QUALITY OF LAST DOCTOR VISIT REPORTS: A COMPARISON OF MEDICAL RECORD AND SURVEY DATA

Gina M. Jay, Robert F. Belli, & James M. Lepkowski, Survey Research Center
Gina M. Jay, Survey Research Center, University of Michigan, 426 Thompson St, Ann Arbor, MI 48106

Key Words: Measurement Error, Health Care Utilization

INTRODUCTION

One means of better understanding how people interact with the health care system, and how such interaction may change over time, is to examine various characteristics of their health care visits (e.g., reason, location,). Since this information would be difficult to assess for all doctor visits, a potentially effective approach to tapping these data is to ask respondents detailed questions only about their last visit. The present report explores this approach by comparing information available in medical records with survey responses on several characteristics of the respondent's last doctor visit. Specifically, reason for last visit, health care professionals seen at last visit, location, lab tests performed at the last visit, and date of last visit, are evaluated in terms of the level of correspondence between the survey reports and the medical records. Subgroup analyses focussing on characteristics of the respondents and the recency of the last visit are also examined.

METHODS

Data derive from the Health Field Study (HFS), a record check study conducted in the metropolitan Detroit area in the spring and summer of 1993. The HFS was a face-to-face survey that covered a variety of health topics, including health care utilization, health insurance coverage, injuries, chronic conditions, health behavior, and mental health. Each interview lasted an average of 60 minutes. There were two versions of the survey, one that was administered to a control group and another to those in an experimental group. The control version was like a standard Survey Research Center survey in format and procedures. The experimental version had several enhancements designed to increase respondent commitment and motivation.

Sample

Respondents were sampled from the membership lists of an HMO in the metropolitan Detroit area. This HMO was selected because it could provide extensive medical record data for comparison with the survey reports. The criteria for sampling was that the individual was a member of the HMO at the time of sampling, and that they had had a health care visit in calendar year 1992. In addition, in order to allow a variety subgroup comparisons, three groups of persons were oversampled: youths aged 14-17, adults aged 65+, and Blacks.

A total of 2006 interviews were completed (66.6% response rate), with 1900 persons (95% of 2006) also providing access to their medical records. These respondents have been disaggregated into three groups, based on whether or not the records for their last visit were expected to be in the electronic database from which the medical record data were gathered (in records, not in records, and no medical access). The analyses presented here focus on the group of 1476 persons whose last visit was expected to be in the medical records. A group of 424 persons were excluded from analyses because they reported their last visit as occurring out-of-system ($\underline{n}$=300), or because their medical record data was expected to be incomplete for other reasons (e.g., routine blood pressure and sugar checks do not consistently get recorded in the electronic records). If all the data were analyzed together, the rates of correspondence could be attenuated by those cases whose last visit did not appear in the records.

Chi-square tests of independence were conducted to determine if there were any differences between the groups on a variety of background variables. The in records and not in records groups differ in age ($\underline{X}^2[6]$=13.274, $\underline{p}$<.05), sex ($\underline{X}^2[1]$=4.629, $\underline{p}$<.05), race ($\underline{X}^2[1]$=15.332, $\underline{p}$<.001), and whether or not respondents checked their personal records when answering the last visit questions ($\underline{X}^2[1]$=5.261, $\underline{p}$<.05). Specifically, young adults are more likely to be in the not in records group, while older adults are more likely to be in the in records group. Men and blacks are more likely to be in the not in records group than women and nonblacks. Persons who check their records when answering the last visit questions are more likely to be in the in records group than out of records group. The in records group differs from the no medical access group in age ($\underline{X}^2[6]$=14.368, $\underline{p}$<.05), race ($\underline{X}^2[1]$=9.912, $\underline{p}$<.01), and education ($\underline{X}^2[2]$=19.283, $\underline{p}$<.001). Older adults (excluding those aged 65+), blacks, and the higher educated are more likely to refuse access to medical records than youths and young adults, whites, and the less educated. No differences between groups were found in income, marital status, experimental condition, or self-rated health.

Measures

Measures used in the present analyses derive from two sources: the HFS survey questionnaire and from medical records. The medical record data were gathered by a medical record abstractor working with the health care system of which the HMO was a part. The medical

data used in the last visit analyses derive from outpatient and lab test data files; both are based on electronically stored data, not the paper medical chart. The visit used in the analyses is the last visit in the medical records.

There are measures of five different last visit characteristics from each of the data sources: reason for last visit, medical professionals seen at last visit, location of last visit, tests or procedures conducted at the last visit, and date of last visit. These measures are described briefly below. Often substantial recoding of both the medical record and survey data were required to make the two sources of information comparable.

Reason for Last Visit. In the survey, respondents were asked what was the reason for their last visit. The open-ended responses were coded using the National Ambulatory Medical Care Survey (NAMCS) Codes. NAMCS codes have seven modules for categorizing open-ended reason for visit responses: a symptom module; disease module; diagnostic, screening and preventive module; treatment module; injuries and adverse effects module; tests results module; and an administrative module.

In the medical record data there are ICD-9 diagnostic codes associated with each visit. The ICD-9 codes and the NAMCS codes were reduced from hundreds of different categories to 19 comparable ones. Eight of the categories refer to body systems (circulatory, respiratory, digestive genitourinary, musculoskeletal, skin, nervous system and sense organs, and endocrine/metabolic); four others refer to general types of illnesses, conditions or problems (neoplasms, infectious diseases, injury/poisoning, mental disorders), and seven others deal with diagnostics, treatments or symptoms that are not specific to a body system (general symptoms, treatments, medical counseling, tests/x-rays, follow-up/progress visits, medical exams, other).

Location of last visit. In the survey, respondents reported an open-ended location of last visit. This response was coded into numeric health site codes by the medical record abstractor. In the medical record data there is a unique site code associated with each health care location in the health system. These codes correspond directly to those that were assigned to the survey open-ended location description.

Professionals seen at last visit. In the survey, respondents were asked to choose from a list what kind of medical doctor or assistant they saw or talked to during their last visit. Multiple mentions of persons seen were coded. These data have been reduced to a primary care/ specialist dichotomy for analysis. Instances where a respondent provided nonspecific information that could not be matched to the medical records was dropped from the analyses ($\underline{n}$=81). These primarily included mentions of seeing a nurse, physician's assistant, or other nonspecific physician. General practitioners, internists, and pediatricians were coded as primary care providers. The remaining professionals were coded as specialists (e.g., allergist, dermatologist, cardiologist).

In the medical records there is information as to the specialty area that was billed for each visit, service or procedure. Examples of these specialty areas include Internal Medicine, Orthopedics, and Radiation Oncology. There may be more than one specialty associated with a given visit, depending on the nature of the visit. There is not a direct one-to-one correspondence between the listing of medical professionals provided in the survey instrument and the specialty areas provided in the medical record data. These data were also reduced to a simple primary care/specialist dichotomy. Family practice, primary care, internal medicine, pediatrics, emergency medicine, and walk-in clinic were coded as primary care. The specialist category included all other specialty categories in the medical record data (e.g., orthopedics).

Tests and procedures conducted at last visit. In the survey, respondents were asked whether four different types or classes of tests/procedures were conducted at their last visit. These four types are blood, urine and other fluid tests; x-rays and related imaging technologies; surgeries; and any other test or procedures. These variables have been coded into four yes/no dichotomies. In addition, a fifth more general dichotomy was derived from these variables indicating whether any test or procedure was conducted at this last visit. In the medical record data there are indications of all of these tests and procedures. They too have been recoded into five yes/no dichotomies.

Date of last visit. In the survey, respondents reported on the month, day and year of their last visit, although not all respondents could provide a complete date. In the medical record data, complete date information (i.e., Month, Day, Year) are provided for each date of service. The date information presented in this paper focus on the month and year of the last visit.

Demographic and Background Variables. Variables used in the subgroup analyses were coded as follows: Age (14-17, 18-24, 25-34, 35-44, 45-54, 55-64, 65+), Race (Black, nonblack), Education (less than high school, high school graduate, some college+), Income (less than 20K, 20-34,999, 35-69,999, 70+), Marital Status (married or living with partner, not married), Experimental group (experimental, control), Self-Rated Health (excellent/very good, good, fair/poor), Checked Records (yes, no), and Visit Recency (within 14 days of the interview, 15-28, 29-180, 181-365, 366+ days).

RESULTS

The results are presented in five sections corresponding to different characteristics of the last visit.

The type of comparison that could be made between the survey reports and the medical record data varied by the different characteristics. The analysis strategy and results for each section are therefore discussed separately. Similarly, the sample available for analysis varied by last visit characteristic such that the n varies across sections.

Reason for Last Visit

There were multiple mentions in both the survey and medical records regarding reason for last visit. In the survey, 56% of the sample had only one mention, 33% had two mentions, 9% had three mentions, and 2% had either four or five mentions. In the medical records, 91% of the sample had one mention, 9% had two mentions, and two people ($<$ 1%) had three mentions.

Table 1 shows the frequency distribution for the 19 reason categories for all mentions. The positive match rate for reason with these 19 categories is 41%. Here the medical reason would have to match the survey reason to be considered a match (e.g., circulatory in medical records and circulatory in survey is a match). If one considers only those respondents mentioning a body system as a part of their reason (the first eight rows of the table), the positive match rate is 52%. The marginal distribution for all mentions are similar for the body system and problem categories, and more dissimilar for the nonspecific diagnostic/treatment categories.

Chi-square analyses were conducted to examine potential subgroup differences in reason for last visit. A dichotomous indicator of match on reason or not was used in the analyses with the demographic and visit recency variables. The results indicate subgroup differences by age ($\underline{X}^2[6]=29.704$, $\underline{p}<.001$); the younger age groups had higher match rates than the older age groups.

Professionals Seen at Last Visit

Table 2 displays the crosstabulation of the medical professional variables derived from the survey and medical record data. The overall match rate between the two data sources for professional seen is 80% (46% primary care match + 34% specialist match). Nonmatches are due primarily to respondents reporting seeing a primary care professional at the last visit when the medical records indicate a specialist (15%). Only 5% of respondents report seeing a specialist when a primary care physician is indicated in the medical records.

In order to correct for chance levels of agreement, a Kappa statistic was computed. The Kappa value for the professional comparison was .599 ($\underline{X}^2[1]=806.06$, $\underline{p}<.001$), indicating moderate agreement between the medical records and survey data in making a primary care/specialist distinction. Subgroup Kappa analyses were conducted with significant differences found for sex ($\underline{X}^2[1]=4.86$, $\underline{p}<.05$), education ($\underline{X}^2[2]=11.56$, $\underline{p}<.01$), income ($\underline{X}^2[3]=9.28$, $\underline{p}<.05$), and

visit recency ($\underline{X}^2[4]=9.66$, $\underline{p}<.05$). Specifically, men have higher levels of agreement on professional seen than women. Those with a college education have a higher level of agreement than those with a high school degree or less. Persons making less than $20,000 a year have a significantly lower level of agreement than those making more than $20,000 a year. Finally, persons whose last visit is within six months of the interview have significantly higher levels of agreement than those whose last visit was more than one year prior to the interview.

Table 1. Reason for Last Visit: All Mentions

Reason Category	Survey ICD-9		Medical Records NAMCS	
	$\underline{n}$	%	$\underline{n}$	%
Body Systems				
Circulatory	93	6	161	10
Respiratory	208	14	194	13
Digestive	73	5	39	3
Genitourinary	171	12	151	10
Musculoskeletal	128	9	124	8
Skin	94	6	111	8
Nervous/Senses	147	10	152	10
Endocrine/Metab.	37	3	68	5
Problems not specific to a body system				
Neoplasms	36	2	47	3
Infectious Diseases	26	2	50	3
Injury/Poisoning	85	6	74	5
Mental Disorders	21	1	47	3
Other not specific to a body system				
General Symptoms	63	4	8	1
Treatments	201	14	0	0
Medical	73	5	6	<1
Counseling	286	19	228	15
Tests & X-rays	100	7	10	1
Follow-up	453	31	145	10
Medical Exams Other	29	2	4	<1
Total	2324	---[a]	1619	---[a]

[a]Percentages > 100 due to multiple mentions.

Location of Last Visit

The physical location of the last visit in the survey data is the open-ended response for location of last visit that was coded to correspond to the numeric site codes that appear in the medical record data. Table 3

Table 2. Professionals Seen at Last Visit: Primary Care and Specialists

Medical Records	Survey					
	Primary Care		Specialist		Total	
	n	%	n	%	n	%
Primary Care	**629**	**46**	64	5	693	50
Specialist	211	15	**471**	**34**	682	50
Total	846	61	535	39	1375	100

Kappa=.599, p <.001.

presents the crosstabulation of the medical record location with that coded from the survey data. There are two major medical centers included, one in the inner city denoted as Urban MMC, and one in the suburbs, denoted as Suburban MMC. The Smaller Suburban Centers category includes all remaining locations. The overall match rate is 92% (22% Urban MMC match, 27% Suburban MMC match, 43% smaller suburban center match). There is no clear pattern to the nonmatches; overall 5% of persons reported that their last visit occurred at a suburban facility, when the medical records indicate the urban facility.

The Kappa value for this comparison is high, .865 ($X^2[1]$=5803.70, p<.001), indicating a substantial level of agreement between the medical record and survey data in location of last visit. Subgroup Kappa analyses indicate significant age ($X^2[6]$=14.17, p<.05), race ($X^2[1]$=14.54, p<.001), and marital status effects ($X^2[1]$=5.49, p<.05) in levels of agreement. Specifically, those aged 18-24 have significantly higher levels of agreement than those aged 35-54 and those aged 65+. The 25-34 year old age group has significantly higher agreement on location than those aged 45-54. Blacks, and persons who are not married, have a higher level of agreement for location than nonblacks, and those who are not married.

Tests and Procedures Conducted at Last Visit

The crosstabulation of the survey reports of the five types of tests and procedures conducted at the last visit and comparable indicators from the medical record data are all presented in Table 4. The sample size is reduced for these analyses because youths, aged 14-17, were not asked lab test questions in the survey. In addition, the medical lab data for fluid and other tests is restricted for VIPs; these cases are also excluded from the fluid, other test, and any test analyses.

Blood, urine, or other fluid tests. As indicated in Table 4, the overall match rate between the medical record and survey data for fluid tests is 70% (20% positive match + 50% negative match). The nonmatches are primarily due to overreports (28%); only 2% of the nonmatches were underreports. The Kappa value for this comparison is .380 ($X^2[1]$=279.54, p<.001), indicating a fair level of agreement for reports of fluid tests conducted at the last visit. Subgroup Kappa analyses indicate significant education ($X^2[2]$=8.69, p<.05) and visit recency effects ($X^2[4]$=30.00, p<.001) on level of agreement. Specifically, persons with at least some college have a significantly higher level of agreement than those with a high school education or less. Persons whose last visit was within four weeks (28 days) of the interview had significantly higher levels of agreement than those whose last visit was more distant (29+ days). In addition, persons whose last visit was anywhere from 29 days to one year prior to the interview had higher agreement levels than those whose last visit was greater than one year prior to the interview.

X-rays and related imaging technologies. The overall match rate between the medical record and survey data for x-rays is 85% (10% positive match + 75% negative match). There were slightly more overreports (10%) than underreports (6%) in the report of x-rays conducted at the last visit. The Kappa value for this comparison is .449 ($X^2[1]$=195.17, p<.001), indicating a moderate level of agreement. Subgroup Kappa analyses indicate significant age ($X^2[5]$=13.06, p<.05) and visit recency effects ($X^2[4]$=21.63, p<.001) on level of agreement. Persons aged 18-34 have significantly higher levels of agreement than those aged 45-54 and those aged 65+. Persons whose last visit was within four weeks of the interview had significantly higher levels of agreement than those whose last visit was 29-180 days prior to the interview, and more than one year prior to the interview.

Surgical procedures. As shown in Table 4, the overall match rate for surgical procedures conducted at the last visit is 96% (2% positive match + 94% negative match). Nonmatches involved 3% overreports and 1% underreports. The Kappa value for this comparison is .479 ($X^2[1]$=62.82, p<.001), indicating a moderate level of agreement for reports of surgical procedures conducted at the last visit. Subgroup Kappa analyses indicate significant age ($X^2[5]$=20.29, p<.01) and marital status effects ($X^2[1]$=5.48, p<.05) on level of agreement. Persons aged 45-54 had significantly higher agreement on whether surgical procedures were conducted at the last visit than those aged 18-24 and 55+. Those aged 35-44 had higher agreement than those aged 18-24, and those aged 55-64. In addition, persons who are not married have a higher level of agreement on surgical procedures conducted than those who are married.

Table 3. Location of Last Visit: Urban Major Medical Center (MMC), Suburban MMC and Other Suburban Centers

Medical Records	Survey							
	Urban MMC		Suburban MMC		Other Suburban Ctrs		Total	
	n	%	n	%	n	%	n	%
Urban MMC	**318**	**22**	34	2	42	3	394	27
Sub MMC	9	<1	**392**	**27**	16	1	417	28
Sub Ctrs	11	<1	15	1	**629**	**43**	655	45
Total	338	23	441	30	687	47	1466	100

Kappa=.865, p<.001.

Table 4. Tests/Procedures Conducted at the Last Visit

Medical Records	Survey					
	Yes		No		Total	
	n	%	n	%	n	%
Fluid						
Yes	**248**	**20**	31	2	279	22
No	346	28	**628**	**50**	974	78
Total	594	47	659	53	1253	100
K=.380*						
X-rays						
Yes	**124**	**10**	79	6	203	16
No	129	10	**971**	**75**	1100	74
Total	253	19	1050	81	1303	100
K=.449*						
Surgical						
Yes	**27**	**2**	18	1	45	3
No	36	3	**1223**	**94**	1259	97
Total	63	5	1241	95	1304	100
K=.479*						
Other						
Yes	**52**	**4**	131	10	183	15
No	131	10	**938**	**75**	1069	85
Total	183	15	1069	85	1252	100
K=.162*						
Any test						
Yes	**507**	**40**	74	6	581	46
No	319	25	**355**	**28**	674	54
Total	826	66	429	34	1255	100
K=.388*						

K=Kappa. *p<.001.

Other tests and procedures. The overall match rate between the medical record and survey data for other tests or procedures conducted at the last visit is 79% (4% positive match + 75% negative match); with equal amounts of both over- and underreporting (10%). The Kappa value for this comparison is .162 ($X^2[1]$=22.19, p<.001), indicating a poor level of agreement. Subgroup Kappa analyses indicate significant race ($X^2[1]$=4.73, p<.05), education ($X^2[2]$=9.01, p<.05), and income effects ($X^2[3]$=8.39, p<.05) on level of agreement. Blacks have a higher level of agreement for other tests and procedures conducted at the last visit than nonblacks. Persons with a high school education had significantly lower agreement than persons with less than a high school education and those with at least some college. Persons whose family income is $20,000-34,999 have significantly lower agreement than persons with a family income of $35,000 or more.

Any test or procedure. The overall match rate for the derived variable, any test or procedure conducted at the last visit, is 68% (40% positive match + 28% negative match). Like fluid tests, nonmatches are primarily due to overreporting (25%); only 6% of the nonmatches were underreports. The Kappa value for this comparison is .388 ($X^2[1]$=266.83, p<.001) indicating a fair level of agreement for reports of any lab test or procedure conducted at the last visit. Subgroup Kappa analyses indicate a significant sex effect ($X^2[1]$=4.04, p<.05) on level of agreement; levels of agreement are significantly higher for women than for men.

Date of Last Visit

The proportion of persons matching exactly on month and year of last visit, and those matching within one month was calculated. Overall, there was a 57% match rate for exact month and year of last visit, and a 75% match rate for visits within one month.

Chi-square analyses were conducted to examine potential subgroup differences in date of last visit. A dichotomous indicator of exact match on date or not was used in the analyses with the demographic and visit recency variables. The results indicate subgroup differences by sex ($X^2[1]$=10.752, p<.001), education ($X^2[2]$=14.532, p<.001), self-rated health status

$(\underline{X}^2[2]=8.195, \underline{p}<.05)$ whether or not respondents checked records when answering the last visit questions $(\underline{X}^2[1]=25.469, \underline{p}<.001)$, and visit recency $(\underline{X}^2[4]=218.991, \underline{p}<.001)$. Specifically, women, those with a higher education, those with better self-rated health, those who checked their records, and those with more recent visits, had higher levels of agreement on the date of the last visit than men, the less educated, those with poorer self-rated health, those who did not check their records, and those whose last visit was more distant.

DISCUSSION

There is variability in the level of agreement between the survey and medical record data by last visit variable. The positive match rate for reason is on the low side if considering the entire sample (41%), but better if examining only those people who reported a condition or symptom referable to a particular body system (52%). There is high agreement for location, moderate agreement for professionals seen, x-rays conducted at the last visit, and surgery conducted at the last visit, fair agreement for fluid tests conducted at the last visit and any test or procedure conducted at the last visit, and poor agreement for other tests and procedures conducted at the last visit.

The survey and medical record marginal distributions for the reason categories are similar for reasons referable to a body system or problem, but lower for general, nonspecific reasons (e.g., follow-ups, exams, treatments). This is in part due to the lack of direct correspondence between the ICD-9 and NAMCS codes. Is may also be linked to discordance in how physician's describe a visit for billing purposes (ICD-9 codes) and how respondents describe the reason they went to see the doctor. It appears that physicians are specific whenever possible, coding a condition or symptoms referable to a body system. Respondents, on the other hand, use more general terminology, like "check-up" or "got a shot," without necessarily stating for what condition the visit occurred.

Agreement on location may be high because many people are reporting on the only place, or the usual place that they go for care. In this sense, any visit in the medical records may have matched on location. Data are not available to test this hypothesis directly, although it does seem a reasonable one. Additional analyses (not reported here) indicate that the subgroup differences in agreement on location are largely due to the Urban MMC. This center is located in the inner city and is likely the location where most urban blacks go for health care.

The moderate agreement level found for professionals seen indicates that respondents are capable of making a distinction between primary care providers and specialists, although some to a better degree than others. Not unexpectedly, the better educated and those with a higher income had higher levels of agreement on this variable.

The moderate agreement levels found for the report of x-rays and surgical procedures may be due to the fact that these procedures are salient to respondents. That is, imaging technologies and surgical procedures may be remembered because they tend to require special equipment, specific appointments, and may also be more invasive than other tests or procedures.

The poor to fair levels of agreement on fluid and other tests may be linked to several factors. It is possible that respondents were unsure how to classify different types of tests. Respondents may also have reported on tests and procedures that were never recorded in the medical records. Consider for example, having one's blood pressure taken, or providing a urine sample for a quick sugar test. In addition, it appears that respondents report anything that was "ordered" at a visit, or for a visit, as occurring at that visit, even if it was ultimately done another day or at a separate location. Among these things are blood tests, and the x-ray or imaging technologies, and any sort of tests that may require fasting or other special preparation. Therefore, tests occurring before or after the actual visit with the physician appear to be reported by the respondent as occurring as a part of that visit. All of these factors work together to increase overreporting and lower agreement levels between the medical and survey data.

A consistent subgroup effect in the agreement analyses was recency of the last visit (recency was also associated with higher match rates for date of visit). For fluid tests and x-rays, agreement levels were significantly higher for last visits occurring within four weeks of the interview, than those occurring at a more distant time. This indicates that data on such tests and procedures may be most accurate for visits occurring within four weeks of the interview.

Limitations to these analyses are related to the characteristics of the HFS sample, as well as characteristics of the data. Respondents to this study all had recent utilization, and received their health care in a single HMO system. As such, generalizability to the general population is severely limited. Persons with less recent utilization, or those who have different types of health insurance coverage, may have different levels of accuracy in reporting the characteristics of their last health care visits. The data reported here also have shortcomings. The most noteworthy being the fact that the electronic medical record system does not include all visits that are covered by the HMO, and the lack of direct correspondence between the medical record data and the survey data for the last visit variables. If the data were more directly comparable, better agreement may have been found on some of the last visit variables examined.