MethodsX 12 (2024) 102673

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/methodsx

Lifetime residential data collection protocol for the Adolescent Brain Cognitive Development (ABCD) Study

Paola Badilla[a], Shermaine Abad[b], Calen Smith[c], Brandon Tsui[b], Carlos Cardenas-Iniguez[d], Megan M. Herting[d,e,*]

[a] Institute for Behavioral Genetics, University of Colorado, Boulder, CO, USA
[b] Department of Radiology, University of California, San Diego, CA, USA
[c] Department of Psychiatry, University of California, San Diego, CA, USA
[d] Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
[e] Department of Pediatrics, Children's Hospital Los Angeles, Los Angeles, CA, USA

ARTICLE INFO

Method name:
Residential address history collection for geolinking exposures

Keywords:
Protocol
Addresses
Lifetime
Retrospective
Prospective
Adolescent Brain Cognitive Development Study
ABCD Study

ABSTRACT

Understanding the impacts of environmental exposures on health outcomes during development is an important area of research for plenty of reasons. Collecting retrospective and prospective residential history can enrich observational studies through eventual linkages to external sources. Augmenting participant health outcome data with environmental data can better inform on the role of the environment, thereby enhancing prevention and intervention efforts. However, collecting the geospatial information needed for this type of research can be difficult, especially when data are collected directly from participants. Participants' residential histories are unique and often complex. Collecting residential history data often involves capturing precise spatial locations along specific timeframes as well as contending with recall bias and unique, complex living arrangements. When trying to assess lifetime environmental exposures, researchers must consider the many changes in location a person goes through and the timeframes in which these changes occur, ultimately creating a multidimensional and dynamic dataset. Creating data collection protocols that are feasible to administer, result in accurate data, and minimize data missingness is a major challenge to undertake. Here, we provide an overview of the protocol developed to collect the lifetime residential address information of participants in the Adolescent Brain Cognitive Development (ABCD) Study.

Specifications table

Subject area:	Environmental Science
More specific subject area:	Environmental Epidemiology
Name of your protocol:	**Lifetime Residential History Data Collection**
Reagents/tools:	Not applicable
Experimental design:	Observational longitudinal cohort study
Trial registration:	Not applicable

(*continued on next page*)

* Corresponding author at: Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.
E-mail address: herting@usc.edu (M.M. Herting).

https://doi.org/10.1016/j.mex.2024.102673
Received 15 January 2024; Accepted 22 March 2024
Available online 3 April 2024

Ethics:	Adolescent Brain Cognitive Development (ABCD) Study procedures are approved under a centralized institutional review board from the University of California, San Diego; each study site also obtained approval from their own institutional review boards. All parents or caregivers provided written informed consent; children provided written assent.
Value of the Protocol:	• Protocol outlines a process through which users without GIS knowledge can easily collect residential history data • Highlights how to handle challenges related to collecting temporally and spatially detailed information on residential history data • Emphasizes the utilization of interactive tools to streamline and enhance staff and family interactions for increased ease and efficiency

Description of protocol

Introduction

The exposome refers to the totality of environmental exposures an individual encounters throughout their lifetime and encompasses external factors such as air quality, socioeconomic status, psychosocial stressors, and chemical and biological agents [22]. With the plethora of external datasets available, researchers can use the spatiotemporal aspects of participants' residential histories as a basis for measuring environmental exposures [10]. Compiling residential histories is a complex task made difficult by both complex life experiences and the participant's ability to recall precise addresses [3]. Other studies utilize health record address histories [24], public records databases [5,12,13,21], or other administrative data [14]. However, relying solely on record or administrative data may introduce bias because data availability is often dependent on age, race/ethnicity, income, and geography [5,11,21,23]. Some studies have collected residential addresses directly from participants, but only for a particular time point, such as interview date, birth, or death [6,7,16,18,20]. However, this method does not account for residential mobility which can result in exposure misclassification [4].

The Adolescent Brain Cognitive Development (ABCD) Study is a national, observational study aiming to understand the factors that contribute to the neurodevelopment of children as they grow from adolescence into young adulthood [19]. Residential history for the ABCD was originally collected for the linkage to the baseline event [8]. To estimate lifetime exposure data for each participant across various domains, the residential history standard operating procedure (SOP) underwent a revision process to gather extensive residential history data. Longitudinal data collection methods from other studies were examined for their potential adoption (Table 1). However, as the complexity of capturing thorough residential histories grew apparent, it became obvious that a tailored solution was necessary.

The original address collection protocol is briefly outlined in [8]. Here, we share an expanded and revised description of the protocol created and adopted by the ABCD Study in April 2023 to capture both retrospective and prospective residential address information from the caregiver. It is important to note that the protocol may not be exactly reproducible since the database interface used in the ABCD Study, known as the Personal Identifiable Information Database (PIID), is not publicly available software because it was created uniquely to fit the ABCD Study design. While the ABCD PIID is not readily available, it was created using open-source frameworks like Django using Python, JavaScript and publicly available APIs. These open-source tools mean that the logic and design of our SOP and PII interface can be adapted and tailored to other studies. The current protocol of collecting residential information, handling special circumstances, and minimizing missing data can be used by other studies planning to collect residential history information of research subjects, and it can serve as a guide for designing a study's database interface.

Table 1
Examples of longitudinal residential data collection methods.

Source	Description	Limitation for use in ABCD
ABCD PIID 1.0 [8]	Electronic form to collect free-text data based on calendar year	Did not allow temporal specificity or ease in collecting longitudinal address history; post hoc geocoding
Health and Retirement Study [1]	Paper form to collect street address information at a yearly resolution	Does not allow for collecting concurrent addresses; post hoc geocoding
NIH CDE Risk Factor Questionnaire (RFQ) [9]	Electronic form to collect city, state, zip, country and duration of residence	Does not collect temporal data needed for precise spatiotemporal linkages; post hoc geocoding
Hanford Thyroid Disease Study [2]	Paper form to collect address data by time period	Does not allow for concurrent addresses; post hoc geocoding
BC Generations Project [17]	Online form to collect street address information at a yearly resolution, proportion of year spent at address	Yearly resolution lacks temporal specificity; post hoc geocoding
Health Cost and Service Utilization Study [15]	Guided questionnaire to collect zip, city and state information from respondents	Temporal specificity tied to health event, lacks geographic and temporal specificity; post hoc geocoding

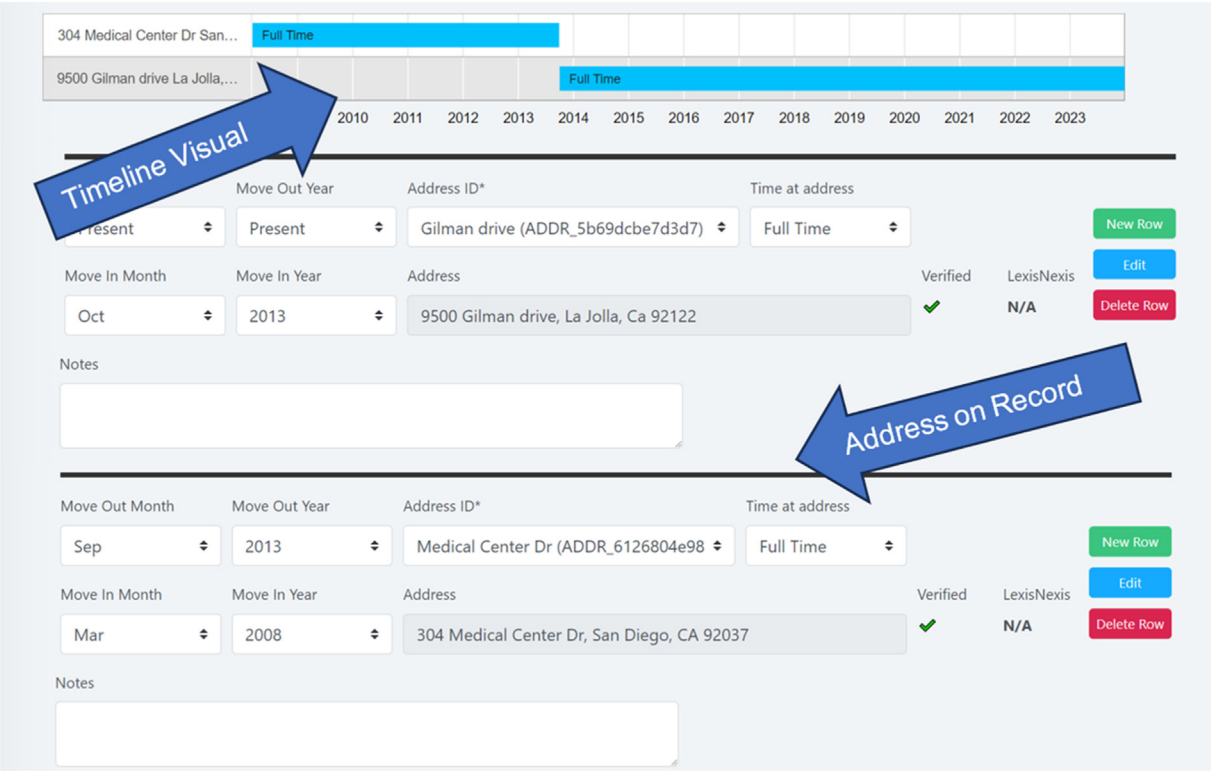

Fig. 1. PIID Interface layout. For any given participant, the interface of the residential history record includes both a visual timeline as well as each address entry provided by the informant.

Protocol description

Context of the protocol

The purpose of collecting address history for a participant in ABCD is to assign information regarding the environmental context surrounding the child's residential locations, such as number of parks or sources of pollution, etc. This allows us to gain a better understanding of the neighborhoods that participants grow up in and how that may impact various health and behavioral outcomes. Because the ABCD Study participants are minors, the residential history information is collected from a guardian or caregiver. The person who is providing information about the youth is referred to as an "informant". The informant is usually a biological parent of the child, but could be an adoptive parent, custodial guardian, or an extended family member who cares for the child. The informant is asked to provide all addresses that the youth has lived at since birth for any amount of time greater than 2 months. We deliberately chose a 2-month minimum based on our previous anecdotal experiences that taught us that capturing every single possible address (i.e. vacations addresses, overnight camps, or transitional housing) was unfeasible from both the data collection side as well as from an informant's recall standpoint. Moreover, the goal of the residential information is to link to many social and physical environments surrounding the home that are mainly static for longer periods of time (i.e. annually).

Data collection interface

Since residential address information is identifiable, the ABCD Study uses in-house software (PIID System) to secure the collection and storage of residential information. Only ABCD Study research staff with IRB training and study approval are given access to this platform to collect these data while working alongside the informant. Participants and informants do not have access to data collected in PIID but may provide visual verification of entered data when working with research staff. Thus, the PIID interface is exclusive to the ABCD Study. It is not publicly available software. However, the design of the PIID interface can serve as a guide for other studies when they are creating their own data collection interface.

The PIID interface has two components: a visual timeline of addresses as well as an Address Entry for each residential location for a given child as provided by the informant (Fig. 1)

A. Visual Timeline: The PIID interface compiles the address entries into a visual timeline to easily indicate to the research staff if there are any errors in the entered data or periods of missing addresses for the participant.

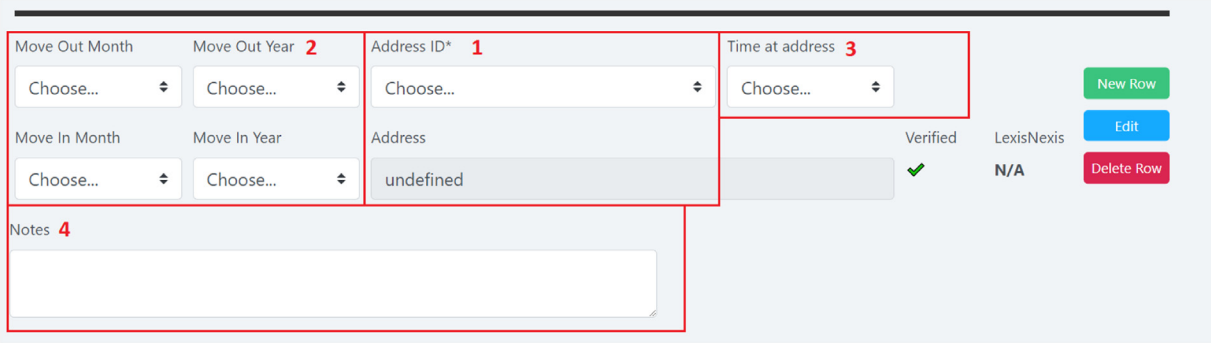

Fig. 2. Data collection fields in PIID system for a single residential address for a specific participant. Each field must be entered to create a complete address entry per participant.

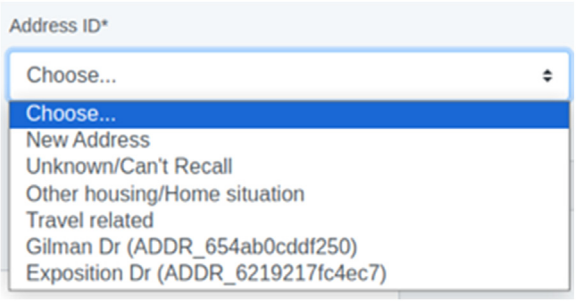

Fig. 3. Options listed in the "Address ID" drop-down menu. This field includes either adding a new address entry, an existing Address ID for a specific address entry, or options for special circumstances (see details below).

B. Address Entry: An address entry is a compilation of information gathered about the address the participant lived at. As needed, each Address Entry is listed in the residential history record of the PIID in chronological order from most recent residential address to the least recent location reported.

In the section below, we outline the information collected for each Address Entry provided by the informant, followed by a more thorough description of how the Visual Timeline integrates information about each Address Entry to help identify gaps in the residential history of a given participant.

Information collected

Each address entry and its associated details are stored in the PIID system as one entry (Fig. 2). The goal is to gather information on: 1) Where the individual lived; 2) When they lived there; and 3) How much time did they spend at that location? This is repeated for as many addresses as the participant resided across their life until the current day of assessment. Each entry is considered 'complete' when the following information is collected in full:

1. Address ID and Verification Markers: A randomly generated ID given to a unique address or most approximate location the participant resided in, or special circumstance regarding residence. This also includes a section for:
   - Address: The automatically generated geocoding ID to validate location.
2. Move In and Move Out Month and Year: The month and year when the participant moved in and moved out of the address.
3. Time at address: Amount of time the participant was living at the address (i.e. Full time, part time, etc.)
4. Notes: space to include additional information about the entry.

For consistency in data collection across all sites, each field in the entry is clearly defined and documented for reference. We now cover each of these inputs in more detail.

Address ID and verification markers

To understand where the individual lived, each Address Entry has an Address ID and built-in verification markers to help reduce human error. The Address ID field is a drop-down menu meant to denote the type of residential circumstance of the youth. This can either be an address the youth currently or previously resided in, or a special circumstance (Fig. 3). An address ID is randomly assigned to each address the caregiver reports for a specific period of time, and then it is added to the drop-down menu as an option.

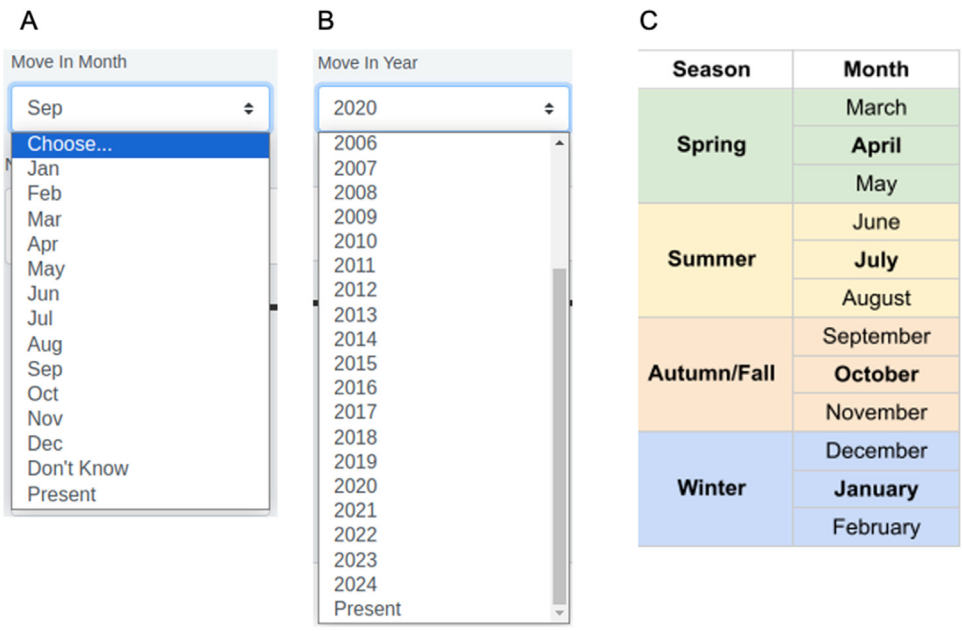

Fig. 4. **(a)** Move In Month, **(b)** Move In Year, **(c)** Seasonal chart.

When a new address is added to the residential record of the participant, the address is geocoded with the integrated Google Place API (see additional details below). It is important that the address is verified through Google Place API because it ensures the address has a latitude and longitude to then link to environmental exposures. Once the address has been geocoded, it is assigned a unique address ID within the database and attached to the participant's record. When a participant is experiencing a special circumstance that cannot be denoted by an address, the following options can be used accordingly:

- Unknown/Can't Recall: The informant does not know the address information or is unable to find this information.
- Other housing/Home situation: The youth is experiencing a period of time without a home or a place to stay or live.
- Travel related: The youth was traveling during the selected period of time and did not have an address.

In the PIID interface, the "Unknown/Can't recall" or "Travel related" options will not require an address to be entered and thus will not have a geocoded verification. Alternatively, the "Other housing/Home situation" option will prompt the research staff to drop a pin on a Google Place API to signify a location at which the participant is most often around in the case where the participant is not residing at any specific address but can provide an approximate location for where the youth was most of their time.

This selection will also verify the location by creating a geocode ID, which allows the location to be used in research on environmental exposure. Additional details of when this option has been found to be especially useful can be found in Section 7: Troubleshooting Google API in real-time.

Move-in/move-out time fields

After an Address Entry has an Address ID specified, it is essential to know when in the child's life they lived there. To accomplish this, there are move-in and move-out fields for each Address Entry. These fields are meant to capture the period of time that the youth lived at the address (Fig. 4a, b).

- Move In Month: Month when youth started living at the address.
- Move In Year: Year when youth started living at the address.
- Move Out Month: Month when youth stopped living at the address.
- Move Out Year: Year when youth stopped living at the address.

If youth is currently living at that address, the move-out month and year are entered as "Present". On the other hand, if known, the first address that the participant lived at should have the birth month and birth year as the move-in month and year. It is often the case that the informant does not remember the exact month the participant moved in or moved out of an address. In this case, it is helpful to broaden the time span from a specific month to a season to help the informant narrow the time window to a general time of year. To implement this strategy, select which seasons or quarters of the year you want to use and classify which months compose each season or quarter. The ABCD Study chose to use the common seasons and defined each one as listed in the chart in Fig. 4c, with the **bolded** month being the month that is entered in the Move-In or Move-Out field accordingly. With this definition in mind, the informant is asked if they remember which season the youth moved into or out of this address. For example, if the informant says

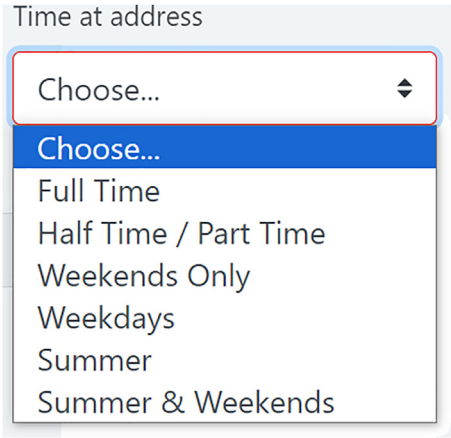

Fig. 5. Options for the "Time at address" drop-down menu.

youth moved into the address in "summer", the research staff would enter "July" for the month. It can also be helpful to prompt the informant: "Do you remember if your child moved in/out during the school year, near the holidays, or if it was particularly hot or cold outside?" This prompt can help the informant recall a narrower window of time from which to pick a month.

There are many complex scenarios in which the informant will not be able to recall or does not know timeframes for when youth lived at a specific address. This can be the case when parents are divorced and do not know each other's housing situation, the informant is not the main caregiver and does not know this information, or if the informant is an adoptive parent and does not know the youth's residential history prior to adoption. In these cases, the option is available to enter "don't know" in the month and year fields to reduce the recall burden on the informant and research staff.

Lastly, the move-out time of one address should match up with the move-in time of the following address. Usually, it is the case that a family will move out of an address at the end of a month and move into the new address at the beginning of the following month. For example, if a family moved out of address "A" in April and moved into address "B" in May, the "move-out" month for "A" should be "April" and the "move-in" month for "B" should be "May". If the family moved out of an address and into the next address within the same month, the move-out month and move-in month are still entered as subsequent months to capture the first *full* month that a participant lived at the address. However, other studies can choose to handle this situation differently based on what information they are interested in capturing.

Time at address

After understanding where and when a child resided, it is useful to capture what percentage of time the participant resided at that address. For the ABCD Study's purpose, this is useful information to have in order to better understand how much the participant was exposed to environmental factors of interest. This is important because there are a lot of complex living arrangements among families (i.e. separated or divorced caregivers, additional family support), and it might be useful to more accurately characterize their environments based on the proportion of time they spent at a given location. Of course, the options and definitions a study chooses for this field can vary and should reflect the type of data the study is aiming to collect. Through trial and error, we found that having individuals report a percentage of time at each location was not useful. Often the percentages across locations did not add up to 100% and the percentages felt arbitrary. Thus, we chose for the options in this field to be defined in general terms so that an informant can more easily categorize the time that the participant typically spends at a given address (Fig. 5).

Each option from the drop-down menu is defined as follows:

- **Full time**: Participant lived at this address for over 80% of their time, meaning the participant did not reside at another address concurrently.
- **Half time/Part Time**: Participant lived at this address in a half-time arrangement. This category is meant to capture living arrangements where the child splits their time between multiple addresses. Common examples of split time are 40/60, 50/50, 30/70.
- **Weekends Only**: Participant lived at this address only on the weekends (2 or 3 days of the week).
- **Weekdays**: Participant lived at this address only on weekdays (5 days of the week, more than half of the week, or "school days", including "after school").
- **Summer**: Participant lived at this address for only the summer months. The "summer" season varies by region, and can be defined as (but not limited to):
- June - August
- When school is out of session (May - August, June - September)
- **Summer and Weekends:** Participant lived at this address for the summer months and the weekends.

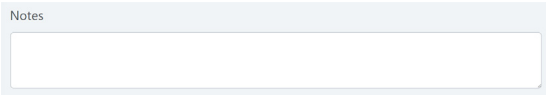

Fig. 6. Notes section for the entry.

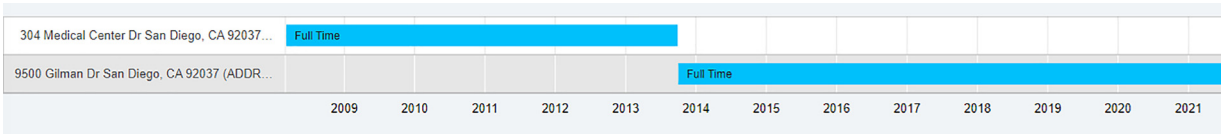

Fig. 7. The visual timeline graph.

Some of these definitions may seem redundant or unnecessary. For example, if the move in/out months are June through August, then there is no need to indicate that the time spent at that address is "Summer". However, there are complex living arrangements in which the child is at an address "Full time" EXCEPT for weekends or summer months. This is different from a situation in which the child is residing at multiple addresses for an evenly split amount of time throughout the year. While the summer is only a quarter of the year, it is important to maintain the distinction between the youth only being at an address for the summer months versus only being at an address a quarter of their time in general (like one week every month). For the ABCD Study, these details can become important when trying to better estimate exposures that follow certain seasonal patterns (i.e. air pollution, climate events, etc.)

Notes

Because there will always exist a situation that cannot be easily captured by the PIID Address Entry interface, there is a "Notes" section that can be used to document any details that the existing fields do not suitably capture (Fig. 6). Each entry has an optional section for notes to add any relevant information. For example, it can be used to include notes on why an address is missing for the entered period of time.

In ABCD, the research assistants are instructed to enter notes following a standardized format: [DATE] [Initials] [NOTE DETAILS].

Visual timeline

The PIID interface compiles the Address Entries for a given participant into a graphical timeline (Fig. 7). Specifically, the respective calendar years of the child's life are shown on the x-axis, whereas each unique Address Entry is listed on the y-axis of the timeline. Bars appear showing the time in which the child lived at each address as well as the time spent at that location. In the example in Fig. 7, the individual lived at 2 residential locations; 1 from 2008 through most of 2013, and then subsequently moved to a second location in which they currently reside. In both cases, the child was reported to be full-time.

This visual timeline was developed to help the research staff and informant verify the validity and completeness of the Address Entries on record. If there are any periods of time that are not accounted for in the record, it will highlight this period in yellow and indicate that there is missing data for a specified time period (Fig. 8a). Additionally, it will sum the amount of time that the participant spent at each address for every month and year of the timeline. In doing so, it will notify the researcher if the calculation exceeds 100% of the time (or "Full-Time") by highlighting these entries in red as well as outlining the Address Entries associated with this error (Fig. 8b, c). Issues with overlapping addresses occur when an address that is marked as "Full Time" overlaps with another address that is marked as either "Half Time" or "Full Time" for the same period.

Data collection process

The current protocol was under development starting in the fall of 2022 and it was deployed in April 2023, which corresponds to the 4 and 5th year study visits. The research staff at ABCD have been trained to review the residential history record of the participant with the informant at every yearly follow-up appointment. During this time, the research assistant verifies all address entries are complete and adds new Address Entries as needed for periods of missing information.

Confirming address entries are complete

To verify an address entry listed in the residential history is complete, the research staff will check if each of the following fields is filled:

1. Address ID and Address
2. Move-in and Move-out month and year
3. Time spent at that address
4. Google Place API verification

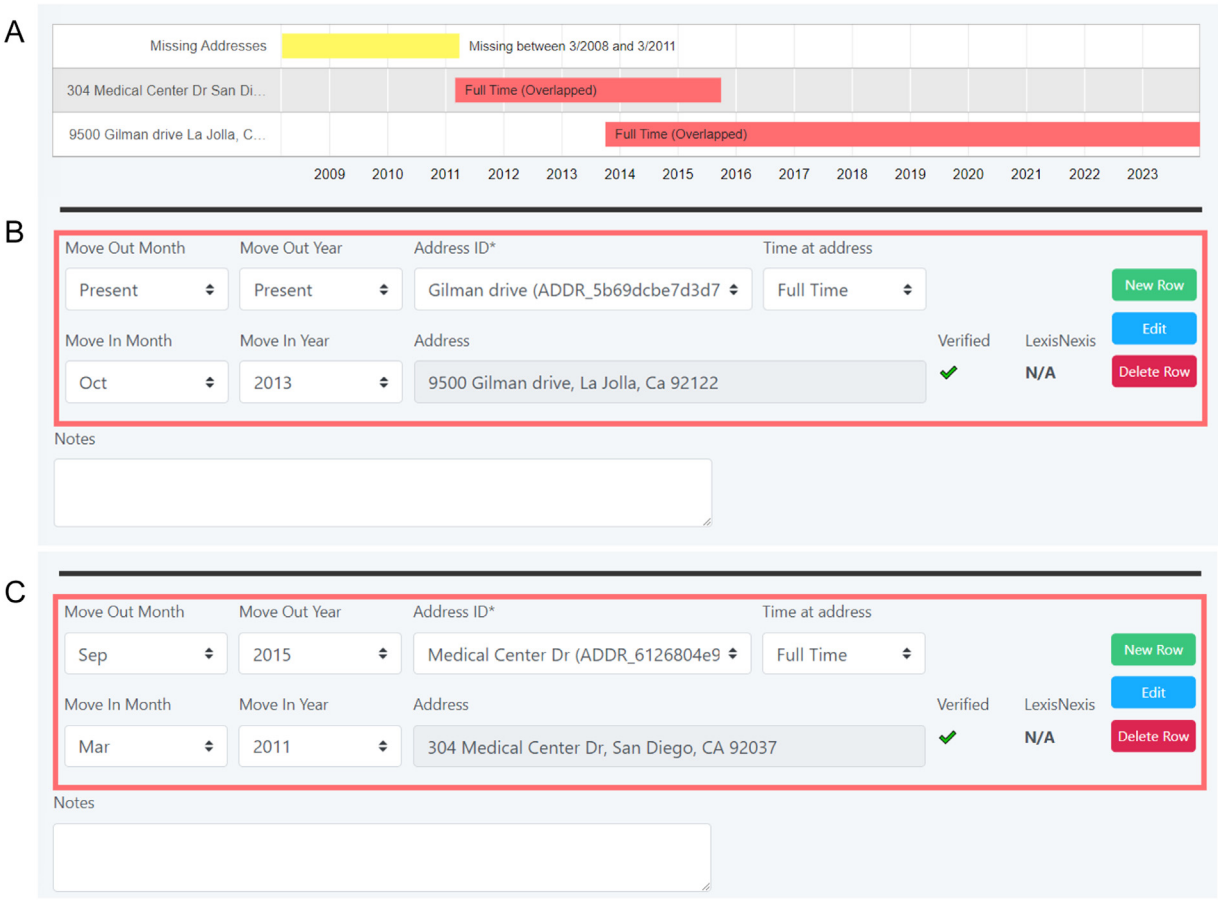

Fig. 8. **(a)** Address entries are missing for the specified time period indicated by the "Missing between" description in yellow. The timeline also displays the address entries that are overlapping and sum to greater than "Full Time" in red. **(b,c)** The address entries that overlap in the visual timeline are also outlined in red on the respective Address Entries to signify data validity issues.

Fig. 9a is an example of what an address entry looks like when it contains all the required information. The research staff review a complete address entry with the informant to ensure data correctness. An example prompt for how the research staff would review the information in Fig. 9a is: "*We have an address entry on record listed as* 304 Medical Center Dr, San Diego, California, 92037 *from* September 2014 *through* September 2015. *The participant was at this address* Full Time.*"* The informant can then confirm this information is correct and the research staff continues to review the next address entry.

As previously mentioned, when an Address Entry is missing information for a given field, the PIID will outline the field in red (Fig. 9b). To fill in missing fields, the research staff will ask the informant to recall this information. An example prompt for how the research staff would collect the information missing in Fig. 9b is: "*We have an address entry on record listed as* 9500 Gilman Drive, San Diego, California, 92093. *We are missing some information regarding this address. Do you remember which month in* 2013 *you moved into this address and which month in* 2014 *you moved out? Was the participant living at this address:* Full-Time, Part-time, Weekends, etc.?" The research staff can then use the additional info to enter the information provided, as well as use the additional protocol on identifying a month based on season if the informant is having difficulty remembering the exact month.

A built-in real-time Google Place API verification

If the address is missing the Google Place API verification, the address entry will have a red "✖" under the "Verified" section (Fig. 9c). The research staff can verify the address through Google Place API by selecting the "Edit" button next to the Address Entry (Fig. 9c, blue button). The "Edit" button will prompt the researcher to enter the address into the Google Place API and automatically create a geocode ID for the address. Once the address has an associated geocode ID, the "Verified" field will show a checkmark "√", and the entry will look like the one in Fig. 9a, b. It is important to note that if Google Place API is used in real-time from the conception of data collection, there won't be a need to retrospectively verify addresses. This is important as initially collecting address information on a text-based form led to addresses provided not being found using Google Place API (i.e., largely due to typos), and thus created a need for verifying addresses retrospectively.

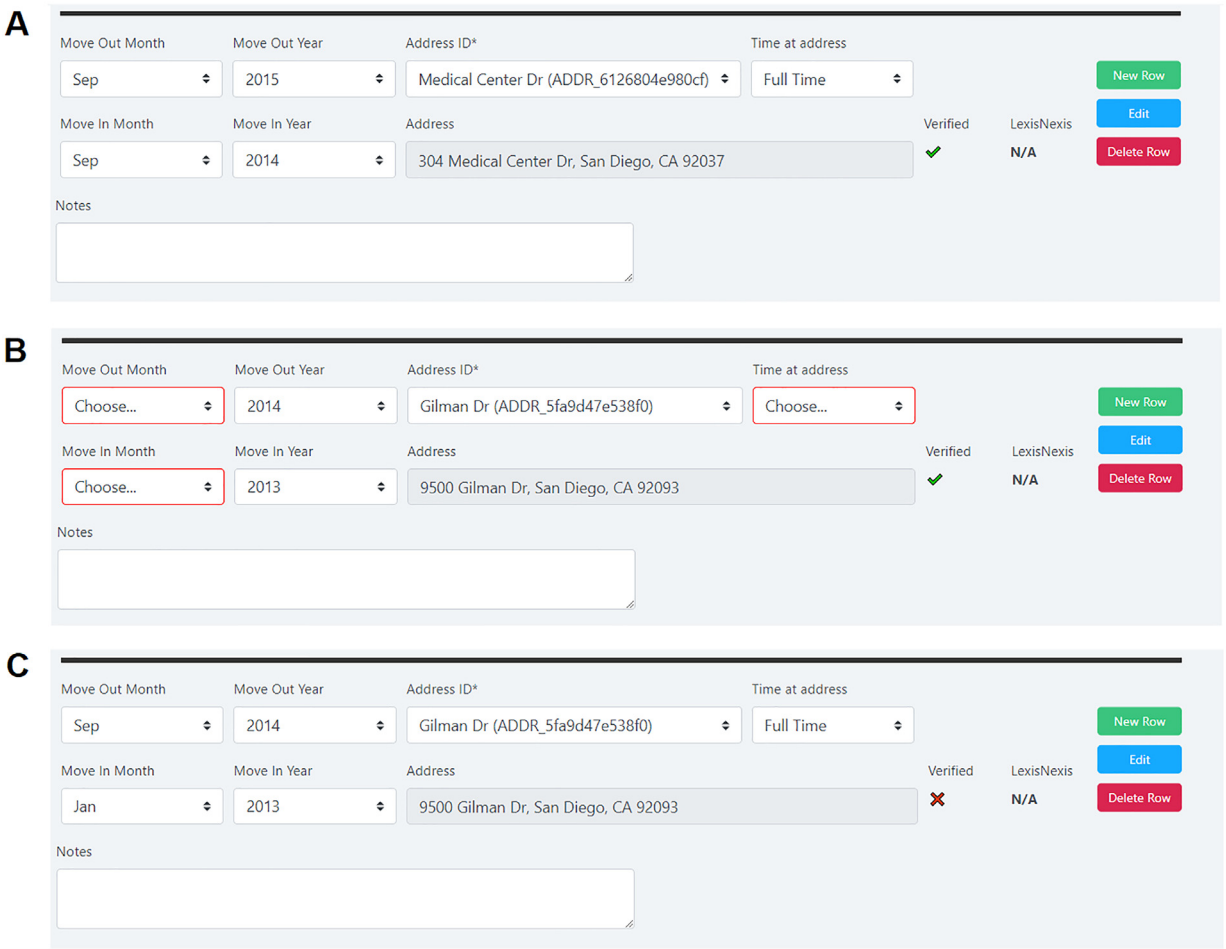

Fig. 9. (a) Top: Address Entry is completed. **(b) Middle:** Address Entry is missing move-in and move-out months, as well as the time at address. These fields are outlined in red. **(c) Bottom:** Address Entry without Google Place API verification.

Error pop-up notifications

A crucial component of the PIID interface during the review of residential addresses with an informant is the error pop-up feature (Fig. 10). This feature was put in place to notify the researcher if there are any errors or missing information in the record. When the researcher clicks the "Save" button at the bottom of the record to save any changes made, the PIID interface will check that all required fields are completed and create a pop up like the one in Fig. 10a if anything is missing.

The PIID interface will also verify the data entered is consistent with the other entries in the record. Because the entries are compiled into one visual timeline graph, the interface will check that the move-in/move-out fields for each entry line up in time (Fig. 10b). Additionally, the interface will check that no entries overlap for an amount of time greater than "Full-time" (Fig. 10c). This ensures that there are no periods of time that are missing addresses and that the amount of time that the participant lived at an address does not exceed 100%. If issues arise with either of these checks, an error message will appear.

Lastly, the error pop-ups have an "OK" button to prompt the researcher to acknowledge the error message. It may seem like a minor detail, but this feature helps indicate if any issues in the record were intentional by the researcher instead of an oversight. This allows for greater confidence that missing information or complicated residential data for a given participant is not a function of human error. Moreover, the different pop-up features are aimed to try to help quickly inform research staff where to focus if a data entry error has occurred.

Adding address entries for periods of missing information

As previously mentioned, the visual timeline graph at the top of the PIID interface will clearly display any periods of time which are missing Address Entries for a given participant (see Fig. 8a). To fix this issue, the research staff must add new Address Entries to fill-in and accurately account for the period of time indicated in the timeline.

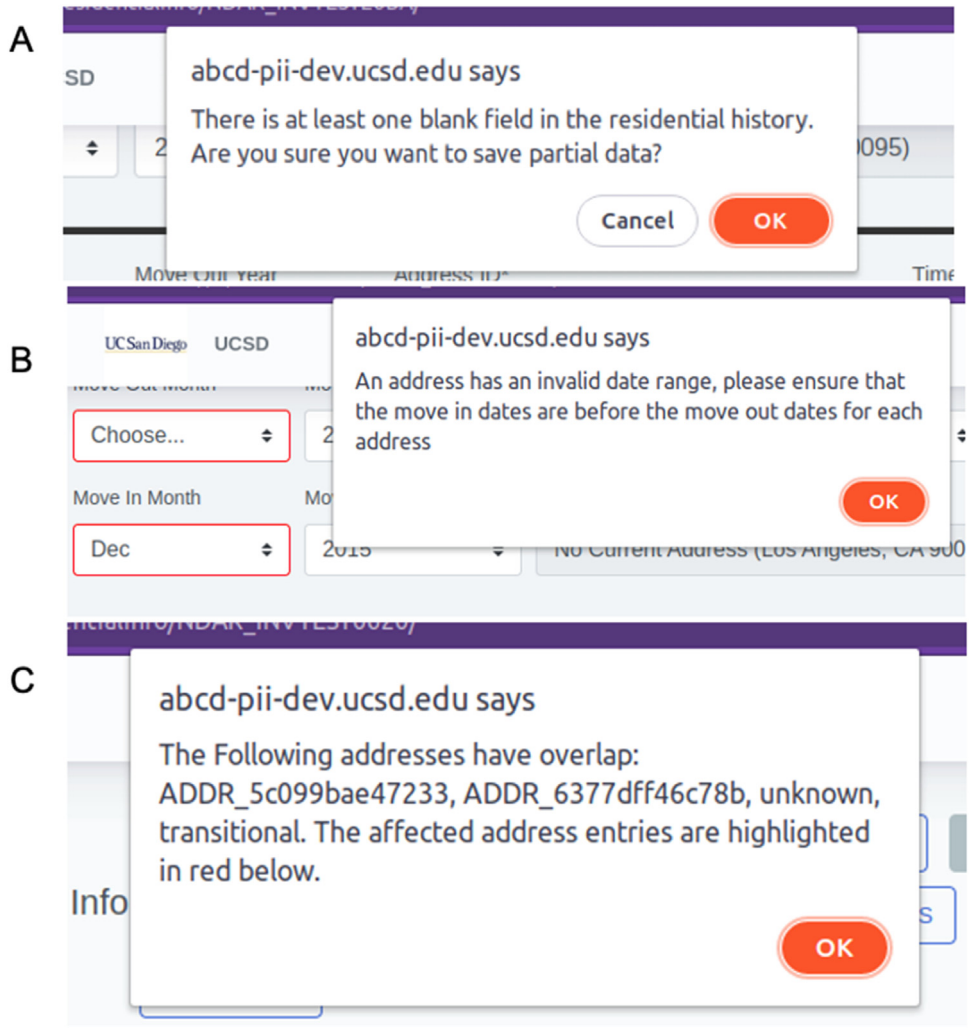

Fig. 10. **(a) Top:** An example of the pop-up that appears when a field is missing information. **(b) Middle:** An example of the pop-up that appears when the move-in and move-out months don't align as expected. **(c) Bottom:** An example of the pop-up that appears when Address Entries overlap and sum to greater than 100 % of time.

AddressID	Street Number	Street Name	Apartment	City	State	Zipcode	LexisNexis	Google PlaceID	Linked
ADDR_6126af4aec477	9500	Gilman Dr	None	San Diego	CA	92037	None	ChIJJ7VRwCcH3IARpOh7CRkO7YQ	Yes
ADDR_6126804e980cf	304	Medical Center Dr	None	San Diego	CA	92037	None	ChIJJY0jn9kG3IARsjrjSbqs6wQ	No

Fig. 11. In linking an address to the participant's record the "Linked" button updates from 'no' to 'yes'.

To add an address entry to the record, the researcher can select the "New Address" option from the "Address ID" drop-down menu. This will take the researcher to the list of all addresses the PIID interface has stored for all participants. It is possible the address of interest exists in the database already because ABCD participants can have siblings, cousins, or other family members in the study. Since those relatives participating in the study would also have a residential history record, addresses in their record exist in the database and can be added to other participant records if relevant. The researcher can search for the address in the database to avoid creating a duplicate address and reduce unnecessary steps involved in creating the address. If the address exists in the database, it can be "attached" to the participant's record by clicking on the button under the "Linked" column of the database (Fig. 11). The button will change from an orange 'no' to a green 'yes', and the Address Entry will now appear in the participant's record. Being able to link

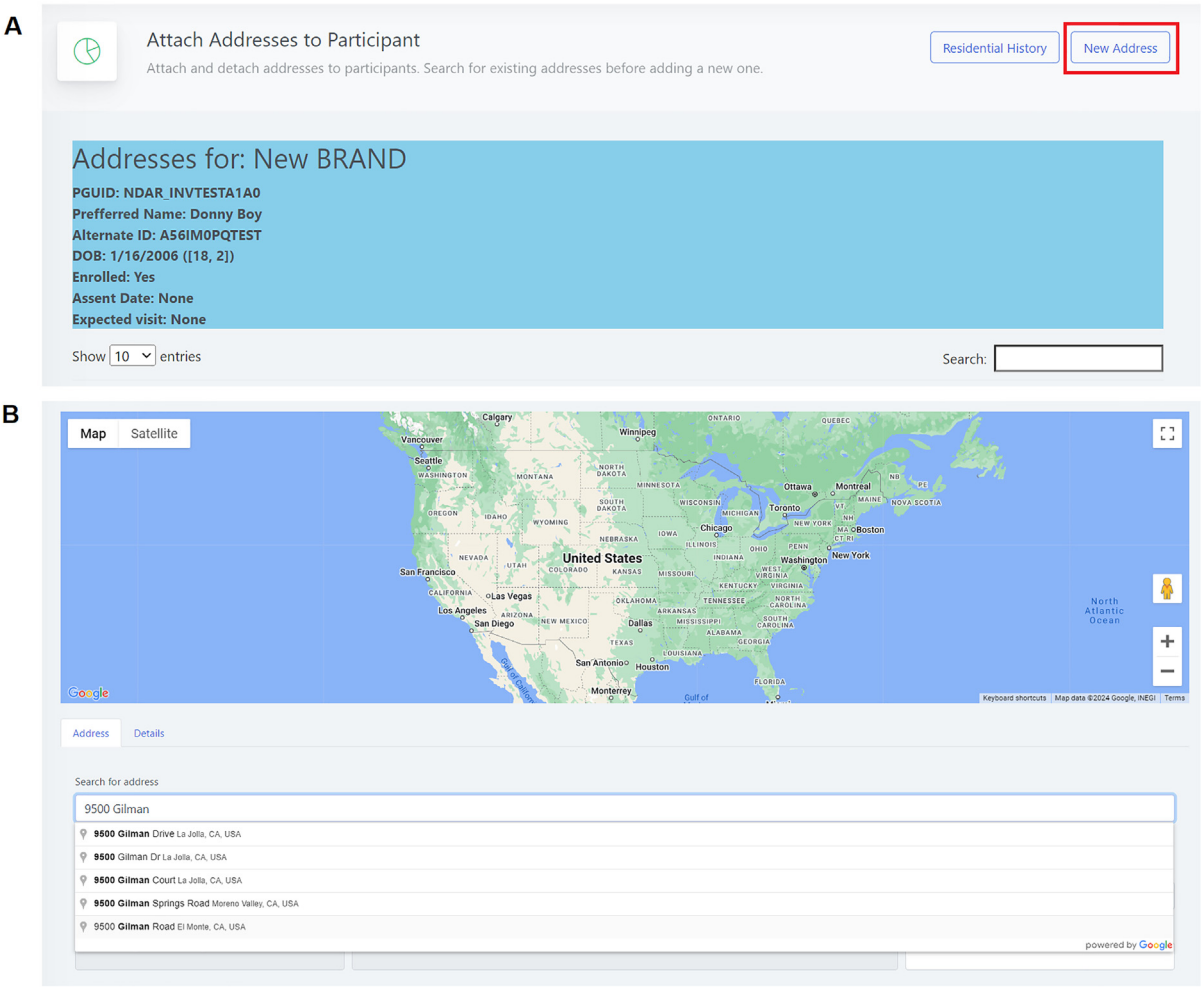

Fig. 12. **(a) Top:** Button layout of the PIID Interface to add a new address. **(b) Bottom:** Google Place API in the PIID interface to add a new address. As the address is being typed into the search bar, a list of addresses appears in a drop-down menu.

an address that exists in PIID to a participant is useful to prevent duplicate addresses from being created as well as reduce research staff burden.

If the address does not already exist in the PIID interface, the researcher can create a new address which will automatically link to the participant's record after creating it. Instead of searching for the address in the database, the researcher can click on the "New Address" button at the top right of the PIID interface page (Fig. 12a). This will prompt the researcher to enter an address onto the Google Place API within PIID (Fig. 12b). As the address is being typed into the search bar, a list of similar addresses will appear in a drop-down menu from which the researcher can choose (Fig. 12b). Once an address is selected, the Google Place API will generate a geocode ID for the address, and the researcher can click on "Save" to finish linking the address to the participant's record. After linking the address to the record, the research staff will then be prompted to complete the address entry by filling out the move-in/move-out fields and the "Time at address" fields (see Section Information Collected).

Special circumstances

Living situations can be highly complex and difficult to document. Understanding the variety of circumstances that a participant can experience is crucial not only for creating a catch-all system for collecting residential history information but simplifying the data collection processes for informants.

Handling approximate locations

A common scenario that happens is the informant is able to recall a nearby location, but not the exact address that the participant resided in. Since the ABCD Study started once the youth were 9 or 10 years of age, there were 9 or 10 years of residential history to recall at the time the study began. It is possible that the informant can remember the street and city, but maybe not the house or

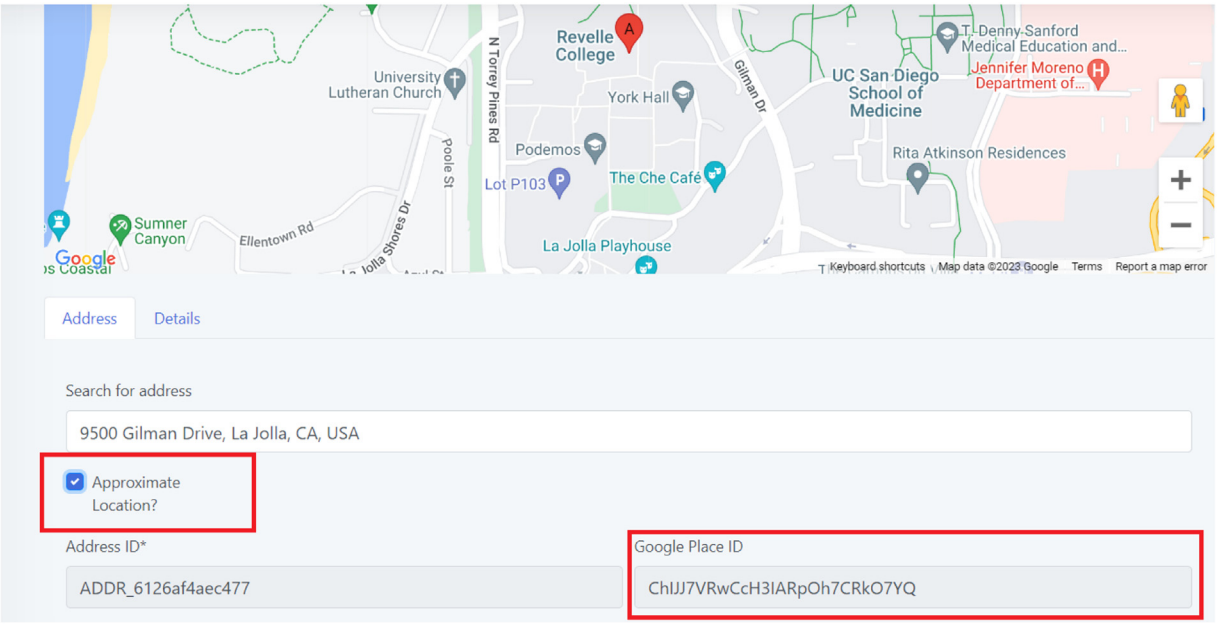

Fig. 13. The approximate location checkbox when marking a nearby location. The location will generate a geocode ID.

building number. In this case, being able to identify a section of the street the participant resided in or even an intersection of the city is enough to generate a geocode ID for the entry. With this in mind, the PIID interface integrated a checkbox that researchers can click to indicate that the address entered on the record is an *approximate location* to where the participant resided. This checkmark can be selected when adding a new entry to the record and it will create an address with a geocode ID (Fig. 13). Approximate locations become extremely useful in troubleshooting several complex situations. For additional details, see Section: Troubleshooting Google API in real-time.

Handling housing instability or varying housing locations

One situation that can be difficult to document is housing instability. If a participant is, for example, sleeping in their car parked on a road or seeking shelter in various community locations, the location may change from night to night. For these situations, there likely won't be a single address to note for where they reside. In an attempt to capture this situation, the researcher can select the "Other housing/home situation" option from the "Address ID" drop-down menu. This option will prompt the researcher to drop a location pin on the Google Place API to identify the area that the participant was around for a considerable amount of time (Fig. 14). This does not require the researcher to type in an address. Instead the researcher can navigate the map to the location that the participant indicates. Latitude and longitude are calculated, and the location is added as an entry on the participant's record. While additional notes will be helpful to determine if this location can be useful for understanding environmental context, having this noted on the record is important in that if this period is left as an "empty" period on the residential visual timeline, study staff might risk thinking it may be missing data and end up asking an informant multiple times about this period of time, which may bring undue negative consequences and burden to the informant.

Handling unknown addresses

Another situation that can arise when collecting residential history information is when the informant does not know or is unable to recall the address information for a specific period of time. In this case, there is no way to generate a geocode ID for the entry as there is no location to note, but an entry can still be added to the record using the "Unknown/Can't Recall" option in the "Address ID" drop-down menu as a placeholder for the period of time that is missing address information (Fig. 15a, b). Again, this feature is to document the reason for missing data as well as reduce the likelihood that the informant is asked repeatedly about this part of their address history at a subsequent follow-up.

Handling different periods of time at the same address

One last situation that can be complicated to document is when a participant lives at an address for a period of time, moves into a different address for another period of time, and then moves back to the previous address. This can be the case when a family moves out of their house during a renovation or construction project and then moves back into the home when it is finished. Similarly, the participant could be at the same address, but the time spent at the provided address drastically changes. An example of this can be when the participant experiences parental separation and goes from living at the address "Full Time" to "Half Time". It can also be the case when the participant moves to college for the school year but returns to a previous address for the "Summer" time.

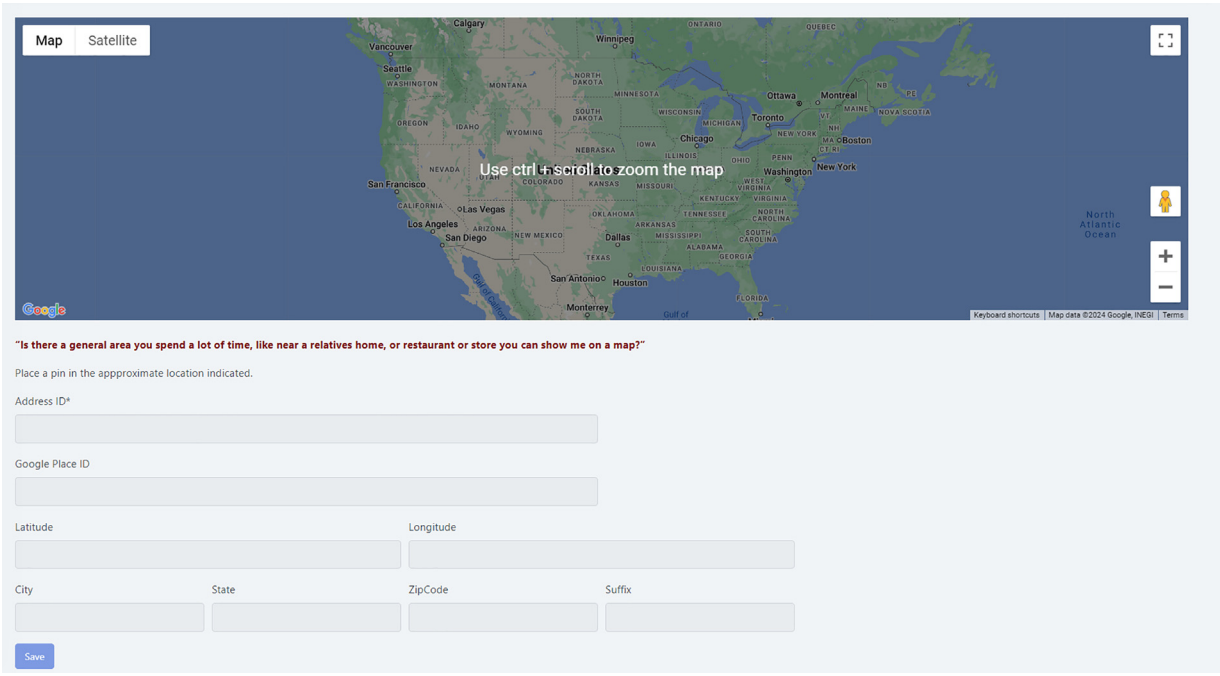

Fig. 14. The interface for placing a location pin on the area of the map at which the participant spent a lot of time at.

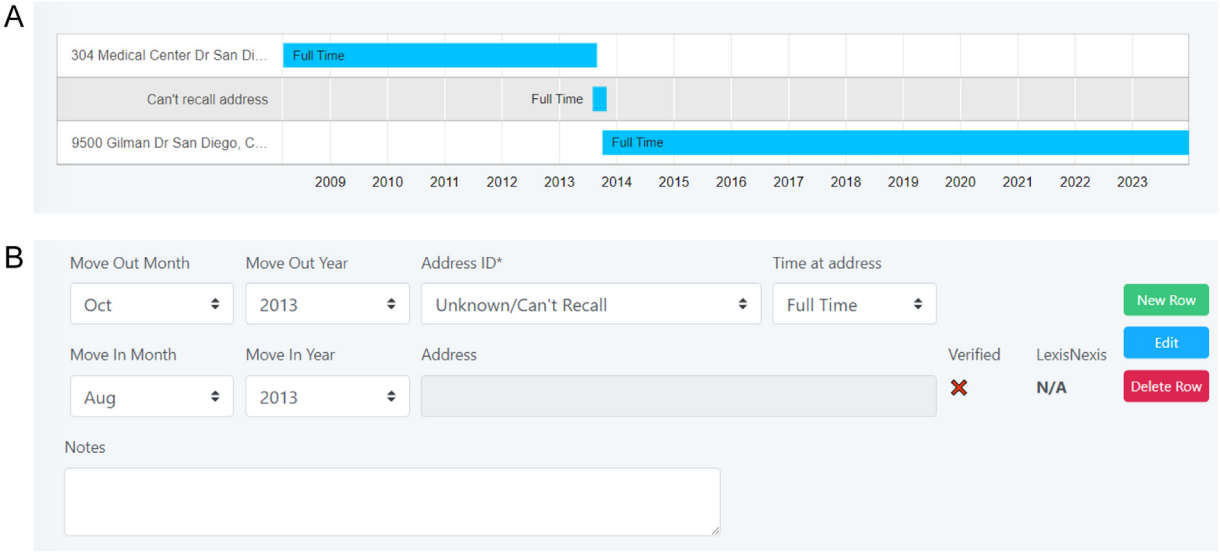

Fig. 15. **(a) Top:** Visual timeline shows a period of time where an address is missing. **(b) Bottom:** The entry is "Unknown" but can explain the gap between the other two addresses in the record instead of displaying as "Missing" on the timeline.

In these scenarios, the researcher can create new Address Entries using an existing address on record, but select the accurate time spent at the address and move-in/move-out for each entry separately to reflect these patterns. To do this, the researcher can click on the "New Row" button to the right of the entry containing the existing address of interest (Fig. 16a), which creates an empty Address Entry so the researcher can edit the information for the new entry accordingly (Fig. 16b). Lastly, Fig. 16c captures how the move-in/move-out fields differ and how the "Time at address" changed from one entry to the other to capture two different periods as well as time spent at a single residential location over the child's lifespan.

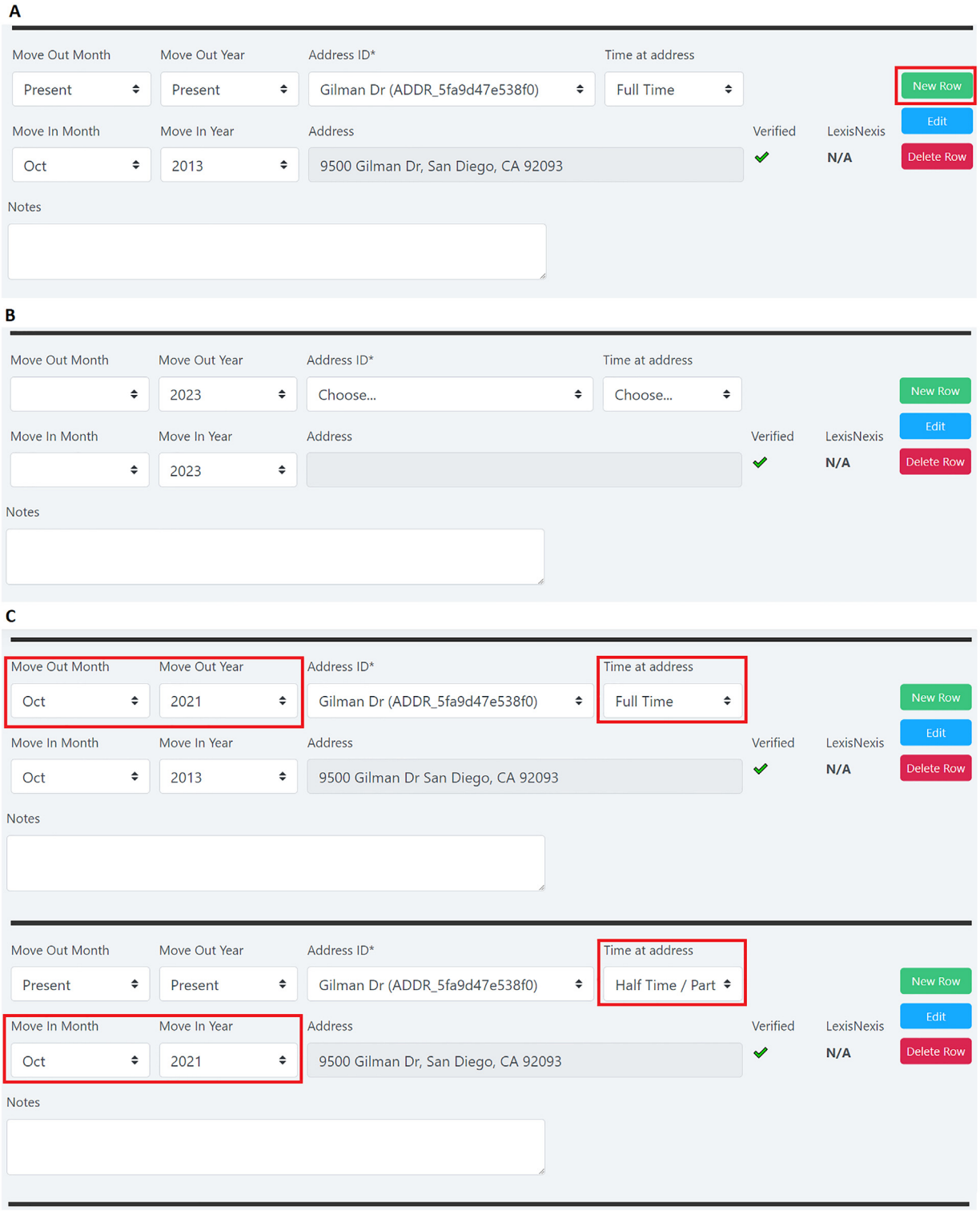

Fig. 16. **(a)** The "New Row" adds an empty new address entry. **(b)** The empty new address entry appears in the record. **(c)** After selecting the existing address from the "Address ID" menu, the move-in/move-out fields and the time spent at that address can reflect the changes of when the individual transitioned from "Full time" to "Half time".

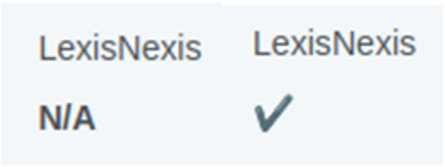

Fig. 17. Indicators for denoting an Address Entry was obtained through LexisNexis.

Using LexisNexis to aid in missing addresses

Having to recall years of residential history can be a difficult task to complete. The risk of missing data is likely, and it is absolutely necessary to establish methods to handle this issue. One method is to use external resources post-data collection to extract missing address data that supplements the data collected from participants [11]. As such, the ABCD Study decided to add to the current protocol the usage of LexisNexis, which is a third-party database that provides access to public information about people, including information on residential addresses. After the informant provides consent for the ABCD Study to implement LexisNexis for this study purpose, a search is conducted on the informant in order to extract address information and create a more complete residential history record for the participant. It is important to note that LexisNexis only pulls address information about the informant who provided consent. Because we want to collect address information *for the participant,* any address collected through LexisNexis needs to be reviewed with the informant to verify the participant indeed lived at that address and can therefore keep the address in the residential record. Additionally, LexisNexis is only used to find addresses for participants that have missing periods of residential history data; if the informant is able to provide complete information about the participant's addresses, LexisNexis is not needed.

If the informant provided consent for LexisNexus usage, data analysts working in the ABCD Study in the Data Analysis, Informatics & Resource Center (DAIRC) use LexisNexus to generate a list of addresses and attach it to the respective address history record as a new Address Entry to be reviewed with the informant at the next follow-up appointment. The PIID interface denotes if an Address Entry was imported from LexisNexis by a checkmark "√" under the "LexisNexis" label. If an address was not obtained through LexisNexis, "N/A" will appear under "LexisNexis" (Fig. 17).

If an address was obtained from LexisNexis but has not been verified with the informant at a previous appointment, the researcher lets the informant know that the address was pulled from LexisNexis and then proceeds to verify the information, add missing details, or remove the address if it is not relevant to the child's residential history record.

Troubleshooting Google API in real-time

While several features were developed in the current protocol to deal with complex situations, there are still a few examples that might prove difficult. We outline a few tricky situations when implementing this protocol when it pertains to identifying an exact address using the Google API interface in real-time.

1. There may be times that the address reported by the informant does not show up on Google API. Examples include:
   A. The address provided by the informant does not match the address that Google geocodes (i.e. the spot on the map looks incorrect to the informant even though the address is correct). This is most often the case when the city boundaries are changing.
   B. If the address provided by the informant is relatively new, Google Maps may not find it. This is most often the case when a new housing complex or development is built, and Google has not updated or added this location as an address to their maps.

Solution: In both cases, the informant verifies the information does not match and the study staff asks the informant to drop a pin on the map where the address should be located.

2. The participant has resided outside of the country.

Solution: Enter the address as accurately as possible. If google is not able to verify the address, enter a city and country. Mark as approximate location (unless you find the exact building in the community).

3. Exact addresses for a location may not be specific enough. Examples include:
   a. The address is for a Military Base and Google API is not geocoding the exact address of the living quarters.
   b. If the address is a large housing/living community and Google is not pulling the correct building in that community.
   c. The participant lives on a college campus.

Solution: Ask the informant to drop a pin on the map where the address should be located. Mark as approximate location (unless you find exact building in the base, community, or college premises on Google API).

Results

Data completeness is actively monitored across the ABCD consortium. Specifically, we compared completeness as defined as the proportion of a participant's lifespan between birth and latest study visit with associated residential history data. Of the participants who have completed a visit since the launch of the currently presented SOP, 87% of participants have residential histories that are at least 80% complete. This is compared to 39% of participants whose last completed visit occurred when the consortium was using the original text-based protocol and 78% of participants who completed their more recent visit after an initial revision of the original SOP that included using the Google Place API. The current SOP leads to less missing data and individual address histories that are temporally and spatially more robust and conducive to research involving studying environmental factors across the lifespan via geolinkages.

Conclusion

The current SOP details the protocol enacted for the longitudinal ABCD study in April 2023. It allows for a standardized method to compile comprehensive data on participants' residential histories while helping to mitigate errors in data collection in real-time. This SOP is useful to both researchers utilizing the ABCD data to better understand the origin and context behind the collected residential data as well as investigators preparing to integrate residential history data collection into their own studies. While the data collection interface does not use a stand-alone software readily available, the ABCD PIID interface used to implement this SOP for the study is created using open-source frameworks like Django using Python, JavaScript and other publicly available APIs. Visual interface aside, the processes, logic, and definitions outlined in the current protocol methods (i.e. definitions of time, seasons, address verification processes in Google Maps, and visual timelines) can be replicated to collect residential history information in other studies.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Paola Badilla: Conceptualization, Methodology, Investigation, Writing – original draft. **Shermaine Abad:** Conceptualization, Methodology, Writing – original draft. **Calen Smith:** Methodology, Investigation. **Brandon Tsui:** Software, Visualization. **Carlos Cardenas-Iniguez:** Conceptualization, Methodology, Supervision. **Megan M. Herting:** Conceptualization, Methodology, Writing – original draft, Supervision, Funding acquisition.

Data availability

The data that has been used is confidential.

Acknowledgments

A special thank you to comments from Michael Rosario, PhD, Ka Ip, PhD and Julia Harris on previous drafts to help improve clarity to outside readers.

This protocol was developed for the Adolescent Brain Cognitive Development[SM] (ABCD) Study (https://abcdstudy.org). This is a multisite, longitudinal study designed to recruit more than 10,000 children aged 9–10 and follow them over 10 years into early adulthood. The ABCD Study® is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/consortium_members/. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators.

Additional support for this work was made possible from NIEHS R01-ES032295 and R01-ES031074.

References

[1] 2017 Fall Life History Mail Survey Residential History State Level Data | Health and Retirement Study [WWW Document], 2024 URL https://hrs.isr.umich.edu/data-products/restricted-data/available-products/12213 (accessed 2.27.24).

[2] Advance Residence History Questionnaire - cdc_142535_DS1.pdf [WWW Document], 2024 URL https://stacks.cdc.gov/pdfjs/web/viewer.html?file=https://stacks.cdc.gov/view/cdc/142535/cdc_142535_DS1.pdf (accessed 2.27.24).

[3] F.P. Boscoe, The use of residential history in environmental health studies, in: Geospatial Analysis of Environmental Health, Geotechnologies and the Environment, Springer Netherlands, Dordrecht, 2011, pp. 93–110, doi:10.1007/978-94-007-0329-2_4.

[4] C. Brokamp, G.K. LeMasters, P.H. Ryan, Residential mobility impacts exposure assessment and community socioeconomic characteristics in longitudinal epidemiology studies, J. Expo. Sci. Environ. Epidemiol. 26 (2016) 428–434, doi:10.1038/jes.2016.10.

[5] M.S. Brooks, A. Bennett, G.S. Lovasi, P.M. Hurvitz, N. Colabianchi, V.J. Howard, J. Manly, S.E. Judd, Matching participant address with public records database in a US national longitudinal cohort study, SSM Popul. Health 15 (2021) 100887, doi:10.1016/j.ssmph.2021.100887.
[6] H.E. Danysh, L.E. Mitchell, K. Zhang, M.E. Scheurer, P.J. Lupo, Differences in environmental exposure assignment due to residential mobility among children with a central nervous system tumor: Texas, 1995–2009, J. Expo. Sci. Environ. Epidemiol. 27 (2017) 41–46, doi:10.1038/jes.2015.63.
[7] A. De Roos, S. Davis, J. Colt, A. Blair, M. Airola, R. Severson, W. Cozen, J. Cerhan, P. Hartge, J. Nuckols, M. Ward, Residential proximity to industrial facilities and risk of non-hodgkin lymphoma, Environ. Res. 110 (2010) 70–78, doi:10.1016/j.envres.2009.09.011.
[8] C.C. Fan, A. Marshall, H. Smolker, M.R. Gonzalez, S.F. Tapert, D.M. Barch, E. Sowell, G.J. Dowling, C. Cardenas-Iniguez, J. Ross, W.K. Thompson, M.M. Herting, Adolescent Brain Cognitive Development (ABCD) Study Linked External Data (LED): protocol and practices for geocoding and assignment of environmental data, Dev. Cogn. Neurosci. 52 (2021) 101030, doi:10.1016/j.dcn.2021.101030.
[9] Form: Risk Factor Questionnaire (RFQ) - Residential History [WWW Document], 2024 URL https://cde.nlm.nih.gov/formView?tinyId=QygWwCmHiX (accessed 2.27.24).
[10] H. Hu, X. Liu, Y. Zheng, X. He, J. Hart, P. James, F. Laden, Y. Chen, J. Bian, methodological challenges in spatial and contextual exposome-health studies, Crit. Rev. Environ. Sci. Technol. 53 (2023) 827–846, doi:10.1080/10643389.2022.2093595.
[11] S. Hurley, A. Hertz, D.O. Nelson, M. Layefsky, J. Von Behren, L. Bernstein, D. Deapen, P. Reynolds, Tracing a path to the past: exploring the use of commercial credit reporting data to construct residential histories for epidemiologic studies of environmental exposures, Am. J. Epidemiol. 185 (2017) 238–246, doi:10.1093/aje/kww108.
[12] G.M. Jacquez, M.J. Slotnick, J.R. Meliker, G. AvRuskin, G. Copeland, J. Nriagu, Accuracy of commercially available residential histories for epidemiologic studies, Am. J. Epidemiol. 173 (2011) 236–243, doi:10.1093/aje/kwq350.
[13] B. Liu, L. Niu, F.F. Lee, Utilizing residential histories to assess environmental exposure and socioeconomic status over the life course among mesothelioma patients, J. Thorac. Dis. 15 (2023) 6126–6139, doi:10.21037/jtd-23-533.
[14] M.C. Long, E. Pelletier, J. Romich, Constructing monthly residential locations of adults using merged state administrative data, Popul. Stud. 76 (2022) 253–272, doi:10.1080/00324728.2022.2085776.
[15] RAND, 2024. Baseline questionnaire [WWW Document]. HCSUS Baseline Quest. Table Contents. URL https://www.rand.org/health-care/projects/hcsus/Base.html (accessed 2.27.24).
[16] N.D. Saenen, T.S. Nawrot, P. Hautekiet, C. Wang, H.A. Roels, P. Dadvand, M. Plusquin, E.M. Bijnens, Residential green space improves cognitive performances in primary schoolchildren independent of traffic-related air pollution exposure, Environ. Health 22 (2023) 33, doi:10.1186/s12940-023-00982-z.
[17] Study details [WWW Document], 2024 URL https://www.bcgenerationsproject.ca/about/study-details/(accessed 2.27.24).
[18] M.C. Turner, D. Krewski, W.R. Diver, C.A. Pope, R.T. Burnett, M. Jerrett, J.D. Marshall, S.M. Gapstur, Ambient air pollution and cancer mortality in the cancer prevention study II, Environ. Health Perspect. 125 (2017) 087013, doi:10.1289/EHP1249.
[19] N.D. Volkow, G.F. Koob, R.T. Croyle, D.W. Bianchi, J.A. Gordon, W.J. Koroshetz, E.J. Pérez-Stable, W.T. Riley, M.H. Bloch, K. Conway, B.G. Deeds, G.J. Dowling, S. Grant, K.D. Howlett, J.A. Matochik, G.D. Morgan, M.M. Murray, A. Noronha, C.Y. Spong, E.M. Wargo, K.R. Warren, S.R.B. Weiss, The conception of the ABCD study: from substance use to a broad NIH collaboration, Dev. Cogn. Neurosci. 32 (2018) 4–7, doi:10.1016/j.dcn.2017.10.002.
[20] P.J. Weyer, J.R. Cerhan, B.C. Kross, G.R. Hallberg, J. Kantamneni, G. Breuer, M.P. Jones, W. Zheng, C.F. Lynch, Municipal drinking water nitrate level and cancer risk in older women: the Iowa Women's Health Study, Epidemiology 12 (2001) 327–338 Camb. Mass, doi:10.1097/00001648-200105000-00013.
[21] D.C. Wheeler, A. Wang, Assessment of residential history generation using a public-record database, Int. J. Environ. Res. Public Health 12 (2015) 11670–11682, doi:10.3390/ijerph120911670.
[22] C.P. Wild, Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology, Cancer Epidemiol. Biomark. Prev. 14 (2005) 1847–1850, doi:10.1158/1055-9965.EPI-05-0456.
[23] W. Xu, M. Agnew, C. Kamis, A. Schultz, S. Salas, K. Malecki, M. Engelman, Constructing residential histories in a general population-based representative sample, Am. J. Epidemiol. kwad188 (2023), doi:10.1093/aje/kwad188.
[24] D. Youens, D.B. Preen, M.N. Harris, R.E. Moorin, The importance of historical residential address information in longitudinal studies using administrative health data, Int. J. Epidemiol. 47 (2018) 69–80, doi:10.1093/ije/dyx156.