American Journal of Epidemiology

Vol. 185, No. 3
DOI: 10.1093/aje/kww108
Advance Access publication:
January 7, 2017

Practice of Epidemiology

Tracing a Path to the Past: Exploring the Use of Commercial Credit Reporting Data to Construct Residential Histories for Epidemiologic Studies of Environmental Exposures

Susan Hurley*, Andrew Hertz, David O. Nelson, Michael Layefsky, Julie Von Behren, Leslie Bernstein, Dennis Deapen, and Peggy Reynolds

* Correspondence to Susan Hurley, Cancer Prevention Institute of California, 2001 Center Street, Suite 700, Berkeley, CA 94704 (e-mail: susan.hurley@cpic.org).

Initially submitted July 1, 2015; accepted for publication February 29, 2016.

Large-scale environmental epidemiologic studies often rely on exposure estimates based on linkage to residential addresses. This approach, however, is limited by the lack of residential histories typically available for study participants. Our objective was to evaluate the feasibility of using address data from LexisNexis (a division of RELX, Inc., Dayton, Ohio), a commercially available credit reporting company, to construct residential histories for participants in the California Teachers Study (CTS), a prospective cohort study initiated in 1995–1996 to study breast cancer (n = 133,479). We evaluated the degree to which LexisNexis could provide retrospective addresses prior to study enrollment, as well as the concordance with existing prospective CTS addresses ascertained at the time of the completion of 4 self-administered questionnaires. For approximately 80% of CTS participants, LexisNexis provided at least 1 retrospective address, including nearly 25,000 addresses completely encompassed by time periods prior to enrollment. This approach more than doubled the proportion of the study population for whom we had an address of residence during the childbearing years—an important window of susceptibility for breast cancer risk. While overall concordance between the prospective addresses contained in these 2 data sources was good (85%), it was diminished among black women and women under the age of 40 years.

data collection; environmental epidemiology; residential history; residential mobility; validation studies

Abbreviations: CTS, California Teachers Study; GIS, geographic information systems.

Recent biomonitoring data have made it clear that humans are exposed to a wide spectrum of environmental contaminants found to have known toxic and carcinogenic effects in animals, raising concerns about the potential human health consequences of such exposures (1–4). Efforts to evaluate the degree to which these exposures pose similar health risks in humans have been stymied by obstacles in exposure ascertainment methods available for epidemiologic studies. Many exposures of concern (e.g., flame retardants, hazardous air pollutants, pesticides) are invisible and are difficult, if not impossible, to self-report. While personal exposure monitors and biomarkers have been developed for some exposures, such methods have been limited by their inability to capture historical exposures, and they are often prohibitively expensive, precluding their use in large-scale epidemiologic studies. Geographic information systems (GIS) enable linkage of existing current and retrospective environmental quality and emissions data to residential locations. While GIS approaches to exposure assessment are well-suited to large cohort studies, the lack of residential history information within these studies has typically limited analyses in such studies to the evaluation of those exposures associated with residence at the time of diagnosis or the time of study entry, limiting their usefulness in studying health outcomes with long latency periods such as cancer (5–9). This may be especially important for studies of breast cancer, given mounting evidence that critical windows of susceptibility occur during specific time periods in life when the breast is especially vulnerable to environmental insults (10–13).

Our purpose in the present study was to evaluate the feasibility of using address data from a commercially available credit reporting company to construct residential histories for a large cohort of California women participating in an ongoing prospective study of breast cancer.

METHODS

Study population

This analysis builds on data collected as part of the California Teachers Study (CTS), a large, ongoing prospective cohort study of female California professional public school employees that was initiated in 1995–1996 specifically to study breast cancer (14). The CTS was created by inviting all recently or currently active and retired female members of the California State Teachers Retirement System to complete a baseline questionnaire at the time of enrollment and to participate in ongoing follow-up activities. The 133,479 women who participate represent a broad age range (22–104 years at enrollment; median age, 53 years), with a wide range of lifestyle experiences and socioeconomic levels, and are geographically dispersed throughout the state (14, 15). Although the CTS cohort is about 85% non-Hispanic white, the cohort contains substantial numbers of Hispanic, African-American, Asian/Pacific Islander, and mixed-race women. The collection and analysis of data from CTS participants and linkage with data resources have been approved by the institutional review board at each institution participating in the CTS and by the California Committee for the Protection of Human Subjects (California Health and Human Services Agency).

Existing address data from routine CTS follow-up

Address information is prospectively collected at various times for all members of the CTS cohort as part of routine CTS follow-up activities, which include updating information on name and mailing address for the purposes of future contact and for determination of continuing California residency. A number of sources are used for address updates, including cohort member feedback to questionnaires and newsletters, linkage to US Postal Service change-of-address forms, and information obtained from major credit reporting agencies such as Experian (Experian Technologies USA, Inc., Addison, Texas). Because these data are maintained primarily for the purpose of contacting study participants, the addresses do not necessarily represent the actual places where participants reside but rather constitute their mailing addresses, which could include nonresidential locations such as workplace addresses and post office boxes. Thus, while the majority of address updates in this file are likely to represent residential moves, they may also include changes in mailing addresses that do not represent changes in residential location. Furthermore, while each address has an associated "valid date," this date does not necessarily capture an accurate "move-in" date; instead it represents the first known date associated with that particular address, reflecting a variety of potential dates, depending on the source of the update (the date on which a participant updated her US Postal Service change-of-address form, the date on which a participant called the CTS study center to provide notification of a change, etc.). While these uncertainties likely introduce some degree of error with regard to ascertainment of longitudinal residential location, such methods of follow-up are standard procedures commonly used in the conduct of large-scale epidemiologic cohort studies (14) and are often the only feasible means available for ascertaining the residential locations of cohort members.

In June 2011, we obtained a file from the central CTS data center that contained all known addresses prospectively identified among CTS participants since their enrollment. After removal of approximately 29,000 duplicate and post-mortem addresses from this file, 245,545 unique addresses for the 133,479 CTS cohort participants remained (Table 1). Slightly more than half of the cohort remained at the same address throughout follow-up. The number of addresses per individual ranged from 1 to 13, with approximately 90% of participants having 3 or fewer unique addresses. Only 16% of the CTS participants were younger than age 40 years at enrollment.

Credit reporting address data (LexisNexis)

Address information for members of the CTS cohort was purchased from LexisNexis (a division of RELX, Inc., Dayton, Ohio), a commercial credit reporting company that, among its services, provides all known addresses for a requested set of individuals. Although the algorithm used by LexisNexis is proprietary, it considers myriad data sources to compile addresses, including: real estate/tax assessor records (current and archived); deed transfers and mortgage records; motor vehicle, boat, and aircraft registrations; driver's license records; court filings, including bankruptcy and Uniform Commercial Code judgment records and federal and state tax liens, jury verdicts, settlements, and arbitrations; professional licenses; voter registrations; Social Security Administration death records from all 50 states; marriage and divorce records from selected states; criminal history records and inmate indexes; business records, including information on incorporation and limited partnership and limited liability companies, fictitious business names, and DBA ("Doing Business As") registrations; and the Office of Foreign Assets Control master list of suspected terrorists. Many of these data sources are likely to be utilized by other credit reporting companies as well, but due to the proprietary nature of these businesses, it is difficult to ascertain the degree of overlap between data sources used by LexisNexis and Experian (which the CTS utilizes as one part of its routine participant tracking efforts).

In August 2013, we provided LexisNexis with a data file containing personal identifiers and last known addresses (as of June 2011) for the 133,479 CTS cohort members. LexisNexis returned a data set containing 358,520 addresses linked to 130,921 CTS participants, along with geographic coordinates (latitude/longitude), the earliest and most recent dates associated with these addresses, and match probability scores which showed how well the names, dates of birth, and Social Security numbers matched their corresponding CTS data record.

Table 1. Existing Prospectively Collected Address Information Available From Routine Follow-up Activities (n = 245,545 Unique Address Records) Carried Out From Enrollment (1995–1996) Through June 1, 2011, for Participants in the California Teachers Study (n = 133,479)

Characteristic	CTS Participants (n = 133,479)	
	No.	%
Total no. of unique addresses per individual		
1	70,187	53
2	35,146	26
3	16,192	12
4	6,899	5
5	2,890	2
6	1,298	1
7	518	<1
8	204	<1
>8	145	<1
Age at enrollment, years[a]		
<20	0	0
20–29	5,548	4
30–39	16,535	12
40–49	33,384	25
50–59	31,845	24
60–69	23,064	17
70–79	15,984	12
≥80	7,119	5

Abbreviation: CTS, California Teachers Study.
[a] Youngest age for which an address was available.

Data processing and cleaning

Both LexisNexis and CTS existing addresses were standardized using US Postal Service Coding Accuracy Support System address correction software (ZP4; Semaphore Corporation, Aptos, California). This software uses US Postal Service databases (Delivery Point Validation and Residential Delivery Index) to identify valid residential street addresses. The Delivery Validation flag was used to eliminate known invalid addresses and unrecognizable addresses. Residential addresses are identified by the Residential Delivery Index flag. A street address location is identified as not a commercial mail-receiving agency or a post office box, rural route, highway contract, or general delivery address. After standardization of the data, both the LexisNexis and CTS existing address data underwent a number of data processing steps to remove duplicates and to allow for appropriate comparisons between the 2 data sets.

For each individual, 2 residential history timelines, one based on CTS existing addresses and one based on LexisNexis data, were created as follows: The date associated with each address (valid date for CTS existing addresses and earliest known date for LexisNexis addresses) was used as the "move-in" date. Because both LexisNexis and the CTS address files were compiled from data from multiple sources, duplicate addresses needed to be removed. To do this, we compared each address to all addresses with the same or earlier "move-in" dates. Addresses that matched were flagged as potential duplicates. Addresses not flagged as duplicates were considered unique. Identical addresses separated by a different address (i.e., a return to a previous address) were also considered unique. Each address was assigned a "move-out" date of the day before the next sequential date; the last address was assigned an artificial move-out date corresponding to the end of follow-up (either June 1, 2011, which was the date on which the CTS address file was created, or, for those who died prior to June 1, 2011, the date of their deaths). Although it was not common, some individuals had multiple unique addresses with the same move-in date; we did not attempt to ascertain which of these addresses was "best" but rather kept them all in the data set. Nonresidential addresses were identified and flagged using the Coding Accuracy Support System ZP4 software as described above.

Analysis

After a preliminary assessment of the full set of addresses received from LexisNexis, we conducted all analyses after excluding nonresidential addresses, duplicate addresses, and addresses for dates falling after the end of CTS follow-up (June 1, 2011) or after the date of death. The extent and scope of the LexisNexis address data were then characterized according to the numbers of participants for whom: no address was provided; a prebaseline address was provided; and an address from an age younger than 40 years was provided. Frequency distributions were generated for all study participants and were stratified by race/ethnicity and age at baseline. To ascertain the statistical significance of differences in distributions by race/ethnicity and age group, we computed Pearson χ^2 statistics and corresponding P values.

Two separate sets of analyses were then performed. A retrospective analysis was conducted to characterize the degree to which LexisNexis could provide information on addresses held prior to cohort enrollment. This retrospective analysis focused on summarizing the number of addresses identified by LexisNexis prior to the enrollment date for each participant and describing the time periods and ages captured by those addresses. The goal of the second set of analyses (our "prospective analyses") was to assess the accuracy of the LexisNexis linkage by comparing the addresses with the existing addresses maintained by the CTS for several points in time during the period after cohort enrollment. We then calculated the concordance rates of the addresses between the CTS existing data and the LexisNexis data. Concordance rates were calculated for the address records in each file for several points in time corresponding to the date on which each CTS questionnaire was completed (baseline (1995–1996); questionnaire 2 (1997); questionnaire 3 (2000); and questionnaire 4 (2005–2006). These records were chosen because we felt the dates and addresses in the CTS

existing address database for these records were the most reliable, as they were generated from (or confirmed by) direct contact with CTS participants and therefore represented the closest thing to a "gold standard" for accurate representation of residential location at a specific point in time. In comparing the address records for these dates, addresses were considered a "match" only if they matched exactly on street number, street name, city, and 5-digit zip code. Concordance rates were calculated as: (number of CTS participants for whom the LexisNexis address exactly matched the CTS existing address for the date on which the questionnaire was completed/number of CTS participants for whom LexisNexis provided an address for the date of questionnaire completion) × 100.

RESULTS

LexisNexis returned a file containing 358,520 address records. An initial assessment of this data set revealed a number of problems with or limitations of the addresses provided (Table 2). In addition to containing 15,694 duplicate addresses for 12,274 participants and 42,577 nonresidential addresses for 30,422 participants, most of the problems were related to limitations regarding the dates associated with addresses, including missing dates and dates that fell after the date of death or before the date of birth. Furthermore, for 7,307 participants, multiple unique addresses were provided for the same point in time. For 2,558 participants, LexisNexis was not able to provide any address.

Table 3 shows the scope and extent of addresses provided by LexisNexis, overall and for specific categories of race/ethnicity and age at baseline. Overall, LexisNexis provided at least 1 address for 98% of CTS participants. For 80% of CTS participants, LexisNexis provided at least 1 address from a time period prior to enrollment in the cohort, and for 42% it provided at least 1 address for an age less than 40 years. The extent and scope of addresses provided by LexisNexis varied significantly by race/ethnicity ($P < 0.001$). Most notable were the smaller proportions of Native Americans for whom LexisNexis was able to provide any address, an address prior to baseline, or an address at an age less than 40 years, compared with other racial/ethnic groups. Additionally, the proportion of participants for whom LexisNexis provided an address for an age less than 40 years was highest among Hispanics and Asian/Pacific Islanders. Some differences in the extent and scope of address data were also noted across categories of age ($P < 0.001$). In general, the success of the LexisNexis linkage seemed to decline with age, particularly among the very old (i.e., those aged ≥70 years). Also notable was a markedly smaller proportion of participants for whom a prebaseline address was provided among women who were aged 20–39 years at baseline.

Table 2. Problems and Limitations of 358,520 Addresses Provided by LexisNexis[a] for 133,479 California Teachers Study Participants Enrolled in 1995–1996

Problem/Limitation[b]	No. of Addresses (n = 358,520)	No. of Participants (n = 133,479)
No address provided		2,558
Duplicate address	15,694	12,274
Nonresidential address	42,577	30,422
Address date fell after date of death	9,068	742
Address date fell prior to date of birth	1,245	1,245
Multiple unique addresses for the same time period	14,991	7,307
Address was missing date	196	16

Abbreviation: CTS, California Teachers Study.
[a] LexisNexis is a division of RELX, Inc., Dayton, Ohio.
[b] Categories are not mutually exclusive.

Retrospective address analysis

Focusing specifically on the degree to which LexisNexis could provide address data prior to the CTS baseline (for which there are no data in the existing CTS address file), we found that LexisNexis was able to provide at least 1 retrospective address (i.e., a "move-in" date prior to enrollment) for nearly 80% of CTS participants. Of the 123,828 retrospective addresses, 24,599 entirely encompassed time frames prior to enrollment (i.e., the "move-out" date was before enrollment), with the remaining 99,229 addresses encompassing some portion of time both prior to and after enrollment (a "move-in" date prior to enrollment and a "move-out" date after enrollment).

Table 4 shows the extent and temporal coverage of these retrospective addresses. While the total number of unique retrospective addresses per individual ranged from 0 to 14, most study participants had only 1 or 2 unique addresses prior to enrollment. For 19% of participants, the earliest address date was during the 1990s; for 23%, the earliest address date was between 1985 and 1989; for 17%, the earliest address date was between 1980 and 1984; for 13%, the earliest address date was during the 1970s; and for approximately 5%, the earliest address date was prior to the 1970s. Also presented in Table 4 is the youngest age captured by the LexisNexis retrospective addresses. Compared with the youngest ages captured by the CTS enrollment addresses (Table 1), LexisNexis was able to provide additional addresses for residences occupied at younger ages. Of particular relevance to breast cancer studies is the identification of addresses used during the childbearing years (<40 years of age): This was provided for approximately 37% of study participants by LexisNexis, whereas only 16% of the CTS cohort was younger than age 40 years at enrollment. For approximately 1% of study participants, the age associated with their earliest address was found to be invalid (i.e., the date associated with the address preceded the date of birth). A closer examination of these records showed that nearly all (99%) had the same, obviously artificially assigned, start date of November 1, 1916. Finally, duration of residence prior to the enrollment date ranged from 0 years to 80 years, with slightly fewer than half of participants having a duration of 10 years or less.

Table 3. Extent and Scope of Address Data Provided by LexisNexis[a] for 133,479 California Teachers Study Participants Enrolled in 1995–1996, by Race/Ethnicity and Age at Baseline[b]

Characteristic	No. of Participants	Participants for Whom Unique Addresses Were Provided by LexisNexis[c]					
		No Address		Prebaseline Address		Address at Age <40 Years	
		No. of Persons	%	No. of Persons	%	No. of Persons	%
All participants	133,479	2,559	2	107,314	80	56,286	42
Race/ethnicity[d]							
White	115,871	2,250	2	92,971	80	46,730	40
Black	3,553	37	1	3,050	86	1,641	46
Hispanic	5,409	37	1	4,344	80	3,551	66
Native American	1,302	110	8	933	72	333	26
Asian/Pacific Islander	4,495	34	1	3,815	85	2,657	59
Age at baseline, years							
20–39	22,083	78	<1	14,318	65	20,626	93
40–49	33,384	153	<1	28,138	84	21,337	64
50–59	31,845	218	<1	27,641	87	10,873	34
60–69	23,064	267	1	19,951	87	2,585	11
70–79	15,984	576	4	12,905	81	506	3
≥80	7,119	1,267	18	4,361	61	359	5

Abbreviation: CTS, California Teachers Study.

[a] LexisNexis is a division of RELX, Inc., Dayton, Ohio.

[b] The distribution of address data varied significantly by race/ethnicity and age at baseline (Pearson χ^2 test: $P < 0.001$).

[c] Data were restricted to nonduplicate residential addresses and excluded addresses with dates that fell after the end of CTS follow-up (June 11, 2011) or after the date of death.

[d] Data for participants with unknown/missing/other information on race/ethnicity ($n = 2,849$) are not shown.

Prospective address analysis

Table 5 shows results from our evaluation of the concordance of LexisNexis addresses with the CTS existing addresses for the 4 snapshots in time corresponding to the fill dates for the first 4 CTS questionnaires. With the exception of the questionnaire 4 address, overall address concordance was quite good, with approximately 85% of the addresses matching at the time of completion of the first 3 questionnaires. For questionnaire 4 (completed in 2005–2006), the concordance was lower (74%). Concordance rates significantly differed ($P < 0.001$) across categories of race/ethnicity, with blacks having notably lower concordance, especially as time since baseline increased. Significant differences were also noted in the address concordance rates across age groups, with rates generally improving with increasing age.

DISCUSSION

The impetus for the present analysis grew from our own need to characterize historical environmental exposures for a study of breast cancer in this cohort of women. Our analysis was inspired by the prior work of Jacquez et al. (16), who undertook a similar approach among participants in a case-control study of bladder cancer in Michigan. Utilizing a more limited approach (compared with ours) in which they purchased information on only the 3 most recent addresses from LexisNexis, Jacquez et al. compared addresses provided by LexisNexis with self-reported lifetime residential histories for approximately 950 study participants living in 11 Michigan counties. Reporting that the LexisNexis addresses accounted for 71% of the lifetime residential histories provided by study participants, the authors concluded that this method held great promise as a cost-effective means of ascertaining residential history for use in environmental epidemiologic studies (16). However, the authors noted a need to evaluate whether their findings would be valid in other, more diverse and mobile study populations. Our findings for the large and geographically diverse CTS cohort generally support those reported by Jacquez et al.

Our results suggest that commercially available credit reporting data may be useful for augmenting existing address information and constructing residential histories for large-scale GIS-based epidemiologic studies of environmental exposures. The linkage service provided by LexisNexis yielded nearly 25,000 addresses completely encompassed by time periods prior to enrollment for which no routinely collected CTS address data currently exist. Furthermore, this approach more than doubled the proportion of the study population for whom we have an address where they resided during their childbearing years. While it is not possible to ascertain the accuracy of the retrospective addresses obtained through LexisNexis with currently available data, our prospective analysis, which demonstrated

Table 4. Results From Retrospective Analysis of Residential Addresses Provided by LexisNexis[a] With Start Dates Prior to the Date of Enrollment (1995–1996) for all California Teachers Study Participants (n = 123,828 Addresses for 133,479 Participants)

Characteristic	No. of CTS Participants	% of CTS Participants
Total no. of unique retrospective addresses per person		
0	29,853	22
1	86,032	64
2	15,400	12
3	1,876	1
4	258	<1
5	41	<1
6–14	19	<1
Earliest calendar year for which an address was available		
1990 or later	26,182	19
1985–1989	30,578	23
1980–1984	22,516	17
1975–1979	11,318	8
1970–1974	6,625	5
1965–1969	2,832	2
Before 1965	3,575	3
No address provided	29,853	22
Youngest age for which a full address was available, years		
≤19	1,527	1
20–29	15,484	12
30–39	31,499	24
40–49	24,052	18
50–59	15,085	11
60–69	10,074	8
70–79	3,911	3
≥80	760	1
Invalid (<0 years)	1,234	1
No address provided	29,853	22
Preenrollment time at baseline address, years[b]		
≤1	4,385	3
>1–5	23,929	18
>5–10	30,422	23
>10–15	19,976	15
>15–20	9,653	7
>20–25	6,089	5
>25–30	2,663	2
>30–40	1,748	1
>40	1,727	1
Missing data[c]	3,034	1
No address provided	29,853	22

Abbreviation: CTS, California Teachers Study.

[a] LexisNexis is a division of RELX, Inc., Dayton, Ohio.

[b] Duration of residence at the participant's baseline address prior to study enrollment.

[c] Preenrollment time at baseline address could not be calculated for 3,034 participants because their CTS enrollment address was not considered a residential address.

Table 5. Results From Prospective Analysis of the Accuracy of LexisNexis[a] Address Data for Participants in the California Teachers Study (CTS), as Captured by the Concordance of LexisNexis Addresses With CTS Addresses Among CTS Participants for Whom LexisNexis was Able to Provide an Address, 1995–2011[b,c]

CTS Study Participants	Address Concordance											
	At Baseline (1995–1996)			At Questionnaire 2 (1997)			At Questionnaire 3 (2000)			At Questionnaire 4 (2005–2006)		
	No. of Matches[d]	Total No.[e]	%[f]	No. of Matches	Total No.	%	No. of Matches	Total No.	%	No. of Matches	Total No.	%
All participants	83,421	97,305	86	66,471	78,114	85	63,460	75,539	84	46,033	62,490	74
Race/ethnicity[g]												
White	72,488	84,031	86	58,407	68,218	86	56,032	66,299	85	41,077	55,210	74
Black	2,248	2,806	80	1,556	1,994	78	1,357	1,766	77	814	1,316	62
Hispanic	3,310	4,071	81	2,380	2,957	80	2,274	2,848	80	1,497	2,247	67
Native American	686	800	86	481	568	85	443	539	82	258	374	69
Asian/Pacific Islander	3,026	3,606	84	2,430	2,903	84	2,214	2,695	82	1,634	2,255	72
Age at baseline, years												
20–39	10,166	13,222	77	8,750	11,113	79	9,418	11,541	82	6,842	9,161	75
40–49	22,192	25,589	87	17,012	19,789	86	16,174	19,147	84	11,475	16,586	69
50–59	21,351	24,989	85	16,472	19,577	84	15,539	18,933	82	12,326	17,315	71
60–69	15,665	17,931	87	12,889	14,929	86	12,367	14,519	85	9,727	12,482	78
70–79	10,503	11,674	90	8,735	9,820	89	7,935	9,072	87	4,990	6,118	82
≥80	3,544	3,900	91	2,613	2,886	91	2,027	2,327	87	673	828	81

Abbreviation: CTS, California Teachers Study.

[a] LexisNexis is a division of RELX, Inc., Dayton, Ohio.

[b] Data were restricted to nonduplicate residential addresses and excluded addresses with dates that fell after the end of CTS follow-up (June 11, 2011) or after the date of death.

[c] The distribution of address concordance varied significantly by race/ethnicity and age at baseline (Pearson χ^2 test: $P < 0.001$).

[d] Number of CTS participants for whom the LexisNexis address exactly matched the CTS address for the date on which the questionnaire was completed.

[e] Number of CTS participants for whom LexisNexis provided an address for the date on which the questionnaire was completed; it varied, because not all participants completed all questionnaires.

[f] (No. of matches/total no.) × 100.

[g] Data for participants with unknown/missing/other information on race/ethnicity (n = 1,991) are not shown.

very good concordance with existing CTS data (approximately 85% for several points in time), provides reassuring evidence of the ability of LexisNexis data to accurately ascertain residential location. However, our findings also suggest a number of caveats to the use of such data.

Our initial assessment of the full set of addresses provided by LexisNexis (Table 2) underscores the importance of carefully evaluating and processing such data prior to their use, especially to ensure the removal of duplicate and nonresidential addresses. Additionally, care should be taken to identify and consider appropriate strategies for handling multiple unique addresses for the same time period, which we identified among several thousand women in our study. We speculate that such instances may represent individuals who have second (vacation) homes or own rental properties, or for whom some nonresidential addresses were not detected by our standardization process and may therefore represent workplace addresses. It may also reflect problems with the dates provided by LexisNexis. Clearly errors in dates exist, as demonstrated by the nearly 1,200 study participants for whom addresses were provided with a date prior to the participant's date of birth, which appeared to be the result of an artificial date assigned by LexisNexis. While it is somewhat reassuring that this constitutes a very small percentage of our study population (<1%), it is likely that additional, less obvious undetected errors exist, the scope of which is impossible to ascertain with the available data. Unaccustomed to data requests for scientific research, LexisNexis provides minimal data documentation, making it difficult to fully ascertain the meaning and source of the dates provided. Jacquez et al. also noted substantial temporal mismatches in their comparison of the LexisNexis data with the self-reported residential histories in their study (16). However, the fact that for approximately 85% of our study participants, the prospective address provided by LexisNexis exactly matched the existing CTS address for 3 of the 4 specific dates evaluated suggests that overall inaccuracies in dates may not be too problematic and that this approach can capture a residential location for a fairly tight time window and precise geographic location for a majority of study subjects.

Because we did not have a full residential history provided by self-report, we could not assess the degree to which LexisNexis linkage could be used to reconstruct lifetime residential histories as did Jacquez et al. in their study (16). While we attempted to construct residential histories during the prospective follow-up using all of the existing CTS address data (i.e., not just limited to addresses at questionnaire completion), we found that given the limitations in the dates available from the CTS (as discussed in the Methods section above), too many speculative assumptions were required, precluding our ability to draw any sound conclusions from comparisons with the LexisNexis data (data not shown).

It is important to note that the effectiveness of the LexisNexis linkage in identifying residential addresses was not entirely uniform for all study participants. In particular, our results suggest that while this approach appeared to be quite successful for most racial/ethnic groups, it may be limited for characterizing the residential locations of Native Americans. Additionally, our findings show that the effectiveness of this approach may be limited for the very young and the very old. In general, the effectiveness of this linkage strategy was markedly diminished among women who were elderly at CTS enrollment, providing a smaller proportion of any address, prebaseline address, or address held during the childbearing years (at an age <40 years). The inability of LexisNexis to provide an address during the childbearing years among older women at baseline is probably explained by the fact that electronic databases that capture the eras during which these older women would have been in this age category do not exist. It remains unclear, however, why the LexisNexis linkage would be less successful in providing any address or an address prior to baseline for this oldest age group of women. The smaller proportion of younger women for which LexisNexis was able to provide a prebaseline address is likely due to the fact that these women have less of a credit history available for the LexisNexis linkages.

The results from our prospective address concordance analysis suggest that overall the accuracy of this linkage strategy for identification of residential location is quite good but may be limited for blacks and younger women. Reasons for lower address concordance among these groups are unclear. While younger women in the CTS constitute the most residentially mobile group, black women are the least residentially mobile. Caution is therefore recommended in applying these methods to such groups of women. Finally, it should be noted that while our study population was larger and more diverse than that of Jacquez et al. (16), it was comprised entirely of professional women. The degree to which these findings are applicable to populations that include men and encompass a broader spectrum of socioeconomic positions is not known.

Despite these limitations, our results support the conclusion of Jacquez et al. (16) and suggest that these methods may provide a feasible and cost-effective strategy for constructing (or augmenting) residential histories for large-scale epidemiologic cohort studies of environmental exposure, offering a reasonable alternative to the expensive, time-consuming, and often infeasible method of collecting residential histories through self-report. We received address data from LexisNexis within a week of submitting our order, and the cost was minimal compared with what it would have cost to collect residential history information from study respondents; and with a large number of records, the cost per subject was substantially lower than that reported by Jacquez et al. (16). Substantial quality assurance/quality control efforts are recommended, however, to ensure removal of nonresidential addresses and duplicate addresses and to evaluate potential inconsistencies due to inaccuracies in dates provided. Finally, caution should be applied when using these methods in populations that include blacks and Native Americans.

ACKNOWLEDGMENTS

Author Affiliations: Cancer Prevention Institute of California, Berkeley, California (Susan Hurley, Andrew Hertz, David O. Nelson, Michael Layefsky, Julie Von Behren, Peggy Reynolds); Department of Population Sciences, Division of Cancer Etiology, Beckman Research Institute of the City of Hope, Duarte, California (Leslie Bernstein); Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California (Dennis Deapen); and Department of Health Research and Policy, Division of Epidemiology, School of Medicine, Stanford University, Stanford, California (Peggy Reynolds).

This research was supported by funding from the Regents of the University of California Breast Cancer Research Program (grant 16ZB-8501) and the National Cancer Institute, National Institutes of Health (grants R01CA170394 and R01CA77398).

We thank Minhthu Le for administrative support. We also thank the members of the California Teachers Study Steering Committee who were responsible for the formation and maintenance of the cohort within which this study was conducted: Drs. Hoda Anton-Culver, Jessica Clague, Christina A. Clarke, Pamela Horn-Ross, James V. Lacey, Jr., Yani Lu, Huiyan Ma, Susan L. Neuhausen, Hannah Park, Rich Pinder, Fredrick Schumacher, Sophia S. Wang, and Argyrios Ziogas.

The opinions, findings, and conclusions herein are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health or the Regents of the University of California or any of its programs.

Conflict of interest: none declared.

REFERENCES

1. American Lung Association. *State of the Air 2014*. Chicago, IL: American Lung Association; 2014. http://www.stateoftheair.org/2014/assets/ALA-SOTA-2014-Full.pdf. Accessed June 5, 2015.
2. Betts KS. Unwelcome guest: PBDEs in indoor dust. *Environ Health Perspect*. 2008;116(5):A202–A208.

3. Centers for Disease Control and Prevention. *Fourth National Report on Human Exposures to Environmental Chemicals*. Atlanta, GA: Centers for Disease Control and Prevention; 2009.
4. Rudel RA, Attfield KR, Schifano JN, et al. Chemicals causing mammary gland tumors in animals signal new directions for epidemiology, chemicals testing, and risk assessment for breast cancer prevention. *Cancer*. 2007; 109(12 suppl):2635–2666.
5. Bell ML, Belanger K. Review of research on residential mobility during pregnancy: consequences for assessment of prenatal environmental exposures. *J Expo Sci Environ Epidemiol*. 2012;22(5):429–438.
6. Boscoe FP. The use of residential history in environmental health studies. In: Maantay JA, McLafferty S, eds. *Geospatial Analysis of Environmental Health*. Dordrecht, the Netherlands: Springer Science+Business Media; 2011:93–110.
7. Jacquez GM, Kaufmann A, Meliker J, et al. Global, local and focused geographic clustering for case-control data with residential histories. *Environ Health*. 2005;4(1):4.
8. Nuckols J, Airola M, Colt J, et al. The impact of residential mobility on exposure assessment in cancer epidemiology. *Epidemiology*. 2009;20(6):S259–S260.
9. Oudin A, Forsberg B, Strömgren M, et al. Impact of residential mobility on exposure assessment in longitudinal air pollution studies: a sensitivity analysis within the ESCAPE project. *ScientificWorldJournal*. 2012;2012:125818.
10. Colditz GA, Bohlke K, Berkey CS. Breast cancer risk accumulation starts early: prevention must also. *Breast Cancer Res Treat*. 2014;145(3):567–579.
11. Shottenfeld D, Fraumeni, J. *Cancer Epidemiology and Prevention*. 3rd ed. New York, NY: Oxford University Press; 2006.
12. Wild CP. The exposome: from concept to utility. *Int J Epidemiol*. 2012;41(1):24–32.
13. Institute of Medicine of the National Academies. *Breast Cancer and the Environment: A Life Course Approach*. Washington, DC: National Academy of Sciences; 2014.
14. Bernstein L, Allen M, Anton-Culver H, et al. High breast cancer incidence rates among California teachers: results from the California Teachers Study (United States). *Cancer Causes Control*. 2002;13(7):625–635.
15. Reynolds P, Hurley S, Goldberg DE, et al. Regional variations in breast cancer among California teachers. *Epidemiology*. 2004;15(6):746–754.
16. Jacquez GM, Slotnick MJ, Meliker JR, et al. Accuracy of commercially available residential histories for epidemiologic studies. *Am J Epidemiol*. 2011;173(2):236–243.