Paper 1675-2014

Validating Self-Reported Survey Measures Using SAS®

Sarah A. Lyons MS, Kimberly A. Kaphingst ScD, Melody S. Goodman PhD
Washington University School of Medicine

ABSTRACT

Researchers often rely on self-reported responses for survey research studies. The accuracy of this self-reported data is often unknown, particularly in a safety-net medical setting that serves a patient population with varying levels of health literacy. We recruited participants from the waiting room of a St. Louis primary care safety net clinic to participate in a survey investigating the relationship between health environments, health behaviors and health outcomes. The survey included questions regarding personal and family history of chronic disease (diabetes, heart disease, and cancer) and self-reported height and weight (used to calculate Body Mass Index). We subsequently accessed the participant's electronic medical record (EMR) and collected physician-reported data on the same variables. We calculated concordance rates between participant answers and information gathered from EMRs using McNemar's test. Logistic regression was then performed to determine the demographic predictors of concordance. Six hundred sixty two patients completed surveys as part of the study; 69% female, 55% African American, 17% with less than high school level education, 74% annual household income less than $20,000, and 27% uninsured. Preliminary findings suggest a 75-95% concordance rate between self-reported and medical record data across outcomes, with the exception of family history of cancer (70%) and heart disease (61%). Determining the validity of self-reported data influences whether self-reported personal and family history of disease and BMI are appropriate proxies for medical record extraction in this patient population.

INTRODUCTION

Researchers often rely on self-reported participant responses for personal and family history of disease when conducting survey research. While a full medical examination is seen as the gold standard, this is rarely feasible. Pulling the desired information from the participant's electronic medical record (EMR) is viewed as the best approximation but it is often too impractical to conduct a medical record review for every study participant in large survey studies. These reviews can be a drain to resources, both with regards to time and manpower. Given the newness of EMRs in primary care settings, they are limited by the amount of information available; with varying dates of implementation, varying levels of physician and medical staff use, and varying amounts of historical archives. The degree to which self-reported participant responses are concordant with the information in their EMR is an important area of investigation. Concordance varies between populations and between health conditions[1]. Health literacy can influence reporting knowledge of medical diagnoses and family disease history, as there might be lower recognition of medical terms and outcomes that are related to specific conditions[2].

SAS® is an excellent tool to analyze the concordance between the two methods of reporting: self-report determined through survey questions and physician-report determined through the EMR. The display of the 2x2 table allows for visualization of the degree (magnitude) and direction of the agreement. The SAS® PROC FREQ and PROC LOGISTIC procedures are built in with various statistical tests for determining the significance of the agreement as well. In order for self-reported responses to be valid in this patient population, we want high agreement between self- and physician-reported data and non-significant p-values when testing for the differences between reporting groups.

STUDY DESCRIPTION

Participants were recruited from the waiting room of a safety net primary care clinic that serves a low-income population living in the St. Louis metro area. Data from 662 patients are used in this analysis. The majority of this population was female (69%), African American (55%) and had an annual household income of less than $20,000 (74%). The written survey contained 120 questions designed to capture each participant's health environments (using both written and visual measures) and health information-seeking behaviors as well as collecting data on health outcomes. Health literacy was assessed using the Rapid Assessment of Adult Literacy in Medicine-Revised[3,4]. After survey completion, each participant's EMR was accessed by trained research assistants; data on personal and family history of disease was extracted.

METHODS

We examine the degree of agreement between participants' self-reports of disease diagnosis and family disease history with physician/medical staff-reported information in their EMR. In order to protect each participant's privacy, survey answers and medical records were linked using a unique numeric identifier and were de-identified of any personal identifiers. It is imperative that the survey variables and the medical record variables are coded consistently through a standardized process. In this dataset, "Never received a diagnosis" was coded as "0" and "Have received diagnosis" was coded as "1".

BMI was calculated from height and weight recorded in the EMR. Self-perception of weight was determined from the survey; participants were asked if they considered themselves to be average weight, overweight or obese. Responses from the survey and the EMR were dichotomized into an outcome variable with '0' being average or underweight and "1" categorizing those who were overweight or obese.

To calculate agreement between the two methods of reporting, the PROC FREQ procedure was utilized. This procedure allows for various methods of determining concordance and its significance. The PROC LOGISTIC procedure was then used to determine if there were any demographic factors that predict concordance rates.

Demographic data collected from the survey included gender, age, education, marital status, race/ethnicity, urban status, health literacy, income, insurance status and familiarity with the concept of a family health history. Age was divided into four categories; <25, 26-35, 36-49, and over 50. Education was categorized into Below High School Education, High School Degree or Equivalent and Above High School. Marital status was dichotomized into "Married" and "Not Married". Race was classified as Non-Hispanic Black, Non-Hispanic White and Other. Urban status was categorized as "Urban" for those living in zip codes within St. Louis City, "Suburban" for those living in zip codes in St. Louis County and "Rural" for those living in surrounding parts of Missouri. Health literacy was approximated using the participant's REALM-R score; a score of 6 or less indicates limited health literacy. Income was dichotomized as an annual household income of less than $20,000 and a household income of greater than or equal to $20,000. Insurance status was classified as being uninsured, having public insurance (Medicaid or Medicare) or having private insurance. On the survey, participants were asked if they had ever heard of a family health history before the survey and the response options were "Yes" or "No".

PROC FREQ – CONCORDANCE RATES

		Electronic Medical Record Report of Heart Disease		
		No	**Yes**	**Total**
Survey Report of Heart Disease	**No**	349 **70.79** 93.57 85.96	24 4.87 6.43 27.59	373 75.66
	Yes	57 11.56 47.50 14.04	63 **12.78** 52.50 72.41	120 24.34
	Total	406 82.35	87 17.65	493 100.00

Figure 1. 2x2 Frequency Table – PROC FREQ Output

Agreement was calculated by adding the concordant percentages; total percent of those who had self- and physician-reported diagnosis of the disease and those who had self- and physician-reported diagnosis of not having the disease (No-No's and Yes-Yes's) as highlighted in Figure 1. For heart disease, agreement is those who have both self-reported history and EMR-reported history of having heart disease (12.78%) and then add those who have no reported history of heart disease, both in the survey and on the EMR (70.79%) to get a total agreement of 83.57%.

To determine the agreement between the survey and EMR, we use McNemar's test instead of a more general Chi-square test to account for the paired responses. McNemar's test takes into account that the two measures (self-report and medical record) are correlated; they are measuring the same participant's history of disease. The null hypothesis for McNemar's test is that there is no difference in the two correlated proportions. We want to fail to reject the null hypothesis (obtain a p-value > than 0.05) as this will give evidence that self-report of personal and family history of disease is an adequate proxy for medical record extraction in this patient population. This test can be called from both the **`tables`** statement or the **`exact`** statement with the **`mcnem`** option. In this case, either method will produce the same results. These statements will also produce the Kappa coefficient which checks for inter-rater reliability. The automatically generated plots were suppressed using the **`plots=none`** option in the **`tables`** statement. These plots would be useful to visualize the direction of agreement.

The PROC FREQ procedure was called using the following code to determine agreement between self- and physician-reported diagnoses of cancer:

```
proc freq data = total;
      tables quest_34*pers_heartdisease / agree plots=none;
run;
```

The output displays the two-by-two table shown in Figure 1. From there, agreement can be calculated. The output also provides a table detailing the results of McNemar's Test (Table 1). The p-value is significant, indicating that there is discordance between what the participant reports on their survey and what is recorded in the EMR for personal history of heart disease. McNemar's test p-value doesn't indicate the direction of the discord, just the degree of significance. To further investigate, we must look at the two-by-two table in Figure 1. We can see that, of those that disagree, more reported heart disease on the survey than were reported in the participant's EMR (n=57; 12%). The Kappa coefficient is 0.4303, indicating Moderate agreement, as defined by Landis and Koch[5].

Statistic (S) **DF** **Pr > S**	13.4444 1 0.0002
Kappa	0.4303

Table 1. McNemar's Test and Simple Kappa Coefficient

More encouraging results were found when investigating the agreement in personal history of diabetes (Figure 2). Substantial agreement (Kappa=0.8840) was found as well as an insignificant p-value (p=0.2568) from McNemar's test (Table 2). The percent concordant was 94.44% (57.54%+36.90%). In this population, the self-reported diabetes survey question is an adequate proxy for disease diagnosis as stated in the EMR.

		Electronic Medical Record Report of Diabetes		
		No	**Yes**	**Total**
Survey Report of Diabetes	**No**	290 **57.54** 96.35 94.46	11 2.18 3.65 5.58	301 59.72
	Yes	17 3.37 8.37 5.54	186 **36.90** 91.63 94.42	203 40.28
	Total	307 60.91	197 39.09	504 100.00

Figure 2. 2x2 Frequency Table – PROC FREQ Output

Statistic (S) **DF** **Pr > S**	1.2857 1 0.2568
Kappa	0.8840

Table 2. McNemar's Test and Simple Kappa Coefficient

PROC LOGISTIC – PREDICTION

To understand what drives concordance rates, logistic regression was performed using an individual's concordance as a dichotomous outcome variable and their demographic factors as predictors. An if/else statement was used to create the dichotomous outcome variable (`fdm_indic`). If the survey and the EMR agreed for the outcome of interest (both '0', or both '1'), then the indicator was coded as '1' for agreement. Otherwise, the indicator was coded as 0 for disagreement.

```
if f_diabetes = . then fdm_indic=.;
      else if quest_13a = . then fdm_indic=.;
      else if f_diabetes = 1 and quest_13b = 1 then fdm_indic = 1;
      else if f_diabetes = 0 and quest_13b = 0 then fdm_indic = 1;
      else fdm_indic = 0;
```

Using the PROC FREQ procedure with the `chisq` option in the `tables` statement we examine if there were differences in concordance rates by demographic factors. Results from bivariate analyses are used to determine the demographic predictors that will be tested in the development of a prediction model for concordance between EMR and self-reported responses. Predictors with a p-value <0.1 in bivariate analyses were included in logistic regression analysis.

Prediction models were created for concordance between self-report and EMR for all outcomes that had significant bivariate associations (BMI, personal history of cancer, and family history of diabetes, heart disease and cancer). The following code creates a model for concordance of family history of diabetes.

```
proc logistic data=total;
      class fdm_indic (ref="0") gender insured income urban / param=ref;
      model fdm_indic = gender insured income urban / selection = stepwise;
run;
```

The **`selection = stepwise`** option allows SAS to build the most significant model using an automated process. This model will include only significant predictors above a certain threshold; the default value is 0.05. This can be changed by using the `slentry` or `slstay` options in the `model` statement which specify the significance for a variable to be entered into the model and the significance for it to stay in the model, respectively. Odds ratios are also produced that will allow for determining demographic subgroups that are most likely to report personal disease diagnosis and family disease history in agreement with their medical record. It will also determine groups for which there are high levels of discordance and for whom targeted interventions should be developed.

RESULTS

	N	% Self- Report	% Physician-Reported	% Agreement	McNemar p-value	Kappa
Personal History						
Diabetes	504	40.28	39.06	94.44	0.26	0.88
Heart Disease	493	24.34	17.65	83.57	<0.001	0.51
Cancer	504	16.47	12.70	92.66	<0.001	0.71
Family History						
Diabetes	369	80.76	63.96	74.8	<0.001	0.40
Heart Disease	325	72.62	44.92	61.23	<0.001	0.36
Cancer	320	70.94	54.69	69.38	<0.001	0.36
BMI						
Obese	435	63.22	83.45	76.55	<0.001	0.43

Table 3. Reported Prevalence and Agreement between Self-Report and EMR

In every case (personal and familial), the prevalence of survey-reported disease is higher than that found in the EMR. These differences are significant in all cases except for personal history of diabetes (Table 3). We can conclude that participant's self-report of diabetes is a reliable way to ascertain a true personal history of the disease (when using the EMR as a gold standard) in this population. Despite the statistically significant rates of discordance there is a fair amount of agreement in all categories, with all personal diagnoses above 80% and all family history above 60%; family history of heart disease has the lowest percent agreement (61.23%).

	Predictor	X^2	P-Value
Personal History			
Diabetes	*None*	---	---
Heart Disease	*None*	---	---
Cancer	Education	5.69	0.06
BMI			
Obese	Gender	10.07	<0.01
	Education	5.63	0.06
	Health Literacy	2.79	0.09
	Insurance Status	9.54	0.01

Table 4a. Bivariate Associations Between Concordance Rates and Demographic Predictors

	Predictor	X^2	P-Value
Family History			
Diabetes	Gender	5.09	0.02
	Urban Status	4.65	0.10
	Income Status	2.99	0.08
	Insurance Status	7.06	0.03
Heart Disease	Health Literacy	2.76	0.10
Cancer	Urban Status	5.19	0.07
	Insurance Status	6.27	0.03
	Health Literacy	2.86	0.09

Table 4b. Bivariate Associations Between Concordance Rates and Demographic Predictors

No bivariate associations (p<0.1) were found for personal history of diabetes and heart disease (results not shown). Significant associations (p<0.1) were found with education, gender, urban status, income, insurance status and health literacy (Tables 4a and 4b). Three models were built, that predicted concordance of BMI and self-perception of weight, family history of cancer, and family history of diabetes. Models were not built for personal history of diabetes and cancer because there were no significant predictors. Results are not shown for personal and family history of heart disease models because no predictors met the entry threshold (`slentry` = 0.05) to be entered in the model.

Models	Predictors	Odds Ratio (95% Wald Confidence Limits)	Wald X^2	P-Value
Model 1. Family History of Diabetes Concordance	Gender			
	(Female vs. Male)	0.43 (0.20, 0.94)	4.45	0.03
	Urban			
	(Rural vs. Urban)	0.24 (0.09, 0.67)	7.44	0.01
	(Suburban vs. Urban)	0.61 (0.31, 1.18)	2.19	0.14
	Income			
	(< $20,000 vs. ≥ $20,000)	0.27 (0.12, 0.80)	5.83	0.02
	Insurance			
	(None vs. Public)	2.68 (1.21,5.95)	5.86	0.02
	(Private vs. Public)	2.09 (0.78, 5.56)	2.16	0.14
Model 2. Family History of Cancer Concordance	Insurance			
	(None vs. Public)	2.04 (1.10, 3.74)	5.25	0.02
	(Private vs. Public)	1.75 (0.80, 3.8)	1.98	0.16
Model 3. BMI and Self-Perception of Weight Concordance	Gender			
	(Female vs. Male)	2.09 (1.28, 3.40)	8.69	<0.001
	Insurance			
	(None vs. Public)	0.87 (0.49, 1.54)	0.23	0.63
	(Private vs. Public)	0.36 (0.20, 0.66)	10.86	0.001

Table 5. Odds Ratios from Logistic Regression Models

As shown in Table 5, Model 1 has the most predictors, with gender, urban status, income and insurance being the significant predictors (p<0.05). Being female, living in rural or suburban Missouri and having low income were associated with higher odds of discordance in family history of diabetes. Having no insurance was associated with higher odds of concordance in family history of diabetes, when compared against having some form of public insurance.

Model 2 only had one significant predictor. In Model 2, having no insurance was associated with higher odds of concordance in family history of cancer, when compared to only public insurance. In Model 3, there were two predictors. Being female was associated with higher odds of concordance between BMI and self-perception of

weight. Having private insurance was associated with lower odds of concordance, when compared to public insurance.

CONCLUSIONS

The methods described here can be used to validate any self-reported survey response against a gold standard. This analysis should be performed on any pilot data to aide in the decision to continue with time-consuming medical record or chart reviews for clinical outcomes. It is important to note a major limitation of this study is the use of EMR as the gold standard for reporting. The results are subject to the validity of the EMR for reporting personal and family disease history. There is a high likelihood that both personal and family history is under-reported on this population's EMRs. Personal history of diabetes does not seem to be under-reported and perhaps is a focus of collection and reporting for this clinic. Further research with a larger sample size should be done to determine what is driving the discordance of the other diseases. This will allow for the identification of certain demographic groups that are more vulnerable to discordance in reporting. Identifying groups that are at high-risk for discordance will allow for targeted interventions that will aim to increase a patient's understanding of personal and family history as well as their importance in prevention and disease management. Providers will also be able to identify which groups should be focused on in terms of collecting and recording the information for their EMR. Despite the statistically significant differences in self- and physician-reported data, we see high (>80%) rates of agreement among personal disease diagnosis and fair (>60%) amount of agreement in the reporting of family disease history.

REFERENCES

1. Bush TL, Miller SR, Golden AL, Hale WE. November 1989. "Self-report and medical record report agreement of selected medical conditions in the elderly." *Am J Public Health*. 79:1554–6.
2. Smith B, Chu LK, Smith TC, Amoroso PJ, Boyko EJ, Hooper TI, Gackstetter GD, Ryan MA. June 2008. "Challenges of self-reported medical conditions and electronic medical records among members of a large military cohort." *BMC Med Res Methodol*. 8:37.
3. Davis WL, Demaio TJ. 1993. "Comparing The Think Aloud Interviewing Technique With Standard Interviewing In The Redesign Of A Dietary Recall Questionnaire." *Proceedings of the Section on Survey Research Methods, American Statistical Association*. 565–570.
4. Davis TC, Crouch MA, Long SW, et al. 1991. "Rapid Assessment of Literacy Level of Adult Primary Care Patients." *Fam Med*. 23(6):433–435.
5. Landis JR, Koch GG. 1977. "The measurement of observer agreement for categorical data." *Biometrics*. 33:159-174

ACKNOWLEDGEMENTS

The project and work of Sarah Lyons is supported by funding from the Barnes Jewish Hospital Foundation, Siteman Cancer Center and Washington University School of Medicine. The work of Dr. Goodman is supported by the Barnes-Jewish Hospital Foundation, Siteman Cancer Center, National Institutes of Health, National Cancer Institute grant U54CA153460, Washington University School of Medicine (WUSM) and WUSM Faculty Diversity Scholars Program. Dr. Kaphingst is supported by funding from the Barnes Jewish Hospital Foundation, Siteman Cancer Center and Washington University School of Medicine.

CONTACT INFORMATION

Name: Sarah Lyons
Organization: Division of Public Health Sciences, Washington University School of Medicine in St. Louis
660 S Euclid Ave, Campus Box 8100 || St. Louis, MO, 63110
Work Phone: 314-747-2568
Email: lyonss@wudosis.wustl.edu