HHS Public Access

Author manuscript

J Expo Sci Environ Epidemiol. Author manuscript; available in PMC 2024 December 06.

Published in final edited form as:

J Expo Sci Environ Epidemiol. 2024 November ; 34(6): 1028–1034. doi:10.1038/s41370-023-00631-0.

Construction of residential histories to estimate long-term environmental exposures in the California Teachers Study cohort

Danielle N. Medgyesi[1], **Emma S. Spielfogel**[2], **Mary H. Ward**[3], **Rena R. Jones**[3], **Kristen E. Savage**[2], **Jennifer L. Benbow**[2,4], **James V. Lacey Jr.**[2], **Tiffany R. Sanchez**[1]

[1]Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY

[2]Division of Health Analytics, Department of Computational and Quantitative Medicine, Beckman Research Institute, City of Hope, Duarte, CA

[3]Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD

[4]Center for Data-Driven Insights and Innovation, University of California Health, Oakland, CA

Abstract

Environmental epidemiologic studies using geospatial data often estimate exposure at a participant's residence upon enrollment, but mobility during the exposure period can lead to misclassification. We aimed to mitigate this issue by constructing residential histories for participants in the California Teachers Study through follow-up (1995–2018). Address records have been collected from the US Postal Service, LexisNexis, Experian, and California Cancer Registry. We identified records of the same address based on geo-coordinate distance (≤250m) and street name similarity. We consolidated addresses, prioritizing those confirmed by participants during follow-up questionnaires, and estimating the duration lived at each address using dates associated with records (e.g., date-first-seen). During 23-years of follow-up, about half of participants moved (48%, including 14% out-of-state). We observed greater mobility among younger women, Hispanic/Latino women, and those in metropolitan and lower socioeconomic status areas. The cumulative proportion of in-state movers remaining eligible for analysis was 21%, 32%, and 41% at 5-, 10-, and 20-years post-enrollment, respectively. Using self-reported information collected 10-years after enrollment, we correctly identified 94% of movers and 95% of non-movers as having moved or not moved from their enrollment address. This dataset provides

Corresponding author: Danielle Medgyesi, MS, Predoctoral Fellow, Columbia University Mailman School of Public Health, Department of Environmental Health Sciences, 722 West 168th St., New York, NY 10032, dm3688@cumc.columbia.edu.

AUTHOR CONTRIBUTIONS: JVL, JLB, KES, and ESS were responsible for participant engagement, which involved managing and gathering residential addresses of study participants. DNM, MHW, RRJ and TRS conceived of the project. All authors contributed to the study's approach. DNM completed the analysis and drafted the manuscript. All authors reviewed and contributed to the content of the manuscript.

COMPETING INTERESTS: The authors declare no competing interests.

ETHICAL APPROVAL: The California Teachers Study has been approved by the institutional review boards at City of Hope, the University of California San Diego, the University of California San Francisco, and the University of California at Irvine, and participants provided informed consent. This study was approved by the institutional review boards of City of Hope and Columbia University.

a foundation for estimating long-term environmental exposures in diverse epidemiologic studies in this cohort.

Keywords

Epidemiology; Geospatial Analyses; Health Studies; Analytic Methods

INTRODUCTION

Epidemiologic cohort studies using spatial environmental data, such as air pollution or water quality, typically estimate a participant's exposure level based on their reported residential address at enrollment. However, exposure misclassification may occur due to changes in environmental levels and participants' residential mobility over time, especially for long-term effects[1–3]. Exposure assessment can be improved by determining an individual's exposure and duration at each address during the exposure period prior to health outcome development.

A residential history is a chronological record of an individual's addresses, including the duration of residency at each location. Residential addresses of participants prior to study have not typically been collected in cohort studies and are subject to potential recall bias and incomplete data when self-reported. Follow-up contact often involves tracing changes of address through the postal service. Commercial databases, such as LexisNexis, have also proven useful in obtaining addresses for prospective cohorts[4–8]. A previous analysis in the California Teachers Study (CTS) cohort reported 85% agreement between LexisNexis-collected addresses and self-reported addresses at enrollment[4].

Constructing residential histories is a challenging task that involves consolidation (i.e., deduplication) of addresses from various sources, estimating move-in and move-out dates, and dealing with potential complications like temporary and family homes or incorrect data linkages. Our objective was to construct residential histories for CTS participants using addresses collected from multiple sources (e.g., US Postal Service, LexisNexis), often confirmed or corrected by participants throughout follow-up. Our goal was to produce a reliable residential dataset that can be used to estimate environmental exposures in CTS studies. This study describes residential mobility within the cohort and differences in mobility by participant sociodemographics and health status, along with the agreement between self-reported and constructed residential duration collected ten years into the study.

METHODS

Study population.

The CTS is a large prospective cohort of 133,477 women, ages 22–104 years (median=53) enrolled in 1995–1996, recruited from members of the California State Teachers Retirement System[9]. Women have been routinely followed for cancer and other health outcomes via linkages with statewide registries (e.g., California Cancer Registry; CCR) and hospital administrative records from the California Department of Health Care Access and Information. Analytic follow-up concludes at the earliest of: date of death, initial relocation

out of California (as outcomes ascertainment mainly relies on statewide data), or end of the current follow-up period (i.e., 12/31/2018). Ongoing follow-up is regularly extended as part of ongoing health linkages and participant engagement.

Participants enrolled by completing a mailed health and lifestyle questionnaire (Q1, 1995–1996). Five follow-up questionnaires were administered via mail, with the exception of the last questionnaire (Q6), which was conducted online for participants with an email address[10, 11]. The response rates among participants alive at the time are: 75% for Q2 (1997–1998), 74% for Q3 (2000–2002), 69% for Q4 (2005–2007), 61% for Q5 (2012–2015), and 43% for Q6 (2017–2019).

Address ascertainment.

To locate participants who moved during follow-up, tracing was primarily conducted through linkages with the U.S. Postal Service National Change of Address [USPS NCOA] and Experian®. These sources were supplemented by annual CCR linkages to identify participants diagnosed with cancer. In 2013, participant data were provided to LexisNexis to retrieve address records, covering the period before enrollment up until 2011 (when LexisNexis data were last available)[4]. Updated addresses were also self-reported during the completion of questionnaires or through other forms of communication, such as newsletters, postcards, phone calls, and the CTS website contact page[11].

This methodology resulted in a pool of addresses from multiple sources, each accompanied by a relevant date which held different meaning based on the source: mail forwarding start date for USPS NCOA records, the date ‘first seen’ from LexisNexis or Experian records, the cancer diagnosis date for CCR records, or the contact date for self-reported addresses. In the final questionnaire (Q6), participants were also asked the year they moved into their current residence.

At each follow-up, participants were asked to confirm or update the address on file. Address records confirmed or updated via CTS questionnaires or other forms of communication were labeled as ‘trusted’. The majority of ‘trusted’ records post-enrollment are from USPS NCOA (58%) or Experian (13%), or are self-reported (29%). Other USPS, NCOA, or Experian records unverified by participants, as well as LexisNexis and CCR records, not typically used to trace participants and therefore unverified, were labeled as ‘untrusted’.

Due to the extensive address search, duplicate addresses from various sources emerged. This underscored the need for address deduplication, prioritizing high-confidence addresses confirmed by participants, and estimating move-in and move-out dates for each unique address. As part of this process, when ‘untrusted’ records matched ‘trusted’ records, we relabeled the former as participant-verified addresses. Remaining unverified addresses were included if they came after a participant’s last contact and were corroborated by at least two sources. Further details are provided herein.

Address data geocoding.

Address records from all sources were geocoded by Texas A&M, USC Geocoding Platform (version 4.1 or 5.0 for Q6 addresses)[12]. There were 124,682 participants (93%) whose

residential address at enrollment was in California. We excluded 5,989 participants whose address at enrollment could not be geocoded to a ZIP code centroid or street address (e.g., PO Boxes, city and county centroids). Post-enrollment address records that were not geocoded to ZIP code centroids or streets addresses were also excluded (n=12,075). Next, we removed identical records that had the same latitude and longitude, address source, and date (n=14,613). After these preprocessing steps, our dataset included 521,447 address records among 118,693 participants.

Identification of unique addresses.

We created a unique identifier for matching address records using a combination of geospatial proximity and phonetic similarity of street names. First, we identified records with coordinates that were within a 250m radius (R package: leaderCluster), a threshold previously described to account for about 95% of positional error[13, 14]. We then identified records whose parsed street name was phonetically similar using the soundex algorithm (R package: stringdist). The algorithm returns a four-character code for each street name that indicates phonetic similarity despite minor spelling differences. We considered street names to match based on the first three characters, ignoring discrepancies in street suffixes. As an example, street names like “park” (phonetic ID: P620), “park lane” (P624) or “park rd” (P626) were matched based on their identical first three characters. Ultimately, our matching criteria incorporating both coordinate proximity and street name similarity resulted in the identification of 290,457 unique addresses. ‘Untrusted’ records that matched ‘trusted’ records were denoted as reflecting the same ‘participant-verified address’.

Construction of residential histories.

We constructed residential histories by consolidating unique residential addresses in California and estimating the respective move-in and move-out dates. The address histories included residences from enrollment until a participant’s end of follow-up (i.e., relocation out of California, death, or December 31st, 2018). Throughout the period up until the date of the last questionnaire response (Q1–Q6), we prioritized and included only participant-verified addresses. After the last response date (i.e., participants stopped responding to CTS questionnaires), we broadened our criteria to include unverified addresses if corroborated by at least two independent sources. This strategy was employed to minimize potential errors arising from incorrect linkages or addresses associated with family residences or secondary homes.

The move-in date for each unique address was estimated as the minimum date among matched records. If participants reported a move-in year for their address at Q6, the move-in date for that address was set to July 1st (mid-year) of that year (n=24,133). The move-in date for the enrollment address was not estimated for 36,712 participants (31%) due to the lack of information predating enrollment, either from LexisNexis[4] or self-reported at Q6 for non-movers. We corrected the move-in date to the birth date for 1,050 participants with move-in dates preceding their birth, assuming these participants never moved in their lifetime before enrollment. The move-out date of each address was set equal to the move-in date of the subsequent unique address. Our constructed residential history dataset included 178,581 addresses for 118,693 participants.

Describing mobility.

We classified women into three mobility categories: non-movers (those with no other addresses besides the enrollment address), within-state movers (those with one or more California addresses post-enrollment but no out-of-state address), and out-of-state movers (those whose follow-up ended due to relocating out of California). We then examined the proportion of women within each of these mobility categories across baseline sociodemographic characteristics and health status. We also assessed mobility by participation status in at least one questionnaire since 2005 (Q4–Q6), about a decade after enrollment, to evaluate potential differences that may arise from challenges in contacting movers. We described the number of years spent at the enrollment address, both prior to and after the start of the study, and total duration. To illustrate mobility trends over time, we computed the annual cumulative proportions of non-movers, once-movers, and twice or more movers within California who were still eligible for follow-up, along with the proportion of out-of-state movers or deceased participants.

Validation.

In Q4 (2005–2007), participants were asked to report the number of years they lived at their current residence, with the options of: <1, 1–4, 5–9, 10–14, 15–19, 20–29 and 30+ years[11]. Because Q4 was about 10 years after enrollment, we assumed women who reported living ≥10 years at their current address were non-movers since enrollment (67%), while those that reported <10 years were classified as movers (34%). We used this information as the gold standard in comparison with the active address at Q4 (i.e., move-in and move-out date overlapped with questionnaire fill date); Figure S1. We assessed whether self-reported non-movers and movers were correctly identified as still residing or not residing at their enrollment address, respectively. We also assessed agreement between self-reported and estimated Q4 address duration, calculated as the year of Q4 completion minus the estimated move-in year for that address. All analyses were conducted in the CTS Researcher Platform, R version 4.0.2[15].

RESULTS

From enrollment to 12/31/2018, 53% of CTS participants never moved from their enrollment address, 34% moved at least once within California, and 14% moved out-of-state (Table 1). The youngest women at enrollment (22–34 years) had the highest in-state mobility (63%) which generally decreased with age. The median age women first moved was 55 years (IQR=43–69 years). Hispanic/Latino women showed the highest in-state mobility (43%), while Black/African-American women exhibited the lowest (27%). Women residing in the highest socioeconomic status (SES) neighborhoods and in suburban areas demonstrated lower mobility compared to those living in the lowest SES neighborhoods and in metropolitan areas. Women reporting health conditions such as diabetes and high blood pressure were less mobile than those without these conditions. Mobility was similar by participation status since 2005.

Among women with an estimated move-in date for their enrollment address (n=81,981), 50% lived at their enrollment address for ≥10 years prior to the start of the study (Table 2).

After the start of the study, 56% of participants lived at their enrollment address for ≥10 years. Combined, 71% of women spent a total of ≥10 years at their enrollment address.

Figure 1 displays mobility patterns in the years after enrollment. Five years after enrollment, 79% of eligible women (those alive living in California) remained at their enrollment address. This proportion dropped to 68% ten-years later, and 59% twenty-years later. Most movers changed addresses only once after enrollment, while a small percentage (6–11%) moved multiple times over the course of follow-up.

Using responses from Q4 as the gold standard, we found that 94% of movers and 96% of non-movers were correctly classified based on whether the address at Q4 was the same as their enrollment address (Table 3). Among non-movers reporting a 10–14 year residency, agreement by mobility status was lowest (88%), which may be explained by misclassification due to participants moving shortly after enrollment. We assessed agreement between self-reported and estimated duration at the Q4 address among those correctly classified by mobility status. Agreement between self-reported and estimated durations was 86% for movers reporting <4 years and 92% for 5–9 years. Duration was generally underestimated for non-movers with a long residency, most notably for those reporting the longest duration of ≥30 years, with an agreement of 58% (median=31, IQR=22–36 years).

DISCUSSION

In a prospective cohort of women living in California, we constructed residential histories during follow-up, using multiple address sources and prioritizing those confirmed by participants during follow-up contact. Our approach to constructing residential histories demonstrated high accuracy (94–96%) compared to self-reported mobility status ten years after enrollment. We found that nearly half of the participants moved from their enrollment address, and that residential mobility during follow-up was higher among younger, Hispanic/Latino women and those from lower SES or metropolitan areas. Finally, we described how mobility rates cumulate over follow-up, highlighting the potential for misclassification in long-term cohort studies without regular address updates.

Our study found that 47% of participants relocated over 20+ years, which is higher than the 32% mobility of older women in rural Iowa (19-year follow-up)[2]. Among women with young children in urban Cincinnati with a 7-year follow-up, 54% had moved[1]. Consistent with patterns described previously[1, 16], we observed higher mobility among younger women and those residing in lower SES neighborhoods and metropolitan areas. Duration at the enrollment address before the start of the study was somewhat lower for our cohort (50% living ≥10 years) compared to a cohort of older participants in Los Angeles (50% living ≥13 years)[5]; both studies used LexisNexis to trace addresses historically. Greater mobility among Hispanic/Latino women and those from lower SES backgrounds could lead to differential exposure misclassification. These findings underscore the importance of integrating residential histories in exposure assessments in the CTS, especially when evaluating disparities.

The cohort residential data will support future environmental studies of drinking water contaminants (e.g., arsenic), air pollution and wildfires and health effects (e.g., cardiovascular disease). Depending on factors like exposure duration and disease latency, these new residential histories provide researchers with the flexibility to estimate exposures at relevant addresses during important exposure windows and over extended periods.

We found similar mobility rates by participation status during follow-up, suggesting that efforts to contact movers using multiple address sources was successful. Each of these address sources has strengths and limitations. The USPS NCOA is a centralized hub for individuals to submit change-of-address requests, thus providing self-reported data and national coverage[17]. However, records are only provided for the last 48-months[17]. Both LexisNexis and Experian gather information from the USPS, as well as other public, proprietary, legal, and financial sources, which can lead to more information, but less certainty[18, 19]. LexisNexis is especially useful for collecting retrospective data[4–8]. The generalizability of our study is limited by address sources specific to the US. Some nations like Sweden may have nationwide population registries with comprehensive residential histories, eliminating the need to consolidate information from multiple sources[20]. In summary, our study documents an approach to de-duplicate addresses by identifying matching records based on geocode distance and street name similarity. Our approach can be readily used by other cohorts wishing to consolidate similar datasets of uncleaned addresses collected over time for participants. However, it should be noted that our matching approach may overlook short moves (e.g., different building on same street).

A major strength of our study is the integration of participant-confirmed addresses and verification against self-reported residential duration. We prioritized high-confidence addresses confirmed by participants during follow-up questionnaires, a strategy strengthened by a high response rate in the first few cycles. However, this approach may miss multiple moves that occurred between questionnaires. For those who did not maintain contact, we used records corroborated by at least two sources to reduce errors from imperfect linkages or family and secondary homes. Our approach showed high accuracy in residential changes and duration compared to self-reported data. However, duration was underestimated for those who stayed at their enrollment address for decades before enrollment, likely due to the limitations of LexisNexis retrieving historical data[4, 5].

In conclusion, our study effectively combined multiple address sources to build residential histories for the CTS cohort, demonstrating high accuracy compared to self-reported data. Approximately half of participants moved, with higher rates among younger women and those living in urban and lower SES areas. The availability of residential histories within established cohorts enables a more comprehensive and accurate assessment of a person’s physical and social environment, thereby enhancing future epidemiologic studies of long-term health.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS:

The authors would like to thank the women who have participated in the study. We also thank the CTS staff and steering committee for their continuous involvement, maintenance, and oversight of the study. A full list of CTS team members is available at https://www.calteachersstudy.org/team.

FUNDING:

The California Teachers Study and the research reported in this publication were supported by the National Cancer Institute (NCI) of the National Institutes of Health (NIH) under award number U01-CA199277; P30-CA033572; P30-CA023100; UM1-CA164917; and R01-CA077398.

This work has also been funded by the NIH Superfund Hazardous Substance Research and Training Program P42-ES033719; NIH National Institute of Environmental Health Sciences (NIEHS) P30 Center for Environmental Health and Justice P30-ES9089; NIH Kirschstein National Research Service Award Institutional Research Training Grant T32-ES007322; NIH Predoctoral Individual Fellowship F31-ES035306; University of California, Breast Cancer Research Program 16ZB-8501; R01-CA170394; and the Intramural Research Program of the NCI.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health. The collection of cancer incidence data used in the California Teachers Study was supported by the California Department of Public Health pursuant to California Health and Safety Code Section 103885; Centers for Disease Control and Prevention's National Program of Cancer Registries, under cooperative agreement 5NU58DP006344; the National Cancer Institute's Surveillance, Epidemiology and End Results Program under contract HHSN261201800032I awarded to the University of California, San Francisco, contract HHSN261201800015I awarded to the University of Southern California, and contract HHSN261201800009I awarded to the Public Health Institute. The opinions, findings, and conclusions expressed herein are those of the author(s) and do not necessarily reflect the official views of the State of California, Department of Public Health, the National Cancer Institute, the National Institutes of Health, the Centers for Disease Control and Prevention or their Contractors and Subcontractors, or the Regents of the University of California, or any of its programs.

DATA AVAILABILITY STATEMENT:

All of the data, code, and resources associated with this publication are available for research use. The California Teachers Study welcomes all such inquiries and encourages individuals to visit https://www.calteachersstudy.org/for-researchers

REFERENCES

1. Brokamp C, LeMasters GK, Ryan PH Residential mobility impacts exposure assessment and community socioeconomic characteristics in longitudinal epidemiology studies. Journal of exposure science & environmental epidemiology 2016; 26: 428–434. [PubMed: 26956935]
2. Medgyesi DN, Fisher JA, Cervi MM, Weyer PJ, Patel DM, Sampson JN et al. Impact of residential mobility on estimated environmental exposures in a prospective cohort of older women. Environ Epidemiol 2020; 4: e110. [PubMed: 33154988]
3. Joseph AC, Fuentes M, Wheeler DC The impact of population mobility on estimates of environmental exposure effects in a case-control study. Statistics in medicine 2020; 39: 1610–1622. [PubMed: 32059071]
4. Hurley S, Hertz A, Nelson DO, Layefsky M, Von Behren J, Bernstein L et al. Tracing a path to the past: exploring the use of commercial credit reporting data to construct residential histories for epidemiologic studies of environmental exposures. American journal of epidemiology 2017; 185: 238–246. [PubMed: 28073765]
5. Medgyesi DN, Fisher JA, Flory AR, Hayes RB, Thurston GD, Liao LM et al. Evaluation of a commercial database to estimate residence histories in the los angeles ultrafines study. Environ Res 2021; 197: 110986. [PubMed: 33689822]
6. Wheeler DC, Wang A Assessment of residential history generation using a public-record database. International journal of environmental research and public health 2015; 12: 11670–11682. [PubMed: 26393626]

7. Jacquez GM, Slotnick MJ, Meliker JR, AvRuskin G, Copeland G, Nriagu J Accuracy of commercially available residential histories for epidemiologic studies. American journal of epidemiology 2011; 173: 236–243. [PubMed: 21084554]
8. Brooks MS, Bennett A, Lovasi GS, Hurvitz PM, Colabianchi N, Howard VJ et al. Matching participant address with public records database in a US national longitudinal cohort study. SSM Popul Health 2021; 15: 100887. [PubMed: 34401464]
9. Bernstein L, Allen M, Anton-Culver H, Deapen D, Horn-Ross PL, Peel D et al. High breast cancer incidence rates among California teachers: results from the California Teachers Study (United States). Cancer Causes Control 2002; 13: 625–635. [PubMed: 12296510]
10. Savage KE, Benbow JL, Duffy C, Spielfogel ES, Chung NT, Wang SS et al. Using Marketing Automation to Modernize Data Collection in the California Teachers Study Cohort. Cancer Epidemiol Biomarkers Prev 2020; 29: 714–723. [PubMed: 32054690]
11. California Teachers Study Webpage. Accessed Oct 17, 2023; https://www.calteachersstudy.org/.
12. Texas A&M University. TAMU GeoServices. GeoInnovation Service Center 2022; https://geoservices.tamu.edu/.
13. Owusu C, Lan Y, Zheng M, Tang W, Delmelle E Geocoding fundamentals and associated challenges. Geospatial data science techniques and applications 2017. 41–62.
14. Fisher JA, Spaur M, Buller ID, Flory AR, Beane Freeman LE, Hofmann JN et al. Spatial Heterogeneity in Positional Errors: A Comparison of Two Residential Geocoding Efforts in the Agricultural Health Study. Int J Environ Res Public Health 2021; 18.
15. Lacey JV Jr., Chung NT, Hughes P, Benbow JL, Duffy C, Savage KE et al. Insights from Adopting a Data Commons Approach for Large-scale Observational Cohort Studies: The California Teachers Study. Cancer Epidemiol Biomarkers Prev 2020; 29: 777–786. [PubMed: 32051191]
16. 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: 429–438. [PubMed: 22617723]
17. United States Postal Service (USPS) National Change of Address NCOALink. Accessed Oct 17, 2023; https://postalpro.usps.com/mailing-and-shipping-services/NCOALink.
18. LexisNexis Public Records. Accessed Oct 17, 2023; https://www.lexisnexis.com/en-us/products/public-records.page.
19. Experian Data-driven solutions for every industry. Accessed Oct 17, 2023; https://www.experian.com/business/index.
20. Oudin A, Forsberg B, Strömgren M, Beelen R, Modig L Impact of residential mobility on exposure assessment in longitudinal air pollution studies: a sensitivity analysis within the ESCAPE project. ScientificWorldJournal 2012; 2012: 125818. [PubMed: 23251098]

IMPACT STATEMENT

Our efforts in constructing residential histories for California Teachers Study participants through follow-up (1995–2018) benefit future environmental epidemiologic studies. Address availability during the exposure period can mitigate misclassification due to residential changes, especially when evaluating long-term exposures and chronic health outcomes. This can reduce differential misclassification among more mobile subgroups, including younger women and those from lower socioeconomic and urban areas. Our approach to consolidating addresses from multiple sources showed high accuracy in comparison to self-reported residential information. The residential dataset produced from this analysis provides a valuable tool for future studies, ultimately enhancing our understanding of environmental health impacts.

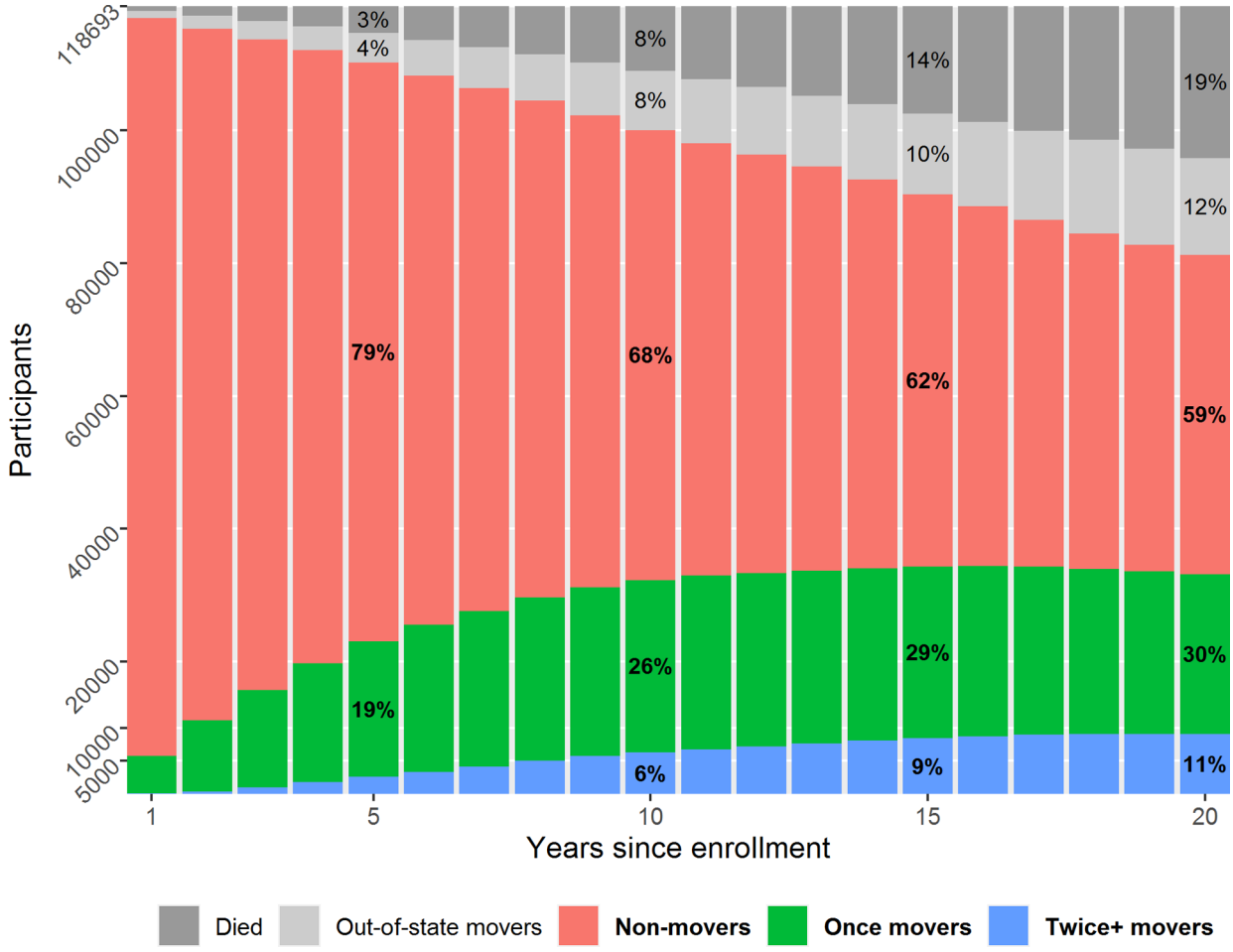

Figure 1.

Cumulative percentage of non-movers, one-time and twice or more movers within California each year since enrollment (i.e., those who remain eligible for each follow-up year). The grey categories show the percentage of the initial participants (n=118,693) who either passed away or moved out of California and therefore are no longer eligible for follow-up.

Table 1.

Percentage of non-movers, movers within California and out-of-state movers from enrollment until the end of follow-up by baseline sociodemographics, health status and follow-up participation in the California Teachers Study.

		Non-movers n (%)	Movers n (%)	Out-of-state movers n (%)
Overall		62329 (53)	40217 (34)	16147 (14)
Age at enrollment	22–34	2665 (21)	7804 (63)	1972 (16)
	35–44	9757 (47)	8362 (41)	2449 (12)
	45–54	19314 (55)	10396 (30)	5276 (15)
	55–64	13368 (60)	5900 (26)	3190 (14)
	65–79	13534 (60)	6365 (28)	2726 (12)
	80–104	3691 (66)	1390 (25)	534 (10)
Race/ethnicity	Hispanic or Latino	2476 (48)	2188 (43)	446 (9)
	Black or African American	1793 (56)	877 (27)	527 (16)
	Asian	2153 (62)	1079 (31)	242 (7)
	White	53677 (52)	34518 (34)	14334 (14)
	Other	1632 (49)	1223 (37)	450 (14)
	Not reported	598 (55)	332 (31)	148 (14)
Neighborhood socioeconomic status quartiles [1]	Q1	2312 (48)	1863 (39)	662 (14)
	Q2	9489 (48)	7416 (37)	2927 (15)
	Q3	19776 (51)	13503 (35)	5395 (14)
	Q4	30459 (56)	17128 (31)	7067 (13)
	Not available	293 (42)	307 (44)	96 (14)
Urbanicity [2]	Rural	7330 (52)	4813 (34)	2084 (15)
	Town	2025 (50)	1459 (36)	597 (15)
	City	10834 (50)	7971 (36)	3049 (14)
	Suburb	35924 (55)	21097 (32)	8583 (13)
	Metropolitan	5938 (48)	4588 (37)	1744 (14)
	Not available	278 (42)	289 (44)	90 (14)
BMI	Underweight:<18.5	1647 (53)	1073 (34)	403 (13)
	Normal:18.5–24.9	33853 (51)	23511 (36)	8863 (13)
	Overweight:25–29.9	15138 (54)	9097 (32)	3986 (14)
	Obese:30+	8557 (54)	4969 (31)	2294 (15)
	Not reported	3134 (59)	1567 (30)	601 (11)
Diabetes	No	60318 (52)	39244 (34)	15736 (14)
	Yes	2011 (59)	973 (29)	411 (12)
High blood pressure	No	50196 (51)	34416 (35)	13492 (14)
	Yes	12133 (59)	5801 (28)	2655 (13)
Smoking status	Never	11111 (49)	8547 (38)	3040 (13)
	Passively exposed	28858 (53)	18500 (34)	7350 (13)
	Former	18190 (54)	10966 (32)	4679 (14)

		Non-movers n (%)	Movers n (%)	Out-of-state movers n (%)
	Current	3354 (56)	1773 (29)	890 (15)
	Missing	816 (57)	431 (30)	188 (13)
	Yes	40620 (50)	28648 (36)	11284 (14)
Participation since 2005 [3]	No	14929 (50)	10295 (35)	4387 (15)

[1] Estimated at the enrollment address. Summary socioeconomic status metric of occupation, education and income among 1990 census block groups in California categorized into quartiles.

[2] Estimated at the enrollment address. Urbanization categories of 1990 census block groups.

[3] Among those still alive in 2005 (n=110,163)

Table 2.

Number of years lived at enrollment address, both prior to and after the start of the study, as well as total duration.

	Duration before enrollment N=81,881[1]		Duration after enrollment N=118,693		Total duration[2] N=118,693	
Years	**N**	**Cumulative N (%)**	**N**	**Cumulative N (%)**	**N**	**Cumulative N (%)**
≥20	14585	14585 (17.8)	48980	48980 (41.3)	65570	65570 (55.2)
19	1918	16503 (20.1)	1108	50088 (42.2)	1568	67138 (56.6)
18	1865	18368 (22.4)	1801	51889 (43.7)	1791	68929 (58.1)
17	1816	20184 (24.6)	1799	53688 (45.2)	1887	70816 (59.7)
16	1766	21950 (26.8)	1859	55547 (46.8)	1887	72703 (61.3)
15	1646	23596 (28.8)	2167	57714 (48.6)	1948	74651 (62.9)
14	1374	24970 (30.5)	2682	60396 (50.9)	2202	76853 (64.7)
13	8504	33474 (40.8)	2015	62411 (52.6)	1805	78658 (66.3)
12	1923	35397 (43.2)	2074	64485 (54.3)	1846	80504 (67.8)
11	2607	38004 (46.4)	2485	66970 (56.4)	1920	82424 (69.4)
10	3306	41310 (50.4)	3298	70268 (59.2)	2238	84662 (71.3)
9	4064	45374 (55.3)	3664	73932 (62.3)	2344	87006 (73.3)
8	5047	50421 (61.5)	3812	77744 (65.5)	2281	89287 (75.2)
7	5433	55854 (68.1)	3985	81729 (68.9)	2515	91802 (77.3)
6	5599	61453 (75)	4150	85879 (72.4)	2908	94710 (79.8)
5	3811	65264 (79.6)	4911	90790 (76.5)	3388	98098 (82.6)
4	3992	69256 (84.5)	5768	96558 (81.4)	4135	102233 (86.1)
3	3842	73098 (89.2)	6275	102833 (86.6)	4590	106823 (90)
2	4325	77423 (94.4)	6985	109818 (92.5)	5183	112006 (94.4)
≤1	4558	81981 (100)	8875	118693 (100)	6687	118693 (100)

[1] Among 81,981 participants (69%) with a matched record predating time of enrollment

[2] Total duration reflects duration after enrollment among those without an estimated duration before enrollment

Table 3.

Agreement in mobility status at follow-up questionnaire 4 (2005–2007) comparing self-reported duration at current residence and whether a move from the enrollment address was observed. Right-hand side the table shows the agreement between self-reported and estimated duration at the Q4 address among those correctly classified by mobility status.

Mobility status				Estimated duration (years)[2,3]				
Self-reported at Q4	**Address at Q4[1]**	**N**	**Agreement (%)**	**Mean**	**Median**	**25th perc.**	**75th perc.**	**Agreement (%)[4]**
Movers: <10 years	Moved	16517	**93.6**	4.7	5	2	7	98.6
	Not moved	1122	6.4					
<4 years	Moved	8470	**93.6**	2.7	3	1	4	85.7
	Not moved	575	6.4					
5–9 years	Moved	8047	**93.6**	6.8	7	6	8	92.4
	Not moved	547	6.4					
Non-movers: ≥10 years	Moved	1552	3.7					
	Not moved	40077	**96.3**	23.6	22	17	29	100.0
10–14 years	Moved	871	12.5					
	Not moved	6120	**87.5**	13.2	13	11	14	84.8
15–19 years	Moved	213	2.7					
	Not moved	7715	**97.3**	17.4	17	16	18	80.0
20–29 years	Moved	251	2.3					
	Not moved	10870	**97.7**	24.0	24	21	27	80.4
30+ years	Moved	217	1.4					
	Not moved	15372	**98.6**	30.6	31	22	36	58.0

[1] Women were classified as moved or not moved if the address at Q4 was different or the same as the enrollment address, respectively

[2] Calculated as the year of Q4 completion minus the estimated move-in year for the address at Q4

[3] Estimated duration among non-movers includes only those with an estimated move-in date at the enrollment address (n=36,750 of 40,077 correctly classified)

[4] Percentage of estimated duration within range of self-reported duration