AMERICAN JOURNAL OF EPIDEMIOLOGY
Vol. 120, No. 4

Printed in U.S.A.

THE EFFICACY OF RANDOM DIGIT DIALING IN SELECTING MATCHED CONTROLS FOR A CASE-CONTROL STUDY OF PEDIATRIC CANCER

ELIZABETH M. WARD,[1,2] SHIRA KRAMER[1] AND ANNA T. MEADOWS[1]

Ward, E.M., S. Kramer (Children's Hospital Cancer Research Center, Philadelphia, PA 19104) and A. T. Meadows. The efficacy of random digit dialing in selecting matched controls for a case-control study of pediatric cancer. *Am J Epidemiol* 1984;120:582-91.

A study of random digit dialing as a method of control selection was conducted as part of a matched case-control study of risk factors for neuroblastoma. The study was conducted in 1981–1982 in the Greater Delaware Valley region. Participation involved parents submitting to a 2–4-hour telephone interview. Potential control telephone numbers were selected by substituting the last two digits of the case family's telephone number with randomly selected digits. Successive random telephone numbers were called until a family with a child matching the case by year of birth plus or minus three years and race who agreed to participate was identified. A total of 3,245 telephone numbers were dialed to select controls for 101 cases, an average of 32.1 telephone numbers per case. Among 1,908 residential households contacted, 25.5% refused to give information about household characteristics. Of households with known characteristics, 45% had children under 25 years of age and 14.9% had a child eligible for the study. Among 181 households invited to participate, 101 agreed. For 18.6% of all calls, residential respondents were randomly offered $10 as an inducement to participate. This inducement did not significantly influence participation rates among families invited to participate.

epidemiologic methods; interviews; telephone; retrospective studies

Random digit dialing has been widely used in marketing and survey research to generate probability samples of residential households for telephone interviews (1). Its application in epidemiologic research is relatively recent. Interest in the random digit dialing method stems from increasing use of the telephone as a means of conducting health surveys, the ability of random digit dialing to generate population samples in case-control studies, and difficulties in traditional methods of selecting population controls for case-control studies.

A sample of the general population is considered the optimal control group when a population-based case series is used (2). Traditional methods of selecting population controls, including constructing sampling frames in a defined geographic area and door-to-door surveys of households have become prohibitively

Received for publication January 26, 1983, and in final form January 20, 1984.

[1] The Children's Cancer Research Center, the Children's Hospital of Philadelphia, 34th and Civic Center Boulevard, Philadelphia, PA 19104. (Reprint requests to Dr. Kramer.)

[2] Current address: Industrywide Studies Branch, National Institute for Occupational Safety and Health, Robert A. Taft Laboratories, Cincinnati, OH.

This work was supported in part by grant number CA 14489 and CA 29275 from the National Institutes of Health, BSRG# RR-05415, and grant number IN-136 from the American Cancer Society.

The authors gratefully acknowledge the dedicated work of Mary Gordon, Helen Pedersen and Carol Symmes in conducting the random digit dialing interviews, and further thank Patricia Jarrett and Ira Caplan for their assistance in the computer analysis.

expensive, especially when controls from a large geographic area are required. Random digit dialing, a method of selecting population controls that reduces time and cost expenditure, may be used in populations where a high proportion of residential households have telephones.

Because of the increasing popularity of random digit dialing in epidemiologic research it is important to examine its efficiency as a method of control selection. Among the issues important to examine are the effort required, the rate of refusals, and the representativeness of participating controls with reference to all possible eligible controls. In order to answer some of these questions, we are reporting the results of our experience in utilizing the random digit dialing method for selecting controls in a case-control study of neuroblastoma, a pediatric tumor.

MATERIALS AND METHODS

This study included 101 cases of neuroblastoma that occurred in a 31-county region in the Greater Delaware Valley over the period 1970–1979, identified through the Greater Delaware Valley Pediatric Tumor Registry, a population-based pediatric tumor registry located at the Children's Hospital of Philadelphia. The 101 cases were predominantly white and under four years of age at the time of diagnosis. All case families interviewed had telephones at the time of the child's diagnosis.

One individually matched control per case was selected through telephone random digit dialing. The sampling frame for each case consisted of the 99 telephone numbers that could be generated by replacing the last two digits of the case family's telephone number at the time of diagnosis with randomly selected digits. Thus, the area code and first five digits of the case telephone number were retained, thereby matching controls to cases on broad residential area. Although telephone exchanges do not always match civil boundaries, telephone exchanges in the Greater Delaware Valley region cover relatively small geographic areas, according to telephone company maps. There was a small probability of selecting controls outside the study area because exchanges on the edge of the Greater Delaware Valley may cover territory outside the region. None of the 101 controls selected posed this problem. Each of the numbers within the sampling frame was called until a family with a control child, matched to the case by date of birth (plus or minus three years) and race who agreed to participate was identified. Random two-digit numbers were generated using a random number generator on a Hewlett-Packard 67 calculator (Corvallis, OR). Each telephone number selected was called until a conclusive outcome was reached or until eight unsuccessful attempts were made. Numbers were called varying time of day and day of week (including nights and weekends) to be sure to include households where all adults work. The interviewers used a standard script in which they identified themselves and the study, verified the telephone number, and established whether the number was residential or nonresidential. Residential respondents were asked the number of adults living in the household and the number of children under age 25. If the respondent had children, he or she was asked to provide the birthdates of the children, race, name and address. The interviewer recorded the date, time and outcome of each attempt as well as other information relating to the call on a precoded form. Classification of outcomes are given in table 1. The sex of the respondent, the presence of a speech or language barrier, and section of the script in which a respondent refused to provide further information were also recorded.

Eligible families were contacted by letter and invited to participate in the

TABLE 1

Classification of results of each dialing and final status for each telephone number selected

Result of each dialing (1–8 attempts)
Nonworking no.
Nonresidential no.
Ring, no answer
Line busy
Unable to verify no.
Residential no., respondent requested that interviewer call back
Residential no., telephone no. verified, interview attempted
Residential no., telephone no. verified, no adult member of the household available for interview after 8 tries or adult member refused to come to the telephone
Other
Final status
Nonworking no.
Nonresidential no.
No answer after 8 attempts, including 1 weekend attempt
Residential with no children
Residential with eligible children
Residential with noneligible children
Residential refusal prior to finding out if children are eligible
Residential refusal prior to finding out if there are any children
Unknown (and interviewer error)

TABLE 2

Outcome of calls to 3,245 telephone numbers dialed for selection of controls

Result	No.	%
Nonworking no.	793	24.4
Nonresidential no.	332	10.2
Not answered in 8 attempts	158	4.9
Residential no.	1,908	58.8
Unknown (and interviewer error)	54	1.7
Total	3,245	100.0

study. One week after the letter was sent, the interviewer called to schedule an appointment for the study interview. If the respondent was reluctant to participate, the interviewer attempted to discuss reasons for refusal and encourage participation. Participation involved a telephone interview of both parents concerning maternal and paternal occupational, residential, medical, drug and reproductive histories, as well as medical histories of all children in the family. Total interview time per family was two to four hours.

RESULTS

A total of 3,245 numbers were called to select controls for 101 cases, an average of 32.1 numbers per case. Table 2 shows the outcome of telephone calls to these numbers, of which 24.4 per cent were nonworking, 10.2 per cent were nonresidential, and 4.9 per cent were never answered. A total of 1,908 numbers dialed (58.8 per cent) were residential. As shown in table 3, 25.5 per cent of residential respondents refused to provide information about household characteristics and their eligibility for the study could not be assessed. Among residential respondents who did provide this information, 45.3 per cent had children and 15.0 per cent had a child who met the eligibility criteria for the study.

The distribution of number of attempts required to reach a final outcome for any individual number is shown in table 4. Among residential numbers, 51.8 per cent answered in the first attempt, while 56.5 per cent of nonresidential numbers answered in the first attempt. Attempts to contact residential households were most likely to be successful when made on weekday evenings or weekends when about 50 per cent of calls could be assigned a final outcome, compared with calls on weekday mornings and afternoons, when about 34 per cent of calls had a definitive outcome (table 5). Approximately 6 per cent of calls made to residential households yielded busy signals, and in about 11 per cent of calls the interviewer was asked to call back at another time, usually because adult members of the household were not at home.

Almost one third (30.7 per cent) of all

TABLE 3

Outcome of calls to residential phone numbers identified by random digit dialing

Outcome	No.	% of residential nos.	% of nos. with known characteristics
Provided household characteristics	1,421	74.5	
No children	777	40.7	54.7
Ineligible children	431	22.6	30.3
Eligible children	213	11.2	15.0
Among the 1,421 who provided household characteristics, those who refused to provide name and address	100	5.2	7.0
Ineligible children	79	4.1	5.6
Eligible children	21	1.1	1.5
Refused prior to determining household characteristics	487	25.5	
Total*	1,908	100.0	

* The totals sum those who provided household characteristics plus those who refused prior to determining household characteristics.

TABLE 4

Number of telephone calls necessary to reach conclusive outcome

No. of calls	Residential phone nos.		Nonresidential phone nos.	
	No.	%	No.	%
1	978	51.8	187	56.5
2	389	20.6	78	23.6
3	230	12.2	39	11.8
4	100	5.3	9	2.7
5	81	4.3	6	1.8
6	49	2.6	5	1.5
7	29	1.5	3	0.9
8	30	1.6	4	1.2
Total*	1,886†	100.0	331†	100.0

* This excludes numbers that were never successfully contacted.

† For 1% of telephone numbers, the number of attempts was unknown due to interviewer error.

respondents failed to provide some information requested by the interviewer. Refusal rates by section of interview are shown in table 6. The largest category of refusals (51.8 per cent) occurred during the first part of the interview in which the purpose of the call was described. Among residential households contacted, 25.5 per cent refused before questions concerning household characteristics were completed and 5.2 per cent answered all questions regarding household characteristics but refused to provide their name and address. If we assume that residential households for which we could not determine household characteristics were similar to those for which we could, 14.9 per cent of the 481 households, or 72 households, would have had an eligible child. The second most common section in which refusals occurred was the section in which respondents were asked to provide their name and address so that they could be sent further information about the study. Of the 100 respondents who provided household characteristics but refused to provide their name and address, 21 had a child who was eligible for the study.

A total of 181 families were invited to participate in the study. Of these, 101 agreed to participate and 58 refused after receiving the letter describing the study. An additional four families expressed willingness to participate in the study but cited specific reasons why they were unable to do so, in most cases related to serious illness or death in the family. Four

TABLE 5

Outcome of telephone calls to residential households by time of day and week

Time of day and week	Ring, no answer (%)	Requested call-back (%)	Busy signal (%)	Conclusive outcome (%)	Other* (%)	No. of calls made
Weekday morning	46.6	11.1	6.0	32.6	3.6	1,121
Weekday afternoon	46.4	11.2	5.0	34.2	3.1	1,187
Weekday evening	24.0	11.8	7.0	51.4	5.8	1,454
Weekend	24.7	10.8	6.1	50.5	7.8	408

* 89% of calls coded under "other" involved inability to verify the telephone number dialed.

families were found to be ineligible after the letter was sent because children in the family were adopted. Twelve respondents initially agreed to participate in the study but repeatedly failed to keep the appointments set for the interview. However, among respondents who agreed to participate, most were fully cooperative. Only two control families began the interview and did not complete it.

A number of strategies were attempted to improve the response rate. Individuals who expressed reluctance to provide information during the initial telephone contact were offered the option of calling a number to validate the legitimacy of the survey. Early in the study we assessed the usefulness of offering a $10 payment as an incentive to participate. A total of 605 telephone numbers were randomly assigned to a payment or nonpayment group. Respondents in the payment group were offered payment after the interviewer verified the telephone number but before questions about household composition were asked. The results of this trial are summarized in table 7. Overall, offering payment did not alter the rate of refusals occurring after the section of the interview in which payment was offered. If anything, offering payment appeared to have a negative effect, with 30 per cent of "paid" respondents refusing to provide their name and address compared with only 8 per cent of "nonpaid" respondents. Among households invited to participate, the participation rates among the two groups were practically identical.

TABLE 6

Section of interview in which telephone respondents refused to give further information

Section content	No. of refusals	% of refusals
Introduction of study	304	51.8
Validation of telephone no.	25	4.3
No. of people living in household	95	16.2
No. of children	9	1.5
Birthdates of children	43	7.3
Race	11	1.9
Name and address	100	17.0

TABLE 7

Comparison of refusal rates among telephone respondents offered and not offered payment for participation

	Paid	Not paid
Total no. dialed	306	299
No. residential	179	182
No. refusing at any time during initial phone call	43	37
% refusals occurring prior to offering payment	39%	65%
% refusals occurring after offering payment	60%	35%
% refusals to give household characteristics	30%	27%
% refusals to give name and address	30%	8%
No. invited to participate	16	19
No. agreeing to participate	9	11

The relatively high refusal rate observed in this study causes two major problems; 1) a measurable decrease in efficiency in the method because a refusal

initiates another round of telephone calls to identify an eligible control, and 2) a potential decrease in the validity of the study because of the potential selective nature of the group who agreed to participate. To illustrate the loss of efficiency caused by refusals, the median number of telephone numbers dialed to identify the first eligible control for each case was six, while the median number of telephone numbers dialed to identify a participating control was 16. For 13 cases, over 50 telephone numbers were dialed to identify a participating control, and for five cases the sampling frame of 99 numbers was exhausted before a participating control was found. For four of these cases, the last three digits of the telephone number were randomized to generate a larger sample. For one case where 150 telephone numbers had been called to identify a control with no success, a control was selected from among eligible children residing in a telephone exchange area adjacent to the telephone exchange area of the case.

In considering the potential decrease in validity related to nonparticipation, 57 of the 101 controls interviewed in the study were the first eligible control selected for the case. For 24 cases, the second identified control was interviewed, and for 10 cases the third identified control was interviewed. For 10 cases, four to six controls were invited to participate before one agreed.

It is difficult to assess any bias that may result from initial telephone refusals because we have so little information about their characteristics. Among respondents who provided household information (74.5 per cent of residential numbers) we compared characteristics of families who were invited to participate and agreed, those who were invited and refused, and those who were eligible for the study but refused to give their name and address (table 8). No differences were found in mean number of children, mean age of children, sex of respondent or race.

Discussion

Two kinds of sampling frames have heretofore been used for telephone surveys: published directories and randomly generated numbers (usually the set of all possible four digit numbers within working exchanges). The primary advantage of using random digit dialing rather than published directories is that over 20 per cent of residential numbers are not listed in current directories, which do not include unpublished numbers or recent listings (1). However, random digit di-

TABLE 8

*Comparison of household characteristics of eligible families who participated in the study and families who refused**

	No. of families	Mean no. of children	Mean age of children, January 1, 1982	% females	% whites
Participated in study	101	2.56	9.4	77.6	96.0
Invited to participate and refused	58	2.74	9.4	82.1	93.1
Refused to give name and address	21	2.09	10.1	73.7	100.0

* Twenty-two families who were contacted to participate in the study are not represented in the categories in this table. Four families were willing but unable to participate due to family illness, four were found to be ineligible, 12 agreed to participate but repeatedly failed to keep appointments and two began, but did not complete, the interview. Eleven families found to be eligible during the screening interview were not contacted, for the most part because they were identified after an eligible control had been selected for the case.

aling has been reported to be more costly than use of published directories because approximately 80 per cent of numbers in the sampling frame are not assigned to households (1). Two refinements of these methods have been proposed. Sudman (4) first samples listed numbers and then replaces their last three digits with three-digit random numbers. For each of the first stage selections, enough random three-digit numbers are used so that a predetermined number of listed household numbers are selected. Waksberg (3) proposed creating primary sampling units from area codes, working exchanges and all possible choices of the next two digits. This requires obtaining a list of all area codes and exchanges in use in the area being sampled. A random selection is made of an eight-digit number from the sampling frame and the next two digits. If the number reached is a residential household, additional numbers are generated within the same primary sampling unit (by random selection of the last two digits). Selection within the primary sampling unit continues until a set number of residential households are reached. This process is repeated until a predesignated number of primary sampling units are chosen. The goal of both methods is to approximate a probability sample of households. However, Sudman's method is not a true probability sample because it excludes unlisted household numbers from bands of 1,000 that do not contain listed numbers. In both methods, households have an equal probability of selection only if each bank or cluster contains at least the desired number of households to be sampled (1).

We chose to employ individual matching in control selection for several reasons. We matched on race and year of birth because they might be related both to the probability of developing the disease and the probability of certain exposures. We wanted to ensure that the geographic distribution of cases and controls was similar because the Greater Delaware Valley region is so diverse. A probability sample of 101 households from among the 2.2 million children in the region might be highly skewed with respect to place of residence by chance alone. Because telephone exchanges cross political boundaries, the control selected for a particular case may not be exactly matched by city, town or county, but matching on telephone exchange does ensure that the distribution of area of residence among cases and controls is similar. Stratified probability sampling by county of residence, which would have achieved a similar result, posed practical difficulties because of the effort that would be required to map working exchanges or listed numbers. Because the characteristics of the cases, including their telephone numbers, were known in advance, it was possible to use the case telephone numbers as a basis for selecting the control sample. The matched method is particularly useful for small, locally based case-control studies because it does not require access to listings of working exchanges or a complicated sampling design.

However, in choosing this type of matched design for a particular study, its potential disadvantages should also be considered. The efficiency of the procedure (number of residential numbers dialed to identify an eligible control) is inversely related to the strictness of the matching rule. If it had been decided to match the cases by county of residence rather than by the blocks of telephone numbers, the number of calls required to identify an eligible family would have been reduced. However, the construction of the sampling frame would have required a great deal more effort. A second concern with this method is the possibility of overmatching of cases and controls for occupational and environmental risk factors by close geographic matching.

In selecting this method of choosing controls, a number of issues were considered. One is the desirability of choosing population controls relative to controls identified through hospitals, friends, neighbors or relatives. In this study, hospital controls were considered inappropriate because the Children's Hospital of Philadelphia, which treats 75 per cent of childhood cancer cases in the Greater Delaware Valley region, is a tertiary referral center and hospital selection bias was a serious concern. We reasoned that children treated at the Children's Hospital of Philadelphia might not be representative of children in the Greater Delaware Valley region in geographic distribution and socioeconomic status, and there was a risk of overmatching in terms of etiologic factors related to childhood illness. We considered that friend, neighborhood or relative controls would lead to overmatching of important study characteristics such as type and place of employment. Birth certificate controls were not available because of state regulations.

We consider this method highly efficient in terms of time and cost. An average of 32.1 numbers per case were called to select controls for 101 cases, with an average of 2.2 dialings required to reach a conclusive outcome for each number. A single interviewer could complete an average of 20 telephone calls per hour, so that an average of 3.5 hours of random digit dialing was required to select a control. Furthermore, interviewers were usually trying to find an average of 15 controls at any given time and could conveniently schedule one to two-hour random digit dialing sessions three times a day on days when the actual study interviews were also scheduled. Although we know of no published estimates of the time involved in selecting hospital controls, which would in any event depend on how the medical records are organized and the selection criteria used, we believe that the average of three and one half hours required to identify a population control by the random digit dialing method is probably no more time-consuming than identifying hospital controls. Given the large area encompassed by the Greater Delaware Valley region, this method was certainly less time-consuming than selecting population controls by classical survey sampling techniques.

The major potential drawback of this method is the refusal rate. Approximately one third of respondents did not provide some important information during the random digit dialing telephone contact, and over one third of families who were invited to participate refused. Of particular concern are the 20–25 per cent of residential numbers dialed that terminated the initial telephone contact before respondent demographic characteristics or eligibility could be ascertained. These refusal rates raise the question of the degree to which participating controls represent a selected population.

The relatively high refusal rates obtained in this study may be related to the fact that many of the calls were made to residents in the highly urbanized Philadelphia-Camden area. In a national telephone survey conducted by the Michigan Institute for Survey Research, the response rate varied by degree of urbanization. The response rate in the 12 largest Standard Metropolitan Statistical Areas (SMSA's) was 65.5 per cent, whereas in rural areas it was 69.7 per cent and in "other SMSA's" it was 74.8 per cent (1). Further evidence that refusal rates vary by area of the country sampled comes from results of a feasibility study which utilized random digit dialing to select controls for a national case-control study of Ewing's sarcoma, using a script and method similar to those reported in this paper. In this study (9), age (plus or minus one year), race and sex-matched controls were identified in an average of 42 working numbers selected, and only

6.8 per cent of residential numbers contacted refused to provide information about household characteristics. However, refusal rates in the northeastern United States were higher than those in other areas of the country. The relatively high refusal rates in our study among families invited to participate may be related to the effort and time required for the interview, and to the detailed and personal nature of the questions asked.

Recent data suggest that improved response rates may be achieved by conducting the actual study interview for eligible subjects during the screening telephone call (5). Screening for eligibility and asking the respondent to provide an address to which a letter explaining the study could be sent resulted in a screening response rate 16 percentage points lower than conducting the actual study interview without asking full name and address. By using the latter method, a response rate of 91 per cent was achieved in this study conducted in the Washington, DC area. The high response rate is probably related to the brevity of the interview (5–10 minutes). This approach is not feasible in a study requiring a longer interview because respondents are unlikely to be available for an hour or more without prescheduling. Another report has suggested that callbacks to households who were not interviewed in the initial survey attempt can improve the overall response rate. Callbacks to 772 households within one to four weeks after the initial refusal resulted in interviews in about 35 per cent of households (6). Other outcomes for which callbacks were made included partially completed household interviews, "ring, no answer" after 11 attempts, and households where subjects were not at home or requested to be called back at another time. Overall, recontacting 61 per cent of potential callbacks increased the response rate from 74 to 81 per cent.

Alternate methods of choosing controls pose similar problems in achieving participation of controls selected. If potential controls are selected from birth certificate or medical records, potential controls may be lost by failure to trace and contact control families as well as by refusals. For example, in a case-control study of risk factors for brain tumors in children, which was conducted in the Baltimore, Maryland area, controls were selected from birth certificates. Of the 364 control children identified by this method, 76 parents were interviewed and 288 parents were not interviewed due to their refusal to participate or inability to locate or contact them (7). In a study of risk factors for rhabdomyosarcoma in childhood conducted in North Carolina, 99 birth certificate controls were initially selected, 70 were found and interviewed, 9 were located and refused to participate, and 20 were untraceable (8). When a substantial proportion of potential controls are not included in a study due to inability to trace them, the resulting biases may be similar to those associated with nonresponse due to refusals.

In conclusion, the method of random digit dialing used in this study is feasible to conduct, does not require listings from the telephone company, and is comparable in efficiency to other telephone sampling methods such as the Waksberg technique (3). This method represents a viable approach to selecting controls in the pediatric age range. The primary challenge is to design and test strategies for improving response rates and evaluating the respresentativeness of controls who choose to participate.

References

1. Groves RM, Kahn RL. Surveys by telephone: a national comparison with personal interviews. New York: Academic Press, 1979.
2. Lilienfeld AM. Foundations of epidemiology. New York: Oxford University Press, 1976.
3. Waksberg J. Sampling methods for random digit dialing. J Am Statist Assoc 1978;73:40-6.

4. Sudman S. The uses of telephone directories for survey sampling. J Marketing Res 1973; X:204-7.
5. Harlow BL, Hartge P. Telephone household screening and interviewing. Am J Epidemiol 1983,117:632-3.
6. Massey JT, Barker PR, Hsiung S. An investigation of response in a telephone survey. Proceedings of the Section on Survey Research Methods. American Statistical Association 1981:426-31.
7. Gold E, Gordis L, Tonascia J, et al. Risk factors for brain tumors in children. Am J Epidemiol 1979;109:309-19.
8. Grufferman S, Wang HH, Delong ER, et al. Environmental factors in the etiology of rhabdomyosarcoma in childhood. JNCI 1982;68: 107-13.
9. Robison LL, Daigle A. Control selection using random digit dialing for cases of childhood cancer. Am J Epidemiol 1984;120:164-6.