Title

Lower agreement on behavioral factors than on medical conditions in self-reported data among pregnant Latina women

Permalink

https://escholarship.org/uc/item/872700gt

Journal

Archives of Medical Research, 35(3)

ISSN

0188-4409

Authors

Hessol, N A
Missett, B
Fuentes-Afflick, E

Publication Date

2004-05-01

Peer reviewed

Assessing Agreement between Self-Reported and Medical Record Data among Pregnant Latina Women

Nancy A. Hessol[1], Brian Missett[2], and Elena Fuentes-Afflick[3-5]

Authors affiliations: [1] Department of Medicine, University of California, San Francisco, CA. [2] Department of Radiation Oncology, University of California, San Francisco, CA. [3] Department of Pediatrics, University of California, San Francisco, CA. [4] Department of Epidemiology and Biostatistics, University of California, San Francisco, CA. [5] Medical Effectiveness Research Center for Diverse Populations, University of California, San Francisco, CA.

Corresponding Author: Nancy A. Hessol, MSPH, Department of Medicine, UCSF, 405 Irving Street, 2nd Floor, San Francisco, CA 94122. Email: nancyh@itsa.ucsf.edu; telephone: 415-502-6281; fax: 415-476-8528.

Running head: Assessing agreement in risk factors

Abstract word count: 218   Text word count: 2,184

ABSTRACT

Background. Agreement between self-reported data and data obtained from medical records is far from perfect and very few studies have analyzed the element of language, when self-reported data are given in one language and this information is recorded in another language in the medical record. Our objective was to assess agreement between self-reported data and medical record data on prenatal risk factors in pregnant Latina women.

Methods. We interviewed 350 Latina women at $\geq$ 20 weeks' gestation regarding alcohol use, tobacco use, use of prenatal vitamins, age, education, use of prenatal care and medical conditions. The Kappa statistic (κ) and 95% confidence intervals were used to calculate agreement between self-reported responses and medical record data. Multiple logistic regression analysis was used to evaluate the effect of maternal characteristics on likelihood of disagreement.

Results. Agreement between self-reported and medical record data was generally lower for behavioral factors (alcohol κ=0.37 and prenatal vitamin use κ=0.09) than for medical conditions (anemia κ=0.63, gestational diabetes κ=0.83 and hypertension κ=0.68). In general, maternal characteristics did not significantly predict patterns of disagreement.

Conclusions. Among pregnant Latina women, self-reported data on behavioral factors had lower agreement than self-reported data on medical conditions. Further study is needed to define the effect of other factors, such as social norms, on the accuracy of self-reported data during pregnancy.

Introduction

Even under the best of circumstances, agreement between self-reported data and data obtained from medical records is far from perfect (1). Numerous studies have shown that several characteristics influence the accuracy of self-reported data, including age, education and gender (2-7). Among pregnant women, who are likely to be counseled not to smoke cigarettes or consume alcohol, there may be greater pressure to misreport if they smoke (8, 9) or if they drink (10, 11). Additionally, medical records may not contain complete or accurate patient information (12), and locating specific information in the medical record is often difficult. Finally, very few studies have analyzed the element of language, when self-reported data are given in one language and this information is recorded in another language in the medical record.

In California, Latina women are the majority of childbearing women (13). Latina and White women have similar rates of low birth weight infants despite the socioeconomic disadvantage of Latina women (14, 15). To explain this phenomenon, childbearing Latina women are hypothesized to have fewer behavioral or nutritional risk factors than White women (16-19). Previous studies have reported that Latinas have lower rates of prenatal tobacco use (20) than White women and non-pregnant Latinas are reported to consume more nutritious diets than White women (16).

Information about risk factors, including behavioral and nutritional factors, can be collected by self-report, medical record review or measured using blood or saliva sampling. Many studies have relied on self-reported data due to limitations of time and money, but it is not clear if the accuracy of self-reported data is consistent across types of variables.

The objective of this study was to evaluate agreement between self-reported data and data obtained from the prenatal medical record in a cohort of pregnant Latina women. We focused on

six behavioral and medical factors that are associated with infant birth weight (21): maternal alcohol use, tobacco use, prenatal vitamin use, anemia, gestational diabetes and hypertension. We also tested whether disagreement between self-reported data and medical record data varied by maternal characteristics.

Materials and Methods

The Latino Health Project is a prospective cohort study based at San Francisco General Hospital, the only public hospital in San Francisco County. From the hospital-based prenatal clinic we recruited 350 Latina women at $\geq$ 20 weeks' gestation. Ethnicity was defined by self-identification. All participants were 15-44 years old, and planned to reside in the San Francisco Bay Area throughout the study period. The women answered 155 questions in a 45-minute structured interview administered by bilingual (Spanish and English) research assistants. The Committee on Human Research at the University of California, San Francisco approved this study and all women participants signed informed consent, in either Spanish or English, prior to enrollment.

For this analysis, we focused on six maternal risk factors: alcohol use, tobacco use, prenatal vitamin use, anemia, gestational diabetes and hypertension. Following the interview, one author (BM, a second year medical student at the time) reviewed the medical record at San Francisco General Hospital and abstracted information onto a standardized data collection form. Many of the providers and staff members working in the Prenatal Clinic at San Francisco General Hospital were bilingual (Spanish-English). The hospital also had an interpreter service available.

To measure agreement between self-reported and medical record data, we calculated the Kappa statistic (κ) and 95 percent confidence interval (95 percent CI). Agreement was defined as concordant responses in self-reported and medical record data. For example, negative responses in both the interview and the medical record were defined as agreement. We did not include the responses for women who did not answer an interview question, did not know the answer to a question or had data missing from the medical record.

To determine whether maternal factors affected the likelihood of disagreement between self-reported and medical record data, we performed multivariate logistic regression analysis. Two dependent (outcome) variables were defined: disagreement on any behavioral characteristic (alcohol, tobacco or prenatal vitamin use) and disagreement on any medical condition (anemia, gestational diabetes or hypertension). Seven potentially confounding variables were included in the multivariate regression models. Maternal age was categorized as <20, 20-26 and >26 years (reference) and education was categorized as <7 and ≥7 years (reference). Marital status was dichotomized as married or non-married. Maternal birthplace was categorized as Central or South America, Mexico (reference), and United States. For foreign-born women, the number of years residing in the United States was categorized as <4 and ≥4 years (reference). Language used for reading and speaking was dichotomized as primarily English (including equal English and Spanish) and primarily Spanish (reference). Prenatal health insurance was dichotomized as insured (MediCal [Medicaid in California], other public insurance, or private [including health maintenance organization or prepaid health plan]) or uninsured (self-pay or none). The insured category was the reference group. All statistical analyses were performed using SAS Institute software (22).

Results

From the original cohort of 350 women recruited for the study, we reviewed 321 medical records. Twenty-nine medical records were not available for review, most likely because they were in storage due to greater than 12 months of inactivity. The women who did not have their medical records reviewed were similar, demographically, to the women whose medical records were reviewed. The mean age for women in both groups was 25 years. The women without medical records available were more likely to be U.S.-born (17 percent) compared to the women with medical records available (5 percent, $P = 0.02$).

The study participants were young adults (mean age 25.3 years) and had low educational attainment (mean 8.7 years, Table 1). The majority of women (57 percent) were born in Mexico. Among foreign-born women, nearly half had resided in the United States less than four years. Nearly half of all women in the study were married. Almost ninety percent of women reported that Spanish was their primary language (75 percent only Spanish, 12 percent mainly Spanish). The vast majority of women (84 percent) received prenatal health insurance through MediCal, the state-run Medicaid program.

Agreement between self-reported and medical record data varied by risk factor. The highest agreement between self-reported and medical record data was found for gestational diabetes (κ 0.83, 95 percent CI 0.66-0.99, Table 2), followed by tobacco use (κ 0.79, 95 percent CI 0.66-0.93), hypertension (κ 0.68, 95 percent CI 0.46-0.90) and anemia (κ 0.63, 95 percent CI 0.50-0.75). Agreement was lowest for maternal alcohol use (κ 0.35, 95 percent CI 0.20-0.50) and use of prenatal vitamins (κ 0.09, 95 percent CI 0.02-0.17).

Among the women with complete data who were included in the multivariate logistic regression analysis for disagreement on any behavioral risk factor, nearly half of women

(133/269, 49 percent) had discordant responses for at least one behavioral factor. After adjusting

for the full set of variables, the only factor that was independently associated with disagreement

on behavioral risk factors was language use. Surprisingly, women who primarily spoke Spanish

were less likely to have disagreement between self-reported and medical record data on any of

the behavioral risk factors than women who primarily used English (Odds ratio [OR] 0.33, 95

percent CI 0.12-0.90, Table 3).

In the multivariate logistic regression analysis for disagreement on any medical

condition, 12.5 percent of women (37/296) had discordant responses for at least one medical

condition. There were no statistically significant predictors for disagreement on medical

conditions (Table 4).

Discussion

In this study of pregnant Latina women in San Francisco, we found variation in

agreement of self-reported and medical record data on prenatal risk factors. As reported in other

studies (1), we found a pattern of lower agreement for most behavioral risk factors compared to

medical conditions. This is particularly noteworthy since self-report of medical risk factors is

conditional on this information being conveyed, in a comprehensive fashion, from the clinician

to the patient and then the patient to the interviewer. Meanwhile, self-report of behavioral

information is directly based on report from the patient to the interviewer.

In previous studies that examined the accuracy of self-reported behaviors with drug test

results among pregnant women, most have found fairly good agreement on usage with slight

under-reporting of amount of alcohol consumption (10, 11) or number of cigarettes smoked (23,

24). In our study, we did not perform laboratory tests measuring drug levels for cigarette smoking or alcohol use and therefore could not assess validity.

In contrast to other studies, however, we did not find that maternal characteristics such as maternal age (2, 5, 25) or education (4, 5) were consistently associated with the likelihood of disagreement. Because our data were collected prospectively, it is unlikely that recall bias affected our results.

Comparing the three behavioral risk factors, the Kappa statistic for tobacco use (0.79) was higher than the Kappa statistic for alcohol use (0.35) or prenatal vitamin use (0.09, Table 2). Interestingly, the percentage of Latina women in our study who had self-reported and medical record data that documented their use of tobacco during pregnancy was approximately the same as for use of alcohol (approximately 6 percent). The higher Kappa statistic for tobacco use was similar to the Kappa statistics for the medical conditions.

The prevalence of smoking during pregnancy in the US has fallen from 17% to 12% over the last decade (20). The low prevalence of maternal tobacco use in our study could be related to factors such as low overall smoking rates in California (26, 27), the high tobacco tax in California and the lower rate of smoking among Latina women compared to non-Hispanic White women (15).

Previous studies on the accuracy of self-reported data on tobacco use in Latinos are conflicting. One study found that Latinos report smoking status less precisely than other ethnic groups (6). On the other hand, among Mexican-American women in the Southwestern United States there was a strong correlation between self-reported smoking status and serum cotinine levels (28). Our study and the study of Mexican-American women in the Southwest suggest that

data concerning tobacco use in Latina women are both accurate (valid using serum cotinine as the biochemical "gold standard")(28) and precise (reproducible across self-reported data sets).

Despite the importance of prenatal vitamins (29), only 58 percent of the Latina women in our study had responses in both the interview and the medical record that documented their use of prenatal vitamins. In our study at San Francisco General Hospital, we found that use of prenatal vitamins was usually documented in the medical record during the initial prenatal visit, while charting during subsequent prenatal visits rarely mentioned use of prenatal vitamins. If women initiated use of vitamins after their first prenatal visit but this question was either not asked during subsequent prenatal visits or not noted in the chart, then this could explain the low kappa. Furthermore, some charts contained a check box questionnaire that prompted the clinician to ask about prenatal vitamins, whereas others did not. We believe that the inclusion of such prompts in all charts would have led to a higher agreement. Our results highlight the need to create an effective system for documenting the use of prenatal vitamins and other medications in the medical record throughout pregnancy.

Previous studies have found that maternal risk factors and the risk of a low birth weight infant vary by maternal birthplace among Latina women (14). Based on the fact that United States-born Latinas have higher educational attainment than foreign-born Latinas (14), we expected that United States-born Latina women would report data more precisely than foreign-born Latina women. Contrary to our expectations, however, we did not find that nativity was independently associated with the accuracy of self-reported data.

Our findings are subject to certain limitations. First, the Latina women for the Latino Health Project were selected from a single hospital-based prenatal clinic and may not be representative of Latina women as a whole. Similarly, most of the women in this study were

foreign-born and Spanish speaking, which may have limited our ability to compare the precision of self-reported data by maternal birthplace and language. Second, the prevalence of certain prenatal risk factors, such as alcohol use (6 percent) and tobacco use (6 percent), was relatively low, which may have affected our results on agreement of self-reported and medical record data. Lastly, while we were able to measure agreement between self-report and medical records, we were unable to assess the accuracy of this information.

In summary, we found that among pregnant Latina women, self-reported data concerning behavioral risk factors had lower agreement than self-reported data on medical risk factors. Although we did not find higher disagreement rates for Spanish speaking women, the influence of non-English speaking patients and the precision with which information in recorded in the medical record is worthy of additional investigation. Further study is also needed to define the most effective way to obtain and record information on prenatal vitamin use, alcohol use and other maternal risk factors. We recommend performing a pilot study that uses a simple, comprehensive standardized questionnaire, in both English and Spanish, to record behavioral and medical risk factors at each prenatal care visit. This would allow for a better assessment of agreement, without the potential confounding of language.

Acknowledgments

References

1. Harlow S, Linet M. Agreement between questionnaire data and medical records: the evidence for accuracy of recall. Am J Epidemiol 1989;129:233-248.

2. Kerber RA, Slattery ML. Comparison of self-reported and database-linked family history of cancer data in a case-control study. Am J Epidemiol 1997;146:244-8.

3. West SL, Savitz DA, Koch G, Sheff KL, Strom BL, Guess HA, Hartzema AG. Demographics, health behaviors, and past drug use as predictors of recall accuracy for previous prescription medication use. J Clin Epidemiol 1997;50:975-80.

4. Bergmann MM, Byers T, Freedman DS, Mokdad A. Validity of self-reported diagnoses leading to hospitalization: a comparison of self-reports with hospital records in a prospective study of American adults. Am J Epidemiol 1998;147:969-77.

5. Bergmann MM, Calle EE, Mervis CA, Miracle-McMahill HL, Thun MJ, Heath CW. Validity of self-reported cancers in a prospective cohort study in comparison with data from state cancer registries. Am J Epidemiol 1998;147:556-62.

6. Navarro AM. Smoking status by proxy and self report: rate of agreement in different ethnic groups. Tob Control 1999;8:182-5.

7. Hessol NA, Schwarcz S, Ameli N, Oliver G, Greenblatt RM. Accuracy of self-reports of acquired immunodeficiency syndrome and acquired immunodeficiency syndrome-related conditions in women. Am J Epidemiol 2001;153:1128-33.

8. Kendrick JS, Zahniser SC, Miller N, Salas N, Stine J, Gargiullo PM, Floyd RL, Spierto FW, Sexton M, Metzger RW, et al. Integrating smoking cessation into routine public prenatal care: the Smoking Cessation in Pregnancy project. Am J Public Health 1995;85:217-22.

9. Klebanoff MA, Levine RJ, Morris CD, Hauth JC, Sibai BM, Ben Curet L, Catalano P, Wilkins DG. Accuracy of self-reported cigarette smoking among pregnant women in the 1990s. Paediatr Perinat Epidemiol 2001;15:140-3.

10. Ernhart CB, Morrow-Tlucak M, Sokol RJ, Martier S. Underreporting of alcohol use in pregnancy. Alcohol Clin Exp Res 1988;12:506-11.

11. Chang G, Goetz MA, Wilkins-Haug L, Berman S. Prenatal alcohol consumption. Self versus collateral report. J Subst Abuse Treat 1999;17:85-9.

12. Mant J, Murphy M, Rose P, Vessey M. The accuracy of general practitioner records of smoking and alcohol use: comparison with patient questionnaires. J Public Health Med 2000;22:198-201.

13. Mathews TJ, Ventura SJ, Curtin SC, Martin JA. Births of Hispanic origin, 1989-95. Mon Vital Stat Rep 1998;46:1-28.

14. Fuentes-Afflick E, Hessol NA, Perez-Stable EJ. Testing the epidemiologic paradox of low birth weight in Latinos. Arch Pediatr Adolesc Med 1999;153:147-53.

15. Hessol NA, Fuentes-Afflick E. The perinatal advantage of Mexican-origin Latina women. Ann Epidemiol 2000;10:516-523.

16. Guendelman S, Abrams B. Dietary, alcohol, and tobacco intake among Mexican-American women of childbearing age: results from HANES data. Am J Health Promot 1994;8:363-72.

17. Camilli AE, McElroy LF, Reed KL. Smoking and pregnancy: a comparison of Mexican-American and non-Hispanic white women. Obstet Gynecol 1994;84:1033-7.

18. Wolff CB, Portis M. Smoking, acculturation, and pregnancy outcome among Mexican Americans. Health Care Women Int 1996;17:563-73.

19. Zambrana RE, Scrimshaw SC, Collins N, Dunkel-Schetter C. Prenatal health behaviors and psychosocial risk factors in pregnant women of Mexican origin: the role of acculturation. Am J Public Health 1997;87:1022-6.

20. Martin JA, Hamilton BE, Ventura SJ, Menacker F, Park MM, Sutton PD. Births: final data for 2001. Natl Vital Stat Rep 2002;51:1-102.

21. Singh GK, Yu SM. Adverse pregnancy outcomes: differences between US- and foreign-born women in major US racial and ethnic groups. Am J Public Health 1996;86:837-43.

22. SAS Institute Inc. SAS/STAT User's Guide, Version 8. Cary, North Carolina: SAS Institute Inc., 1999.

23. Klebanoff MA, Levine RJ, Clemens JD, DerSimonian R, Wilkins DG. Serum cotinine concentration and self-reported smoking during pregnancy. Am J Epidemiol 1998;148:259-62.

24. Ross JA, Swensen AR, Murphy SE. Prevalence of cigarette smoking in pregnant women participating in the special supplemental nutrition programme for Women, Infants and Children (WIC) in Minneapolis and Saint Paul, Minnesota, USA. Paediatr Perinat Epidemiol 2002;16:246-8.

25. West SL, Savitz DA, Koch G, Strom BL, Guess HA, Hartzema A. Recall accuracy for prescription medications: self-report compared with database information. Am J Epidemiol 1995;142:1103-12.

26. Fichtenberg CM, Glantz SA. Association of the California Tobacco Control Program with declines in cigarette consumption and mortality from heart disease. N Engl J Med 2000;343:1772-7.

27. Siegel M. The Effectiveness of State-level Tobacco Control Interventions: A Review of Program Implementation and Behavioral Outcomes. Annu Rev Public Health 2002;23:45-71.

28. Perez-Stable EJ, Benowitz NL, Marin G. Is serum cotinine a better measure of cigarette smoking than self-report? Prev Med 1995;24:171-9.

29. Scholl TO, Hediger ML, Bendich A, Schall JI, Smith WK, Krueger PM. Use of multivitamin/mineral prenatal supplements: influence on the outcome of pregnancy. Am J Epidemiol 1997;146:134-41.

Table 1. Participant characteristics: Latino Health Project, San Francisco, CA.

		Number	Percent
Age, years	Mean: 25.3, Range: 15-43		
Education, years	Mean: 8.7, Range: 0-21		
Birthplace	Central America	122	36.5%
	Mexico	190	56.9%
	South America	4	1.2%
	United States	18	5.4%
Marital Status	Married	149	44.6%
	Non-married	185	55.4%
Time in the United States, foreign-born women, years	<4	151	45.4%
	≥4	182	54.6%
Language for speaking / reading	Only Spanish	252	75.4%
	Mainly Spanish	39	11.7%
	Both English/Spanish	27	8.1%
	Mainly English	11	3.3%
	Only English	5	1.5%
Prenatal Health Insurance	MediCal	269	83.5%
	Other public	1	0.3%
	Private	4	1.2%
	Self pay	6	1.9%
	None	42	13.0%

Table 2. Agreement (Kappa statistic, 95% confidence interval) between self-reported (SR) and medical record (MR) data for behavioral and medical risk factors: Latino Health Project, San Francisco, CA.

	SR yes MR yes (%)	SR no MR no (%)	SR yes MR no (%)	SR no MR yes (%)	Kappa Statistic	95% Confidence Interval
Behavioral factor						
Alcohol use	17 (5.7)	235 (79.4)	25 (8.5)	19 (6.4)	0.35	0.20-0.50
Tobacco use	17 (5.9)	265 (91.4)	3 (1.0)	5 (1.7)	0.79	0.66-0.93
Use of prenatal vitamins	181 (58.0)	15 (4.8)	108 (34.6)	8 (2.6)	0.09	0.02-0.17
Medical condition						
Anemia	32 (10.2)	253 (80.3)	21 (6.7)	9 (2.9)	0.63	0.50-0.75
Gestational diabetes	10 (3.2)	299 (95.5)	4 (1.3)	0 (0)	0.83	0.66-0.99
Hypertension	8 (2.5)	300 (95.2)	5 (1.6)	2 (0.6)	0.68	0.46-0.90

SR: self report; MR: medical record

Table 3. Odds ratios (95% confidence intervals) for disagreement on any behavioral risk factor: Latino Health Project, San Francisco, CA.

Variable	Odds Ratio	95% Confidence Interval
Age, years		
<20	0.84	0.39-1.82
20-26	0.78	0.45-1.36
>26	1.00 (referent)	
Education, years		
<7	0.67	0.39-1.15
≥7	1.00 (referent)	
Birthplace		
Central or South America	1.15	0.68-1.95
Mexico	1.00 (referent)	
United States	0.67	0.14-3.17
Marital status		
Unmarried	1.21	0.73-2.03
Married	1.00 (referent)	
Time in the United States, years		
<4	1.15	0.68-1.96
≥4	1.00 (referent)	
Language for speaking / reading		
Primarily Spanish	0.33	0.13-0.89
Primarily English	1.00 (referent)	
Prenatal health insurance		
Private, MediCal, or other public	1.00 (referent)	
None or Self-pay	0.96	0.48-1.94

Table 4. Odds ratios (95% confidence intervals) for disagreement on any medical condition: Latino Health Project, San Francisco, CA.

Variable	Odds Ratio	95% Confidence Interval
Age, years		
<20	0.92	0.30-2.86
20-26	1.15	0.51-2.55
>26	1.00 (referent)	
Education, years		
<7	0.98	0.45-2.12
≥7	1.00 (referent)	
Birthplace		
Central or South America	0.99	0.47-2.12
Mexico	1.00 (referent)	
United States	0.88	0.14-5.65
Marital status		
Unmarried	1.25	0.59-2.65
Married	1.00 (referent)	
Time in the United States, years		
<4	0.88	0.40-1.92
≥4	1.00 (referent)	
Language for speaking / reading		
Primarily Spanish	0.75	0.22-2.52
Primarily English	1.00 (referent)	
Prenatal health insurance		
Private, MediCal, or other public	1.00 (referent)	
None or Self-pay	0.93	0.33-2.60