Title
Clustering Life Course to Understand the Heterogeneous Effects of Life Events, Gender, and Generation on Habitual Travel Modes

Authors
Jin, Ling
Lazar, Alina
Sears, James
et al.

Publication Date
2020

DOI
10.1109/access.2020.3032328

Peer reviewed

Clustering Life Course to Understand the Heterogeneous Effects of Life Events, Gender, and Generation on Habitual Travel Modes

Authors:

Ling Jin[1], Alina Lazar[2], James Sears, Annika Todd-Blick[1], Alexander Sim[1], Kesheng Wu[1], Hung-Chia Yang[1], C. Anna Spurlock[1]

[1]Lawrence Berkeley National Laboratory

[2]Youngstown State University

[3]University of California at Berkeley

Energy Analysis and Environmental Impacts Division
Lawrence Berkeley National Laboratory

Sustainable Transportation Initiative

October 2020

This is the author's final version of a paper published in *IEEE Access*. The published version of the article can be found here:

https://doi.org/10.1109/ACCESS.2020.3032328. Please cite the published version.

Jin, L., Lazar, A., Sears, J., Todd-Blick, A., Sim, A., Wu, K., Yang, H.C. and Spurlock, C.A., 2020. Clustering Life Course to Understand the Heterogeneous Effects of Life Events, Gender, and Generation on Habitual Travel Modes. *IEEE Access*, 8, pp.190964-190980.

This work was supported by the Energy Efficient Mobility Systems (EEMS) Program of the U.S. Department of Energy's Vehicle Technologies Office under Lawrence Berkeley National Laboratory Contract No. DEAC02-05CH11231.

Clustering Life Course to Understand the Heterogeneous Effects of Life Events, Gender, and Generation on Habitual Travel Modes

LING JIN[1], ALINA LAZAR[2], (Member, IEEE), JAMES SEARS[3], ANNIKA TODD[1], ALEXANDER SIM[1] (Senior Member, IEEE), KESHENG WU[1] (Senior Member, IEEE), HUNG-CHIA YANG[1], C. ANNA SPURLOCK[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA 94720
[2]Youngstown State University, Youngstown, OH 44555
[3]University of California at Berkeley, Berkeley, CA 94120

Corresponding author: C. Anna Spurlock (e-mail: caspurlock@lbl.gov).

ABSTRACT Daily transportation mode choice is largely habitual, but transitions between life events may disrupt travel habits and can shift choices between alternative transportation modes. Although much is known about general mode switches following life event transitions, less is understood about differences that may exist between subpopulations, especially from a long-term perspective. Understanding these differences will help planners and policymakers introduce more targeted policy interventions to promote sustainable transportation modes and inform longer-term predictions. Extending beyond existing literature, we use data collected from a retrospective survey to investigate the effects of life course events on mode use situated within different long-term life trajectory contexts. We apply a machine-learning method called joint social sequence clustering to define five distinct and interpretable cohorts based on trajectory patterns in family and career domains over their life courses. We use these patterns as an innovative contextual system to investigate (1) the heterogeneous effects of life events on travel mode use and (2) further differentiation between gender and generation groups in these life event effects. We find that events occurring relatively early in life are more strongly associated with changes in mode-use behavior, and that mode use can also be affected by the relative order of events. This timing and order effect can have lasting impacts on mode use aggregated over entire life cycles: members of our "Have-it-alls" cohort—who finish their education, start working, partner up, and have children early in life—ramp up car use at each event, resulting in the highest rate of car use occurring the earliest among all the cohorts. Women drive more when having children primarily when their family formation and career formation are intertwined early in life, and younger generations rely relatively more on car use during familial events when their careers have a later start.

INDEX TERMS Life cycle, mode use, gender, generation, joint social sequence clustering, machine learning

I. INTRODUCTION

THE transportation modes (driving, biking, walking, etc.) we choose every day can have a significant aggregate impact on the transportation system and environment over the long term. As of 2016, the transportation sector accounted for about a third of total U.S. energy use and greenhouse gas emissions, of which, 56% was from light-duty vehicles used largely for passenger transport [1]. These day-to-day mode choices are largely habitual and develop over time [2]–[5]. Understanding key factors associated with the habit formation and change process is therefore important for helping policymakers design effective approaches to promoting sustainable transportation choices, and for helping planners better anticipate long-term patterns and transitions in travel mode use.

Traditional travel behavior studies have mostly relied on use of cross-sectional survey data, and they have related mobility decisions such as mode choices to individual characteristics, attitudes, and the built environments at a point in time (as discussed in [6]–[8]). This static perspective ignores the dynamics of individual travel behavior and can obscure important factors that contribute to long-term mobility decisions. More recent studies, using what are often referred to as "mobility biography" approaches, have emerged to analyze the dynamics of travel choices over the life course with a focus on the influence of life events on individual travel behavior [2]–[4], [6], [9]–[13]. This approach builds on the notion of continuity of travel behavior over a lifetime, with routinized travel habits being interrupted only by events that involve major changes in other life domains.

Major life course events, such as attending school, becoming employed, getting married, and having a child, can impact routine travel behavior and potentially cause shifts in routine choice between available transportation modes. Key life events investigated in the mobility biography framework generally fall into three hierarchical domains [14], [15]: (1) lifestyle, including household and family events (e.g., marriage and childbirth) [3], [16]–[19] and career events (e.g., attending school and getting a job) [5], [20], [21]; (2) accessibility, such as work or residential relocation [18], [22]; and (3) mobility, such as car or transit pass ownership and commute distances [20]. Family and career life events at the top level of the hierarchy represent the longest-term decisions that may directly alter travel behaviors and/or serve as fundamental triggers of other related events in the accessibility and mobility domains. For example, employment changes may trigger changes in commute distances following work relocation [23]. Childbirth may change household maintenance tasks [16] and generally increases both car acquisition and car use [24].

Although the mobility biography and related literature has investigated the role of family and career life events on *aggregate* changes in travel modes across a broad population, less is understood about differences that may exist between subpopulations, especially from a long-term perspective. For example, [19] has shown that, despite the highly car-oriented nature of travel among families with children on average, shifting towards car-based mobility did not happen to all new parents. Whether and how people change their travel mode in response to a given life event depend on the contexts within which they make these decisions, and these contexts vary across the population.

To improve the understanding of subpopulation differences in the role of family and career events, we employ a data-driven approach to derive archetypal life course cohorts and examine the effects of different life events on mode use situated within different life trajectory contexts defined by these cohorts. We use a machine-learning approach called joint social sequence clustering [25] based on patterns and timing of events in both family and career dimensions to derive life course cohorts. These cohorts are easily interpretable in terms of their life history context. We then estimate the marginal effects on the probability of different travel mode use (driving, using public transportation, and walking or biking) associated with the following life event eras: attending school, being employed, living with a partner, and having a child. We investigate how these impacts differ across the distinct life history contexts defined from the sequence clustering. We further disaggregate these impacts by gender and generation to understand how the prevalence of different life course trajectories changes based on these factors as well as how mode use is impacted by these differences.

Our study makes several important contributions to the literature. First, research that has explored this type of heterogeneity in the population has tended to define subpopulation segments using individual characteristics that do not change over time, such as gender or generation [26], [27]; previous mode preferences and attitude characteristics [19], [28], [29]; or life cycle stages or sociodemographic characteristics (e.g., age and household structure) concurrent with life events [26], [30]. However, long-term mobility decisions depend on one's current situation, past experiences, and future plans [11], and such a dynamic and multidimensional decision context can only be revealed in a long-term life history, rather than in life cycle stages defined statically at the moment an event occurs. Our analysis uses longitudinal data and brings this critical dynamic perspective.

Second, machine-learning methods such as cluster analysis have been applied to travel behavior studies to uncover heterogeneous patterns. However, such methods have more commonly been applied to short-term trip data, especially with increasingly available global positioning system (GPS) data tracking daily human mobility. Specifically, cluster or classification based pattern mining machine-learning methods have been used to identify short-term dynamics and correlations between people's daily lives, events, and the built environment [31]. In contrast, research applying machine learning to the study of habitual travel behavior using long-term observations (from a life history perspective) has just started to emerge. A review of the literature on mobility biography research using long-term longitudinal data sets (Table S1 in the Supporting Information for a summary) shows that most studies attempting this rely on traditional regression analysis. A few more advanced methods are applied in recent literature, such as dynamic Bayesian network models [32], [33] and latent transition analysis [34]. However, these studies do not explicitly identify or have an easy interpretation of the subpopulations by their long-term dynamics from multiple life dimensions, which is an innovative contribution of our study.

Third, although joint social sequence clustering has been used in sociological life course research [35], our study represents its first application to the mobility biography research on habitual mode use. By explicitly grouping the long-term life trajectory dynamics in family and career dimensions, it is possible to discover not only representative patterns based on the overall life trajectory of a given individual's characteristics but also the pathways through which individuals arrive at a given mobility decision. We demonstrate the innovation and value of this method by applying it to data from the life history calendar portion of the WholeTraveler Transportation Behavior Study survey conducted in the San Francisco Bay Area in 2018 [36]–[38]. Finally, this paper is the first mobility biography study using retrospecive longitudinal data collected in the United States.

The remainder of the paper is organized as follows: Section II describes the data and analysis methods, Section III describes the results and provides some discussion, and Section IV concludes.

II. DATA AND METHODS

A. SURVEY DATA DESCRIPTION

We use data collected as part of the WholeTraveler Transportation Behavior Study, which is part of the U.S. Department of Energy's Systems and Modeling for Accelerated Research in Transportation (SMART) Mobility Consortium [36]–[38]. The WholeTraveler survey was administered in the nine California counties that make up the San Francisco Bay Area. Resource Systems Group (RSG), a prominent transportation survey firm, administered the survey with oversight by the WholeTraveler team. Invitations to respond to an online survey were sent to a random sample of 60,000 active residential addresses via a mailed letter followed by a reminder postcard. The letter requested that a single household member over the age of 18 respond to the survey. To encourage as balanced a sample as possible, especially with respect to gender, the letter requested that the eligible member with the most recent birthday be the one to respond. To complete the survey, the respondent went online on a desktop or laptop computer. Each respondent received a $10 Amazon gift card for completing the full survey. Of the 60,000 households that were sent invitations, 997 completed the entire survey. The representation of the sample is discussed in [36]. Overall, the response rate of 1.7%, although low, is consistent with other implementations using similar unsolicited mailings to recruit and similar incentive payment levels [39]. The survey achieved a gender balance [36], but those who answered the survey were disproportionately highly educated and higher income, even within the San Francisco Bay Area, a phenomenon that is commonly found in previous retrospective survey-based mobility biography studies in other countries [9], [40]. The results should therefore be interpreted with this level of representation in mind.

The life history calendar portion of the WholeTraveler survey asked respondents to indicate when certain key life events occurred and what other factors pertained to their lives on an annual basis starting at age 20 and up to age 50. Retrospective surveys have been shown to cover relatively long periods of life course reasonably accurately when properly designed [12], [40]. We followed the recommendation in [12] on the survey structure, administration, sampling, and range of life course and mobility events considered. Respondents less than 20 years of age were not asked to answer the life history calendar questions. Table 1 itemizes all of the information requested in the life history calendar. Data from questions 1, 2, 4, and 6 are the primary source for the analysis in this paper.

TABLE 1. Life History Calendar Data

1	**Significant Events Affecting Travel Needs—the individual years in which each of the following types of events occurred, if applicable:**
1a	Children were born, adopted, or joined your household
1b	You moved or your place of school or work changed
1c	You completed a level of education (e.g., bachelor's, master's, PhD, etc.)
2	**Household—all the years when your household included the following:**
2a	A partner, spouse, or significant other
2b	At least one child 7 years old or younger
3	**Household size—all the years when your household size (including any adults or children) was as follows:**
3a	1 member
3b	2 members
3c	3 members
3d	4 members
3e	5 or more members
4	**Employment and Education—all the years when:**
4a	You were working at least 35 hours per week on average
4b	You were enrolled in school or a training program (e.g., college, trade school, internship, medical school, law school, city college, etc.)
5	**Transportation Modes Available—all the years when each of these modes was available to you to use, whether or not you did use it:**
5a	Public mass transit (bus, BART, MUNI, train, ferry) was available.
6	**Transportation Modes Used—all the years when you used each of these modes for your commute to work, school, or other primary destination regularly (two or more times per week):**
6a	Public mass transit (bus, BART, MUNI, train, ferry)
6b	Uber, Lyft, or similar app-based rideshare service
6c	Your own vehicle
6d	Walk or bike
7	**Vehicle Ownership—all the years when your household had each of the indicated numbers of vehicles:**
7a	No vehicle
7b	1 vehicle
7c	2 vehicles
7d	3 vehicles
7e	4 vehicles
7f	5 or more vehicles

We worked with RSG to design the life history calendar response interface to enable smooth, fast, and intuitive completion with a short instructional video and several logic checks to ensure respondents understood what was being asked. An example image can be found in the SI, Figure S1, showing how the questions relating to children entering the home and transportation use would appear to users of the survey when responding online. The years available for response were customized to reflect the years when the respondent was between 20 and 50 years old. Cells with boxes could be responded to using a tick mark. Cells without the inset boxes could be selected and unselected by clicking and dragging the cursor over ranges of years. This facilitated a faster response to this portion of the survey than if each cell had to be selected separately.

B. DATA PREPROCESSING

Data for this analysis are drawn from the 997 respondents who completed the entire survey, including the life history calendar. Prior to analysis, we cleaned the data to account for missing observations and erroneous responses. First, following [?], we dropped 19 respondents with a response time of less than 12 minutes. This is to screen out respondents who completed the survey only to receive the Amazon gift card—and therefore clicked through responses without reading questions or answering meaningfully. Then, we dropped observations for respondents who chose "prefer not to answer" for all years in the life history calendar in response to key variables used in this study including those related to having a child or partner, or being in school or employed (questions 1, 2, and 4), and also those related to regular mode use (question 6). Because we use information on household size and number of cars owned to verify the validity of answers to questions related to having a child and partner and using one's own car mode, we also dropped observations for respondents who chose "prefer not to answer" for all years in response to household size (question 3) and number of cars (question 7). In total, these cleaning steps removed 110 respondents. Next, to account for incongruous responses to two variables, we dropped respondents who reported using their own car but owned no cars throughout the life history or who reported living with a child and/or partner yet reported a household size of one person throughout the life history. These removed 62 respondents. We restrict the analysis to the 17,777 annual observations from the 569 respondents who were age 35 or older at the time they took the survey (in 2018) (71% of the data remaining after the above-described cleaning steps). This selection ensures that we observe responses for ages between 20 and 35 to be used in the clustering analyses, which captures a life period that presents the greatest heterogeneity among the population [20], [41]. The number of remaining respondents used in this analysis (569) is in the range of the sample sizes from previous life history calendar studies in transportation behavior research (see Table S1 in the SI; sample sizes of 66-1799 with a median of 414) for deriving insights with a similar level of confidence.

C. SEQUENCE ANALYSIS TO DERIVE LIFE COURSE COHORTS

The objective of the sequence analysis is to define archetypal life trajectories that represent different development dynamics in family and career dimensions of the life course. We construct cohorts of respondents who share similar life course trajectories using sequence analysis.

Given the many factors that affect life trajectories, and their interdependencies—such as education, employment, and family planning—we employ joint sequence analysis, formalized by Pollock [25] and Gauthier et al. [42], which simultaneously considers life sequences of multiple life dimensions. This approach is able to differentiate longitudinal experiences represented by multiple variables and hence account more realistically for the inherent complexity of life trajectory patterns [43]–[46].

For the analysis, we consider life course sequences described by family and career status captured annually. The family status is defined by having a partner or not, and having children or not. The career status is defined by attending school or not, and being employed or not. We align the family and career trajectories by age and focus on the age range between 20 and 35 to determine the similarity among respondents and their subsequent cluster membership. This age range is chosen because demographic events that cause the most variable status changes in life trajectories occur relatively early in life [20], [41]. Furthermore, this age range has no missing data in the analysis sample, so we can avoid biased clustering results due to treatment of missing values in computing the similarity metrics as documented in our previous work [47], [48].

As life trajectory sequences are categorical time series (as opposed to continuous numeric), we use Optimal Matching (OM) to determine an edit-distance-based dissimilarity measure. We follow the joint sequence analysis approach by [25] to compute dissimilarities between sequences describing life trajectories of multiple life dimensions.

Let's consider $A = (a_1, a_2, a_3, ...a_N)$ a sequence of states a_i and $\mathbf{A}$ the sequence space, where $A \in \mathbf{A}$. To explain OM, we first need to define the three transformation operators (op) for sequences: insert, delete and substitution (as shown in equations 1).

$$\begin{aligned} Ins(A, i, a^{'}) &= (a_1, a_2, a_3, ...a_i, a^{'}, a_{i+1}, ..., a_N) \\ Del(A, i) &= (a_1, a_2, a_3, ...a_i, a_{i+1}, ..., a_N) \\ Sub(A, i, a^{'}) &= (a_1, a_2, a_3, ...a_{i-1}, a^{'}, a_{i+1}, ..., a_N) \end{aligned} \quad (1)$$

A sequence of the operators defined above can be used to define a transformation between two sequences A_s and A_r.

$$A_r = op_1 \diamond op_2 \diamond ... \diamond op_n(A_s) \quad (2)$$

where $op_1 \diamond op_2$ denotes one transformation operation.

The goal of OM is to find the transformation with the minimum number of operations. To be able to fine tune the algorithms the concept of operation cost was introduced [49]. The cost of insertion and deletion is usually referred as *indel* and separately the substitution cost is defined. The total transformation cost between two sequences can be formalized has follows:

$$d(A_r, A_s) = \min_{opSeq}\{c(opSeq), A_r = opSeq(A_s)\} \quad (3)$$

where $opSeq = (op_1, op_2, ...op_q)$ and $c(opSeq)$ is the total cost of the transformation. Consequentially, the goal of OM becomes the minimization of $d(A_r, A_s)$, for a chosen *indel* and *substitution* cost.

In practice the balance between the *indel* and the *substitution* cost influences the resulting patterns. When the ratio between the *substitution* and *indel* cost is low OM becomes the Hamming distance, based mainly on substitutions. The higher the above mentioned ratio, the closer OM changes to the Levenshtein II distance, that equivalent to the length of the longest common subsequence. The Hamming distance is very sensitive to timing and in order to allow sequence alignments with time delays, *indels* should be used alongside *substitutions*.

The method used for our application is implemented in the TraMineR package version 2.0-6. We follow the implementation reported previously in our pilot work [47], [48]. The *indel* cost is set to 1. The substitution costs are generally set to 2 as it takes one insert and one delete to substitute. The substitution costs used in our study are instead determined and set independently for family and career trajectories. A data-driven approach is employed to adjust our substitution costs according to the transitional frequency between given states [50]. To implement this approach a substitution-cost matrix, with dimension $ns * ns$, where ns is the number of states in the alphabet of the sequence object, is computed based on the 4 formula. An element (i, j) of this matrix is defined as the cost of substituting state i with state j. The substitution cost between states i and j in any given sequence is obtained with the following formula:

$$SC(i, j) = substitution\_cost - P(i, j) - P(j, i) \quad (4)$$

where $P(i, j)$ is the transition rate from state i to j and $substitution\_cost$ is the generic constant substitution cost value 2.

Our application requires us to consider multiple sequences in the same time. Computing the multichannel distance for multiple sequences, can be done following the strategy proposed by Pollock [25]. First, all the sequences from each channel in consideration builds are combination together to form a new sequence. Second, the data-driven substitution cost matrix is derived by averaging the costs of substitution over all channels.

Lastly, the cohorts are derived by Ward hierarchical clustering of the sequences following [25], [51], [52] based on these OM distances. The Ward method was shown to produce relatively evenly distributed clusters so that a more diverse and balanced set of life history patterns can be derived [53].

D. CLUSTER VALIDITY MEASURES

When clustering life-course sequences from the WholeTraveler survey, there are no predefined categories to be used as the ”ground truth” for the quality evaluation of the resulting clusters. In order to determine the appropriate number of clusters, an iterative process based on two different internal clustering validity measures has been proposed. This process provides a quantitative quality measure for the resulting cluster partitions. Internal clustering validity measures [54], as opposed to the external validity measures, not only evaluate the quality of the returned clustering structure without the ”ground truth”, but can be used to determine the optimal number of clusters and the best clustering algorithm and for a given problem. We evaluate the cluster quality results by varying the number of clusters k taking values in the $[1, 10]$ interval and using the Point Biserial Correlation (PBC) [55] and the Average Silhouette Width Index (ASW) [56] internal cluster validity indexes. PBC is an easy measure of the resemblance between the distance matrix and the resulting hierarchical clustering dendrogram. ASW validates clustering performance based on the pairwise difference of between- and within-cluster distances. These metrics are further explained in details below.

The ”Point Biserial Correlation” (PBC) (equation 5), proposed by Hennig and Liao [55] uses the Pearson’s correlation to evaluate and compare cluster solutions. PBC measures the resemblance between the proximity matrix and the resulting hierarchical clustering dendrogram. Quantitatively, it determines the correlation between the proximity matrix d and a binary (zeros and ones) matrix d_{bin} indicating whether two objects are in the same cluster or not. The PBC is computed as follows:

$$PBC = \frac{s_{d,d_{bin}}}{s_{d_{bin}} \cdot s_{d_{bin}}} \quad (5)$$

Where s_d and $s_{d_{bin}}$ are the standard deviation of d and d_{bin} respectively, and $s_{d,d_{bin}}$ is the covariance between d and d_{bin}.

The Average Silhouette Width (AWS) was initially proposed by Kaufman and Rousseeuw [56]. This measure is based how similar the points in a given cluster are versus the points in other clusters. ASW is calculated by comparing the average distance of a point to the other members of its group with the average weighted distance to the closest cluster. Let NC be the number of clusters, n_i be the number of objects in cluster i, C_i denote cluster i, and $d(x, y)$ be the distance between x and y. Given above notation, ASW is computed as follows:

$$ASW = \frac{1}{NC}\sum_i\left(\frac{1}{n_i}\sum_{x\in C_i}\frac{b(x) - a(x)}{max(b(x), a(x))}\right) \quad (6)$$

$$a(x) = \frac{1}{n_i - 1}\sum_{y\in C_i, y\neq x} d(x, y) \quad (7)$$

$$b(x) = min_{j,j\neq i}\frac{1}{n_j}\sum_{y\in C_i} d(x, y) \quad (8)$$

These two measures provide an objective way to choose the number of clusters. Once the optimal number of clusters is selected, it is used to generate the clustering groups.

E. EMPIRICAL ANALYSIS SPECIFICATIONS TO UNDERSTAND LIFE EVENT EFFECTS

The goal of the empirical analysis is to understand the following. (1) To what extent do life events alter the uses of different travel modes? (2) Are the marginal changes in the probability of using a travel mode associated with a life event era (living with a partner, living with a child, being in school, or being employed) different across subpopulations with different life course trajectories? (3) Are these differences further elucidated by differences across gender and generation?

To address these questions, we use data to test four sets of hypotheses motivated largely by the role of individual life events on travel mode use previously identified in the literature.

- Hypothesis set 1: Living with a partner, living with a child, being in school, and being employed each changes mode use and has a lasting effect at the population level.
- Hypothesis set 2: There is differentiation across the life history cohorts in the extent to which living with a partner, living with a child, being in school, and being employed each changes mode use.
- Hypothesis set 3: Gender and generational differences exist in these relationships at the population level.
- Hypothesis set 4: Gender and generational differences exist in these relationships at the subpopulation level.

There are four separate outcome variables of interest (Y_{igt}), which are defined for each person i of age g in year t: three binary choices of regularly using (more than twice per week on average) one of three different transportation modes (drive own car, use public transit, walking/biking) and one value for the total number of modes used.

We run fixed-effect panel regressions with robust clustered standard errors for each outcome variable to estimate the average marginal effect on that outcome variable of being in the following life event eras (estimated both with and without differentiating across life course cohorts): being in school, being employed, having a partner, and preparing for or having a child. The life event era of having a child includes what we refer to as the “nesting” period, which is defined as the 2 years prior to the year in which the first child enters the household, as well as all the years in which at least one child under the age of 18 is in the household. We include the nesting period because our data indicate people tend to move in anticipation of parenthood within the 2 years prior to having their first child (see SI, Table S2). In contrast to other studies that focus on the before and after changes of a given event (e.g., [11]), our approach estimates the marginal effects averaged over the whole period (or era) of life status defined by these specific life events in order to derive the long-term effects.

The initial specification (Equation 9) quantifies the marginal effect of each of the life event eras on the outcome variables over all the respondents without considering heterogeneity across the life history cohorts.

$$Y_{igt} = \alpha_i \ + \varphi \cdot X_{it} + \delta_g + \varepsilon_{it} \tag{9}$$

In Equation 9 X_{it} indicates one of the binary variables $school_{it}$, $employ_{it}$, $partner_{it}$, $nesting/children_{it}$, each equal to 1 when respondent i is in that life event era during year t, zero otherwise. Note that nesting/children equals 1 during the years when there are any children in the home between the ages of 0 and 18, and it equals 1 during the 2 years prior to the year the first child enters the home, equaling zero otherwise. We include a person fixed effect α_i, which controls for everything that differs across individuals but does not change over time, and an age fixed effect δ_g that controls for age-specific factors that are shared across all individuals. To account for serial correlation across time observations within individuals, we cluster the standard errors of the estimates at the individual level. The error terms are therefore assumed to be independent and identically distributed (IID) normal across individuals, but correlation within an individual over time is accounted for.

Then we estimate the effects differentiated across the life course cohorts defined by the clustering analysis (Equation 10).

$$Y_{igt} = \alpha_i \ + \sum_c \varphi_c \cdot X_{it} \cdot cohort_{c,i} + \delta_g + \varepsilon_{it} \tag{10}$$

All terms in Equation 10 and defined as in Equation 9, with the addition of $cohort_{c,i}$, which is an indicator variable equal 1 if individual i is in cohort c, 0 otherwise. The cohorts are those defined by the joint social sequence clustering analysis.

To differentiate potential gender and generational effects within these life trajectory cohorts, we further divide the population into two gender categories (female or not) and two generational categories (born after 1964 or not). The year 1964 is used to divide baby boomers (born between 1946 and 1964) and Generation X (born between 1965 and the early 1980s). Because our sample is restricted to people who were 35 or older in 2018, the respondents were born in 1983 or earlier. For simplicity, we refer to the younger generation cohort (born 1965 to 1983) as GenX. We estimate the gender and generational differences of the effects of life event eras on outcome variables using the specifications in Equations (11) and (12).

$$\begin{gathered} Y_{igt} = \alpha_i \ + \sum_c \varphi_c \cdot X_{it} \cdot cohort_{c,i} + \\ \sum_c \gamma_c \cdot X_{it} \cdot cohort_{c,i} \cdot fem_i + \delta_g + \varepsilon_{it} \end{gathered} \tag{11}$$

$$
\begin{aligned}
Y_{igt} = \alpha_i \ + \sum_c \varphi_c \cdot X_{it} \cdot cohort_{c,i} + \\
\sum_c \eta_c \cdot X_{it} \cdot cohort_{c,i} \cdot genX_i + \delta_g + \varepsilon_{it}
\end{aligned} \tag{12}
$$

Where fem_i is an indicator variable equal to 1 if respondent i identifies as female, 0 otherwise, and $genX_i$ is an indicator variable equal to 1 if respondent i was born in or after 1965, 0 otherwise. Here the parameters γ_c and η_c capture the life event effects of the female and GenX respondents in life trajectory cluster $cohort_c$ relative to their male and older generation counterparts in the same cohort, respectively.

F. SUMMARY STATISTICS

Table 2 summarizes the data used in our analysis. The top part of the table summarizes data from the panel provided by the life history calendar, while the bottom part summarizes characteristics of individual respondents that do not change over time. Because not all respondents were over age 50 when they completed the survey, the average age represented in the data is below the midpoint between 20 and 50, at 33.5 years. Most of the respondents included in this analysis were born in the 1960s (25%) and 1970s (28%), with GenX accounting for 57% of the respondents. Female and non-female respondents are roughly equally represented.

The life history calendar shows that 37% of observations come from individuals who had at least one child between the ages of 20 and 50 and 52% of the observations from individuals who lived with a partner between ages 20 and 50. The respondents on average were employed during most of the period (75%) covered in the life history calendar, while the period of attending school accounted for 21% of the observations.

For approximately 77% of the observations, respondents indicated that public transit was available for regular use, whether or not it was used. Most respondents used driving as a regular transportation mode; respondents reported driving regularly (defined as two or more times per week) in 65% of observations. Public transit and walking or biking were used regularly during 27% and 21% of the annual observations, respectively. Interestingly, respondents reported relatively low levels of multi-modality; on average, individuals used 1.1 modes regularly in a given year. However, substantial variation exists, with some respondents using none of the modes regularly and others using as many as four.

At the population level (Figure 1), individuals generally finish school and achieve full-time employment relatively before age 30. Among all respondents, the percentage of people living with a partner slowly increases between ages 25 and 35, and children arrive on average 7 years after people first report living with a partner. The percentage of households with children peaks around age 40. Among all respondents, the percentage of people driving regularly increases with age, while the percentages regularly using public transportation and walking/biking decrease.

TABLE 2. Summary Statistics of Life History and Individual-Level Variables

Life History Calendar Summary	Mean	SD	Min	Max
Nesting or has a child (< 18 yr old) in house	0.37	0.48	0	1
Has partner in the house	0.52	0.5	0	1
Employed ($\geq$ 35hr/wk avg)	0.75	0.43	0	1
Enrolled in school or training	0.21	0.41	0	1
Drove regularly	0.65	0.48	0	1
Took public transit regularly	0.27	0.44	0	1
Walked or biked regularly	0.21	0.4	0	1
Number of modes used	1.14	0.65	0	4
Public transit available	0.77	0.42	0	1
Age in lifecycle calendar	33.53	8.53	20	50
Number of observations	15,381			
Individual-Level Summary	Mean	SD	Min	Max
Born 1930s	0.02	0.12	0	1
Born 1940s	0.1	0.3	0	1
Born 1950s	0.19	0.39	0	1
Born 1960s	0.25	0.44	0	1
Born 1970s	0.28	0.45	0	1
Born 1980s	0.17	0.37	0	1
Born in/after 1965	0.57	0.5	0	1
Born before 1965	0.43	0.5	0	1
Female	0.49	0.5	0	1
Male	0.51	0.5	0	1
# of respondents	569			

III. RESULTS AND DISCUSSION

A. ARCHETYPAL LIFE COURSE PATTERNS DERIVED FROM SEQUENCE CLUSTERING

Life trajectory clusters are derived through the Ward's linkage hierarchical algorithm applied to multichannel OM distance matrices computed from the family (having a partner or not, having a child or not) and career (in school or not, employed or not) sequences (Figure 2a). We evaluate the cluster quality using the Point Biserial Correlation (PBC) [55] and the Average Silhouette Width Index (ASW) [56] cluster validity indexes. The number of clusters k taking values in the [1,10] interval. For both PBC and ASW indexes, higher value of the index indicates better cluster quality. The PBC index increases with the number of clusters and stabilizes around five, while the ASW index overall decreases with number of clusters, especially after five. Based on these observations, we choose five clusters when both index values are reasonably large while the number of clusters is relatively small for interpretation (Figure 2b).

Figure 3 summarizes the dynamic patterns across age of the percentage of people in each life event era separately for each of the five clusters. The clusters are mostly driven by the timing of partner and children in the family sequences, whereas career trajectories are more homogeneous except for the last two smaller clusters, which exhibit patterns distinctive from the rest of the population. Based on their observable life trajectory patterns, the five clusters are referred to as "Singles," "Couples," "Have-it-alls," "Late Bloomers," and

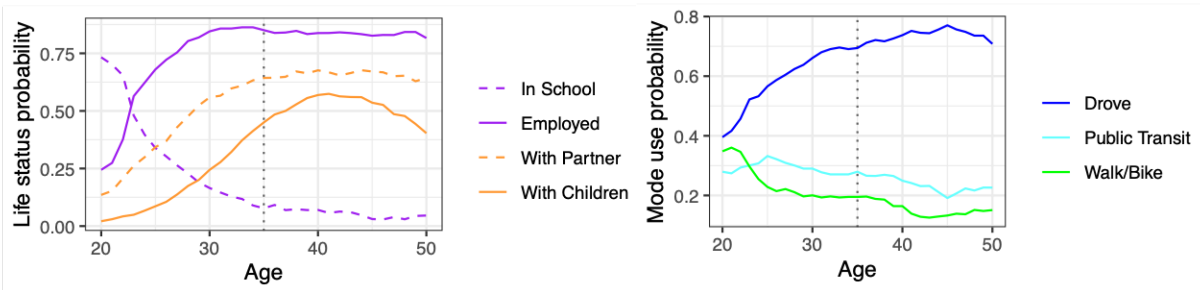


FIGURE 1. Population-level dynamics of family and career status (left) and mode-use probability (right) over the life course.

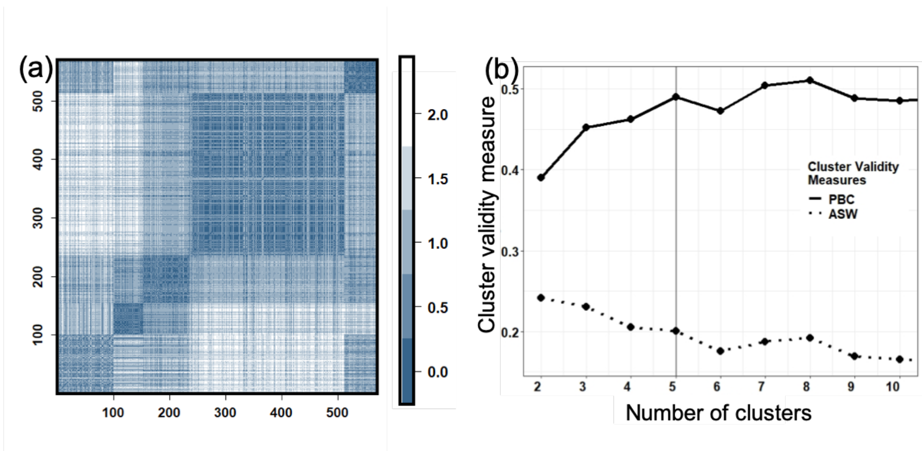


FIGURE 2. OM distance matrix with smaller values indicating more similarity (a); and cluster validity measures vs. number of clusters (b).

"Family First" (as named in [46]) and serve as the five subpopulations (referred to as "life history cohorts") for the subsequent analysis. The life history calendar variables used in the analysis as well as gender and generational information are summarized in Table 3 for each of the cohorts. Here we describe each of these five life history cohorts in more detail.

- Singles is the largest cohort in our sample (40% of the sample). Members of this cohort tend to finish school and enter the workforce early and delay or forgo having a partner or children.
- Couples is the second largest cohort in our sample (27%). Cohort members tend to finish school, work, and partner up early but delay (on average 11 years after coupling up), or forgo having children.
- The Have-it-alls cohort accounts for 18% of respondents. Cohort members finish school and start to work early in life, and they partner up and have children only slightly later (on average 4.6 years after coupling up). Similar to those in the Singles and Couples cohorts, they generally (more than 85% of them) finish school around 25–30 years old and are employed full time by around 30. Similar to those in Couples, but in contrast to those in Singles, these people partner up relatively early; 80% of them already live with a partner before age 27. However, in contrast to those in Couples, more than 90% of them begin to have children between the ages of 26 and 35.
- Members of the Late Bloomers cohort, making up only 8% of the sample, generally delay school, work, partnering, and children until much later in life, if at all.
- Family First is the smallest cohort, making up 7% of the sample. These people tend to partner up and have children early and delay school and/or work.

B. DESCRIPTIVE LIFE HISTORY COHORT ANALYSIS

The life history calendar variables used in the analysis as well as gender and generational information are summarized in Table 3 for each of the cohorts. Although the clustering to generate the five cohorts is based on the pattern observed during ages 20–35, we observe those who were older than 35 when surveyed for periods after the 20–35 age range. The summary statistics in Table 3 encompass all of the data observed for those included in the clustering analysis, including for ages beyond 35 and up to age 50.

Overall, members of the Singles, Couples, and Have-it-alls cohorts have the highest percentage of observations in

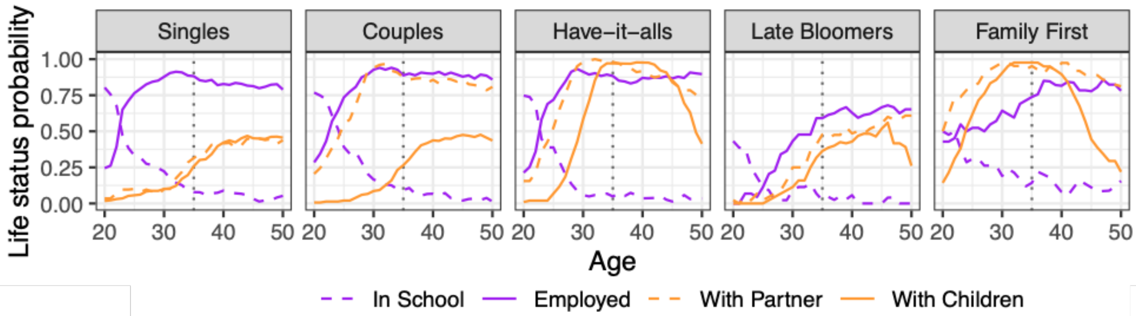


FIGURE 3. Life course patterns of family and career status in five life course cohorts.

which they report being employed, and members of the Late Bloomers cohort have the lowest percentage. For around 70% of their observations, the Have-it-alls and Family First cohorts have children, and for more than 70% of the observations between ages 20 and 50, they are living with a partner. Despite those in the Couples cohort being in the nesting stage or having children in only 25% of the observations, they are living with a partner in 74% of the observations.

The gender and generational compositions vary by life history cohorts (Table 3). Gender splits are roughly equal in Singles and Couples. Those in the Have-it-alls and Late Bloomers cohorts are more likely to be male (60% and 56%, respectively), whereas the Family First cohort is dominated by female respondents (60%). Consistent with the whole sample average, GenX dominates (at around 60%) most cohorts except for Family First, in which the older generation dominates (62%). At a finer resolution (Figure 4), people born in the 1930s tend to follow the life trajectories of Family First and Have-it-alls. The composition of life trajectories experiences a rapid shift around the 1940s birth years and stabilizes after the 1950s birth years, after which the Singles, Couples, and Have-it-alls cohorts dominate. The Singles and Couples cohorts are largely defined by having children much later in life, if at all, whereas the Family First and Have-it-alls cohorts generally have children much earlier. This distinction explains much of the distributional difference in generation between life course trajectories.

The overall mode use over the individuals' life courses varies across the cohorts as well (Table 3, also visualized in the SI, Figure S2). Those in the Have-it-alls cohort rely most on regular car use (73% of observations) and regularly use walking/biking (16%) and public transportation (24%) less compared with the whole-sample averages. The Couples cohort also exhibits more frequent regular car use compared with the sample average; however, this cohort is the most multi-modal, and exhibits higher rates of regularly walking/biking and using public transit compared with most other cohorts. Those in Couples and Singles overall regularly walk/bike and use public transit the most. In addition, those in Singles regularly drive less over their lifetimes (63% of observations) relative to the population average (65%). The

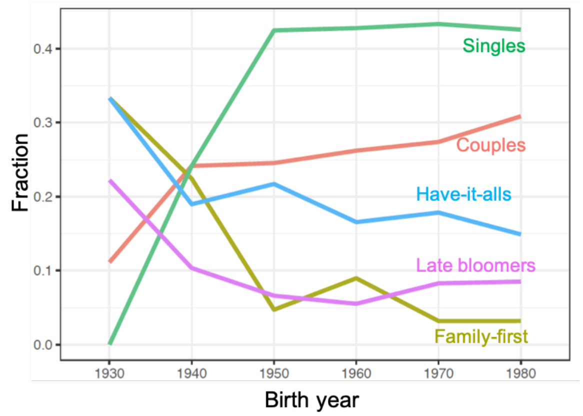


FIGURE 4. Generational evolution of cohort composition.

Family First cohort has only a slightly lower level of regular car use compared with the sample average and, along with Have-it-alls, has the lowest share of observations for which walking/biking and public transit are regularly used. Late Bloomers depend least on driving their own cars but also have relatively low use of walking/biking and public transit.

The dynamics of mode use across age (Figure 5) are such that the percentage of people driving regularly generally increases with age, whereas the percentages of regular public transit use and walking/biking decrease. The magnitude and timing of the trends, however, vary across the cohorts. For example, the increasing trend in driving is much steeper for those in Have-it-alls between ages 20 and 30 than for those in Singles, whereas a sharp rise in driving for those in Late Bloomers comes after age 30. The association of life events with these observed trends in travel mode use is determined quantitatively in our regression analysis presented in the next section.

C. EFFECTS OF LIFE EVENTS ON MODE USE

The marginal effects (φ and φ_c) estimated from Equations (9) and (10) are the change of mode choices averaged over

TABLE 3. Summary by Life Course Cohorts

Variables	Singles 229 40%	Couples 151 27%	Have-it-alls 103 18%	Late Bloomers 44 8%	Family First 42 7%
	Cohort and Number of Respondents in Cohort (% of Total Sample)				
Life History Calendar Summary					
Nesting or has child (< 18 yr old) in house	0.25	0.25	0.65	0.27	0.73
Has partner in the house	0.24	0.74	0.76	0.3	0.87
Employed ($\geq$ 35hr/wk avg) in the year	0.77	0.8	0.79	0.42	0.68
Enrolled in school/training program in the year	0.22	0.24	0.18	0.1	0.22
Drove regularly	0.63	0.69	0.73	0.45	0.64
Took public transit regularly	0.32	0.27	0.24	0.22	0.15
Walked or biked regularly	0.22	0.23	0.16	0.2	0.16
Number of modes used	1.16	1.19	1.13	0.88	0.95
Public transit available	0.77	0.82	0.79	0.62	0.67
Age in life cycle calendar	33	33	34	33	34
Individual-Level Summary					
Born 1930s	0	0.01	0.03	0.05	0.07
Born 1940s	0.06	0.09	0.11	0.14	0.31
Born 1950s	0.2	0.17	0.22	0.16	0.12
Born 1960s	0.27	0.25	0.23	0.18	0.31
Born 1970s	0.3	0.28	0.27	0.3	0.12
Born 1980s	0.17	0.19	0.14	0.18	0.07
Born in/after 1965	0.6	0.57	0.56	0.61	0.38
Born before 1965	0.4	0.43	0.44	0.39	0.62
Female	0.51	0.49	0.4	0.44	0.6
Male	0.49	0.51	0.6	0.56	0.4

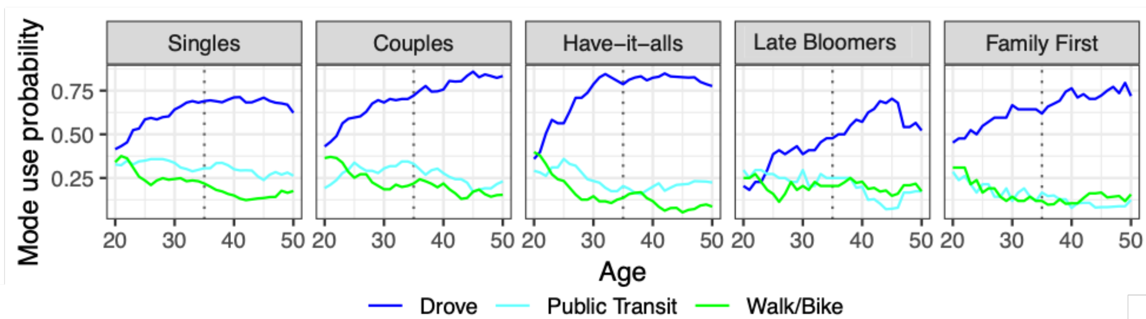


FIGURE 5. Life course patterns of mode-use probability in the five cohorts.

a specific life event era relative to not being in that era, and they can be interpreted as the percentage point change in the probability of choosing this mode during this life event era relative to outside that era. We explicitly account for the age effect by controlling for it in the panel regression as fixed effects, so our results are meant to represent the effects of life events above and beyond the underlying overall socioeconomic status evolution with aging. Figure 6 summarizes the regression results and the detailed results are shown in SI Tables S3–S6.

The *aggregated* level (leftmost column "Ave Effects" in Figure 6) results represent answers to Hypothesis set 1. First, attending school is associated with a lower probability of regularly driving and a higher probability of regularly using public transit and walking/biking relative to not attending school. In addition, a higher overall number of modes is regularly used during this era. Second, being employed full time is associated with a higher probability of regularly driving and regularly using public transit as well as a lower probability of walking/biking, relative to not being employed full time. Third, living with a partner is associated with an increased probability of regularly driving and a decreased probability of regularly walking/biking, relative to not living with a partner. Finally, parenting is associated with a decreased probability of regularly using public transit and walking/biking as well as an overall reduction in multi-modality, relative to not being a parent. The directions of the sample average life event effects seen here are largely consistent with the mobility biography studies conducted in other countries, e.g., [3], [57].

Consistent with Hypothesis set 2, these life event effects

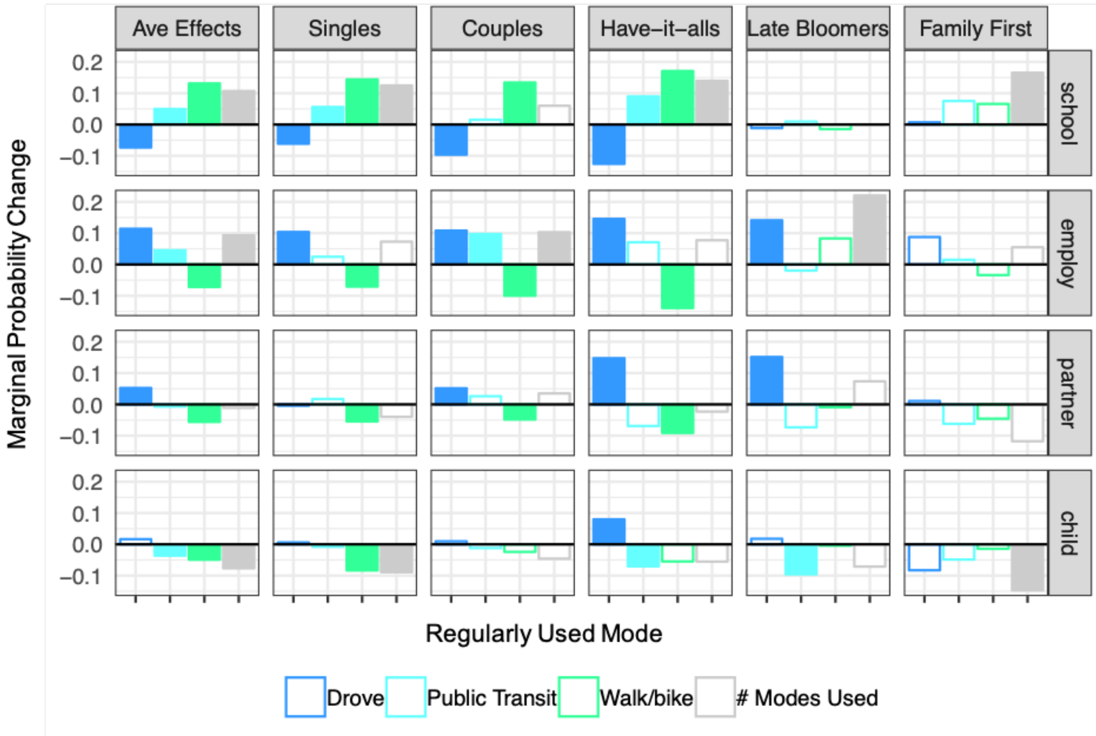


FIGURE 6. Marginal effects of life events (indicated by row facet) on mode use by cohorts (indicated by column facet). Solid bars indicate values are statistically different from zero at the 10% level.

vary by life trajectory cohort (i.e., when the events are situated within different life trajectories) as shown in the other five columns of Figure 6. Attending school is associated with a lower probability of regularly driving and a higher probability of regularly walking/biking relative to not being in school, particularly among those in the Singles, Couples, and Have-it-alls cohorts, all of whom finish school early in life. However, school has minimal effects on regular car use and walking/biking for those in the Family First and Late Bloomers cohorts, who generally delay their education or are not in school during the observed period (ages 20 to 50). In addition, being in school increases the number of modes used for the Singles, Have-it-alls, and Family First cohorts.

Employment is associated with a higher probability of regularly driving and a lower probability of regularly walking/biking among the Singles, Couples, and Have-it-alls cohorts, relative to not being employed for those same cohorts. The overall number of modes used increases with employment among the Couples and Late Bloomers cohorts. Employment appears to have minimal effect on regular mode use for those in Family First, who generally delay their employment and focus on family development before age 35.

In general, living with a partner is mostly associated with shifting modes, often from regularly walking/biking to more regular driving, rather than a change in the overall number of modes regularly used.

Having children has the most heterogeneous impact across cohorts: mode use is minimally impacted for Couples, the number of modes regularly used decreases for Family First, probabilities of regularly walking/biking and overall multimodality decrease for Singles, the probability of regular driving increases and the probability of regular public transit use decreases for Have-it-alls, and the probability of regular public transit use decreases for Late Bloomers. All these effects are relative to not being in the child-rearing era within each of these cohorts. One mechanism to explain the decrease in regular public transit use associated with having children is transit availability, which may vary if residence or frequent destinations change during this same period. To test this hypothesis, we further use public transit availability as the outcome variable in Equation 10 and estimate the marginal effects of life events on public transit availability in the residence area. We find that only the Have-it-alls cohort exhibits significant and similar directional differences in public transit availability as in regular public transit use (SI Tables S6 and S7) during the parenting era. This indicates that those in the Have-it-alls cohort might be more likely than others to transition to a more suburban—and therefore more transit-poor and car-dependent—lifestyle when having children.

D. GENDER AND GENERATIONAL EFFECTS

The literature generally suggests that travel needs and constraints differ between men and women, especially with regard to familial events [20], [27], [58] over the life cycle. As a result, we focus on discussing our results related to the "partner" and "child" events (Figure 7(a)) differentiated not only between life history cohorts, but also further differentiated by gender. Complete results can be found in SI Tables S8–S11.

Mobility patterns also differ by generation owing to different residence preferences, travel time allocation, economics, and community infrastructure changes among generations [41]. We do not find significant differences among younger (GenX) versus older generations regarding the effects of attending school or being employed. Therefore, we only discuss the differential impacts of familial events across generations (Figure 7b and complete results presented in SI Tables S12–S15).

Gender and generational gaps are observed at the population level for some of the mode uses, confirming Hypothesis set 3. At the *aggregate* level (Figure 7(a), leftmost column "Ave Effects"), women appear to regularly use less public transit (around 7 percentage points) and regularly use less numbers of modes (-0.12) during parenthood relative to men for the same comparison. This is consistent with the literature, which suggests that women are more likely to be primary caregivers at home on a daily basis and make more adjustments following childbirth [11], [27]. Overall, no gender differences are found in car use ("Drove" mode) for the "child" event, which is also consistent with previous findings in [27].

For the generation gap at the population level (Figure 7b, leftmost column "Ave Effects"), having a child is associated with a higher probability of regularly driving and a lower probability of regularly using public transit among the younger generation compared with the older generation. This is consistent with the common perception that U.S. car use has increased as residents have migrated from central cities to suburbs over the past several decades [59]. These results suggest that such a migration is more likely to happen when people are anticipating or having a child.

However, counter to Hypothesis set 4, across the life history cohorts, gender differentiation related to familial events and mode usage is only pronounced for those in the Family First and Have-it-alls cohorts which have children relatively early. Women in these cohorts tend to have higher probabilities of regularly driving relative to men in the same cohorts when living with a partner (Family First) or having a child (Have-it-alls). The cohort-specific average analysis shows children associated with higher probabilities of regular car use in the Have-it-alls cohort (Figure 6), and the gender-based analysis suggests that this increase is primarily due to women in this subpopulation regularly driving more. However, our analysis adds nuance to previous research by suggesting that women drive more when having children primarily when their family formation and career formation are intertwined (i.e., only in our Have-it-alls cohort) and therefore are more likely to be time-poor. In contrast to findings from existing literature, our results suggest that having children does not necessarily produce differential changes in regular mode use for women relative to men in subpopulations characterized by children arriving later in life.

At the same time, across cohorts, having a child has a differential impact by generation only for the Family First cohort, which includes a disproportionate number of people born in and before the 1940s (Table 3); the higher probability of regularly driving among the younger generation relative to the older generation in the child-rearing era is likely related to the effects of increasing suburbanization over time discussed above. Living with a partner is associated with a higher probability of regularly driving and a lower probability of regularly walking/biking for the younger generation in the Late Bloomers cohort relative to the older generation. Among the three cohorts that dominate the younger generation (Singles, Couples, and Have-it-alls), there is no significant generational difference estimated in regular car use during these familial events, whereas the younger generations in the Singles and Have-it-alls cohorts are less likely to regularly use public transit during parenthood relative to older generations.

IV. DISCUSSION AND CONCLUSIONS

Our study expands the life course research in transportation by clarifying the life course dynamics of travel choices and their heterogeneous responses to family and career life events across subpopulations. We propose and apply a methodology to derive interpretable archetypal life course cohorts and use the life course itself as a contextual system, enabling mobility decisions and their gender and generational differences to be evaluated for life events situated within different continua of past and future experiences and decisions.

The innovation of our study is the employment of a machine-learning method to identify heterogeneous subpopulations. Given the many factors that affect life trajectories—such as education, employment, and family planning—and their interdependencies, we employ joint sequence analysis [25], [42], which simultaneously considers life sequences of multiple life dimensions. Although joint social sequence clustering has been used in life course research in the sociology field [35], our study represents its first application to the mobility biography research on habitual mode use. By explicitly grouping the long-term life trajectory dynamics in family and career dimensions, it is possible to discover not only representative patterns based on the overall life trajectory of a given individual's characteristics but also the pathways through which individuals arrive at a given mobility decision. We demonstrate the innovation and value of this method by applying it to data from the life history calendar portion of the WholeTraveler Transportation Behavior Study survey conducted in the San Francisco Bay Area in 2018 [36]–[38].

Existing published retrospective survey studies cover Eu-

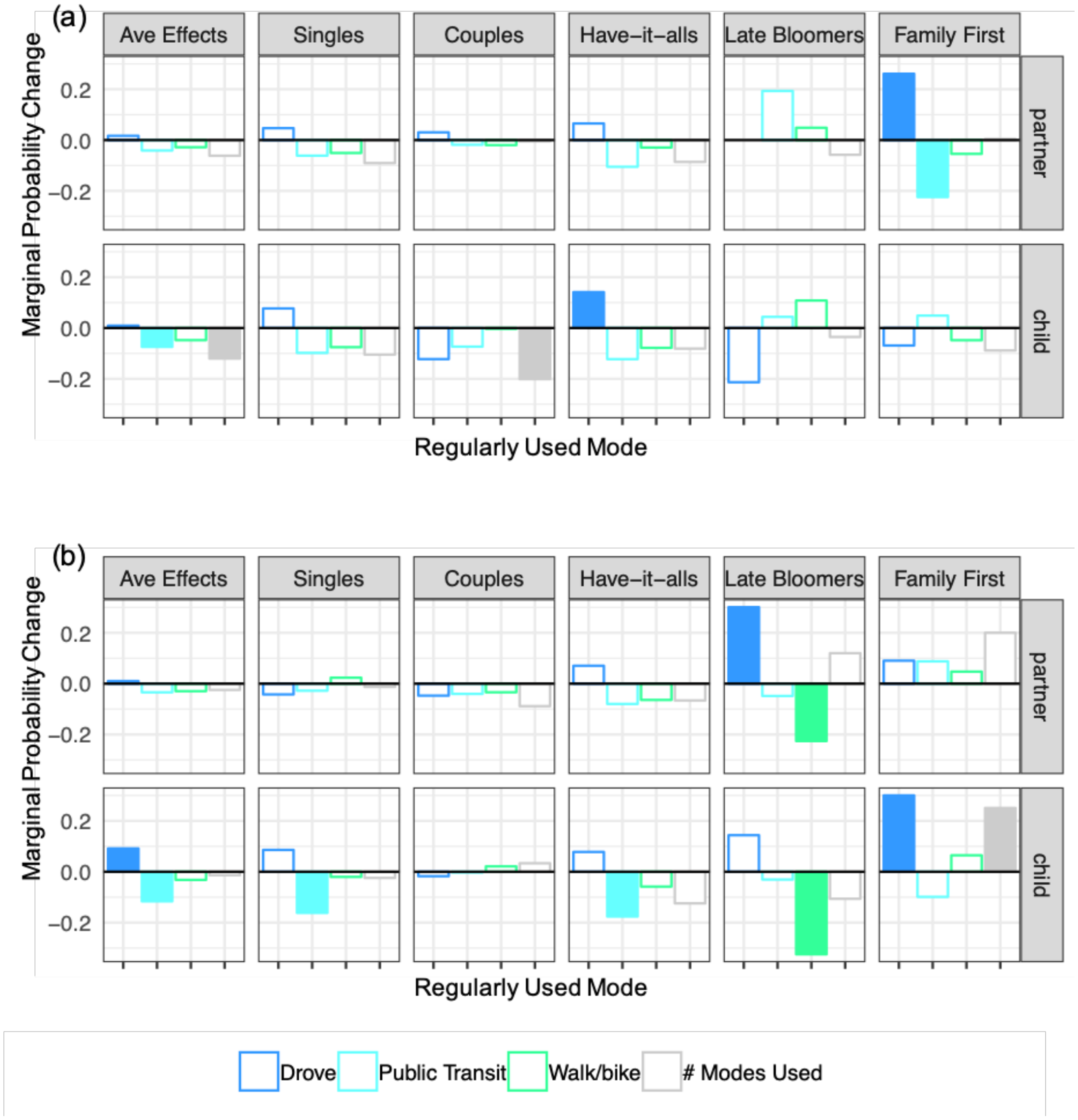


FIGURE 7. Difference in the marginal effects of life events (indicated by row facets) on women relative to men (a), and on GenX (born in and after 1965) relative to the older generation (born before 1965) (b). Solid bars indicate values statistically different from zero at the 10% level.

rope and Australia, with more recent studies covering Africa and China (see the review in Table S1). Ours is the first study of the unique life history calendar survey collected in the United States. Our study corroborates existing knowledge about the general population in the other countries—confirming their similarity with parts of the U.S. population—while providing important additional insights on subpopulations.

Our aggregate results—averaged across our entire sample of Bay Area survey respondents—are largely consistent with results from the literature with regard to the associations between regular mode use and select family and career events. However, the results from our life trajectory cohort analysis suggest that such aggregate associations cannot fully capture subpopulation heterogeneity, and they enable us to offer novel insights based on a life course perspective.

Events that occur relatively early in life are more strongly associated with changes in mode-use behavior compared with events that occur later. This is exemplified by comparing the association of driving with life events for the Singles, Couples, and Have-it-alls cohorts, which share important life event similarities and thus enable specific associations to be isolated. The literature suggests that, in aggregate, regular driving behavior changes owing to attending school (less driving) and working, partnering, and having children (more driving). However, for these three cohorts, only life events that initiate early (before age 35) are associated with changes in driving. All three cohorts have school as their earliest life event, and in all cases school is associated with relatively less driving. All three also have a similar pattern of early full-time employment associated with increased driving. Yet only two—Couples and Have-it-alls—are characterized by early partnering, and for only these two is partnering associated with increased driving. Finally, among these three cohorts, only the Have-it-alls cohort is characterized by having children early, and only for this cohort is having children associated with increased driving.

Beyond the timing of events, mode use can also be affected by the relative order of events. For example, those in the Couples and Singles cohorts have children relatively late: when they first have children, members of both cohorts would largely be described as at a life stage of "middle-aged and living with a partner." However, from a life course perspective, those in the Couples cohort have lived with a partner for a much longer period than those in the Singles cohort have when they have children, and having children has minimal impact on their travel modes. However, for the Singles cohort, having children reduces the probability of regularly walking/biking and reduces multi-modality. In another example, those in the Singles, Couples, and Have-it-alls cohorts are employed relatively early in life (before their 30s), but regular public transit use only increases with full-time employment for those in the Couples cohort, who typically live with a partner but have no children before age 35.

The timing and order of life events can have lasting effects on mode use aggregated over entire life cycles. For example, only the Have-it-alls cohort has driving habits that are affected by all the life events considered here. This cohort generally follows a life trajectory of attending school → becoming employed → partnering up → having children, all before age 35. Car use increases for those in this cohort at each of these life stages after finishing school, and as a result they reach the highest level of regular car use (80%) earlier (by age 30) compared with all other cohorts; on average across the whole sample, the highest rates of regular car use of 70% do not occur until age 33. There is also suggestive evidence in our data that those in the Have-it-alls cohort may be relatively more likely to move to locations where public transit is less available when having children, which likely contributes to their dependence on cars. Conversely, those in the Singles cohort only increase their car use when becoming employed, and their car use is minimally affected by family formation later in life. As a result, those in the Singles cohort have relatively low car dependence accumulated over their life cycle. Those in the Late Bloomers cohort are the overall least dependent on driving, probably because they have the highest life cycle unemployment rate, which may reduce their need and/or resources to drive.

Our gender analysis suggests that women drive more when having children primarily when their family formation and career formation are intertwined early in life (i.e., only in the Have-it-alls cohort) and therefore are particularly time-poor. On the other hand, having children does not necessarily produce differential changes in regular mode use for women relative to men in cohorts characterized by children arriving later in life. This observation points to gender gaps for a more targeted subpopulation.

Lastly, we find generational differences in the choice of regularly used modes at the aggregate level and associated with heterogeneous life event effects across cohorts. In general, younger generations rely more on regular car use than older generations do during familial events when they have a late start to their careers. In contrast, in cohorts with careers starting before age 30, no significant generational difference in car use is estimated during these familial events.

Some limitations in our study suggest the need for further research. First, emerging transportation technologies and services such as ride hailing and micro-mobility expand the mode choice set and could affect both short- and long-term travel decisions. Although ride hailing data were collected in the life history calendar, the observations were limited owing to the short overlap between service availability and respondents' age range. A more focused data-collection effort on mode-use patterns of ride hailing and micro-mobility among subpopulations would be valuable. Second, similar to existing mobility biography literature (Table S1), our study focuses on a single city. Further data collection from more regions or application of a similar method to a different country would enhance the generalizability of the results. Third, the results around the two smaller cohorts, Later Bloomers and Family First, need to be further confirmed once data across survey samples with a wider spectrum of income and education levels becomes available.

Our results highlight the role of long-term life contexts on the dynamics of regularly used travel modes in response to life course events. The clear distinction in mode-use changes across the life trajectory cohorts signifies the importance of both past experiences and future expectations in mobility decision making. Understanding these differences can help planners and policymakers better understand the tendencies and constraints faced by different individuals and design policies to targeted subpopulations (for example, the Have-it-alls women) upon life events that are likely to have lasting effects on car dependency. This understanding might inform better predictions of mode use, and it might aid in designing policies related to commuting, public transit, or other travel behaviors.

ACKNOWLEDGMENT

This paper and the work described were sponsored by the U.S. Department of Energy (DOE) Vehicle Technologies Office (VTO) under the Systems and Modeling for Accelerated Research in Transportation (SMART) Mobility Laboratory Consortium, an initiative of the Energy Efficient Mobility Systems (EEMS) Program, under Lawrence Berkeley National Laboratory Contract No. DE-AC02-05CH11231. The following DOE Office of Energy Efficiency and Renewable Energy (EERE) managers played important roles in establishing the project concept, advancing implementation, and providing ongoing guidance: David Anderson, Prasad Gupte, Rachael Nealer, and Jake Ward. The authors would also like to thank Ted Kwasnik, Scott Carmichael, Terry Chan, and Morgan Faulkner for their help with technical and administrative support and data collection; Andrew Duvall, Margaret Taylor, Saika Belal, Gabrielle Wong-Parodi, Victor Walker, and Anand Gopal for contributing to the discussion surrounding the survey design and analysis plan; and Jarett Zuboy for technical editing.

REFERENCES

[1] U.S. Environmental Protection Agency., "Inventory of u.s. greenhouse gas emissions and sinks: 1990-2016. washington d.c. https://www.epa.gov/ghgemissions/inventory-usgreenhouse- gas-emissions-and-sinks-1990-2016."

[2] J. Zhang and V. Van Acker, "Life-oriented travel behavior research: An overview," vol. 104, pp. 167–178. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0965856417306912

[3] B. Clark, K. Chatterjee, and S. Melia, "Changes to commute mode: The role of life events, spatial context and environmental attitude," vol. 89, pp. 89–105. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0965856416303871

[4] J. Scheiner, "Why is there change in travel behaviour? in search of a theoretical framework for mobility biographies," vol. 72, no. 1, pp. 41–62. [Online]. Available: https://www.erdkunde.uni-bonn.de/archive/2018/why-is-there-change-in-travel-behaviour-in-search-of-a-theoretical-framework-for-mobility-biographies

[5] K. Chatterjee, B. Clark, and C. Bartle, "Commute mode choice dynamics: Accounting for day-to-day variability in longer term change," vol. 16, no. 4, number: 4. [Online]. Available: https://journals.open.tudelft.nl/ejtir/article/view/3167

[6] R. Schoenduwe, M. G. Mueller, A. Peters, and M. Lanzendorf, "Analysing mobility biographies with the life course calendar: a retrospective survey methodology for longitudinal data collection," *Journal of Transport Geography*, vol. 42, pp. 98–109, 2015.

[7] J. Guo, T. Feng, and H. J. P. Timmermans, "Time-varying dependencies among mobility decisions and key life course events: An application of dynamic bayesian decision networks," vol. 130, pp. 82–92. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0965856418309194

[8] N. J. Klein and M. J. Smart, "Life events, poverty, and car ownership in the united states: A mobility biography approach," vol. 12, no. 1, number: 1. [Online]. Available: https://www.jtlu.org/index.php/jtlu/article/view/1482

[9] A. T. M. Oakil, D. Ettema, T. Arentze, and H. Timmermans, "Changing household car ownership level and life cycle events: an action in anticipation or an action on occurrence," vol. 41, no. 4, pp. 889–904. [Online]. Available: https://doi.org/10.1007/s11116-013-9507-0

[10] J. Zhang, B. Yu, and M. Chikaraishi, "Interdependences between household residential and car ownership behavior: a life history analysis," vol. 34, pp. 165–174. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0966692313002469

[11] S. Beige and K. W. Axhausen, "The dynamics of commuting over the life course: Swiss experiences," vol. 104, pp. 179–194. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0965856416302609

[12] H. Rau and R. Manton, "Life events and mobility milestones: Advances in mobility biography theory and research," vol. 52, pp. 51–60. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0966692316000326

[13] B. Clark, K. Chatterjee, S. Melia, G. Knies, and H. Laurie, "Life events and travel behavior: Exploring the interrelationship using uk household longitudinal study data," *Transportation Research Record*, vol. 2413, no. 1, pp. 54–64, 2014.

[14] I. Salomon, "Life styles–a broader perspective on travel behaviour," *Recent advances in travel demand analysis*, pp. 290–310, 1983.

[15] M. Lanzendorf, "Mobility styles and travel behavior: Application of a lifestyle approach to leisure travel," *Transportation Research Record*, vol. 1807, no. 1, pp. 163–173, 2002.

[16] ——, "Key events and their effect on mobility biographies: The case of childbirth," *International Journal of Sustainable Transportation*, vol. 4, no. 5, pp. 272–292, 2010.

[17] A. T. M. Oakil, D. Manting, and H. Nijland, "Dynamics in car ownership: the role of entry into parenthood," vol. 16, no. 4. [Online]. Available: https://journals.open.tudelft.nl/ejtir/article/view/3164

[18] J. Prillwitz, S. Harms, and M. Lanzendorf, "Impact of life-course events on car ownership," *Transportation Research Record*, vol. 1985, no. 1, pp. 71–77, 2006.

[19] L. McCarthy, A. Delbosc, G. Currie, and A. Molloy, "'transit faithfuls' or 'transit leavers'? understanding mobility trajectories of new parents," vol. 78, pp. 105–112. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0967070X18308291

[20] S. Beige and K. W. Axhausen, "Interdependencies between turning points in life and long-term mobility decisions," vol. 39, no. 4, pp. 857–872. [Online]. Available: https://doi.org/10.1007/s11116-012-9404-y

[21] A. Busch-Geertsema and M. Lanzendorf, "From university to work life–jumping behind the wheel? explaining mode change of students making the transition to professional life," *Transportation research part A: policy and practice*, vol. 106, pp. 181–196, 2017.

[22] N. Morgan, "The stickiness of cycling: residential relocation and changes in utility cycling in johannesburg," *Journal of transport geography*, vol. 85, p. 102734, 2020.

[23] J. Prillwitz, S. Harms, and M. Lanzendorf, "Interactions between residential relocations, life course events, and daily commute distances," *Transportation Research Record*, vol. 2021, no. 1, pp. 64–69, 2007.

[24] J. Scheiner and C. Holz-Rau, "A comprehensive study of life course, cohort, and period effects on changes in travel mode use," *Transportation Research Part A: Policy and Practice*, vol. 47, pp. 167–181, 2013.

[25] G. Pollock, "Holistic trajectories: a study of combined employment, housing and family careers by using multiple-sequence analysis," vol. 170, no. 1, pp. 167–183. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1111/j.1467-985X.2006.00450.x/full

[26] Y. O. Susilo, C. Liu, and M. Börjesson, "The changes of activity-travel participation across gender, life-cycle, and generations in sweden over 30 years." [Online]. Available: https://doi.org/10.1007/s11116-018-9868-5

[27] J. Scheiner, "Gendered key events in the life course: effects on changes in travel mode choice over time," *Journal of Transport Geography*, vol. 37, pp. 47–60, 2014.

[28] J. H. Lee and K. G. Goulias, "A decade of dynamics of residential location, car ownership, activity, travel and land use in the seattle metropolitan region," vol. 114, pp. 272–287. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0965856417310510

[29] S. Haustein and M. Hunecke, "Identifying target groups for environmentally sustainable transport: assessment of different segmentation approaches," *Current Opinion in Environmental Sustainability*, vol. 5, no. 2, pp. 197–204, 2013.

[30] W. Li and M. Kamargianni, "Investigating the mode switching behavior from different non-car modes to car: the role of life course events and policy opportunities," *Transportation research record*, vol. 2673, no. 3, pp. 676–685, 2019.

[31] A. L. Alfeo, M. G. C. Cimino, S. Egidi, B. Lepri, A. Pentland, and G. Vaglini, "Stigmergy-based modeling to discover urban activity patterns from positioning data," in *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*. Springer, 2017, pp. 292–301.

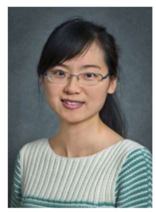
[32] J. Guo, T. Feng, and H. J. Timmermans, "Time-varying dependencies among mobility decisions and key life course events: An application of dynamic bayesian decision networks," *Transportation research part A: policy and practice*, vol. 130, pp. 82–92, 2019.

[33] B. Wang, S. Rasouli, H. Timmermans, and C. Shao, "Relationships between consecutive long-term and mid-term mobility decisions over the life course: a bayesian network approach," *Transportation research record*, vol. 2672, no. 47, pp. 159–170, 2018.

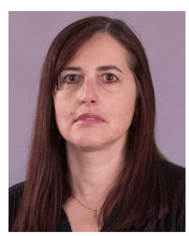
[34] M. De Haas, C. Scheepers, L. Harms, and M. Kroesen, "Travel pattern transitions: applying latent transition analysis within the mobility biographies framework," *Transportation Research Part A: Policy and Practice*, vol. 107, pp. 140–151, 2018.

[35] G. Ritschard and M. Studer, *Sequence analysis and related approaches: Innovative methods and applications*. Springer, 2018, vol. 10.

[36] C. A. Spurlock, J. Sears, G. Wong-Parodi, V. Walker, L. Jin, M. Taylor, A. Duvall, A. Gopal, and A. Todd, "Describing the users: Understanding adoption of and interest in shared, electrified, and automated transportation in the san francisco bay area," *Transportation Research Part D: Transport and Environment*, vol. 71, pp. 283–301, 2019.
[37] DOE, "SMART mobility: Mobility decision science capstone report." United States Department of Energy, Washington DC (United States), Tech. Rep., 2020.
[38] C. A. Spurlock, A. Todd-Blick, G. Wong-Parodi, and V. Walker, "Children, income, and the impact of home delivery on household shopping trips," *Transportation Research Record*, p. 0361198120935113, 2020.
[39] California Energy Commission, "2015-2017 california vehicle survey consultant report. CEC-200-2018-006."
[40] C. Calastri, R. C. dit Sourd, and S. Hess, "We want it all: experiences from a survey seeking to capture social network structures, lifetime events and short-term travel and activity planning," *Transportation*, pp. 1–27, 2018.
[41] A. Enam and K. C. Konduri, "Time allocation behavior of twentieth-century american generations: GI generation, silent generation, baby boomers, generation x, and millennials," vol. 2672, no. 49, pp. 69–80. [Online]. Available: https://doi.org/10.1177/0361198118794710
[42] J.-A. Gauthier, E. D. Widmer, P. Bucher, and C. Notredame, "Multichannel sequence analysis applied to social science data," vol. 40, no. 1, pp. 1–38. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9531.2010.01227.x/full
[43] P. Johnson, "Making social science useful," vol. 55, no. 1, pp. 23–30. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1111/j.1468-4446.2004.00003.x/full
[44] H. Lauder, P. Brown, and A. H. Halsey, "Sociology and political arithmetic: some principles of a new policy science," vol. 55, no. 1, pp. 3–22. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1111/j.1468-4446.2004.00002.x/full
[45] P. Wiles, "Policy and sociology," vol. 55, no. 1, pp. 31–34. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1111/j.1468-4446.2004.00004.x/full
[46] A. Lazar, A. Ballow, L. Jin, C. A. Spurlock, A. Sim, and K. Wu, "Machine learning for prediction of mid to long term habitual transportation mode use," in *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, pp. 4520–4524.
[47] A. Lazar, L. Jin, C. A. Spurlock, A. Todd, K. Wu, and A. Sim, "Data quality challenges with missing values and mixed types in joint sequence analysis," in *Proceedings of the Big Data (Big Data), 2017 IEEE International Conference on. IEEE, 2017.* IEEE.
[48] A. Lazar, L. Jin, C. A. Spurlock, K. Wu, A. Sim, and A. Todd, "Evaluating the effects of missing values and mixed data types on social sequence clustering using t-SNE visualization," vol. 11, no. 2, pp. 7:1–7:22. [Online]. Available: http://doi.acm.org/10.1145/3301294
[49] L. Lesnard, "Setting cost in optimal matching to uncover contemporaneous socio-temporal patterns," *Sociological Methods & Research*, vol. 38, no. 3, pp. 389–419, 2010.
[50] R. Piccarreta and F. C. Billari, "Clustering work and family trajectories by using a divisive algorithm," vol. 170, no. 4, pp. 1061–1078.
[51] M. Studer and G. Ritschard, "A comparative review of sequence dissimilarity measures." [Online]. Available: https://archive-ouverte.unige.ch/unige:78575
[52] R. Piccarreta, "Joint sequence analysis: Association and clustering," vol. 46, no. 2, pp. 252–287. [Online]. Available: https://doi.org/10.1177/0049124115591013
[53] L. Jin, D. Lee, A. Sim, S. Borgeson, K. Wu, C. A. Spurlock, and A. Todd, "Comparison of clustering techniques for residential energy behavior using smart meter data," in *Workshops at the Thirty-First AAAI Conference on Artificial Intelligence*. [Online]. Available: https://www.aaai.org/ocs/index.php/WS/AAAIW17/paper/view/15166
[54] M. Studer, "WeightedCluster library manual: A practical guide to creating typologies of trajectories in the social sciences with r." [Online]. Available: https://archive-ouverte.unige.ch/unige:78576
[55] C. Hennig and T. F. Liao, "Comparing latent class and dissimilarity based clustering for mixed type variables with application to social stratification." [Online]. Available: http://www.cnmd.ac.uk/statistics/research/pdfs/rr308.pdf
[56] L. Kaufman and P. J. Rousseeuw, "Partitioning around medoids (program pam)," pp. 68–125. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/9780470316801.ch2/summary
[57] R. Kitamura, "A dynamic model system of household car ownership, trip generation, and modal split: model development and simulation experiment," vol. 36, no. 6, pp. 711–732. [Online]. Available: https://doi.org/10.1007/s11116-009-9241-9
[58] T. Cresswell, *Gendered Mobilities*. Routledge, google-Books-ID: 0XMG-DAAAQBAJ.
[59] E. National Academies of Sciences, *Understanding How Individuals Make Travel and Location Decisions: Implications for Public Transportation*. [Online]. Available: https://www.nap.edu/catalog/23124/understanding-how-individuals-make-travel-and-location-decisions-implications-for-public-transportation



LING JIN has multidisciplinary training in air-quality engineering, statistics, and resource economics. She applies data science (statistical, machine-learning, and econometric techniques) to the domains of climate/atmospheric science, electricity markets, and transportation. She has led projects on air-quality modeling, spatial pattern and time series mining, and social sequence analysis, with work published by AGU, ACS, AAAI, IEEE, and ACM. She is currently a research scientist at Lawrence Berkeley National Laboratory. She holds a PhD in energy and resources and an MA in statistics from UC Berkeley as well as a BS in physical geography from Peking University.



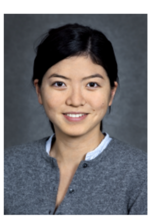
ALINA LAZAR received a BS in computer science from Western University of Timisoara, Romania, and a PhD in computer science from Wayne State University, Detroit, MI, USA. She is a professor of computer science in the Department of Computer Science and Information Systems at Youngstown State University. Her current research interests include applying machine learning to large scientific and software engineering related datasets.

JAMES SEARS is a PhD candidate in the Department of Agricultural and Resource Economics at UC Berkeley. He conducts research in environmental and energy, public, and health economics as well as applied microeconometrics. Some of his recent work focuses on behavioral responses to residential water conservation policies, the interactions between local and federal nutritional assistance programs, and the mobility and health impacts of COVID-19 stay-at-home policies.

ANNIKA TODD-BLICK is a principal scientific engineering associate in the Electricity Markets and Policy Department at Lawrence Berkeley National Laboratory. Annika is an experimental and behavioral economist and conducts research and analysis on energy efficiency, demand response, and smart grid topics, including Smart Grid Investment Grant (SGIG) dynamic pricing projects; evaluation, measurement, and verification of energy efficiency programs; and technical assistance to states on energy efficiency programs. Annika has a PhD in economics from Stanford University and holds a BA in molecular and cell biology as well as a BA in economics from UC Berkeley.

ALEXANDER SIM is currently a senior computing engineer at Lawrence Berkeley National Laboratory. He has authored and coauthored more than 250 technical publications and has released software packages under open source license. His current research and development activities include data modeling, data analysis methods, learning models, distributed resource management, and high-performance data systems. He is a senior member of IEEE.

KESHENG WU is a senior scientist at Lawrence Berkeley National Laboratory. He works extensively on data management, data analysis, and scientific computing topics. He is the developer of a number of widely used algorithms including FastBit bitmap indexes for querying large scientific datasets, the Thick-Restart Lanczos (TRLan) algorithm for solving eigenvalue problems, and IDEALEM for statistical data reduction and feature extraction.



HUANG-CHIA YANG is a senior scientific engineering associate in the Energy Efficiency Standards Department at Lawrence Berkeley National Laboratory. As a member of the Economic Research sub-team, Hung-Chia studies a wide-range of economic issues associated with the impact of energy conservation standards for the U.S. Department of Energy. In addition, Hung-Chia applies her data analytics skills to study the innovation process of renewable energy resources, the cost-effectiveness of zero net energy homes in California, the decision-making process around demand response in commercial buildings, and people's behavioral patterns around transportation decision-making. Hung-Chia holds a master's degree in economics from Duke University and a bachelor's degree in statistics from National Chengchi University, Taiwan.

C. ANNA SPURLOCK is an environmental and behavioral economist conducting research at Lawrence Berkeley National Laboratory. Anna co-founded the Behavior Analytics research initiative at the laboratory, in which advanced data science and machine learning are coupled with rigorous statistical and econometric techniques to generate insights from high-resolution smart meter data regarding consumer behavior and decision-making. In addition, Anna leads the Economics Research sub-team within the Energy Efficiency Standards Group. This team advises and supports the minimum efficiency standards program, and it conducts primary research on the market impacts of minimum efficiency standards policy. Anna is the principal investigator (PI) of the WholeTraveler research project under the U.S. Department of Energy's SMART Mobility Initiative. This project is an integral part of the Mobility Decision Science Pillar of the Initiative, focusing on a dynamic and holistic understanding of individual traveler behavior and decision-making around transportation. Finally, Anna is the pillar lead of the Mobility Decision Science Pillar, managing coordination of six to seven separate research projects with collaborators from five different national laboratories.