Published in final edited form as:

Maturitas. 2013 January ; 74(1): 95–98. doi:10.1016/j.maturitas.2012.09.016.

Reliability of Self-Report on Basic Health Conditions In Mexican Americans Age 80 and Older

Carlos Siordia^{1,*}

¹Preventive Medicine and Community Health, University of Texas Medical Branch, Galveston, Texas, USA.

Abstract

Reliability of self-reports in survey research is important. Little is known about the reliability of self-reports in the very old (aged 80+). The objective of this study is to evaluate the level of reliability in self-reports by assessing the level of agreement between a “respondents” and their “informants” (children) as they rate the respondent’s: hearing ability; vision capacity; and nine health conditions. Descriptive statistics, observed agreement, and kappa values are computed on 389 pairs using HEPSE Wave-7 data (2010-2011). Although further studies are required, there is evidence that self-reports on basic health conditions in the very old are reliable.

Keywords

reliability; self-report; proxies; level of agreement; aging; health; Mexicans

1. INTRODUCTION

The validity of any research enterprise is founded on the quality of data being used in the analysis. Epidemiologic studies frequently rely on information collected through questionnaires where study subjects are asked to self-report. Reliability of self-reports is most volatile with fragile study subjects—such as very young children, mentally challenged individuals, and the very old. Determining if self-reports are trustworthy is of great importance for academic researchers and clinicians. This brief report contributes to this line of research by investigating the level of agreement between a *respondent* (i.e., the aged adult) and their *informant* (i.e., a child) as both report on respondent’s basic health conditions.

Research has explored the reliability of self-reports in health-related investigations [1-5]. However, an extensive review of the literature failed to turn up a single publication focusing on the reliability of self-reports on basic health-related items in very old (age 80+) Mexican-origin Latinos/as (hereafter simply refer to as Mexican Americans: MAs). Clinicians, epidemiologist, and social scientist at large would benefit from having some evidence that

© 2012 Elsevier Ireland Ltd. All rights reserved.

*Correspondence to author: Community Health and Preventive Medicine, University of Texas Medical Branch, Galveston, Texas, 77555. Phone: (409) 772-1128, Fax: (409) 772-5272, Carlos.Siordia@utmb.edu.

The author has no competing interest or conflict of interest with *Maturitas*

The single author completed all items on the manuscript

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

very old patients/study participants are a reliable source of information. This project investigates the reliability of self-reports on 18 health items (see Appendix A). Reliability was measured comparing the aged adult's responses and the response of his/her child.

The question being explored is: What is the level of agreement between respondents and their informants as they report on respondent's basic health conditions? Previous work has delineated the various scenarios under which disagreement between a respondent and informant may arise [6]. In this paper, if agreement between the two reports is high, then respondent's self-report on his/her health conditions is said to be reliable.

Given previous research on levels of agreement between respondents and informants [7-10], it is expected that: as the level of complexity (or degree of privacy) in the health item increases, the level of disagreement will increase. For example, asking if the aged adult can hear with a hearing-aid would be consider more complex and private than asking whether the subject has diabetes. The argument, as more eloquently stated by another [9], is that potential for disagreement between respondent and informant rises as questions are less concrete, observable, and involve private behaviors.

2. METHODS

2.1 Data Source

Determining the reliability of self-reports on basic health conditions in the very old with direct methods requires the creation of new data—since no study I am aware of provides test and re-test data on health condition survey items in aged minority adults. The second best approach, a proximal approach, is to investigate the level of agreement between the aged adult's self-report and a report by somebody who is close to them—in our case a child. This rare data is available in Wave-7 (2010-2011) from the Hispanic Established Population for the Epidemiological Study of the Elderly (HEPESE), a longitudinal dataset of aged Latinos/as of Mexican heritage [11]. The HEPESE study has collected data on 3,952 community-dwelling MAs who were over the age of 64 and who resided in one of the five United States (US) Southwestern US [11]. Ethical approval for the study was received from the internal review board at the University of Texas Medical Branch in Galveston, Texas, US (IRB Protocol # 92-85). Both respondents and informants gave oral assent after a statement about the study and their participation was read to them before commencing the administration of the questionnaire.

2.2 Sample

During Wave-7, 1,078 “respondents” (i.e., parents) aged 80 and over were interviewed in person, at the survey participant's home, by trained interviewers. “Informants” (i.e., children) were used for the first time in Wave-7. Informants were identified during Wave-7 data collection by asking respondents to provide the name of a child who provided them with the most advice and knew them best. This paper is unique in that it is the first to ever measure the level agreement on reports of basic health conditions between aged MAs and their children.

Wave-7 has 925 respondent-informant pairs. The analytic sample excluded non-child-parent pairs; dyads where the respondent completed the questionnaire with the help of a “proxy” (i.e., a person assisting respondent in answering questions and for whom there is no information on); and those who belong to a “couple” in the study—because some spouses are also respondents in the study and in a few cases are the informant. From a potential 925 set of pairs, the sample being used includes 389 dyads—the sample size for the agreement analysis varies by health conditions because missingness varies by item. Please note that

because the analytic sample experienced selectivity in many instances, the findings from this report are not generalizable to all MAs in the United States.

2.3 Agreement Analysis

The agreement analysis uses data from 389 respondent-informant pairs (a total of 778 individuals). Response comparability is evaluated by calculating various Kappa statistics [12]. Cohen's Kappa coefficient [13] is the most widely used method for assessing the degree of agreement between two raters—in this case the agreement on respondent's basic health conditions as reported by the respondent and his/her informant. The Kappa (k) statistic represents the level of agreement between the respondent-informant reports—after eliminating the congruency which would occur by chance. Previously established Kappa threshold categories are used to label the different levels of agreement [14-15]. By following what others have done [9], “strong agreement” is said to be present with k values of 0.80 and above, “moderate agreement” with k values between 0.41 to 0.79, and “low agreement” with anything at or below 0.40 [16].

Because *Kappa paradox* (the fact that observed proportion of agreement is high but the value of the kappa statistic is low) may be present [17], the advice of others [18] is followed and the “observed agreement” is reported and discussed. Observed agreement is the percentage of all the responses for which the respondent and informant agree (whether yes or no) divided by the total number of responses. More technically, it is the sum of quadrant-1 (both agree respondent *does not have* health condition) and quadrant-4 (both agree respondent *has* health condition) divided by the total n in a two-by-two table [19]. Please see Appendix B for more details on Quadrants. All data coding, analyses, and outputs for this paper are generated using SAS 9.2 software (Copyright, SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA).

3. RESULTS

3.1 Level of Agreement

For a complete discussion on the descriptive statistics of the analytic sample see Appendix C. In the first part of Table 1, the counts for how each of the dyads responds to each of the basic health items under investigation are given (Please see Appendix B for a detail expiation of each quadrant). Since misreporting (regardless of the information source) is a possibility when disagreement is present, we use observed agreement as a proxy measure for the reliability of a respondents self-report. In truth, when disagreement occurs, misreporting could be coming from the respondent, informant, or both [6]. Since we are unable to untangle these important difference, we simple interpret high levels of observed agreement as signaling that a respondent's self-report is reliable and that low levels of observed agreement signal a questionable report by the respondent.

Observed agreement, the main point of the article, is now examined. As per Kappa scores and in general, moderate agreement is present in most items. Please note that an instance of a Kappa paradox [17] is present with the hypertension health item. Using “OA” (i.e., the observed agreement score: Quadrant 1 plus Quadrant 4 divided by the total of all quadrants), we see that respondent and informants agree most (97% of the time) when reporting on respondents diabetes status. This is followed by cancer (agree 96% of the time) and stroke (agree 95% of the time). Observed agreement is lowest (71%) when reporting on whether the respondent can hear with a hearing-aid—please note there are only 58 dyads reporting on this condition. Pain when walking has amongst the lowest level of agreement (with dyads

disagreeing about 21% of the time), followed by an aged adults ability to hear with no aid (with dyads disagreeing 20% of the time).

Please note that previous work has found that respondents self-report health conditions more favorably than their informant [6,9,10]. For example, respondent are said to “self-report more favorably” when they report being able to see across the street and their child reports the opposite. When Quad-2 (respondent self-reports not having the condition and informant disagrees) is compared to Quad³ (respondent self-reports having the condition and informant disagrees), we see that in most instances, respondents rate themselves more favorably than their informant (i.e., the Quad-2 number is larger than the Quad-3 number).

4. DISCUSSION

This project has investigated the level of agreement between an aged adult and his/her informant-child. The findings are that an aged adult’s self-reports on basic health conditions are reliable. Of notable importance here is the fact that agreement was not at 100% on any of the health items examined. Because self-reports in those aged 80 and so frequently used in research and in clinical settings, future studies are needed to explore how other sources of information can help evaluate and aged adults self-reports on his/her basic health conditions.

Acknowledgments

This study was supported by the Longitudinal Study of Mexican American Elderly Health (PI-KM), funded by the National Institute of Health (NIH) through the National Institute on Aging (NIA)(R01 AG10939-18).

Appendix A: HEPSE questions

Health

Has a doctor or other health care professional ever told your (respondent) that he/she had any of the following conditions?

1. Had or suspect diabetes?
2. Cancer or malignant tumor?
3. Had or suspect a stroke, blood clot in the brain, or brain hemorrhage?
4. Had high blood pressure?
5. Had or suspect a heart attack, coronary, myocardial infarction, or coronary thrombosis?
6. Had broken or fractured any other bone other than hip?
7. Has fallen in the past 12 months?
8. Arthritis or rheumatism?
9. Noticed any pain or discomfort when they stood or walked in the past month?

Hearing

Can (respondent) usually hear and understand what a person says without seeing his face if that person talks in a normal voice to you in a quiet room?

10. Without a hearing aid

11. With a hearing aid

Vision

Can (respondent) recognize a friend or family member (when wearing glasses/contacts if applicable)?

12. Across the street

13. Across the room

14. Arm's length away

Appendix B: Appendix 4

Detail Explanation of Quadrants in Level of Agreement Analysis

- Quadrant 1
Count of dyads where both agree that respondent does not have the condition
- Quadrant 2
Count of pairs where the respondent says he/she doesn't have the condition and the informant disagrees—here the respondent is said to rate themselves “more favorably” than the informant
- Quadrant 3
Count of pairs where the respondent says he/she has the condition and the informant disagrees—in this instance, the respondent is said to rating his/herself “less favorably” relative to the informant
- Quadrant 4
Count of dyads where both respondent and informant agree that respondent has the health condition.

Appendix C:

Discussion on the Descriptive Statistics of Analytic Sample

The table below displays respondent and informant characteristics. The number of observations, the mean, and standard deviation is given for each health item under review along with some additional information. From the respondent column, we see that: 71% of them are female; 46% are over the age of 85; 59% are overweight as per body mass index measures; 33% of them have two or more limitations with basic activities of daily living (BADL); and 74% have two or more limitations with instrumental activities of daily living (IADL). From the informant column in the table below, we see that: 69% of them are female; none are over the age of 85; 74% are overweight as per body mass index measures; 31% of them have two or more limitations with BADLs; and 78% have two or more limitations with IADLs. As a group and when compared to the respondents (i.e., the parents), the children (i.e., the informants) are more overweight and report about the same amount of limitations with BADLs and IADLs.

From the table below, we note that amongst respondents: 39% are diabetic; 11% report experiencing cancer; 7% have had a stroke; 98% are hypertensive; 10% report having had a heart attack; 23% have broken a non-hip bone; 41% felled in the last twelve months; 68% of them have been told they have arthritis; and 56% say they feel pain when walking. When informants report on the same conditions for the respondent, as a group, they report the about the same level of: diabetes; stroke; non-hip fractures; falls in the last twelve months; and arthritis. The most notable difference is on hypertension, where informants as a group report that respondents as a group have a 78% prevalence rate. While both respondents and informant report that almost 80% of respondents can hear with no hearing aid, there is a big disparity when the two groups report on respondent's ability to hear with an aid. On vision items, both respondent and informant groups seem to report the same level of prevalence amongst respondents.

	Respondents ¹			Informants ²		
	N	Mean	SD	N	Mean	SD
Demographics						
Female	388	0.71	0.45	393	0.69	0.46
Age > 85	388	0.46	0.50	393	0.00	0.00
Overweight	388	0.59	0.49	393	0.74	0.44
BADLs ³ 2	388	0.33	0.47	393	0.31	0.46
IADLs ⁴ 2	388	0.74	0.44	393	0.78	0.42
Health						
Diabetes	389	0.39	0.49	388	0.39	0.49
Cancer	389	0.11	0.32	389	0.14	0.35
Stroke	389	0.07	0.25	388	0.08	0.28
Hypertension	303	0.98	0.13	387	0.78	0.42
Heart attack	389	0.10	0.30	388	0.13	0.34
Non-hip fracture	389	0.23	0.42	389	0.24	0.43
Fallen in last year	389	0.41	0.49	386	0.41	0.49
Arthritis	389	0.68	0.47	383	0.69	0.46
Pain when walking	389	0.56	0.50	386	0.60	0.49
Hearing						
Hear with no aid	389	0.78	0.41	380	0.77	0.42
Hear with aid	389	0.08	0.27	59	0.49	0.50
Vision						
See across street	389	0.86	0.35	389	0.83	0.38
See across room	389	0.93	0.26	389	0.94	0.24
See at arm's length	393	0.93	0.25	389	0.94	0.23

¹ Adults from 80 to 102 years of age

² Children of aged adults with an age range of 26-74

³ Basic activities of daily living

⁴ Instrumental activities of daily living

REFERENCES

- [1]. Nadalin V, Bentvelsen K, Kreiger N. Reliability of self-reports: data from the Canadian multi-centre osteoporosis study (CaMos). *Chronic Diseases in Canada*. 2004; 25:1–4.
- [2]. Gruenewald, Paul J.; Johnson, Fred W. The stability and reliability of self-reported drinking measures. *Journal of Studies on Alcohol*. 2006; 67:738–745. [PubMed: 16847543]
- [3]. Renzi C, Mastroeni S, Mannooranparampil TJ, et al. Reliability of self-reported information on skin cancer among elderly patients with squamous cell carcinoma. *Annals of Epidemiology*. 2011; 21:551–554. [PubMed: 21524593]
- [4]. Soubelet A, Salthouse TA. Influence of social desirability on age differences in self-reports of mood and personality. *Journal of Personality*. 2011; 79:741–762. [PubMed: 21682727]
- [5]. Brown JL, Sales JM, DiClemente RJ, et al. Predicting discordance between self-reports of sexual behavior and incident sexually transmitted infections with african american female adolescents: results from a 4-city study. *Aids and Behavior*. 2012 DOI 10.1007/s10461-012-0163-8.
- [6]. Siordia, C. On the different scenarios of disagreement between self- and informant-reports on an aged adults ability to perform activites of daily living. *Maturitas*. 2012. <http://dx.doi.org/10.1016/j.maturitas.2012.06.012>
- [7]. Jonas C, Schiffczyk C, Lahmeyer C, et al. Staging dementia using proxy-reported activities of daily living. *Dementia and Geriatric Cognitive Disorders*. 2011; 32:111–117.
- [8]. Sharderl M, Alley DE, Hicks GE, et al. Comparing reports from hip-fracture patients and their proxies: implications on evaluating sex differences in disability and depressive symptoms. *Journal of Aging and Health*. 2012; 24:367–383. [PubMed: 22210805]
- [9]. Magaziner J, Bassett SS, Hebel JR, et al. Use of proxies to measure health and functional status in epidemiologic studies of community-dwelling women aged 65 years and older. *American Journal of Epidemiology*. 1988; 143:283–292. [PubMed: 8561163]
- [10]. Singer E. Agreement between “inaccessible” respondents and informants. *Public Opinion Quarterly*. 1973; 36:603–611.
- [11]. Markides, KS.; Stroup-Benham, CA.; Black, SA.; Satish, S.; Perkowski, LC.; Ostir, G. The health of Mexican American Elderly: Selected findings from the Hispanic EPESE. In: Wykle, ML.; Ford, AB., editors. *Serving minority elders in the 21st century*. 1st edn. Springer; 1999. p. 72-90.
- [12]. Cunningham, M. More than just the kappa coefficient: a program to fully characterize inter-rater reliability between two raters. *SAS Global Forum*; Pittsburgh: PA: 2009.
- [13]. Cohen JA. Coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 1960; 20:37–46.
- [14]. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychological Bulletin*. 1971; 76:378–382.
- [15]. Fleiss, JL. *Statistical methods for rates and proportions*. John Wiley; New York: 1973.
- [16]. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159–174. [PubMed: 843571]
- [17]. Feinstein AR, Cicchetti DV. High agreement but low kappa: the problems of two paradoxes. *Journal of Clinical Epidemiology*. 1990; 43:543–9. [PubMed: 2348207]
- [18]. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical Therapy*. 2005; 85:257–268. [PubMed: 15733050]
- [19]. Viera AJ, Garret JM. Understanding interobserver agreement: the kappa statistic. *Family Medicine*. 2005; 37:360–363. [PubMed: 15883903]

Table 1

Level of agreement analysis on chronic health items, hearing, and vision

	Counts				Level of Agreement				
	Quad-1 ^a	Quad-2 ^b	Quad-3 ^c	Quad-4 ^d	Total	OA ^e	Kappa	LL ^f	UL ^g
Health									
Diabetes	59%	1%	2%	38%	388	0.97	0.94	0.90	0.97
Cancer	86%	3%	1%	11%	388	0.96	0.82	0.74	0.91
Stroke	90%	4%	2%	5%	388	0.95	0.62	0.46	0.77
Hypertension	0%	1%	5%	93%	298	0.94	0.07	−0.10	0.24
Heart attack	85%	5%	2%	8%	388	0.93	0.67	0.55	0.79
Non-hip fracture	71%	6%	6%	17%	389	0.88	0.66	0.57	0.75
Fallen in last year	68%	7%	8%	44%	302	0.88	0.75	0.69	0.82
Arthritis	24%	8%	7%	62%	380	0.86	0.66	0.58	0.74
Pain when walking	31%	12%	8%	48%	386	0.79	0.57	0.49	0.66
Hearing									
Hear with no aid	11%	8%	11%	69%	379	0.80	0.41	0.30	0.53
Hear with aid	47%	24%	5%	24%	58	0.71	0.41	0.19	0.62
Vision									
See across street	11%	3%	6%	80%	388	0.91	0.68	0.58	0.78
See across room	4%	3%	2%	91%	389	0.95	0.63	0.47	0.79
See at arm's length	3%	4%	3%	91%	389	0.94	0.47	0.29	0.65

^a Quadrant 1: Percent of dyads from "total" where both respondent and informant agree that respondent does not have the condition^b Quadrant 2: Percent of dyads from "total" where respondent says he/she doesn't have the condition and the informant disagrees^c Quadrant 3: Percent of dyads from "total" where respondent says he/she has the condition and the informant disagrees^d Quadrant 4: Percent of dyads from "total" where both respondent and informant agree that respondent has the condition^e Observed agreement: (Quadrant 1 + Quadrant 4) ÷ (Total)^f Lower limit of 95% confidence interval^g Upper limit of 95% confidence interval