Validity of Self-reported Tooth Counts During a Telephone Screening Interview

Gregg H. Gilbert, DDS; R. Paul Duncan, PhD; Andrew M. Kulley, PhD

Abstract

Objectives: *Telephone screening has become a common method used in health services research to identify efficiently persons in specific populations of interest. In this research, we used a large-scale telephone screening survey to assess: (1) the effectiveness of the telephone method in gathering tooth count information by measuring response rate (cooperation) to specific questions and (2) the validity of subjects' reports of the number of remaining natural teeth.* ***Methods:*** *We used a telephone screening methodology to identify dentate persons (at least one natural tooth remaining) who were 45 years old or older and resided in one of four counties of north Florida. At a second stage, a sample of the telephone screening participants was selected for further study, which consisted of a baseline in-person interview and a clinical examination. We compared the number of remaining teeth reported during the telephone interview with the number determined at baseline examination.* ***Results:*** *The telephone method was effective at gathering tooth count information, although response rates varied with the level of specificity required. Almost all subjects reported the number of teeth at least at the nominal and ordinal levels, but fewer than three-fourths reported the number at the interval level. When the unit of analysis was the overall sample, self-reported number of teeth was a valid measure of the true number. When the unit of analysis was the individual subject, validity was associated with certain clinical and sociodemographic factors.* ***Conclusions:*** *When the unit of analysis is the overall sample, these results suggest that self-reported tooth counts during a telephone interview are sufficiently valid to meet all but the most stringent data requirements. When the unit of analysis is the individual subject, these tooth counts may not be valid, depending upon the degree of specificity required and subject characteristics. [J Public Health Dent 1997;57(3):176-80]*

Key Words: epidemiologic methods, health surveys, research design, telephone, dental.

Telephone screening surveys are now a common method used in health research to identify efficiently subjects in the population of interest (1,2). Although eligibility for various research designs focuses typically on sociodemographic factors, the presence and number of remaining teeth are clearly relevant for dental health studies.

Prior research has documented varying degrees of validity of subjects' self-reported number of remaining natural teeth (as compared to the true number), using telephone (3), mail (4-8), and in-person (9-12) survey methods. However, the literature contains little information comparing the validity of the self-reported number of teeth across important sociodemographic groups. This void exists despite the fact that certain groups, such as racial minorities and persons who reside in poor households, have been targeted increasingly in research designs because of their increased risk for disease. Furthermore, to our knowledge, no information has been reported on how response rates to questions about the number of remaining teeth vary with the level of specificity required.

In this paper, we report findings from the Florida Dental Care Study (FDCS), which was designed to develop a risk assessment model of oral health outcomes in middle-aged and older adults, with a special interest in understanding such risk in poor individuals, blacks, and residents of rural areas. The purpose of this report is to assess the effectiveness of the telephone method in gathering tooth count information by measuring response rate (cooperation) to specific questions, and to assess the validity of the number of remaining natural teeth as reported during a large-scale telephone screening survey used to identify subjects for the FDCS.

Methods

A telephone screening methodology was designed to identify persons who met these eligibility criteria: (1) resided in one of four counties of north Florida; (2) for one of these counties, resided in one of 21 urban zip codes; (3) spoke English; (4) were capable of engaging in a coherent telephone conversation; (5) resided in a private household; (6) were 45 years old or older; and (7) reported race as either black or non-Hispanic white. The rationale and procedures for this research design have been described in detail previously (13). A pool of 5,254 subjects who met these eligibility criteria was identified, of whom 3,998 subjects met an additional eligibility criterion of having at least one remaining natural tooth.

During an initial telephone screening interview, up to 30 questions were asked querying sociodemographic variables, nondental health problems, recency of last dental visit (as a transition to questions about number of teeth), and the number of remaining natural teeth. The question sequence regarding the self-reported number of remaining natural teeth started with:

Send correspondence and reprint requests to Dr. Gilbert, Claude Pepper Center for Research on Oral Health in Aging, Division of Oral Medicine, College of Dentistry, PO Box 100416, University of Florida, Gainesville, FL 32610-0416. Internet: gilbert@nervm.nerdc.ufl.edu. Dr. Duncan is with the Department of Health Services Administration, College of Health Professions, University of Florida. Dr. Kulley is with the Response Analysis Corporation, Princeton, NJ. Manuscript received: 2/20/96; returned to authors for revision: 5/29/96; accepted for publication: 12/20/96.

"Do you have at least one of your own natural teeth left in your mouth?" Those who responded "yes" or "don't know" were asked: "How many natural teeth do you have in your upper jaw?" Subjects who responded "don't know" to this question were asked: "Everyone does not get the same number of natural teeth. Please take a moment to count the number of natural teeth you have in your upper jaw." For those who said "don't know" to that question, "We realize that you may not know the exact number, but would you guess there are no natural teeth, one to six teeth, or more than six natural teeth in your upper jaw?" Similar questions were asked about the lower arch.

At a second stage of the research, a stratified random sample of 1,800 subjects, all of whom reported having at least one natural tooth, was selected for a baseline in-person interview and clinical dental examination. From these 1,800 subjects, the project successfully recruited 873 subjects for baseline data collection. To estimate interexaminer reliability, six dentist examiners conducted replicate examinations with 82 of these 873 subjects. Examiner pairs agreed on the number and location of remaining teeth for all but one subject, for whom the examiners disagreed about one tooth only. Other sampling and methodologic procedures have been described previously (13-16).

The validity of subjects' reports of the number of remaining natural teeth was assessed by comparing the reported number to the number determined by actual clinical examination. A total of 104 (weighted value) subjects from the baseline examination were excluded from this assessment because they reported at baseline that they had had at least one tooth removed since their telephone screening interview or within the previous year.

Three levels of specificity for reporting number of teeth were possible: (1) nominal (at least one tooth or not, for either arch or for the entire mouth); (2) ordinal (no teeth in a given arch, 1–6 teeth in a given arch, or 7 or more teeth in a given arch); or (3) interval (a specific number for each arch; technically, this is a "ratio" scale, but for ease of presentation we refer to this as an interval scale). Because we did not examine subjects who reported during the telephone screening that they were edentulous, we can make no inferences regarding the validity of self-reported number at the nominal level (no teeth/at least one tooth for the entire mouth). However, it is of interest that one subject was excluded from the study because she reported 28 remaining natural teeth, but upon examination had full dentures, i.e., no remaining natural teeth (a "false positive"). All other subjects who came for the baseline data gathering session reported correctly that they had at least one tooth in the entire mouth. Because we only examined subjects who reported having one or more teeth, we had no opportunity to assess validity among those who reported having no teeth.

We assessed validity at the ordinal level by measuring concordance between the self-reported ordinal category and the category as determined by clinical examination. We used the Mantel-Haenszel chi-square trend test to test differences in concordance between groups. We assessed validity at the interval level by comparing the number of teeth that the subject reported with the number determined clinically. We used the intraclass correlation coefficient (ICC) and linear regression to test differences at the interval level. Conforming to the Shrout and Fleiss designation of ICCs (17), we used the "ICC (2,1)." In most estimates of reliability and validity, the ICC is preferable to the Pearson's correlation coefficient because it quantifies the similarity of the actual scores by raters, not merely the similarity of their relative ratings (18). In the multivariate regression described in the "Results" section below, we included clinical variables (number of remaining teeth, number of retained root fragments, number of fixed prosthetic pontics or cantilevers) and sociodemographic variables (area of residence, age group, sex, race, poverty status, and highest level of formal education) in an exploratory fashion, hypothesizing that these variables might be associated with validity.

Analyses reported here were done using SAS System for Windows® (19). Comments about statistical significance refer to probabilities of less than .05. Multicollinearity was measured using a procedure described by Belsley, Kuh, and Welsch (20); ultimately, no multicollinearity was observed. Results were weighted using the sampling proportions to reflect the population in the counties studied and stratification thereon, using procedures described previously (13). For example, although 35 percent of the sample of 873 subjects was poor, the weighted percentage was 16 percent to reflect the percent of 45-year-old or older persons in these counties who were actually poor.

Results

Response Rates at Three Levels of Specificity. Response rates were lower as the degree of reporting specificity increased. Almost all subjects answered the question at the nominal level (99.6%) and almost all subjects (98.8%) responded to the questions at the ordinal level. Seventy percent (weighted value) of dentate subjects reported a number at the interval level for the upper arch, while 76 percent (weighted) did so for the lower arch. When analysis was limited to subjects who participated for the baseline session (weighted n=769, excluding those who reported at baseline that they had at least one tooth removed since their telephone screening interview or within the previous year), the number of remaining teeth, as determined by clinical examination, was strongly associated with self-reporting at the interval level: 93 percent of subjects with 1–8 teeth reported at the interval level, compared to 84 percent of subjects with 9–16 teeth, 61 percent of subjects with 17–24 teeth, and 61 percent of subjects with 25 or more teeth (Mantel-Haenszel chi-square trend test, P<.001).

Validity of Self-reported Number of Teeth. *Ordinal-level Reports.* When the level of report was based at the ordinal level, 85 percent of subjects reported the correct ordinal category for both the upper and lower arches, 10 percent reported the correct category for only one of the two arches, and 5 percent gave incorrect reports for both arches. Subjects with more teeth, whites, subjects who were not below 100 percent of the US poverty level, and high school graduates were more likely to have valid self-reports (Table 1). No statistically significant differences in concordance were observed based on age group, sex, or area of residence.

The actual number of teeth was strongly associated with concordance in the upper arch: 96 percent of sub-

jects with no teeth in the upper arch were concordant, as were 92 percent of subjects with seven or more teeth in the upper arch, but only 56 percent of subjects with 1–6 teeth ($P<.001$; Mantel-Haenszel chi-square trend tests; not shown in Table 1). This trend was not evident in the lower arch: 85 percent of subjects with no teeth in the lower arch were concordant, compared to 90 percent of subjects with seven or more teeth in the lower arch, and 90 percent of subjects with 1–6 teeth.

Interval-level Reports. When the unit of analysis was the entire baseline sample, self-report at the interval level was a valid measure of the true number: the mean (SD) number of teeth reported during the telephone screening interview was 21.1 (9.1), and 21.2 (7.8) when determined by clinical examination. Similar results were observed for specific arches: for the maxillary arch, self-report was 10.2 (5.5) and by clinical examination, 10.0 (4.9); for the mandibular arch, self-report was 11.0 (4.4) and by clinical examination, 11.3 (3.5). When the unit of analysis was the individual subject, correlation was high, but not perfect: ICC (standard error) between subjects' self-reported number of teeth and the number as determined clinically was 0.89 (0.01) for the upper arch, 0.78 (0.02) for the lower arch, and 0.87 (0.02) for the total number of teeth.

We hypothesized that certain clinical characteristics (retained root fragments, fixed prosthetic pontics or cantilevers, and number of teeth) would be associated with the validity of self-reports. Our rationale was that root fragments and prosthetic teeth might have confused some subjects as to whether these should be included in their self-reported number, and some subjects may not have asked for clarification. We also hypothesized that subjects with many teeth would give less valid reports. Because of our interest in using the telephone screening methodology to target high-risk groups, we tested whether validity differed among subjects based on sociodemographic factors [age group (45–64 years old, 65+ years old), sex, race, poverty status, area of residence, and level of formal education].

Root fragments and prosthetic pontics and/or cantilevers were associated with validity of self-reported total number of teeth, although not in a monotonic fashion. The ICC between number of teeth by self-report and by examination was 0.88 (0.02) for subjects with no pontics, 0.69 (0.10) for subjects with one pontic, 0.73 (0.12) for subjects with two pontics, and 0.64 (0.14) for subjects with three or more pontics. The ICC between number of teeth by self-report and by examination was 0.88 (0.02) for subjects with no root fragments, 0.77 (0.09) for subjects with one root fragment, 0.68 (0.27) for subjects with two root fragments, and 0.87 (0.04) for subjects with three or more root fragments.

There was no tendency to overre-

TABLE 1
Ordinal-level Concordance Between Self-reported Category of Number of Teeth and Category Determined by Clinical Examination, by Number of Teeth, Race, Poverty Status, Level of Formal Education, Age Group, Sex, and Area of Residence (Weighted n=769)

Characteristic	Weighted n	% Persons with Characteristic Who Reported Correct Category in: Both Arches	1 Arch Only	Neither Arch
No. teeth in entire mouth, by clinical examination				
1–16	152	70	24	6*
17–24	224	80	12	9
25–32	389	94	3	2
Missing	4			
Race				
Black	211	75	15	10*
White	555	89	8	3
Missing	3			
Poverty status				
At or above poverty level	617	89	8	4*
Below poverty level	108	73	17	10
Missing	44			
Level of formal education				
Did not graduate high school	156	74	16	10*
Did graduate high school	612	88	8	4
Missing	1			
Age group (years)				
45–64	452	87	8	5†
65 or older	317	83	12	5
Sex				
Male	341	88	7	5†
Female	427	83	12	5
Area of residence				
Rural	385	84	11	5†
Urban	384	87	8	5

104 subjects were excluded because they reported at baseline that they had had at least one tooth removed since their telephone screening interview or within the previous year. Some characteristic totals do not add to 769 because of rounding after weighting and some row percentages do not add to 100 percent due to rounding. Subjects were asked to report the number of teeth in each arch separately, and to report within these categories: (1) no teeth in the arch, (2) 1–6 teeth in the arch, (3) seven or more teeth in the arch.

Results of Mantel-Haenzel chi-square trend tests:
*$P<.05$.
†Not statistically significant.

port or underreport, judging from a statistically nonsignificant ICC between the number of teeth determined clinically and the magnitude of differences between self-report and actual number. Nor was there a tendency for subjects to be less valid as the actual number of teeth increased, judging from a statistically nonsignificant ICC between the number of teeth determined clinically and the absolute value of the differences between self-report and actual number. Most subjects reported the correct number of remaining teeth in their entire mouths within three teeth (Table 2).

Some statistically significant differences in validity across relevant sociodemographic groups were found. Correlations between the number of teeth determined clinically and self-report were 0.90 (0.02) for whites and 0.81 (0.04) for blacks ($P<.05$); ICC was 0.89 (0.02) for subjects who were not below 100 percent of the US poverty level, and 0.74 (0.06) for subjects below the poverty level ($P<.05$). Differences between ICCs were not statistically significant for these groups: 0.87 (0.02) for subjects 45–64 years old and 0.86 (0.03) for subjects 65 years old or older; 0.88 (0.02) for rural residents and 0.86 (0.03) for urban residents; 0.88 (0.03) for males and 0.86 (0.02) for females.

A single ordinary least squares multivariate linear regression (REG procedure; SAS) of the absolute value of the difference between the self-reported number of teeth and the actual number was done. Because the distribution of the outcome variable, absolute value of the difference, was highly skewed, a $\log_{10}$ transformation was done. Three clinical variables were included in the model (actual number of teeth, number of prosthetic pontics and/or cantilevers, number of retained root fragments), along with the relevant sociodemographic variables. With the sociodemographic variables taken into account, each of the three clinical variables were significantly associated with the absolute value of the difference (number of teeth $\beta=0.005$, $SE\beta=0.002$; number of pontics $\beta=0.078$, $SE\beta=0.020$; number of root fragments $\beta=0.085$, $SE\beta=0.027$). With the three clinical variables taken into account, only one of the sociodemographic variables was significantly associated with the absolute value of the difference (age group $\beta=0.070$, $SE\beta=0.032$; 0=45–64 years old, 1=65+ years old). However, overall model fit was poor (adjusted $R^2=8\%$), suggesting that much of the variance in validity was unexplained and/or random.

TABLE 2

Percent Distribution of Subjects According to Absolute Value of Differences Between Interval-level Number of Remaining Natural Teeth Reported by Subject and Interval-level Number Determined by Clinical Examination, by Category of Number of Teeth Determined by Clinical Examination (Weighted n=506)

Absolute Value of Difference	Weighted n	Number of Teeth in Entire Mouth Determined by Clinical Examination (%)		
		1–16	17–24	25–32
0	120	31	9	28
1	112	29	22	18
2	90	13	19	20
3	45	5	12	9
4	34	7	7	6
5	33	5	6	7
6	20	4	3	4
7	16	1	5	4
8	8	2	4	1
9	8	1	3	1
10 or more	21	2	9	2

104 subjects were excluded because they reported at baseline that they had had at least one tooth removed since their telephone screening interview or within the previous year, and an additional 263 subjects were excluded because they did not provide interval-level tooth counts. Column total does not add to 506 because of rounding after weighting and some column percentages do not add to 100 percent due to rounding. Subjects were asked to report the number of teeth in each arch separately, and these two numbers were added to total the number of teeth in the entire mouth reported by the subject.

Discussion

When the unit of analysis is the overall sample, these FDCS results suggest that self-report is sufficiently valid to meet all but the most stringent of data requirements for measuring number of teeth. When the unit of analysis is the individual subject, this will likely not be the case.

Regarding tooth count estimation, the report in the literature most comparable to the FDCS is a study of 50 adults 70 years old and older in the Boston area (3). Douglass and colleagues (3) conducted a telephone interview that queried nominal- and interval-level reports of number of remaining teeth. Pearson correlation coefficients were higher for the Boston area sample (0.95 or greater) than the ICCs in the FDCS. (Although not reported, Pearson correlations and ICCs for the FDCS differed by less than 10 percent in those instances where the FDCS and the Boston study made similar comparisons.) Like subjects in the FDCS, there was no tendency for subjects in the greater Boston study to overreport or underreport the number of teeth; however, unlike the FDCS interval-level results, validity was less among subjects who had more teeth. Findings from studies using in-person interviews or mailed questionnaires have reported similar or greater validity of self-reports (4-12). Conclusions from these same studies regarding over- or undercounting, and the influence of the age of the subject, have been mixed (4-12).

Axelsson and Helgadóttir (7) have noted that most studies on the validity of self-reported number of teeth have been conducted in Western Europe, and they suggested that replication among populations with dissimilar socioeconomic backgrounds is needed. These FDCS results suggest that race, poverty status, and level of formal education are associated with validity of self-reports at the bivariate level, but not when clinical conditions are taken into account using multivariate methods. The multivariate model that we presented also suggests that much

of the variance in validity is unexplained and/or random.

Validity of self-reported tooth counts may be improved if the wording of the question explicitly distinguishes retained root fragments and prosthetic pontics/cantilevers, and does so in terms understandable to all respondents. Although the FDCS telephone screening interviewers were trained to distinguish root fragments and prosthetic replacements, we did not alter the wording of the questions to reflect that distinction. When the actual number of teeth in dentate persons is of primary interest in a study, these results suggest that validity would be improved if this distinction is emphasized in the text of the questions.

The lower response rates observed with questions that required higher levels of specificity is a salient issue for dental research conducted by telephone. Asking subjects to provide interval-level tooth counts during the context of a telephone screening interview apparently imposes a response burden that some subjects are unwilling or unable to bear. The primary interest in the FDCS was to gather information at the nominal level, because the objective of the telephone screening interview was to identify persons who had one or more teeth (as well as other eligibility criteria). We asked the ordinal- and interval-level questions because of our interest in ultimately assessing validity and because we speculated that this knowledge might make baseline appointment scheduling more efficient. That is, ordinal- and interval-level reports were of secondary interest, and telephone screening interviewers were instructed accordingly to weigh the information needs against the need to obtain an otherwise complete interview. Higher response rates to interval-level questions might be achievable in large-scale telephone interviews when interval-level reporting is of primary interest, and questioning methods and interviewer training reflect that primary interest. Our experience with continuation of the FDCS longitudinally suggests that this is the case. As we have continued to interview the 873 baseline participants by telephone every six months (currently at 48 months after baseline), we have had no refusals to answer interval-level tooth loss incidence queries.

Our results suggest that subjects' reported number of remaining teeth during a large-scale telephone screening interview is typically a valid measure of the true number, although the validity varies with the degree of specificity required and important characteristics of the subject reporting that number. The association between validity and these attributes should be taken into consideration during the design of dental survey research to be conducted by telephone.

Acknowledgments

This investigation was supported by a USPHS contract (DE-12587), with partial support from grants DE-11020, DE-00392, DE-08845, and AG-11183. Additional support was provided by funds from the University of Florida. The authors wish to thank D. E. Antonson, T. A. Dolan, U. Foerster, D. W. Legler, and M. L. Ringelberg, who, in addition to G. H. Gilbert, served as dentist examiners for the study. We are also grateful to J. L. Earls and B. A. Ringelberg, who served as regional coordinators for the baseline phase of the study.

References

1. Marcus AC, Crane LA. Telephone surveys in public health research. Med Care 1986;24:97-112.
2. Corey CR, Freeman HE. Use of telephone interviewing in health care research. Health Serv Res 1990;25:129-44.
3. Douglass CW, Berlin J, Tennstedt S. The validity of self-reported oral health status in the elderly. J Public Health Dent 1991;51:220-2.
4. Helöe LA. Comparison of dental health data obtained from questionnaires, interviews, and clinical examination. Scand J Dent Res 1972;80:495-9.
5. Könönen M, Lipasti J, Murtomaa H. Comparison of dental information obtained from self-examination and clinical examination. Community Dent Oral Epidemiol 1986;14:258-60.
6. Palmqvist S, Söderfeldt B, Arnbjerg D. Self-assessment of dental conditions: validity of a questionnaire. Community Dent Oral Epidemiol 1991;19:249-51.
7. Axelsson G, Helgadóttir S. Comparison of oral health data from self-administered questionnaire and clinical examination. Community Dent Oral Epidemiol 1995;23:365-8.
8. Lahti S, Tuutti H, Honkala E. Comparison of numbers of remaining teeth from questionnaires and clinical examination. Proc Finn Dent Soc 1989;85:217-23.
9. Hand J, Kohout F. Tooth loss estimates relying on subjects' tooth counts [Abstract]. J Dent Res 1992;71(Spec Iss):166.
10. Gilbert GH, Branch LG, Longmate J. Eligible veterans' choice between VA-covered and non-VA-covered dental care. J Public Health Dent 1992;52:277-87.
11. Gilbert GH, Duncan RP, Crandall LA, Heft MW, Ringelberg ML. Attitudinal and behavioral characteristics of older Floridians with tooth loss. Community Dent Oral Epidemiol 1993;21:384-9.
12. Widström E, Nilsson B. Dental health and perceived treatment needs in Finnish immigrants in Sweden. Scand J Soc Med 1984;12:129-36.
13. Gilbert GH, Duncan RP, Kulley AM, Coward RT, Heft MW. Evaluation of bias and logistics in a survey of adults at increased risk for oral health decrements. J Public Health Dent 1997;57:48-58.
14. Gilbert GH, Antonson DE, Mjör IA, et al. Coronal caries, root fragments, and restoration and cusp fractures in US adults. Caries Res 1996;30:101-11.
15. Ringelberg ML, Gilbert GH, Antonson DE, et al. Root caries and root defects in urban and rural adults. J Am Dent Assoc 1996;127:885-91.
16. Dolan TA, Gilbert GH, Ringelberg ML, et al. Behavioral risk indicators of attachment loss in adult Floridians. J Clin Perio 1997;24:223-32.
17. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420-8.
18. McDowell I, Newell C. Measuring health: a guide to rating scales and questionnaires. 2nd ed. New York: Oxford University Press, 1996:39.
19. SAS Institute. SAS/STAT user's guide. Version 6. 4th ed. Cary, NC: SAS Institute, Inc., 1989.
20. Belsley DA, Kuh E, Welsch RE. Regression diagnostics: identifying influential data and sources of collinearity. New York: Wiley, 1989.