Original Paper

Reliability of Information Collected by Proxy in Family Studies of Alzheimer's Disease

Serkalem Demissie[a]   Robert C. Green[c,d]   Lorelei Mucci[b]   Sophia Tziavas[c]
Kathy Martelli[f]   Kaitlin Bang[e]   Lisa Coons[g]   Sylvie Bourque[m]
Dina Buchillon[h]   Kris Johnson[i]   Tamisson Smith[j]   N. Sharrow[k]
Nicola Lautenschlager[n]   Robert Friedland[l]   L. Adrienne Cupples[b]
Lindsay A. Farrer[b–d]

[a]Data Coordinating Center, and Departments of [b]Epidemiology and Biostatistics, [c]Medicine (Genetics Program) and
[d]Neurology, Boston University Schools of Medicine and Public Health, [e]Department of Neurology, Harvard
University School of Medicine, Boston, Mass., [f]University of Southern California, Los Angeles, Calif., [g]Department
of Epidemiology, University of Washington, Seattle, Wash., [h]University of Miami School of Medicine, Miami, Fla.,
[i]Department of Health Sciences Research, Mayo Clinic and Foundation, Rochester, Minn., [j]Department of
Neurology, University of Kansas Medical Center, Kansas City, Kans., [k]Department of Neurology, University of
Alabama, Birmingham, Ala., [l]Department of Neurology, Case Western Reserve University, Cleveland, Ohio, USA;
[m]Department of Medical Genetics, University of British Columbia, Vancouver, Canada; and [n]Psychiatrische Klinik
der Technischen Universität München, Germany

Abstract
The study evaluated the reliability of data obtained from
proxy informants. The index subjects in this study were
81 nondemented participants in the Multi-Institutional
Research in Alzheimer Genetic Epidemiology (MIRAGE)
study. These index subjects and 159 proxy informants,
identified by the index subjects, participated in the study.
The κ statistic with multiple raters per subject (for dichotomous variables) and the intraclass correlation coefficient (for continuous variables) were used to measure
reliability. Among proxy respondents who provided answers, there was excellent agreement between proxy
responses and the responses of the index subjects (0.7 ≤
κ ≤ 0.9), with the exception of questions about head injury (κ = 0.4). A large proportion (>90%) of the proxy informants in this study were able to provide information on
most items. Higher nonresponse rates (as high as 30%)
were observed for medication history and women's
health questions. This study supports the reliability of
proxy responses for most categories of questions that
are elicited in typical epidemiological studies, including
the MIRAGE study.

Copyright © 2001 S. Karger AG, Basel

Introduction

Investigators using information from family studies of
Alzheimer's disease (AD) have contributed significantly
to our understanding of the genetic contribution to this
prevalent disorder [1]. But persons with AD have progressive memory loss and cognitive deficits, so it is necessary
for researchers to rely upon relatives or friends to provide

Dr. Lindsay Farrer
Genetics Progam, Boston University School of Medicine
715 Albany St., L-320, Boston, MA 02118 (USA)
Tel. +1 617 638 5393, Fax +1 617 638 4275
E-Mail farrer@neugen.bu.edu

proxy information about the family and medical history about the affected subjects. Misclassification of exposure status and higher nonresponse rate are two major concerns in study designs where proxy respondents are the primary sources of exposure data.

Data provided by a proxy respondent can be unreliable through loss of precision (random measurement error), or through under- or overreporting of exposure status [2]. Unreliable data can lead to biased effect estimates or to decreased ability to control for confounding variables or detect interactions between genetic and environmental factors [3]. Another important aspect of proxy-based data collection is whether proxy informants are able to provide any information. In general, proxy respondents are known to have a greater nonresponse rate than index subjects [4, 5]. This is usually due to the fact that proxy respondents do not always know the information requested. 'Do not know' responses and other types of non-response are generally treated as missing information. Besides the statistical power reduction, missing information may also seriously affect study inferences due to non-response biases [6].

This study focuses on reports of health behavior, medical outcomes, head injury, medication history and women's health variables using proxy informants. We performed a reliability assessment of proxy data using nondemented participants in the MIRAGE study. (Multi-Institutional Research in Alzheimer Genetic Epidemiology) [7]. We gathered and compared direct and proxy information on nondemented persons who were family members of AD probands by using other family members as informants. The goals of this study were to evaluate the reliability of data obtained from proxy informants, and to determine whether some groups of family members (e.g. spouse, sibling or adult child) provide better proxy information than others.

Methods

Subjects and Study Variables

The MIRAGE Study is an ongoing multicenter genetic epidemiological study of AD that relies on data typically collected from proxy informants about demented individuals or deceased family member [7]. In most cases, the proxy informant for the proband is the principal caregiver who is usually a spouse, adult child or sibling. The most direct means of assessing reliability would require the comparison of information from the proxy respondent with information provided by the AD proband. However, since the probands are cognitively impaired and thus unable to provide reliable information, the degree of agreement cannot be assessed directly.

Table 1. Characteristics of index subjects and proxies

	Index subjects (n = 81)	Proxies (n = 159)
Demographic characteristics		
Gender		
Male, n (%)	17 (21.0)	53 (33.3)
Female, n (%)	64 (79.0)	106 (66.7)
Years proxy has known index		39.7 ± 15.6
Age, years	61.1 ± 14.6	50.5 ± 14.1
Relation of proxies to index subjects, n (%)		
Spouse	20 (12.6)	
Sister	39 (24.5)	
Brother	18 (11.3)	
Daughter	32 (20.1)	
Son	14 (8.8)	
Other	36 (22.6)	

At ten centers participating in the MIRAGE project, investigators invited 1 cognitively normal relative of 10 consecutive AD probands to serve as the index subject for this study. A total of 81 subjects agreed to participate and completed a brief questionnaire describing their own medical history, medication use, health behaviors and cognitive abilities. The questionnaire was an abbreviated version of the personal history form used in the MIRAGE study. Each of the 81 index causes was asked to identify 2 relatives (proxy informants) to complete the same questionnaire about them without assistance.

The health behavior items included the questions 'Did you (Did your relative ...) ever smoke?' and 'Have you (Has your relative ...) ever been a social, occasional or frequent user of alcohol?'. For medical history, the question 'Have you (Has your relative) had any of the following conditions?' was asked about conditions such as arthritis, cancer, diabetes, heart disease, high blood pressure and thyroid disease. The question 'Have you (Has your relative ...) ever had a serious head injury, which required medical care and/or caused unconsciousness?' was used to obtain information on head injury. For all these questions, the possible responses were 'yes', 'no' or 'don't know'. For smoking, 'number smoked per day' (summarized as pack years) and for alcohol, the number of drinks consumed at ages 16–39, 40–64 and 65+ were also asked. A summary measure for alcohol was calculated as the average of these three measures.

Specific questions were also asked about (1) cholesterol-lowering medication; (2) birth control pills or estrogen replacement therapy (in female index cases only); (3) nonsteroidal anti-inflammatory medication, and (4) steroidal anti-inflammatory medication. The questions on medication history were in the form 'Did you/your relative ever take ... (medication) on a daily basis for more than 6 months?'. The other health information items asked female index subjects about menopausal status (pre, post, don't know) and whether the subject had had a hysterectomy.

Data Analysis

For dichotomous variables the κ statistic with multiple raters per subject [8, 9] was used to measure reliability. Ninety-five percent

Table 2. Percent 'don't know' responses among proxy informants

	All proxies	Proxy: by gender		Proxy: relationship to index					
		males	females	spouse	brother	sister	daughter	son	other
Health behavior									
Smoker	2.50	0.00	3.77	0.00	0.00	2.56	6.25	0.00	2.78
Alcohol use	1.25	0.00	2.89	0.00	0.00	0.00	3.13	0.00	2.78
Medical outcomes									
Arthritis	7.50	1.92	9.43	0.00	0.00	5.13	15.63	7.14	11.11
Cancer	1.88	0.00	2.83	0.00	0.00	0.00	3.13	0.00	5.56
Diabetes	1.88	0.00	2.83	0.00	0.00	0.00	3.13	0.00	5.56
Heart disease	3.75	1.92	4.72	0.00	5.56	2.56	6.25	0.00	5.56
Hypertension	9.38	5.77	11.3	0.00	11.11	2.56	12.50	7.14	19.44
Thyroid disease	3.75	7.69	1.89	10.00	11.11	0.00	0.00	0.00	5.56
Head injury	6.88	1.92	9.43	0.00	5.56	7.69	9.38	0.00	11.11
Medication history									
NSAID	24.38	23.08	25.47	15.00	33.33	12.82	31.25	21.43	33.33
Steroids	17.50	13.46	19.81	5.00	22.22	15.38	18.75	14.29	25.00
Cholesterol med.	18.13	9.62	22.64	0.00	16.67	23.38	15.63	14.29	27.78
Women's health									
Estrogen use	29.92	30.95	28.92	18.75	46.15	21.88	30.43	40.00	31.25
Postmenopause	9.60	16.67	6.17	6.25	7.69	0.00	4.55	50.00	12.90
Hysterectomy	12.00	14.29	9.88	0.00	23.08	0.00	18.18	30.00	12.90

confidence intervals were computed using bootstrap percentiles, based on 1,000 bootstrap samples [10]. For continuous variables, the intraclass correlation coefficient (ICC) was calculated [11] to take into account agreement on the mean level of exposure [12]. For qualitative interpretation of the κ values, we used the classification 'poor' reliability or agreement if κ < 0.2; 'fair' 0.21–0.40; 'moderate' 0.41–0.60; 'substantial' 0.61–0.80, or 'almost perfect' 0.81–1.00 [9]. In this report, the term agreement refers to agreement among the proxy informants and their index subject unless indicated otherwise.

The κ and the ICC statistics were estimated both after including and after excluding 'don't know' responses, since the index subject was reporting on himself/himself and rarely responded 'don't know'. To describe the ability or inability of subjects to respond to specific questions, the percentage of 'don't know' responses was calculated. We computed the overall (crude) estimates, stratified by the proxy's gender and estimates stratified by proxy's family relationship to the index subject.

Results

Demographic Characteristics

Tables 1a and b present demographic characteristics for the index subjects and their proxy informants. Completed questionnaires were collected from 81 index subjects and 159 proxy informants. Three index subjects pro-

vided only 1 proxy. Seventy-seven percent of proxies were first-degree relatives of the index subjects. The remaining 23% of proxy informants were cousins, friends and stepparents. Only 12.6% of proxies were related as a spouse. Proxies were on average 10 years younger and had known the index subject for 40 years.

Nonresponse

The index subjects were able to provide answers for most questions. The question that elicited the maximum nonresponse rate among index subjects (4.8%) was the menopausal status questions. Table 2 presents the response rate for the proxy informants. Among proxies, the nonresponse rate was dependent upon the relationship and sex of the informant and type of question. Overall, the nonresponse rate among proxies was highest for estrogen use (30%) followed by medication history (18–24%). Relatively lower nonresponse rates were observed for medical outcome and health behavior questions.

For most variables, the nonresponse rate was higher among female proxies, with the exception of women's health and thyroid disease questions. The greater nonresponse rate among male proxies on issues related to wom-

Table 3. Reliability, stratified by gender of the proxy

	κ (Bootstrap percentile –95% confidence level)		
	overall (crude)	males	females
Health behavior			
Smoker	0.87 (0.77, 0.95)	0.88 (0.73, 1.00)	0.83 (0.71, 0.93)
Alcohol use	0.74 (0.58, 0.87)	0.84 (0.62, 1.00)	0.68 (0.47, 0.85)
Medical outcomes			
Arthritis	0.68 (0.51, 0.83)	0.76 (0.48, 0.95)	0.69 (0.48, 0.85)
Cancer	0.72 (0.52, 0.88)	1.00	0.64 (0.39, 0.84)
Diabetes	0.89 (0.49, 1.00)	1.00	0.85 (0.43, 1.00)
Heart disease	0.82 (0.60, 0.95)	0.79 (0.37, 1.00)	0.77 (0.51, 0.93)
Hypertension	0.72 (0.52, 0.88)	0.74 (0.39, 1.00)	0.74 (0.52, 0.90)
Thyroid disease	0.75 (0.52, 0.91)	0.65 (0.19, 1.00)	0.80 (0.57, 0.96)
Head injury	0.43 (0.04, 0.69)	0.56 (0.00, 1.00)	0.42 (–0.12, 0.78)
Medication history			
NSAID	0.70 (0.40, 0.90)	0.88 (0.49, 1.00)	0.59 (0.17, 0.86)
Steroids	0.88 (0.20, 1.00)	1.00	0.82 (0.11, 1.00)
Cholesterol med.	0.84 (0.59, 1.00)	0.85 (0.07, 1.00)	0.75 (0.34, 1.00)
Women's health			
Estrogen	0.69 (0.51, 0.83)	0.69 (0.40, 0.93)	0.69 (0.46, 0.88)
Postmenopause	0.74 (0.56, 0.88)	0.48 (0.14, 0.77)	0.84 (0.63, 0.97)
Hysterectomy	0.87 (0.73, 0.97)	1.00	0.82 (0.63, 0.96)

en's health was mainly due to 'don't know' responses from sons and brothers. Brothers and sons had a higher nonresponse rate on women's issues compared to all other respondents.

Reliability

Estimates of κ and the 95% confidence intervals are shown in tables 3 and 4. Table 3 presents the crude estimates and the estimates stratified by the proxy's gender, and table 4 presents the estimates stratified by the proxy's family relationship to the index subject.

Health Behavior. The overall agreement on smoking behavior was almost perfect; for pack years the ICC was 0.90. The overall agreement on the amount of alcohol consumed at ages 16–39 and 40–64 was substantial (0.62 and 0.71); however, the agreement on alcohol consumed at age 65 and older was only fair (0.31). Agreement on the average amount of alcohol consumed at the three stages of life (average of the amount of alcohol consumed at age 16–39, 40–64 and 65 and older) was substantial (ICC = 0.64). Male proxies had almost perfect agreement and female proxies had substantial agreement on alcohol use. Sons and spouses had almost perfect agreement on smoking and alcohol use. Daughters had substantial agreement

with the indexes on both smoking and alcohol use. Variable results were observed for siblings and for the other proxies (table 4).

Medical Outcomes and Head Injury. Overall agreement was almost perfect on diabetes and heart disease; substantial on thyroid disease, hypertension, cancer and arthritis, and moderate on head injury. Agreement on cancer responses was almost perfect for male proxies and substantial for female proxies, no differences were observed with respect to the other medical outcome items. Sibs and children had substantial or almost perfect agreement with their index subjects on most medical outcome items. Spouses had almost perfect agreement on arthritis, cancer and diabetes questions.

Medication History. Overall agreement on medication history was substantial for NSAID and almost perfect for steroids and cholesterol medications. Gender-stratified analyses indicated that male proxies agreed almost perfectly on all of these items, while females agreed only moderately on NSAID, substantially on steroids and almost perfectly on cholesterol medications. Due to a high nonresponse rate for the medication history, it was not possible to estimate reliably the κ statistics stratified by proxies' relationship to the index subjects. However, the

Table 4. Reliability (κ) stratified by relationship of proxy to the patient

	Spouse	Brother	Sister	Daughter	Son	Other
Health behavior						
Smoker	1.00	0.67*	0.71*	0.80*	1.00	1.00
Alcohol use	0.86*	0.56	0.58*	0.71*	1.00	0.61*
Medical outcomes						
Arthritis	1.00	0.54*	0.73*	0.73*	1.00	0.58*
Cancer	1.00	1.00	0.52*	0.85*	1.00	0.70*
Diabetes	1.00	1.00	0.68*	1.00	1.00	1.00
Heart disease	0.77	0.82*	0.80*	1.00	1.00	0.61
Hypertension	0.61	0.82*	0.74*	0.72*	1.00	0.60
Thyroid disease	0.77	1.00	0.91*	1.00	−0.01	−0.08
Head injury	0.48	1.00	0.40	0.55	0.03	−0.29
Medication history						
NSAID	0.77	0.76	0.66	0.66	1.00	0.53
Steroids	1.00	1.00	1.00	1.00	1.00	0.67
Cholesterol med.	1.00	0.64	0.80*	1.00	1.00	0.04
Women's health						
Estrogen use	0.85*	0.42	0.72*	0.67*	−0.34	0.70*
Postmenopause	0.21	−0.14	1.00	0.90*	1.00	0.70*
Hysterectomy	1.00	1.00	0.93*	1.00	1.00	0.54*

*Statistically significantly different from zero at α = 0.05.

available data suggested that spouses and children had almost perfect agreement with index subjects on cholesterol medications; and all first-degree relatives had almost perfect agreement on steroid use.

Women's Health. Overall agreement was substantial for menopausal status and estrogen use, and almost perfect for hysterectomy. On menopausal status, male proxies agreed only moderately and female proxies agreed almost perfectly with their female index subject. Sons had almost perfect agreement and spouses and brothers had poor agreement for menopausal status. Agreement on hysterectomy was almost perfect for all first-degree relative proxies and moderate for other proxies. On estrogen use, spouses had almost perfect agreement.

Discussion

Among proxy respondents who provided answers, there was excellent agreement between proxy responses and the responses of the index subjects. This high reliability was present across all classes of questions, with the exception of questions about head injury, where reliability was only fair. A large proportion of the proxy infor-

mants in this study was able to provide information about the index subjects on health behavior (>97%), medical outcome and head injury (>90%). Higher nonresponse rates, as high as 30%, were observed for medication history and women's health questions.

In line with findings of several other proxy-based studies, the nonresponse rate in our study was dependent on both proxy-index family relationship and the types of questions [3, 4, 13]. Spouses were able to provide information more often than other types of proxy informants, whereas proxy informants that were not first-degree relatives were less able to do so. Pickle et al. [4] reported similar findings based upon a large study surrogate respondents using a case-control study design. Although these authors recommended spouses as preferred respondents or offspring when no spouse was available, our findings indicate that sibs, particularly sisters, are the second ablest respondents. On the other hand, Pickle et al. [4] reported that sibs were the best respondents for events that occurred during the subject's early life. Since we do not have information in our study about where the event occurred in the lifespan of the index subjects, we did not have data to substantiate this finding.

Types of questions and proxy-index relationships have been shown to be associated with nonresponse rates [3, 14], and several strategies have been proposed to minimize nonresponse rates and to address the reduction in power due to high nonresponse rate [3, 4]. For example, investigators have been urged to collect data on large samples and to select proxy informants, based upon family relationship, who are less likely to have high nonresponse rates [3, 15].

In most analytic situations, 'don't know' responses are treated as missing data. This practice raises the potential for selection biases, but we did not have data to evaluate how serious this possibility might be. In actual practice, experience suggests that when the index subject is cognitively impaired, the likely informant (usually a spouse or adult child serving as the primary caregiver) will be *more* knowledgeable than the informants in our reliability study about such things as medications, and perhaps even the medical history, of the family member. Therefore, in the context of the AD studies, where proxy information is typically gathered on impaired persons, the nonresponse rate is likely to be lower. For example, in this reliability study, the highest overall nonresponse rates were for questions about NSAID medication information (24.4%) and for estrogen use among female index cases (30.0%). However, examination of these two questions in the MIRAGE data set indicates that the nonresponse rates for proxy informants on these two questions were actually 5.4 and 10.8%, respectively.

In this study, reliability estimates indicated excellent agreement between proxy and index subject's responses across all classes of questions, with the exception of questions about head injury. These findings are consistent with those from 4 studies which evaluated the reliability of information about AD patients obtained from proxies [16–19] and reported highest levels of agreement on questions about lifestyle, medical interventions and clear-cut medical disorders, and poorest agreement for questions about head injury. A review of these earlier studies [20] postulated that the low agreement on head injury could be due to the wording of the head injury question, for example, asking the respondents to make their own decision as to what 'severe' head injury meant. In our study, even though the agreement for head injury ($\kappa = 0.43$) was the lowest among all medical history variables, it was higher than that reported in other studies, perhaps because the wording of our question on head injury was comparatively less ambiguous.

In this study, there were no substantial differences between the reliability of male and female proxy respondents, except that agreement on menopausal status was greater among female respondents. Examining reliability within specific categories of family relationships revealed that spouses had almost perfect agreement on variables like arthritis, cancer, diabetes, cholesterol medications, estrogen use and hysterectomy. Brothers, sons and daughters of index subjects demonstrated greater agreements on most of the medical outcome variables. In the medication history category, NSAID response appeared to be the least reliable. For the other two medications (steroids and cholesterol medication), the low nonresponse rate and the perfect agreement with the indexes by spouses could be an indication that spouses are more reliable informants of these items. Similarly, for head injury information, brothers had almost perfect agreement and a low nonresponse rate.

In summary, our study supports the reliability of proxy responses for most categories of questions that are elicited in typical epidemiological studies of AD, including the MIRAGE study.

References

1 Farrer LA: Genetics and the dementia patient. Neurologist 1997;3:13–30.
2 Nelson LM, Longstreth WT, Koepsell TD, Checkoway H, Belle GV: Completeness and accuracy of interview data from proxy respondents: Demographic, medical and lifestyle factors. Epidemiology 1994;5:204–217.
3 Nelson LM, Longstreth WT, Koepsell TD, Belle GV: Proxy respondents in epidemiologic research. Epidemiol Rev 1990;12:71–78.
4 Pickle LW, Brown LM, Blot WJ: Information available from surrogate respondents in case control interview studies. Am J Epidemiol 1983;118:99–108.
5 Blot WJ, McLaughlin JK: Practical issues in the design and conduct of case-control studies: Use of next-of-kin interviews; in Blot WJ, Hirayama T, Hoel DG (ed): Statistical Methods in Cancer Research. Hiroshima, Radiation Effects Research Foundation, 1984, pp 49–62.
6 Little RJA, Rubin DB: Statistical Analysis with Missing Data. New York, Wiley, 1987.

7 Lautenschlager NT, Cupples LA, Rao VS, Auerbach SA, Becker R, Burke J, Chui H, Duara R, Foley EJ, Glatt SL, Green RC, Jones R, Karlinsky H, Kukull WA, Kurz A, Larson EB, Martelli K, Sadovnick AD, Volicer L, Waring SC, Growdon JH, Farrer LA: Risk of dementia among relatives of Alzheimer's disease patients in the MIRAGE study: What is in store for the oldest of the old. Neurology 1996;46:641–650.
8 Fleiss JL: Statistical Methods for Rates and Proportions. New York, Wiley, 1981.
9 Landis JR, Koch GG: The measurement of observer agreement for categorical data. Biometrics 1977;33:159–174.
10 SAS Institute Inc: http://ftp.sas.com/techsup/download/stat/jackboot.sas, 1999.
11 SAS Institute Inc: http://ftp.sas.com/techsup/download/stat/intracc.sas, 1999.
12 Bartko JJ: The intraclass correlation coefficient as a measure of reliability. Psychol Rep 1966;19:3–11.
13 Tepper A, Connally LB, Haltmeier P, Smith E, Sweeney MH: Knowledge of medical history information among proxy respondents for deceased study subjects. J Clin Epidemiol 1993;46:1243–1248.
14 Wynder EL: Epidemiological issues in weak associations. Int J Epidemiol 1990;19:S5–S7.

15 Herrmann N: Retrospective information from questionnaires: Comparability of primary respondents and their next-of-kin. Am J Epidemiol 1985;121:937–947.
16 Broe GA, Henderson AS, Creasey H, McCusker E, Korton AE, Jorm AF, Longley W, Anthony JC: A case-control study of Alzheimer's disease in Australia. Neurology 1990;40:1698–1707.
17 Graves AB: A case-control study of Alzheimer's disease. Unpublished PhD thesis, University of Washington, 1988.
18 Heyman A, Wilkinson WE, Stafford JA, Helms MJ, Sigmon AH, Weinberg T: Alzheimer's disease: A study of epidemiological aspects. Ann Neurol 1984;15:335–341.
19 Rocca WA, Fratiglioni L, Bracco L, Bracco L, Pedone D, Groppi C, Schoenberg BS: The use of surrogate respondents to obtain questionnaire data in case-control studies of neurologic diseases. J Chronic Dis 1986;39:907–912.
20 Korten AE, Jorn AF, Henderson AS, McCusker E, Creasey H: Control-informant agreement on exposure history in case-control studies of Alzheimer's disease. Int J Epidemiol 1992;21:1121–1131.