Reliability of Random Digit Dialing Calls to Enumerate an Adult Female Population

Sally L. Glaser and Cynthia B. Stearns

Challenges to random digit dialing have been documented, but the reliability of random digit dialing outcomes from telephone number calling, household identification, and enumeration has never been addressed, despite its potential to bias population representativeness by affecting completeness of coverage. The authors explored interobserver reliability of calls to numbers generated by random digit dialing for a 1990–1996 population-based case-control study in San Francisco, California, area women, using data from a quality control effort in which 122 of 4,890 random digit dialing numbers were assigned to a second interviewer for recontacting within 4 months. The 34 numbers discrepant between the first and second calls did not differ from the 88 unchanged outcomes, and reliability was good (kappa = 0.65, 95% confidence interval: 0.55, 0.75). Eligibility (an adult woman in the household) was confirmed for nine of 11 eligible households. However, six of 29 households originally ineligible because of gender were eligible on recontact, and eligible residences rose from 24% to 39% between the two calls, although the two groups of eligible women did not differ in age or race. This underenumeration of women by random digit dialing confirms prior observations, although interviewer differences or changes in respondents or household composition between the first and second calls may have contributed. Recontact of gender-ineligible households may improve completeness of random digit dialing coverage for female populations. Am J Epidemiol 2002;155:972–5.

epidemiologic methods; random digit dialing; reliability

Random digit dialing has been widely used in population-based health research, including cross-sectional surveys and case-control studies. Its utility is premised on reaching a true random sample of the population at risk through complete population coverage (achieved by accessing telephone numbers that cover this population, screening (contacting) all or a random sample of these numbers, and enumerating members of each household) and having acceptable response rates for all contacts, including recruitment of eligible persons for participation in the research study (1–5). Although many of the difficulties using random digit dialing to reach a random sample of the population have been documented, the reliability of screening by random digit dialing (telephone number calling, household identification, and enumeration) has never been addressed, to our knowledge. Yet, this aspect of the random digit dialing process has the potential to introduce error into study findings if it adversely affects the completeness and representativeness of the households being enumerated.

For a population-based case-control study of reproductive factors in Hodgkin’s disease etiology in women conducted in the San Francisco, California, area in 1990–1996 (6–8), we evaluated several aspects of random digit dialing as a strategy to recruit controls. Here we report on the interobserver reliability of screening calls made by two different interviewers 2–4 months apart to random digit dialing numbers generated by simple random sampling.

MATERIALS AND METHODS

Study cases were English-speaking women aged 19–79 years with incident Hodgkin’s disease from July 1988 to December 1994 reported to the population-based Greater Bay Area Cancer Registry. Consistent with incidence patterns for this lymphoma, the interviewed cases were predominantly White (81 percent) and young adult (79 percent less than 45 years). Controls were English-speaking women without Hodgkin’s disease, identified by random digit dialing and frequency matched to cases on specific race/ethnicity and 5-year age group. Study participation involved a standardized, in-person interview with no financial compensation.

For the first 2 study years, random digit dialing was based on simple random sampling of the complete frame of telephone numbers generated from all 960 working area code/prefixes in the nine-county study area. The numbers were updated and consolidated when one new area code was introduced. Trained, experienced female interviewers were assigned batches of 100–125 randomly selected telephone numbers to dial up to 10 times, at different times of the day and days of the week. They confirmed the number reached before describing the study’s purpose (“to help pre-
vent a cancer that affects women”). After three calls answered by machine, they left a short message. Once a number was identified as residential, the interviewer attempted to enumerate all adult household females by age and race/ethnicity, explain the possibility of an interview in the future, and request a first name and address for further contact.

One batch of random digit dialing numbers (n = 122) called by a single interviewer in November and December of 1991 was reassigned for quality control purposes to a second interviewer, who called them in February 1992 following the identical protocol. The resulting data allowed us to assess interobserver reliability of random digit dialing screening and household enumeration. We used the chi-square statistic to test for significant differences (p < 0.05) between outcomes of the first and second calls and the kappa statistic to measure agreement beyond chance (kappa values of 0.40–0.75 are considered to indicate fair to good agreement, and kappa values above 0.75 are considered to indicate very good agreement) (9, 10).

RESULTS

The 122 numbers assigned twice represented a 2.5 percent random sample of the 4,890 simple random digit dialing telephone numbers called for the study. Table 1 shows that the overall distributions of screening outcomes did not differ significantly between the first and second calls (p = 0.56). While specific outcomes were discrepant for 34 numbers (28 percent), the outcome distribution of these 34 calls did not differ from that of the 88 unchanged outcomes (p = 0.46). Reliability of screening was good both overall (kappa = 0.65, 95 percent confidence interval: 0.55, 0.75) and among calls to presumed residences to enumerate household members (kappa = 0.59, 95 percent confidence interval: 0.37, 0.79). Of the five initially unanswered numbers, two were answered on recontact, with one an eligible household. Among the 72 numbers originally coded as nonhouseholds (businesses, not working, electronic), four were found on the second call to be households, one with an eligible woman. Of the 11 initially eligible households, eligibility was confirmed for nine (82 percent); for these nine, the specific ages (range, 32–63 years; mean, 43.1 years) and races/ethnicities (seven Whites, one Black, one Hispanic) of the eligible women agreed across both calls except for a single 2-year age difference. Among the 29 households originally ineligible because of gender, six (21 percent) were found on recontact to have eligible women; all the women were White, with ages ranging from 32 to 68 years (mean age, 50.4 years). Thus, between the first and second contacts, the number of eligible households increased from 11 to 18, and the proportion of presumed residences with eligible women rose from 24 percent to 39 percent. However, comparison of the available characteristics of the eligible women identified by the first and second interviewers showed no significant differences (mean ages, 44.1 vs. 42.4 years, p = 0.73; percent White, 72.7 vs. 81.8, p = 0.55).

DISCUSSION

One potential source of error in random digit dialing is inaccuracy of screening outcomes, particularly identification and enumeration of households. Such error could affect the validity of study results by impacting the completeness of the identified study population even where telephone coverage is high, as in the Greater Bay Area (98 percent), and thus possibly the representativeness of the population at risk. Although the completeness of random digit dialing population enumeration has been evaluated previously by comparison with other standards (4, 11–14), we found no data published about the interobserver reliability of random digit dialing outcomes. Therefore, we used quality control data for our study targeting an adult female population to explore random digit dialing reliability, finding that it was good across all outcomes of our small sample of telephone numbers. However, we detected underreporting of eligible households and, of particular concern, misclassification of eligibility based on gender for approximately one fifth of the

<table>
<thead>
<tr>
<th>Initial random digit dialing call†</th>
<th>No answer, 10 tries</th>
<th>Business/pay phone</th>
<th>Not working/disconnect</th>
<th>Electronic/fax/pager</th>
<th>Refused enumeration</th>
<th>Language problem</th>
<th>Refused to be on hold</th>
<th>Ineligible for gender</th>
<th>Eligible</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No answer, 10 tries</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>19</td>
<td>122</td>
</tr>
<tr>
<td>Business/pay phone</td>
<td>1</td>
<td>14</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>19</td>
<td>1</td>
<td>48</td>
<td>122</td>
</tr>
<tr>
<td>Not working/disconnect</td>
<td>3</td>
<td>6</td>
<td>38</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>56</td>
<td>122</td>
</tr>
<tr>
<td>Electronic/fax/pager</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>44</td>
<td>122</td>
</tr>
<tr>
<td>Refused enumeration‡</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>44</td>
<td>122</td>
</tr>
<tr>
<td>Refused to be on hold‡</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ineligible for gender‡</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
<td>122</td>
</tr>
<tr>
<td>Eligible‡</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
<td>122</td>
</tr>
<tr>
<td>Total</td>
<td>7</td>
<td>22</td>
<td>43</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>20</td>
<td>18</td>
<td>11</td>
<td>122</td>
</tr>
</tbody>
</table>

† Called in late 1991 by one study interviewer.

‡ Presumed residences defined as including enumeration refusal, language problems, refusal to be placed on hold, ineligible because of gender, and eligible outcomes.
households. For our simple random-sampling random digit
dialing, a total of 224 households were classified as having
no eligible women; if 20 percent of these did in fact have
adult females, the number of eligible households would
have risen from 428 to 473, an 11 percent increase. This
difference raises concerns regarding bias due to systematic
underenumeration in the original calls, although we found
no evidence of age or race differences in the small group of
eligible females enumerated by our interviewers. Such
underreporting also would have affected study cost, as
excess screening would have taken place to generate the
required number of eligible households for the study.

The underrepresentation of women by random digit dial-
ing has been reported by others, most recently Brogan et al.
(11), who found underenumeration of approximately 11 per-
cent of females aged 20–54 years in Atlanta, Georgia,
Seattle, Washington, and New Jersey in 1990–1992, com-
pared with area probability sampling and census data. As
Brogan et al. determined that underascertainment of the
female population in their study could not be attributed to
households that were unscreened, underreporting of house-
hold women in our data may reflect a purposeful attempt to
maintain privacy and security. Both our interviewers were
female and White, and having race-matched callers
increases participation among study participants (15). Never-
theless, because the first and second calls were made
by different interviewers, the enumeration disagreement
may represent a response to subtle differences between
them, such as their respective persuasiveness (16) or the
lower voice of the first interviewer. Such effects may have
prompted hesitation about enumerating household women
in some respondents (3), and the extent of underascertain-
ment might have differed if the first calls had been made by
the second interviewer.

Disagreement between the first and second random digit
dialing calls may have explanations unrelated to interviewer
activity, including different respondents on the second call or
an influence of the first random digit dialing contact on the
respondent. Response to random digit dialing targeting
females has been shown to be higher following prior contact
by mail (12) or in person (11); thus, in our study, the higher
proportion of households with eligible women in the second
calls may represent a similar effect due to the repeated con-
tact. In addition, phone numbers could have been reassigned
between random digit dialing calls, especially in a mobile
region like the Greater Bay Area, although Voigt et al. (4)
found minimal conversion of nonresidential to residential
telephone numbers after 1 year, and we further minimized the
possibility of reassignment by timing the second calls within
between 2 and 4 months of the first. The magnitude of under-
reporting of females in our data (20 percent) also makes true
household change a somewhat unlikely explanation.

This first information on interobserver reliability of ran-
dom digit dialing screening of English-speaking households
for adult women suggests that random digit dialing reliability
is good. The interobserver differences we did note are consis-
tent with prior observations of underenumeration of women
but did not seem to affect the representativeness of the eligi-
bale households. However, a larger study with more detailed
information about household change and with interviewers’
undertaking a mix of first and second calls would be required
to confirm our findings. In the meantime, recontact of house-
holds ineligible on the basis of gender, possibly using a sec-
ond interviewer, may be one simple, cost-effective tactic for
improving the completeness of random digit dialing coverage
for studies targeting female populations.

ACKNOWLEDGMENTS

This work was supported in part by grant NCI R29
CA50381. Cancer incidence data were collected by the
Northern California Cancer Center under contract N01-CN-
65107 with the National Cancer Institute, National Institutes
of Health, and with support of the California Cancer
Registry, a project of the Cancer Surveillance Section, Cali-
fornia Department of Health Services, under subcon-
tact 050M-8701/8-S1522 with the Public Health Institute.

The authors thank Scarlett Lin, Rebecca Nugent, and Dr.
Christina Clarke for their contributions to this project.

REFERENCES

1. Potthoff RF. Telephone sampling in epidemiologic research: to
reap the benefits, avoid the pitfalls. Am J Epidemiol 1994;139:
967–78.
2. Wacholder S, Silverman DT, McLaughlin JK, et al. Selection
of controls in case-control studies. II. Types of controls. Am J
3. Hartge P, Brinton LA, Rosenthal JF, et al. Random digit dial-
ing in selecting a population-based control group. Am J
Epidemiol 1984;120:825–33.
4. Voigt LF, Davis S, Heuser L. Random digit dialing: the potential
effect on sample characteristics of the conversion of nonresi-
5. Davis S, Ross A, Voigt LF, et al. Qualitative and quantitative
assessment of geographic clustering of population samples
selected using different methods of random digit dialing. Am J
Hodgkin’s disease risk factors in older women: evidence from
a population-based case-control study. Leuk Lymphoma 2001;
7. Glaser SL, Dorfman RF, Clarke CA. Expert review of the diag-
nosis and histologic classification of Hodgkin’s disease in a
population-based cancer registry: interobserver reliability and
impact on incidence and survival rates. Cancer 2001;92:
218–24.
8. Glaser SL, Clarke CA, Nugent RA, et al. Social class and risk of
measurement in epidemiology. Oxford, United Kingdom:
10. Landis JR, Koch GG. The measurement of observer agreement
phone sampling and area sampling: response rates and within-
12. Funkhouser E, Macaluso M, Wang X. Alternative strategies for
selecting population controls: comparison of random digit
dialing screening of English-speaking households. Am J Epidemiol
Vol. 155, No. 10, 2002