g., overreport). Negative values imply that self-reported measures of utilization are less than the measures of utilization in the MOH data (e.g., underreport). The same independent variables used in the above-mentioned logistic regression appear in each of the four difference linear regression equations. In addition, in each of the four difference equations (e.g., difference in length of stay), the respective volume utilization measure from the MOH data was included as an independent variable (e.g., length stay from MOH data). Inclusion of volume measure utilization from the MOH data as an independent variable controlled for estimation bias (Glandon, Counte, and Tancredi 1992).

RESULTS

Demographic Characteristics

The sample size consisted of 1,038 seniors with a mean age of 73 years ($SD = \pm 6.3$). Table 1 shows that 51 percent of respondents were females, 53 percent had some form of high school or higher education, and 53 percent had gross household incomes below $25,000 per year in 1991. The majority of respondents were married (64 percent), resided with at least one other person

Table 1: Distribution of Demographic Characteristics of Non-Institutionalized Community Dwelling Seniors Aged 65 Years and Older ($n = 1{,}038$), 1992

Variables	*n (%)*
Gender	
Female	529 (51)
Male	509 (49)
Age Groups	
65–69	353 (34)
70–74	332 (32)
75–79	187 (18)
80 & older	166 (16)
Household Income*	
Less than $11,999	119 (11)
$12,000–$24,999	441 (42)
$25,000–49,999	331 (32)
$50,000 & greater	103 (10)
Education	
Below high school	488 (47)
High school and above	550 (53)
Marital Status	
Married	664 (64)
Widowed	291 (28)
Other (e.g., never married/divorced/separated)	83 (8)
Living Arrangement	
Alone	343 (33)
Living with someone	695 (67)
Type of dwelling	
House	716 (69)
Apartment	197 (19)
Other (e.g., trailer, townhouse, and other types of dwellings)	125 (12)
Smoking	
Current smokers	104 (10)
Past smokers	477 (46)
Never smokers	455 (44)
Alcohol†	
Never drinkers	280 (27)
Sometimes drinkers	466 (45)
Regular drinkers	290 (28)

*Missing information accounted for the remaining 5% of the data.

†Status of alcohol drinking: never drinkers (did not drink alcohol at all), sometime drinkers (drank one to three times a month or less than once a month), and regular drinkers (drank at least once a week).

Table 2: Summary of Exact Agreement between Survey and Routinely Collected Data on Health Care Utilization Measures among Seniors Aged 65 Years and Older, 1992

Contact Utilization Measure (n = sample size)	*Observed Agreement: n (%)*	*Kappa (95% Confidence Intervals)*
Any stay in hospital (n = 1,036)	970 (93.6)	0.771 (0.718, 0.824)
Any contact with general practitioner (n = 1,028)	815 (79.3)	0.195 (0.126, 0.264)
Any contact with physiotherapist or chiropractor (n = 1,028)	945 (91.9)	0.685 (0.622, 0.748)
Any contact with other medical specialists (n = 1,028)	793 (77.1)	0.351 (0.286, 0.417)
Volume Utilization Measure (n = sample size)	*Observed Agreement: n (%)*	*Intraclass Correlation Coefficient (95% Confidence Intervals)*
Length of stay in hospital (n = 1,036)	865 (83.5)	0.50 (0.45, 0.55)
Number of contacts with general practitioner (n = 1,028)	140 (13.6)	0.34 (0.25, 0.42)
Number of contacts with physiotherapist or chiropractor (n = 1,028)	844 (82.1)	0.41 (0.36, 0.46)
Number of contacts with other medical specialists (n = 1,028)	197 (19.2)	0.25 (0.13, 0.35)

(67 percent), and lived in a house (69 percent). Furthermore, 56 percent of respondents reported that they were either current or past smokers, and 72 percent drank alcohol either ''sometimes'' or ''regularly.''

Agreement between Survey and Routinely Collected Data

Table 2 shows the observed agreement between the survey and routinely collected data for the health-care utilization measures. The proportion of observed agreement for the contact utilization measures (77.1 percent to 93.6 percent) was higher than the agreement for the volume utilization measures (13.6 percent to 83.5 percent). According to Landis and Koch's scale (1977), the strengths of observed agreement for contact utilization measures ranged from substantial to almost perfect compared with poor to moderate strengths of observed agreement for the volume utilization measures. When κ was calculated, however, poor to substantial levels of agreement were seen for the contact utilization measures. The substantial level of beyond chance agreement

was found for any stay in hospital ($\kappa = 0.771$, 95 percent Confidence interval [CI]: 0.718 to 0.824) and any contact with PT or chiropractor ($\kappa = 0.685$, 95 percent CI: 0.622 to 0.748). However, the poor to fair agreement was observed for any contact with a GP ($\kappa = 0.195$, 95 percent CI: 0.126 to 0.265) and any contact with other medical specialists ($\kappa = 0.351$, 95 percent CI: 0.286 to 0.417).

Poor to moderate strengths of beyond chance agreement was shown for the volume utilization measures. The ICC showed moderate agreement for length of stay in hospital (ICC = 0.50, 95 percent CI: 0.45 to 0.55) and number of contacts with PT or chiropractor (ICC = 0.41, 95 percent CI: 0.36 to 0.46); fair agreement was found for the number of contacts with GP (ICC = 0.34, 95 percent CI: 0.25 to 0.42). The poorest agreement was found for the number of

Table 3: Comparison between Survey Responses and Routinely Collected Data on the Contact Utilization Measures among Seniors Aged 65 Years and Older, 1992

Contact Utilization Measure (*n = sample size*)	*Survey Response**	*Database Indication**	*Frequency*	*Percent*
Any stay in hospital (n = 1,036)				
Agreement	No	No	830	80.1
	Yes	Yes	140	13.5
Disagreement: Overreporting in survey	Yes	No	34	3.3
Underreporting in survey	No	Yes	32	3.1
Any contact with general practitioner (n = 1,028)				
Agreement	No	No	44	4.3
	Yes	Yes	771	75.0
Disagreement: Overreporting in survey	Yes	No	169	16.4**
Underreporting in survey	No	Yes	44	4.3**
Any contact with physiotherapist or chiropractor (n = 1,028)				
Agreement	No	No	832	80.9
	Yes	Yes	113	11.0
Disagreement: Overreporting in survey	Yes	No	62	6.0**
Underreporting in survey	No	Yes	21	2.0**
Any contact with other medical specialists (n = 1,028)				
Agreement	No	No	114	11.0
	Yes	Yes	679	66.1
Disagreement: Overreporting in survey	Yes	No	75	7.3**
Underreporting in survey	No	Yes	160	15.6**

*''No'' indicates no stay or contact reported in the data source, ''yes'' indicates that at least one stay or contact was reported in the data source.

**Indicates a significant difference between the two types of disagreement ($p \leq 0.05$).

Table 4: Comparison between Survey Responses and Administrative Database on the Volume Utilization Measures among Seniors Aged 65 Years and Older, 1992

	Length of stay in hospital (n = 1,036)		*Number of contacts with general practitioner (n = 1,028)*		*Number of contacts with physiotherapist or chiropractor (n = 1,028)*		*Number of contacts with other medical specialists (n = 1,028)*	
	Freq	%	*Freq*	%	*Freq*	%	*Freq*	%
Agreement*								
On responses of zero	830	80.1	44	4.3	832	80.9	114	11.1
On responses of one or more	35	3.4	96	9.3	12	1.2	83	8.1
Disagreement								
Overreporting in survey by 1	22	2.1	93	9.0	17	1.6	59	5.7
Overreporting in survey by 2	10	1.0	51	5.0	9	0.9	25	2.4
Overreporting in survey by 3 or more	47	4.5	133	12.9	75	7.3	15	1.4
TOTAL OVERREPORTING	79	7.6	277	26.9**	101	9.8	182	9.6**
Underreporting in survey by 1	31	3.0	106	10.3	14	1.4	123	12.0
Underreporting in survey by 2	15	1.4	82	8.0	11	1.1	98	9.5
Underreporting in survey by 3 or more	46	4.4	423	41.1	58	5.6	511	49.7
TOTAL UNDERREPORTING	92	8.9	611	59.4**	83	8.1	732	71.2**

*Agreement of the volume utilization measures included zero responses (e.g., respondents reported zero for length of stays at a hospital and the database was also zero) and responses of one or more (e.g., respondents of the survey reported one or more nights of stay at a hospital and database also reported one or more nights stay).

**Indicates a significant difference between the two types of disagreement ($p < 0.0010$).

contacts with other medical specialists (ICC = 0.25, 95 percent CI: 0.13 to 0.35).

Comparison of Contact and Volume Utilization Measures

The levels of under- and overreporting between survey and routinely collected data among seniors is shown in Tables 3 and 4. Among the contact utilization measures, the proportion of over- and underreporting for any stay in a hospital (3.3 percent and 3.1 percent respectively) were relatively small and balanced each other out (Table 3). However, respondents significantly overestimated their contact with a GP, PT, or chiropractor, and underestimated their contact with other medical specialists ($p \leq 0.05$). More specifically, 16.4 percent of seniors overreported contact with GPs in the past year compared to 4.3 percent who underreported. For any contact with a PT, 6.0 percent of seniors overreported visits compared to 2.0 percent who underreported. As well, 15.6 percent of seniors underreported visits with other medical specialists, while 7.3 percent overreported. Therefore, the net reporting error ranged from 4.0 percent for contact with a PT to 12.1 percent for contact with a GP.

Among the volume utilization measures (Table 4), there was no significant difference between the over- and underreporting for length of stay at a hospital and number of contacts with a PT or chiropractor. However, respondents significantly underreported their contact with GPs and other medical specialists (Table 4). Among patients who reported to have at least one contact with their GP in the past year, 9.0 percent overreported contact by one, 5.0 percent overreported contact by two, and 12.9 percent overreported contact by 3 or more. Similarly, 10.3 percent underreported contact with a GP by one, 8.0 percent underreported contact by two, and 41.1 percent underreported contact by three or more. Overall 26.9 percent of respondents overreported the number of contacts with GPs compared with 59.4 percent who underreported their contact ($p < 0.001$).

Among respondents who reported to have at least one contact with other medical specialists in the past year, 5.7 percent overreported contact by one, 2.4 percent overreported contact by two, and 1.4 percent overreported contact by three or more. Similarly, 12.0 percent underreported contact with a medical specialist by one, 9.5 percent underreported contact by two, and 49.7 percent underreported contact by three or more. Overall, 9.6 percent of respondents overestimated the number of contacts with other medical specialists compared with 71.2 percent who underestimated their contact ($p < 0.001$).

Table 5: Logistic Regression Results for Agreement/Disagreement* on the Contact Utilization Measures among Seniors Aged 65 Years and Older, 1992

	Odds Ratio (95% Confidence Interval)			
*Independent Factors***	*Any stay in hospital (n = 981)*	*Any contact with GP (n = 974)*	*Any contact with PT or chiropractor (n = 974)*	*Any contact with other medical specialists (n = 974)*
Gender:				
Female	Ref	Ref	Ref	Ref
Male	0.57 (0.33, 0.98)	0.81 (0.58, 1.11)	1.20 (0.74, 1.93)	1.05 (0.77, 1.45)
Age:				
65–74 years	Ref	Ref	Ref	Ref
75+ years	0.55 (0.32, 0.95)	1.20 (0.85, 1.72)	1.00 (0.61, 1.64)	0.81 (0.59, 1.12)
Income:				
≤$25,000	Ref	Ref	Ref	Ref
>$25,000	1.60 (0.89, 2.87)	1.60 (1.14, 2.24)	1.12 (0.68, 1.83)	1.40 (1.01, 1.94)
Self-Assessed Health Status:				
Poor, fair, or good	Ref	Ref	Ref	Ref
Very good or excellent	2.51 (1.18, 5.34)	0.84 (0.41, 1.70)	1.55 (0.71, 3.41)	0.68 (0.34, 1.36)
Self-Assessed Emotional Health Status:				
Poor, fair, or good	Ref	Ref	Ref	Ref
Very good or excellent	2.11 (0.94, 4.73)	0.68 (0.30, 1.53)	0.58 (0.19, 1.73)	0.94 (0.46, 1.91)

Number of chronic conditions:				
0	Ref	Ref	Ref	Ref
1	0.43 (0.05, 3.71)	0.78 (0.42, 1.45)	0.43 (0.05, 3.74)	1.32 (0.72, 2.43)
2	0.19 (0.03, 1.49)	1.19 (0.64, 2.21)	0.13 (0.02, 0.99)	1.60 (0.88, 2.90)
3 or more	0.22 (0.03, 1.65)	1.54 (0.86, 2.74)	0.12 (0.02, 0.91)	1.77 (1.02, 3.05)
Hosmer and Lemeshow Goodness-of-Fit	7.85 (df = 8) sig = 0.4481	14.08 (df = 8) sig = 0.0798	8.39 (df = 8) sig = 0.3965	6.50 (df = 8) sig = 0.5911

*Agreement/disagreement includes all responses from survey and administrative data (e.g., agreement = ''yes/yes'' and ''no/no'' and disagreement = ''no/yes'' and ''yes/no'').

**The variables education, marital status, living arrangement, type of dwelling, and current status of smoking and alcohol drinking were omitted from the analyses because they were not significantly associated with agreement between the two sources of data ($p > 0.05$) and did not contribute to the overall model.

Logistic and Linear Regression

The logistic regression results for the contact utilization measures are shown in Table 5. Due to missing data for the independent factors, sample sizes for the logistic regression were decreased by 55 cases for the hospital contact and volume utilization measures, and 54 cases for the health professional contact and volume utilization measures. Most models had a good overall goodness-of-fit on Hosmer and Lemeshow statistics. Among statistically significant results, the odds of males agreeing on any stay in a hospital were lower than that of females (Odds Ratio [OR] = 0.57, 95 percent CI: 0.33 to 0.98). The odds of seniors aged 75 years and older agreeing on any stay in a hospital were less than those aged 65–74 years old (OR = 0.55, 95 percent CI: 0.32 to 0.95). Seniors who reported incomes of greater than $25,000 had significantly greater odds of agreeing on any contact with both a GP (OR = 1.60, 95 percent CI: 1.14 to 2.24) and other medical specialists (OR = 1.40, 95 percent CI: 1.01 to 1.94) than seniors with incomes of $25,000 or less. Seniors who reported very good or excellent self-assessed health status had more than twice the odds of agreeing on any stay in hospital (OR = 2.51, 95 percent CI: 1.18 to 5.34) than those who reported poor, fair, or good health. As well, seniors with two or more chronic conditions had lower odds of agreeing on any contact with a PT or chiropractor (OR for 2 chronic conditions = 0.13, 95 percent CI: 0.02 to 0.99 and OR for 3 or more chronic conditions = 0.12, 95 percent CI: 0.02 to 0.91) than seniors who did not report any chronic conditions. However, respondents with three or more chronic conditions had greater odds of agreeing on any contact with a medical specialist (OR = 1.77, 95 percent CI: 1.02 to 3.05).

The linear regression results for the volume utilization measures are shown in Table 6. For the length of stay, statistical significance ($\beta = -7.73$, p-value < 0.0001) was found only for the self-reported health status. Seniors who had reported a lower health status tended to overreport length of stay in the self-reported data. For number of contacts with GPs, the significant variables were again self-reported health status ($\beta = -1.91$, p-value < 0.0004) and number of reported chronic conditions ($\beta = 0.89$, p-value < 0.0001). An increase in chronic condition results in more overeporting. For number of contacts with PTs or chiropractors, significant relationships were found for gender ($\beta = -1.06$, p-value < 0.0255) and self-reported health status ($\beta = -1.92$, p-value < 0.0425).

Table 6: Linear Regression for Predictors of Differences between Self-reported and MOH Data for Volume Measures among Seniors Aged 65 years and Older, 1992

	Dependent Variables											
	Length of stay in hospital (n = 981)			*# of contacts with GP (n = 974)*			*# of contacts with PT (n = 974)*			*# of contacts with medical specialists (MS) (n = 974)*		
Independent Factors	β	*SE*	*P-Value*	β	*SE*	*P-Value*	β	*SE*	*P-Value*	β	*SE*	*P-Value*
Constant	6.49	2.13	0.0024	4.44	0.90	0.0001	2.57	1.52	0.0467	0.59	0.72	0.4152
Gender (0 = Female & 1 = Male)	−0.80	0.65	0.2247	−0.05	0.27	0.8130	−1.06	0.47	0.0255	−0.16	0.24	0.4960
Age (0 = 65–74 & 1 = 75+)	−1.08	0.69	0.1196	−0.27	0.29	0.3572	−0.24	0.50	0.6368	0.41	0.25	0.1068
Income (0 =≤ 25,000 & 1 => 25,000)	0.76	0.67	0.2588	−0.43	0.28	0.1244	−0.74	0.48	0.1288	0.21	0.24	0.3853
Self-Assessed Health (0 = poor/good & 1 = very good/excellent)	−7.73	1.31	0.0001	−1.99	0.55	0.0004	−1.92	0.95	0.0425	−0.22	0.48	0.6415
Self-Assessed Emotional Health (0 = poor/good & 1 = very good/excellent)	1.91	1.43	0.1824	−1.07	0.60	0.0741	0.20	1.04	0.8413	−0.44	0.52	0.3947
# of Chronic Conditions	0.43	0.33	0.2044	0.89	0.14	0.0001	0.46	0.24	0.0572	0.22	0.12	0.0762
Length of Stay in MOH data	−0.54	0.03	0.0001		—			—			—	
# of contacts with GP in MOH data		—		−0.80	0.01	0.0001		—			—	
Any of contacts with PT in MOH data		—			—		−0.16	0.49	0.0011		—	
# of contacts with MS in MOH data		—			—			—		−0.84	0.10	0.0001

**The variables education, marital status, living arrangement, type of dwelling, and current status of smoking and alcohol drinking were omitted from the analyses since they were not significantly associated with agreement between the two sources of data ($p > 0.05$) and did not contribute to the overall model.

DISCUSSION

The purpose of this study was to examine the level of agreement between self-reported and routinely collected health-care utilization data, as well as to explore the salient factors that may be associated with agreement between the two data sources. Our findings showed substantial to almost perfect observed agreement between survey and routinely collected data for the contact utilization measures. This supports findings by Glandon and colleagues (1992), who studied a group of 234 adults aged 62 years and older and found that 91.6 percent of the sample accurately reported contact with a physician within the past 6 months. As well, Wallihan and colleagues (1999) also reported that 88.4 percent of adults aged 60 years and older from a clinical study population accurately reported a hospitalization in the past 12 months.

A wide range of observed agreement, from poor to almost perfect observed agreement, was found for the volume utilization measures. More specifically, poor observed agreement was found for the number of contacts with a GP and other medical specialists in the past year, and almost perfect observed agreement was found for length of stay in a hospital and for number of contacts with a PT or chiropractor. In the latter case, however, the agreement was mainly the result of a large proportion of respondents reporting no health service use. We can assume that it is easier for respondents to accurately report no use of health services over the past year compared with respondents who had many encounters. When looking only at respondents who had contact with the health-care system, observed agreement for the exact length of stay at a hospital was found to be 17.0 percent (35/206), a value slightly lower than a recently published result of 26.5 percent by Wallihan and colleagues (1999). Similarly, our observed agreement of 9.8 percent (96/984) for number of contacts with a GP is lower than the value of 28.2 percent found by Glandon and colleagues (1992). Lower levels of observed agreement in our findings compared with previous studies might be largely explained by the differences in study populations. Our study used community-based noninstitutionalized seniors rather than patients who received medical care from a clinic or HMO (Wallihan, Stump, and Callahan 1999; Glandon, Counte, and Tancredi 1992). Compared with a clinical group, our community sample of seniors may have been more likely to be healthier and less likely to see a physician or to stay overnight at a hospital. Therefore, they may also have been less likely to

remember the frequency of their contacts with health professionals compared with clinical seniors patients.

Our study also found that the beyond chance agreement measured by κ for any contact with a GPs was much lower than expected from our observed agreement. The κ showed poor agreement of 0.195 compared with substantial observed agreement of 79.3 percent. This paradoxical result may have occurred because high observed values can often be associated with low levels of κ when marginal totals are symmetrically unbalanced in a 2 by 2 table (Feinstein and Cicchetti 1990). In this situation, therefore, the observed agreement is a closer approximation of agreement than κ.

More detailed analyses between survey and administrative data related to the contact utilization measures showed no difference between the proportion of overestimation and underestimation for any stay at a hospital. However, respondents significantly overestimated any contact with a GP and PT or chiropractor, and significantly underestimated any contact with other medical specialists. Lack of difference between the proportion of over- and under-reporting for any stay in a hospital may be because hospitalization is a highly salient event and likely to be remembered. Roos and Havens (1991) have also found that more salient events like surgical admissions may be more likely to be remembered than routine medical admissions.

As expected, the volume utilization measures showed larger disagreement than the contact utilization measures. No differences were found between the proportions over- and underestimated for length of stay at a hospital and for number of contacts with a PT or chiropractor over the past 12 months. However, respondents significantly underestimated the number of contacts with GPs and other medical specialists. A number of studies have shown that the inaccuracy of recalling an event may be a function of the number of services used. Both Wallihan and colleagues (1999) and Roberts and colleagues (1996) found that the inaccuracy of estimations increased as the volume of services increased. Among respondents who reported at least one hospitalization in the past 12 months, 26.5 percent were exact in their estimation of length of stay at a hospital, 43.8 percent either over- or underestimated their stay by 1 to 3 days, and another 29.7 percent either over- or underestimated their stay by 4 or more days (Wallihan, Stump and Callahan 1999). Therefore, while individuals may remember whether they visited a doctor or were admitted a hospital, they rarely remember them as countable series (Raina 1995).

Significant disagreements occurred in both contact and volume utilization despite a previous study by Raina (1995), who found high test–retest

reliability related to the any contact and frequency of contact with health professionals. Kappa values of 0.83 to 0.91 were found for any contact with a GP and other medical specialists, and κ of 0.69 to 0.90 were found for the frequency of contact with a GP and other medical specialists. At least four other reasons may explain the significant over- and underreporting of both contact and volume utilization measures in our study. First, confusion in the definition of survey questions or terminology may have occurred. Related to the OMH administrative databases, if a health professional billed for health-care services that was not identical to the respondents understanding of the term ''contact,'' then our analysis would show disagreement. For example, a respondent's interpretation of contact with a health professional may have included phone calls to the physician's office, compared with the administrative meaning of contact, which is a face to face meeting that results in a bill to the OMH. In addition, there is also a possibility that elderly respondents did not clearly understand the difference between contact with a GP versus other medical specialists in providing primary care services. Although 16.4 percent of respondents overreported contact in the survey for a GP while a similar number (15.6 percent) underreported for a medical specialist, further analysis (not shown) did not support the notion that respondents are mistaking specialists for GPs. Respondents who overreported GP contact were actually more likely to overreport, rather than underreport, specialist contact. Therefore respondent confusion between contact with a GP and other medical specialists was not evident in our study. From the perspective of the physician, discrepancies may have occurred from inaccurate billing (e.g., specialists providing primary care services and therefore billing as a GP). However, as reported earlier, these discrepancies were minimal because the differences between the two databases occurred in less than 2 percent of the listed clinicians.

Second, discrepancies between survey and routinely collected data may be due to differences in reporting systems. Related to PT and chiropractic services, in Ontario, Canada, only some physiotherapy or chiropractic clinics are allowed to bill OHIP. All other clinics are privately-owned and bill patients directly. Their data may not necessarily be collected and included in the administrative database. As well, if visits to a PT occur under a hospital outpatient physiotherapy department, they would not appear in the routinely collected administrative data but rather under a global hospital budget. Discrepancies related to contact with a GP might result when individuals visit their physician for follow-up care, but the data does not get recorded appropriately. Discrepancies related to reporting of overnight stays at a hospital are influenced by the time an individual was admitted to a hospital. For

example, a patient admitted at 3:00 A.M. may believe that he or she is spending that night in the hospital. However, because the admission was missed on a hospital's midnight census, the patient would not be counted as having stayed that night in the administrative database.

Third, the discrepancy between self-reported and routinely collected data may also be affected by the quality of the recorded health data. Romm and Putnam (1981) compared medical records with verbatim transcripts of outpatient visits and found that only 59 percent of information present in the transcripts was found in the medical record. Problems related to the recording of medical data include inadequate recording of information by physicians, improper or incomplete recording of discharge form information, and miscoding of the data (Rawson, Malcolm, and D'Arcy 1997; Marrie, Duant, and Sealy, 1987; Doremus and Michenzi 1983; Demlo, Campbell, and Brown 1978). Rawson and D'Arcy (1991) noted that there are many opportunities for biases and errors to occur from first contact of a patient with a health professional to the recording of this event in a database. The reliability of the information may be affected by potential problems ranging from minor inaccuracies to complete misrepresentation of data (Romm and Putnam 1981).

Fourth, the accuracy of reporting health-care utilization may be affected by factors that decrease recall memory among seniors. A number of studies in the literature have shown that problems of recall memory increase among older adults who were advanced in age (Wiederholt et al. 1993; Spiers and Kutsik 1995). Wiederholt and colleagues (1993) studied the performance of neuropsychological tests on 1,692 community-dwelling subjects aged 55 to 94 years and found that the short-term and long-term recall ability decreased progressively with age. In addition, they found that women performed better than men on memory tests even with advancing age. As well, both men and women with some college education performed better on most tests than men and women with high school education (Wiederholt et al. 1993). Although recall bias may have played an important role in our results, our sample of older adults was a randomly selected group of community-dwelling seniors who were more likely to be younger in age and have a higher household income than nonrespondents. Therefore, our findings may overestimate the proportion of agreement compared with the general population seniors. The length of our recall period, 12 months, may also have had an effect on the respondents' ability to remember the number of contacts they had with health-care professionals.

Unlike previous studies in the area, we have not assumed a gold standard for either the administrative or self-reported data. Our findings suggest that, depending on the source of data, contact utilization measures may differ by almost one-quarter. Disagreement ranged from 6.4 percent for any stay in a hospital to 22.9 percent for contact with other medical specialists. Similarly, volume utilization measures between the two data sources may differ by as much as 85 percent. Failure to recognize these discrepancies between survey and administrative data may lead to the establishment of inappropriate health-care policies.

The logistic regression analyses in our study suggested an important association between sociodemographic and health status variables, and in the level of agreement between survey data and routinely collected data. Agreement generally decreased if respondents were male, aged 75 years and older, had incomes of less than $25,000, and had a lower (poor/fair/good versus very good/excellent) self-assessed health status. The linear regression suggested that poor or fair health status and male gender tended to be associated with underreporting on volume measures in the surveys. This is consistent with previous reports that also found greater inaccuracies in self-reports occurring in respondents with increasing age and disability, fewer social supports, less education, and lower income (Wallihan, Stump, and Callahan 1999; Gladdon, Counte, and Tancredi 1992).

These large differences in estimates suggest that previous studies, which relied only on one source of data to estimate health-care use, may be seriously biased in one direction or another. Therefore, investigators seeking to study the future use of health-care services among community-dwelling older adults can not solely rely on self-reported data or routinely collected data. Self-reported and administrative data are considered useful sources, however, investigators need to confront the validity and quality of their data when attempting to interpret findings. Future studies should examine the magnitude of error in measuring the contact and volume of health services utilization, as well as other factors that may more accurately predict health-care use among community-based seniors using these data sources.

REFERENCES

Branch, L., A. Jette, and C. Evashwick. 1981. ''Towards understanding elders' health service utilisation.'' *Journal of Community Health* 7 (2): 80–92.

Bush, T. L., S. R. Miller, A. L. Golden, and W. E. Hale. 1989. ''Self-report and Medical Record Report Agreement of Selected Medical Conditions in the Elderly.'' *American Journal of Public Health* 79: 1554–6.

Chappell, N., L. Strain, and A. Blandford. 1986. *Aging and Health Care: A Social Perspective*. Toronto: Holt, Rinehart and Winston of Canada.

Cummings, S. R., M. C. Nevitt, and S. Kidd. 1988. ''Forgetting Falls: The Limited Accuracy of Recall of Falls in the Elderly.'' *Journal of American Geriatric Society* 36 (7): 613–6.

Demlo, L. K., P. M. Campbell, and S. S. Brown. 1978. ''Reliability of Information Abstracted from Patient's Medical Records.'' *Medical Care* 16 (12): 995–1005.

Doremus, H. D., and E. M. Michenzi. 1983. ''Data Quality. An Illustration of its Potential Impact upon a Diagnosis-related Group's Case Mix Index and Reimbursment.'' *Medical Care* 21 (10): 1001–11.

Feinstein, A. R., and D. V. Cicchetti. 1990. ''High Agreement but Low Kappa: I. The Problems of Two Paradoxes.'' *Journal of Clinical Epidemiology* 43: 543–9.

Fleiss, J. L. 1986. *The Design and Analysis of Clinical Experiments*. New York: John Wiley.

Glandon, G. L., M. A. Counte, and D. Tancredi. 1992. ''An Analysis of Physician Utilisation by Elderly Persons: Systematic Differences between Self-report and Archival Information.'' *Journal of Gerontology* 47(5): S245–52.

Holland, C. A., and P. M. Rabbitt. 1991. ''Ageing Memory: Use Versus Impairment.'' *British Journal of Psychology* 82 (part 1): 29–38.

Landis, J. R., and G. G. Koch. 1977. ''The Measurement of Observer Agreement for Categorical Data.'' *Biometrics* 33: 159–74.

Mackenbach, J. P., C. W. N. Looman, and J. B. W. Van der Meer. 1996. ''Differences in the Misreporting of Chronic Conditions, by Level of Education: The Effect on Inequalities in Prevalence Rates.'' *American Journal of Public Health* 86: 706–11.

Marrie, T. J., H. Durant, and E. Sealy. 1987. ''Pneumonia. The Quality of Medical Records Data.'' *Medical Care* 25 (1): 20–4.

Mossey, J. M., B. Havens, and F. D. Wolinsky. 1989. ''The Consistency of Formal Health Care Utilisation: Physician and Hospital Utilisation.'' In *Aging and Health Care*, edited by M. G. Ory and K. Bond, pp. 81–98. London: Routledge Press.

Muller, C. 1986. ''Review of Twenty Years of Research on Medical Care Utilization.'' *Health Services Research* 21 (2P+1): 129–44.

National Center for Health Statistics. 1987. ''Reporting Chronic Conditions in the National Health Interview Survey: A Review of Tendencies from Evaluation Studies and Methodological Tests.'' *Vital Health Statistics* Series 2 (105): 105–13. DHHS Pub. No. (PHS) 87-1379. Public Health Service. Washington, DC: Government Printing Office.

Ory, M. G., and K. Bond. 1989. ''Introduction: Health Care for an Aging Population.'' In *Aging and Health Care*, edited by M. G. Ory and K. Bond, pp. 1–21. London: Routledge Press.

Prigerson, H. G., P. K. Maciejewski, and R. A. Rosenheck. 1999. ''The Effects of Marital Dissolution and Marital Quality on Health and Health Service Use among Women.'' *Medical Care* 37 (9): 858–73.

Raina, P., B. Bonnett, D. Waltner-Toews, C. Woodward, and T. Abernathy. 1999. ''How Reliable Are Selected Scales from Population-based Health Surveys? An Analysis among Seniors.'' *Canadian Journal of Public Health* 90 (1): 60–4.

Raina P., D. Waltner-Toews, B. Bonnett, C. Woodward, and T. Abernathy. 1999. ''Influence of Companion Animals on the Physical and Psychological Health of Older People: An Analysis of a One-Year Longitudinal Study.'' *Journal of American Geriatric Society* 47: 323–29.

Raina, P. 1995. ''The Impact of Pet Ownership and Attachment on the Health and Health Care Use of the Elderly in Wellington County, Ontario.'' Dissertation. Ontario: University of Guelph.

Rawson, N. S. B., E. Malcolm, and C. D'Arcy. 1997. ''Reliability of the Recording of Schisophrenia and Depressive Disorder in the Saskatchewan Health Care Datafiles.'' *Social Psychiatry and Psychiatric Epidemiology* 32 (4): 191–9.

Roberts, R. O., E. J. Bergstralth, L. Schmidt, and S. J. Jacobsen. 1996. ''Comparison of Self-reported and Medical Record Health Care Utilisation Measures.'' *Journal of Clinical Epidemiology* 49: 989–95.

Romm, F. J., and S. M. Putnam. 1981. ''The Validity of the Medical Record.'' *Medical Care* 19 (3): 310–5.

Roos, N., and B. Havens. 1991. ''Predictors of Successful Aging: A Twelve-Year Study of Manitoba Elderly.'' *American Journal of Public Health* 81 (1): 63–8.

Spiers, M. V., and D. M. Kutsik. 1995. ''Self-reported Memory of Medication Use by the Elderly.'' *American Journal of Health-System Pharmacology* 52 (9): 985–90.

Statistics Canada. 1994. *National Population Health Survey 1994–1995*. Ottawa: Statistics Canada.

Strout, P. E., and J. L. Fleiss. 1979. ''Intraclass Correlations: Uses in Assessing Rater Reliability.'' *Psychology Bulletin* 86: 420–8.

Wallihan, D. B., T. E. Stump, and C. M. Callahan. 1999. ''Accuracy of Self-reported Health Services Use and Patterns of Care among Urban Older Adults.'' *Medical Care* 37 (7): 662–70.

Wan, T. T. H. 1989. ''The Behavioral Model of Health Care Utilisation by Older People.'' In *Aging and Health Care*, edited by M. G. Ory and K. Bond, pp. 52–71. London: Routledge Press.

Wiederholt, W. C., D. Cahn, N. M. Butters, D. P. Salmon, D. Krits-Silverstein, and E. Barrett-Connor. 1993. ''Effects of Age, Gender, and Education on Selected Neurophychological Tests in an Elderly Community Cohort.'' *Journal of American Geriatric Society* 41 (6): 639–47.