Data in Brief 44 (2022) 108507

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Careful: An administrative child welfare and electronic health records linked dataset

Sarah J. Beal[a,b,*], Katie Nause[b], Robert T. Ammerman[a,b], Eric S. Hall[c], Constance A. Mara[a,b], Mary V. Greiner[a,d]

[a] Department of Pediatrics, University of Cincinnati College of Medicine, 3333 Burnet Avenue, Cincinnati, OH 45229-3036 USA
[b] Division of Behavioral Medicine and Clinical Psychology, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229-3036 USA
[c] Department of Research Informatics and Innovation, Geisinger Health Systems, 100 North Academy Ave., Danville, PA 17822 USA
[d] Division of General and Community Pediatrics, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229-3036 USA

ARTICLE INFO

Article history:
Received 3 April 2022
Revised 27 July 2022
Accepted 28 July 2022
Available online 2 August 2022

Dataset link: CAREFul (Reference data)

Keywords:
Child protective services
Health services research
Foster care
Kinship care
Out-of-home care
Data linking

ABSTRACT

Between 2012 and 2017, $N = 2814$ youth between the ages of 4 and 20 were in child protective services (CPS) custody in Hamilton County, Ohio, and placed in out-of-home care. Child welfare administrative records were extracted and linked to electronic health records for all encounters at Cincinnati Children's Hospital Medical Center, with $n = 2787$ (99.1%) of records successfully linked prior to de-identifying the data for research purposes. Child welfare administrative data fields in the dataset include demographics, dates of entry into and exit from protective custody and out-of-home care, reasons for entry into custody, dates of placement changes, reasons for placement changes, and types of placement (e.g., foster home, kinship home, group home, residential treatment, independent living). Electronic health records (EHR) data fields include demographics, all inpatient and outpatient encounters with medications, diagnoses, screening

DOI of original article: 10.1016/j.chiabu.2022.105592
* Corresponding author at: Behavioral Medicine and Clinical Psychology, Cincinnati Children's Hospital, 3333 Burnet Avenue, MLC 7029, Cincinnati, OH 45229-3036 USA.
E-mail address: sarah.beal@cchmc.org (S.J. Beal).
Social media: @sarahbealphd (S.J. Beal), @RobertTAmmerma1 (R.T. Ammerman), @EricHallPhD (E.S. Hall), @checkdr (M.V. Greiner)

https://doi.org/10.1016/j.dib.2022.108507

results, laboratory test results, flowsheet data, and problem list entries. Data have been coded to capture broader categories of health needs and encounter details, medications, and other health concerns. Due to the high representation of children in CPS custody and out-of-home care who are also represented in the EHR data, this dataset provides a comprehensive view of the medical needs and health concerns for school-aged children in CPS custody in an entire county. As a result, these data can be useful for understanding the emergence of global and specific health concerns, frequency of healthcare use, and placement stability for all youth in CPS custody in this community, accounting for variation due to other health and child welfare factors. These data are likely generalizable to other mid-sized urban communities where academic medical centers provide healthcare for children in CPS custody. De-identified data may be made available to other researchers with approved data transfer agreements between academic institutions in place.

Specifications Table

Subject	Perinatology, Paediatrics and Child Health; Applied Psychology; Social Science; Health
Specific subject area	Exposure to maltreatment, child protective services involvement, out-of-home care, healthcare, and related health concerns in a paediatric population
Type of data	Table
How the data were acquired	This study uses secondary data collected for healthcare and service delivery purposes and not for research. Archival administrative child welfare records were extracted from the Ohio Statewide Automated Child Welfare Information System for children in Hamilton County Job and Family Services Protective Custody between 2012 and 2017 (https://sacwis.ohio.gov/sacwislogin/). Electronic health records over the same time period were extracted from Epic© at Cincinnati Children's Hospital Medical Center (https://www.cincinnatichildrens.org/careers/demand/epic). A data transfer agreement between the two institutions was established to support this collaborative effort, and institutional review boards/committees at both organizations reviewed and approved the study. An honest broker protocol was implemented to temporarily provide identifiable data to an informatics team at Cincinnati Children's Hospital to link the two data sources. Once linked, a limited datasets containing a unique shared identifier and key variables of interest from both data sources was created and provided to the research team for cleaning and analysis. At that point, the research team used dates to calculate time between events (e.g., entry into foster care and primary care visits) and to calculate counts of healthcare encounters over specified periods of time. Once fully coded, dates were removed, retaining only year, and the data was then fully de-identified. Data cleaning and coding was conducted using SAS version 9.4 software (SAS Institute, Inc.).
Data format	Filtered, Analysed
Description of data collection	Data were extracted from the child welfare administrative record for all children ages 4–20 years (inclusive) in child protective services custody for $\geq$1 day between July 1, 2012 and October 30, 2017 (N = 2814). The vast majority (n = 2787, 99.1%) of records were linked to electronic health records data at Cincinnati Children's Hospital. Most (n = 2679; 96.1%) youth with linked records experienced $\geq$1 healthcare encounters at Cincinnati Children's Hospital during that time.

(*continued on next page*)

Data source location	Cincinnati Children's Hospital Medical Center (electronic health records data; e.g., Epic ©) and Hamilton County Job and Family Services (child welfare administrative records data) Cincinnati and Hamilton County, Ohio United States of America The CAREFul Project, including instructions for data access, is provided at https://www.cincinnatichildrens.org/research/divisions/f/foster-care-research-lab/projects/careful-data-dashboard The CAREFul Project, including instructions for data access, is registered at https://fairsharing.org/4203 Hamilton County, Ohio is the third most populous county in the state of Ohio (N = 830,639). It hosts a large urban city, Cincinnati, along with 18 smaller suburban cities, 19 villages, and 12 townships. Approximately 23% of the county population is under the age of 18. Census data indicate that 27% of the population are African American, 65% are White and non-Hispanic, 4% are Hispanic or Latinx, and 3% are multiracial. Hamilton County Job and Family Services oversees the care of approximately 3400 children in child protective services custody and out-of-home care annually, and a daily census between 1900 and 2100 over the study period. Of those, approximately 60% are school-aged. The majority (>70%) are in custody for fewer than 2 years. Consistent with other communities, children of color are over-represented in Hamilton County child protective services custody, where 50% of children in protective custody in the county were identified as African American, 38% were identified as white and non-Hispanic, and 10% were identified as multiracial. Cincinnati Children's Hospital supports approximately 1.3 million patient encounters annually and has 670 registered beds, including 140 inpatient and residential mental health beds. The majority (83%) of encounters are for outpatient care, followed by emergency and urgent care (12%) and hospital admissions (2%). Cincinnati Children's Hospital established a contract with Hamilton County Job and Family Services in 2012 to provide comprehensive medical screenings for all children in Hamilton County child protective services custody at the time of entry into foster care and with each placement change [1]. As a result, most children in Hamilton County child protective services custody receive healthcare at Cincinnati Children's Hospital. In addition, Cincinnati Children's Hospital is the only paediatric healthcare system for inpatient admissions (including for behavioural health concerns), the largest paediatric healthcare system for emergency, urgent, and outpatient specialty care, and offers six outpatient primary care clinics for publicly-insured children. As a result, the majority of outpatient healthcare services provided to children residing in Hamilton County who are publicly insured are provided by Cincinnati Children's Hospital. One important exception is outpatient behavioural health services, where historically agencies unaffiliated with Cincinnati Children's Hospital provide behavioural health care in outpatient settings, and those data are not included in this study.
Data accessibility	Repository name: CAREFul Data identification number: 10.25504/FAIRsharing.dab53d Direct URL to data: https://www.cincinnatichildrens.org/research/divisions/f/foster-care-research-lab/projects/careful-data-dashboard This study uses secondary data collected for healthcare and service delivery purposes and not for research. The data were linked, de-identified, and made available to the researcher through an established data use agreement between Cincinnati Children's Hospital and Hamilton County Job and Family Services (Data Use Agreement [DUA] template included in supplemental materials, instructions for data access also located here: https://www.cincinnatichildrens.org/research/divisions/f/foster-care-research-lab/projects/careful-data-dashboard). Researchers interested in using these de-identified data for research purposes should adapt the DUA and contact the corresponding author to fully execute the DUA between Hamilton County Job and Family Services, Cincinnati Children's Hospital, and the academic institution the researcher is affiliated with. An approved institutional review board protocol describing the research being conducted will also be required. Please see https://www.cincinnatichildrens.org/research/divisions/f/foster-care-research-lab/projects/careful-data-dashboard for details.

(*continued on next page*)

Related research article	[2] S. J. Beal, R. T. Ammerman, C. A. Mara, K. Nause, M. V. Greiner, Patterns of Healthcare Utilization with Placement changes for Youth in Foster Care, Child Abuse and Neglect, 128 (2022) in press. https://doi.org/10.1016/j.chiabu.2022.105592

Value of the Data

- These linked child welfare administrative and electronic health records data provide a complete summary of children's physical health (as captured by the healthcare system serving this population) where children are confirmed to be in child protective services custody and details about children's health are documented. This is useful for studying health risks, health needs, patterns of healthcare use, and how child protective services influences child health.
- Any researcher interested in understanding specific aspects of health (e.g., immunization rates, medication use, diagnoses, healthcare use patterns, health risk behaviours) for children in child protective services custody or the interplay between child welfare involvement and health could benefit from access to these data.
- Further, because healthcare and child welfare policy in the United States is evolving, findings uncovered from these data may directly influence how services are delivered to this population in the future, benefiting all children in child protective services custody in the future.
- Future studies could examine how access to one type of service influences other healthcare or child welfare factors; how diagnoses or treatment plans differ by type of maltreatment, type of placement, with placement changes, or related to other health or child welfare factors for this population; or to examine specific health concerns in the general population or with specific subgroups
- This data could also be combined with data from another community or time-period to examine differences by geographic location, policy around healthcare and/or child welfare (in instances where policy differs between communities or changed over time) or generalizability of health concerns and needs to a population outside of this community.

1. Data Description

Data for these analyses are derived from the electronic health record (EHR) at Cincinnati Children's Hospital Medical Center, and from the Statewide Automated Child Welfare Information System (SACWIS). Datasets were extracted and linked based on shared identifiers. Specifically, our team adapted the process described by Dexheimer et al. [3] to apply similar procedures to link child welfare and EHR data for research purposes, using archival data from 2012-2017. We developed a set of rules for linking data that mimicked the logic laid out in Dexheimer et al. to identify deterministic matches. Those rules included the following, modified to include insurance type and exclude social security number, which Dexheimer et al. reported led to erroneous matches.

Deterministic:

1. First name AND/OR alias first name match from both data sources

   AND

2. Last name AND/OR alias last name match

   AND

3. Date of birth matches

   AND

4. gender matches

   AND

5. Race matches

   AND

6. Insurance type matches

Rules were applied to $n = 2{,}812$ child welfare records and $n = 438{,}461$ electronic health records for children between the ages of 4 and 20 (inclusive) between 2012 and 2017, when the study took place. Most ($n = 2{,}619$ child welfare records; 93%) had identifying data in both systems that deterministically matched, such that we were confident that child welfare and electronic health records were from the same person. Our deterministically matched sample rate was consistent with that reported by Dexheimer et al. [3]. A random review of 10% if the data revealed that this matching process performed well also, where all deterministic matches were appropriate.

For the remaining $n = 199$ child welfare records that remained unmatched after deterministic rules were applied, a set of non-deterministic rules were applied to identify matching records among the $n = 435{,}649$ electronic health records that remained. Those non-deterministic rules included the following.

Non-deterministic:

1. First name AND/OR alias first name match from both data sources

   OR

2. Last name AND/OR alias last name match

   OR

3. Date of birth matches

   OR

4. Maternal first name match

   OR

5. Maternal last name match

   AND

6. gender matches

   OR

7. Race matches

   OR

8. Insurance type matches

Most ($n = 168$ child welfare records; 84%) of records that were not deterministically linked had a non-deterministic potential match that, upon review by a study coordinator who had access to the electronic health record, were approved and linked, such that we were confident that child welfare and electronic health records were from the same person. Our non-deterministically matched sample rate exceeded that reported by Dexheimer et al. [3]. A second member of the research team independently reviewed all 168 non-deterministic and determined matches were appropriate, further validating the accuracy of this process.

With these procedures, we were successful at linking $n = 2787$ (99.1%) of records. The remaining $n = 25$ child welfare records did not have matching electronic health records that we could identify. This matching rate is on-par with that reported by Dexheimer et al. [3].

The dataset for these analyses was created from the linked raw data files with some variables represented in their raw form, and some variables transformed from the raw data. See Table 1

Table 1
List of variables from source data and transformed variables.

Electronic Health Record			
Variable	Description	Derived From	Coding
Month	Month during the study period	Created based on calendar month during the study period	0 (July 2012) – 63 (October 2017)
DOB*	Date of birth	Source data	Date (mdy)
Age	Age in years	Calculated from date of birth and the first day of each month during the study period	Numerical- continuous
Gender	Biological sex	Source data	0 = Male, 1 = Female
Race	Race	Source data	0 = White, 1 = Black/African American, 2 = American Indian and Alaska Native, 3 = Asian, 4 = Hispanic/Latino, 5 = Native Hawaiian/Other Pacific Islander, 6 = Multi-racial, 7 = Other, 8 = Unknown
Race_recode	Comparison of white, black, and other races	Recoded from source data race	0 = White, 1 = Black/African American, 2 = All other races
Race_wh	Comparison of white race to all other races	Recoded from source data race	1 = White, 0 = All other races
Race_bl	Comparison of black race to all other races	Recoded from source data race	1 = Black, 0 = All other races
Ethnicity	Ethnicity	Source data	0 = Non-Hispanic, 1 = Hispanic, 2 = Other/Unknown
Ethnicity_r	Comparison of Hispanic and other ethnicities to non-Hispanic	Recoded from source data race	0 = Non-Hispanic, 1 = All other ethnicities
Enc_date*	Date of encounter	Source data	Date (mdy)
Visit_type*	Type of visit	Source data (text)	Office visit, Telemedicine, Specimen collection, Testing, Education, Social work, Surgery
Base_class*	Designation of outpatient, emergency department, or inpatient healthcare encounter	Source data (text)	Outpatient, Inpatient, Emergency/Urgent
Appt_status*	Appointment status	Source data (text)	Completed, Canceled, No-show, Left without being seen
Department*	Department or clinic affiliated with the encounter	Source data (text)	Primary care, Specialist, Emergency department, Urgent care
Dx_name*	Diagnoses on the problem list	Source data	ICD-9/ICD-10 codes (see Tables 4 and 5 for coding)
Planned	Number of planned healthcare days per month	Time-varying count of planned healthcare days each month coded from encounter dates	Numerical- continuous
Planned_sum	Cumulative sum of planned healthcare days during the study period	Time-varying cumulative count of planned healthcare days during the study period	Numerical- continuous
Unplanned	Number of unplanned healthcare days per month	Time-varying count of unplanned healthcare days each month coded from encounter dates	Numerical- continuous

(*continued on next page*)

Table 1 (*continued*)

		Electronic Health Record	
Variable	Description	Derived From	Coding
Unplanned_sum	Cumulative sum of unplanned healthcare days during the study period	Time-varying cumulative count of unplanned healthcare days during the study period	Numerical- continuous
Mandated	Number of mandated healthcare days per month	Time-varying count of mandated healthcare days per month coded from encounter dates	Numerical- continuous
Mandated_sum	Cumulative sum of mandated healthcare days during the study period	Time-varying cumulative count of mandated healthcare days during the study period	Numerical- continuous
Missed	Number of missed healthcare days per month	Time-varying count of missed healthcare days per month coded from encounter dates	Numerical- continuous
Missed_sum	Cumulative sum of missed healthcare days during the study period	Time-varying cumulative count of missed healthcare days during the study period	Numerical- continuous
Totalhcd	Number of total healthcare days per month	Time-varying count of total healthcare days per month coded from encounter dates	Numerical- continuous
Total_sum	Cumulative sum of total healthcare days during the study period	Time-varying cumulative count of total healthcare days during the study period	Numerical- continuous
Mandate_b	Time from placement entry to mandated healthcare encounter	Number of days between placement begin and mandated healthcare day categorized as 3 days or less, 4-8 days, 9-31 days, or 32 days or more	0 = 3 days or less, 1 = 4–8 days, 2 = 9–31 days, 3 = 32 days or more
Totalccdx	Total number of chronic medical diagnostic categories	Coded from problem list diagnosis codes	Range 0–21
Ccdx	Binary indicator of at least one chronic medical condition diagnosis	Coded from problem list diagnosis codes	0 = No chronic medical conditions, 1 = At least one chronic medical condition
Allergy	Presence of allergy diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Asthma	Presence of asthma diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Cardiology	Presence of cardiology diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Dermatology	Presence of dermatology diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Endocrine	Presence of endocrine diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Gastroenterology	Presence of gastroenterology diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Genetics	Presence of genetics diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Gynecology	Presence of gynecology diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category

(*continued on next page*)

Table 1 (*continued*)

Electronic Health Record			
Variable	Description	Derived From	Coding
Headache	Presence of headache diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Hematology	Presence of hematology diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Hepatology	Presence of hepatology diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Id	Presence of infectious disease diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Neonatology	Presence of neonatology diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Nephrology	Presence of nephrology diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Neurology	Presence of neurology diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Oncology	Presence of oncology diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Otolaryngology	Presence of otolaryngology diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Podiatry	Presence of podiatry diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Pulmonology	Presence of pulmonology diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Rheumatology	Presence of rheumatology diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Urology	Presence of urology diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Totalmhdx	Total number of mental health diagnostic categories	Coded from problem list diagnosis codes	Range 0 -18
Mhdx	Binary indicator of at least one mental health condition diagnosis	Coded from problem list diagnosis codes	0 = No mental health conditions, 1 = At least one mental health condition
Adjustment	Presence of adjustment disorder diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Anxiety	Presence of anxiety disorder diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Attention	Presence of attention disorder diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Bipolar	Presence of bipolar disorder diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Depression	Presence of depressive disorder diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Disruptive	Presence of disruptive, impulse control, or conduct disorder diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Dissociative	Presence of dissociative disorder diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Eating	Presence of feeding and eating disorder diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Identity	Presence of gender identity disorder diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Mood	Presence of mood disorder diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category

(*continued on next page*)

Table 1 (*continued*)

Electronic Health Record			
Variable	Description	Derived From	Coding
Neurodevelopmental	Presence of neurodevelopmental disorder diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Ocd	Presence of obsessive-compulsive disorder diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Personality	Presence of personality disorder diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Psychotic	Presence of psychotic disorder diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Sleep	Presence of sleep-wake disorder diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Somatic	Presence of somatic disorder diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Suicidality	Presence of suicidality diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Trauma	Presence of trauma and stressor-related disorder diagnosis	Coded from problem list diagnosis codes	0 = No diagnosis in category, 1 = At least one diagnosis in category
Child Welfare Administrative Data			
Variable	Description	Derived From	Coding
Custody_begin*	Date of entry into CPS custody	Source data	Date (mdy)
Custody_end*	Date of exit from CPS custody	Source data	Date (mdy)
Lifetimeepisodes	Lifetime number of CPS custody episodes	Source data	Numerical- continuous
Priorlos	Number of days in CPS custody prior to study start	Days between date of entry into CPS custody and date of exit from CPS custody for all episodes starting before the study period	Numerical- continuous
Duringlosti	Number of days in CPS custody during the study period	Days between date of entry into CPS custody and date of exit from CPS custody for all episodes during the study period	Numerical- continuous
Placement_begin*	Date of entry into placement	Source data	Date (mdy)
Placement_end*	Date of exit from placement	Source data	Date (mdy)
Placechange	Indicator of placement change within the month	New placement initiated coded from source data placement entry dates	0 = No placement change/Not in CPS custody that month, 1 = At least one placement change
Placespermonth	Number of placements entered per month	Time-varying count of placements beginning each month coded from source data placement entry dates	Numerical- continuous
Placesduringti	Total number of placements during the study period	Number of placements during the study period coded from source data placement entry dates	Numerical- continuous

(*continued on next page*)

Table 1 (*continued*)

		Electronic Health Record	
Variable	Description	Derived From	Coding
Placesduringtv	Cumulative number of placements during the study period	Time-varying count of placements beginning during the study coded from source data placement entry dates	Numerical- continuous
Priorplacements	Number of placements before the study period	Time-invariant count of placements beginning before the study period coded from source data placement entry dates	Numerical- continuous
Placementlos	Number of days spent in each placement	Days between date of entry into placement and date of exit from placement for all placements	Numerical- continuous
Adoptive	Currently in pre-adoptive placement	Dummy-coded from source data placement type	0 = Not in pre-adoptive placement, 1 = In pre-adoptive placement
Foster	Currently in foster home	Dummy-coded from source data placement type	0 = Not in foster placement, 1 = In foster placement
Grouphome	Currently in group home	Dummy-coded from source data placement type	0 = Not in group home placement, 1 = In group home placement
Incarceration	Currently incarcerated in detention facility	Dummy-coded from source data placement type	0 = Not currently incarcerated, 1 = Currently incarcerated
Indliving	Currently in independent living	Dummy-coded from source data placement type	0 = Not in independent living placement, 1 = In independent living placement
Kinship	Currently in kinship placement	Dummy-coded from source data placement type	0 = Not in kinship placement, 1 = In kinship placement
Ownhome	Currently with family of origin (Before first CPS custody, during trial home placement, and after return to family of origin)	Dummy-coded from source data placement type	0 = Not in own home, 1 = In own home/Not in CPS custody that month
Residential	Currently in residential placement	Dummy-coded from source data placement type	0 = Not in residential placement, 1 = In residential placement
Shelter	Currently in emergency shelter	Dummy-coded from source data placement type	0 = Not in emergency shelter, 1 = In emergency shelter
Fambase	Family-based placement setting	Coded based on current placement type	1 = In family setting (foster, kinship, adoptive, or own home placement types), 0 = Not in family setting (all other placement types)
Monitoring	Level of parental monitoring (larger values indicate higher levels of parental monitoring)	Coded based on current placement type	-99 (missing) = Residential facility/Incarceration 0 = Independent living, 1 = Emergency shelter, 2 = Group home, 3 = Trial Home Visit/Adoptive placement, 4 = Kinship, 5 = Foster home

Table 2
Planned healthcare days.

Visit Types	Labels from Source Data
Exams and consultations	Office visit, Appointment, Telemedicine, Hospital encounter, Nurse only, Therapy visit
Testing and Procedures	Cardiology testing, EEG, Specimen collection, Exercise, Endo Stim testing, Urodynamic testing, Surgery
Education and Services	Social work, Education, Nutrition
Departments	
Primary care services	All completed healthcare encounters at primary care clinics
Specialty care services	All completed outpatient healthcare encounters at specialist clinics (cardiology, psychology, pulmonology, etc.)
Testing and procedures	All outpatient laboratory testing (MRI, EEG, Specimen collection) and outpatient surgeries

Table 3
Unplanned healthcare days.

Visit Types	Labels from Source Data
Exams and consultations	Office visit, Appointment, Hospital encounter, Therapy visit
Testing and Procedures	Cardiology testing, Specimen collection, Endo Stim testing
Departments	
Emergency care services	Emergency department and urgent care encounters and associated encounters (testing, imaging, specialist consultations)
Inpatient care services	Inpatient admissions and associated encounters during admission (testing, imaging, specialist consultations)

for complete list of variables and descriptions. The analytic dataset was structured so that each calendar month during the study period was a unique observation for all participants included. The month variable served as the time metric for the dataset, with Month 0 (July 2012) being the first observation in the data set, and Month 62 (October 2017) being the last observation. Additionally, some variables are time-variant (i.e., values change from one observation to the next), and some are time-invariant (i.e., values are constant across observations).

Variables represented in both raw datasets included biological sex, race, ethnicity, and date of birth. In the event of discrepancies in these shared variables, the values from the presumed most accurate dataset were used. Biological sex was determined using the EHR, and date of birth, race, and ethnicity were determined using child welfare administrative data. In the analytic dataset, biological sex was included in its raw form, race was recoded to Black (1) or all other races (0), and ethnicity was recoded to non-Hispanic (0) or all other ethnicities (1). Biological sex, race, and ethnicity were time-invariant. Date of birth was used to create the age at each observation (the first day of each calendar month) during the study period and was time-variant.

Raw variables from the EHR used to create the analytic dataset included the date of each healthcare encounter, visit type, department or clinic affiliated with the encounter, appointment status, diagnoses from the problem list, and designation of outpatient, inpatient, or emergency/urgent care.

Variables that were created from raw EHR data included healthcare day type (planned, unplanned, mandated, missed, and total), number of each healthcare day type per month, and the cumulative sum of each healthcare day type during the study period. All healthcare day types were time-variant. See Table 2 for coding of planned healthcare days and Table 3 for coding of unplanned healthcare days. Diagnostic codes from the problem list were used to create time-invariant total number of chronic medical conditions (Table 4) and mental health conditions (Table 5) and were coded as the count of unique diagnostic categories present for each participant.

Raw variables from the child welfare administrative records used to create the analytic dataset included date of entry into CPS custody, date of exit from CPS custody, date of entry

Table 4
Chronic medical condition diagnostic categories.

Diagnostic Category	ICD 10 Codes
Allergy/Immunology	D84, H10, J30-J31, L50, T78.1, Z88
Asthma	J45
Cardiology	I05-I99, Q20-Q28, R00-R09, R94.3
Dermatology	B07, D22-D23, I78.1, L00-L99, Q80-Q82
Endocrinology	D35-D35, E00-E89, L68, Q56
Gastroenterology/ Metabolic disorders	E73, K00-K95, Q39-A45, R10-R19
Genetics	D55.0, D82.1, E70-E78, E88.4, G60.0, G71.2, Q04, Q13.4, Q44.7, Q79.6, Q85, Q90-Q99
Gynecology	D24, E28.2, N70-N73, N80, N91-N94, Q51
Headache	G43-G44, R51
Hematology	D18, D50-D69, E61.1, I82.90
Hepatology	B18-B19, E88.01, K70-K77, Q44, R76.8
Infectious disease	A15-A19, M86, R76.1
Neonatology	O35, Q67.6
Nephrology	E83.5, N00-N07, N17-N19, N25-N29, Q60-Q63, Z90.5
Neurology	D17, D36, F07.81, F44.7, G10-G14, G25, G35-G37, G40, G51, G71, G81-G82, G90-G91, G93, G95, G98, I67.83, P91, Q01-Q06, Q28.2, Q75.3, R25, R27, R48-R49, R56
Oncology	C49, C69, C71-C72, C74, C85, C91-C92, C96, D14, D17, D21, D49
Otolaryngology	H65-H66, H83, H90-H91, J32, J35, J38-J39, R04.0,
Podiatry	M67.0
Pulmonology	E84, J44, J47, J82, J84, J98, P25, P27, Q33, R06, R91, Z14.1
Rheumatology	M05, M24-M25, M33-M35, M79
Urology	F98.0, I86.1, N13, N31, N35, N39, N43, N48, N50, Q53-Q55, Q62, Q64, R32, R39

Table 5
Mental health condition diagnostic categories.

Diagnostic Category	ICD 10 Codes
Adjustment	F43
Anxiety	F40-F41, F94.0
Attention	F90, F98.8
Bipolar	F31, F34.0
Depression	F32-F34, F53, R45.89
Disruptive, Impulse Control, and Conduct	F52, F63, F91, F98.4, R45.4, R45.87, R45.89, X97, Y09, Z72.81, Z72.89, Z73.81
Dissociative	F44.9
Feeding and Eating	F50, F98.21, R63.0, Z72.4
Gender Identity	F64
Mood	F06.3, F30, F34, F39, F48.9
Neurodevelopmental	F09, F70-F72, F79, F80-F82, F84, F88, F95, F98.5, G31.84, G80, I69.91, I69.92, P04.41, P04.9, Q86.0, R20.9, R29.818, R41, R47-R48, R62.50
Obsessive-Compulsive	F42, F63.3, R46.81, R46.89,
Personality	F60
Psychotic	F06.1, F20, F22-F23, F25, F29, F39, H53.16, R44.0-R44.3
Sleep-wake	F51, G47, R40.0,
Somatic	F44.4, F44.7, F44.9, F45
Suicidality	F48.9, R45.851, R46.89, T39.012, T43.202, T45.0 × 2, T46.4 × 2, T50.902, T65.92, T71.162, X78-X80, X83, Z72.89
Trauma	F43.1, F43.9, F51.5, F93.0, F94.1, F94.2

into placement, date of exit from placement, and placement type. Lifetime number of CPS custody episodes before the end of the study period was also included in the analytic dataset in its raw form and did not vary across observations.

Variables in the analytic dataset created from raw child welfare administrative data that were time-invariant included the total number of days in CPS custody prior to the study, total number of placements before the study, total number of placements during the study, and total number of days in CPS custody during the study. If a participant was in CPS custody at the time the

study began, the number of days in CPS custody prior to the study was calculated using the date of entry into custody and the first day of the study period, July 1, 2012. If a participant was in CPS custody at the time the study ended, the number of days in CPS custody during the study was calculated using the date of entry into custody and last day of the study, October 31, 2017.

Variables transformed from the raw child welfare administrative data that were time-variant included an indicator of placement change occurring that month, the number of placement changes occurring that month, cumulative number of placements during the study period, number of days in each placement, placement types, whether or not the current placement was a family-based setting, and the level of parental or caregiver monitoring based on placement type. Placement types included pre-adoptive placements, certified foster home, group home, detention facility (incarceration), independent living, kinship (certified approved relative or non-relative), residential facility, emergency shelter, and own home (trial home visit, at home before CPS custody, returned home after CPS custody). Family-based placement settings included foster home, kinship, adoptive placement, and own home. Non-family-based placement settings included group home, independent living, residential facility, emergency shelter, and detention facility. Level of parental monitoring was determined using the placement type and ranged from 0 (low level of monitoring) to 5 (highest level of monitoring).

The number of days between entry into placement and mandated healthcare encounter was created from placement entry dates in the raw child welfare data, and healthcare encounter dates in the raw EHR data. Of note, some variables described in Table 1 include identifying data (e.g., dates) or other data that required coding to ensure that there were more than 10 youth in a category. To ensure that data are de-identified, those variables are not available in the de-identified dataset, and are marked with a * in Table 1.

2. Experimental Design, Materials and Methods

Our research team was provided with a linked, limited dataset that contained encounter-level data for all children who received care from Cincinnati Children's Hospital Medical Center and their corresponding child welfare administrative data (see Table 1 for complete list of data fields). To accomplish the goal of examining patterns of healthcare use as children experienced changes in placement while in child welfare protective custody [2], data provided to our team had to be cleaned and new variables were derived from existing data fields. Below we detail the methods used to identify each data field in the original data source, how it was transformed (when necessary), and how it is represented in the data analysed for our companion publication [2].

1. Electronic Health Records Data: Each child included in this study ($N = 2787$) had a linked child welfare and electronic health record (EHR). EHR data could be stored in structured fields (e.g., radio buttons, drop-down menu options, check boxes) or unstructured fields (e.g., free-text notes written by clinicians). While most data fields are associated with an encounter (e.g., diagnoses made by the provider, department or clinic associated with the encounter) some data elements apply to the child and span multiple encounters (e.g., demographic data, problem list diagnoses). As a result, for all patients with linked data ($N = 2787$) some data from the EHR were available for use in this study. However, the most complete data occurs when a child experienced one or more healthcare encounters during the study period. This was the case for most children included in this study ($n = 2679$; 96.1%).
   a) Data associated with encounters. At each healthcare encounter, fields were provided to indicate the following:
      - Date of the encounter (which has since been removed from the data file)
      - Department or clinic affiliated with the encounter
      - Type of visit
      - Diagnoses tied to the encounter
      - Diagnoses added or removed from the problem list at that encounter

- Medications prescribed in the encounter
- Medications reconciled during the encounter

b) Variables calculated to span multiple encounters. A total of 8 variables were calculated from the EHR for use in this study.

Healthcare days. Frequently, children are seen by multiple healthcare providers or clinical staff on the same day. This can happen for a number of reasons, including multidisciplinary clinics (e.g., having social work, psychology, and paediatric care as part of the same encounter), needing laboratory testing or a secondary procedure in conjunction with a visit (e.g., completing an EKG in cardiology as part of a neurology visit, a telehealth visit combined with laboratory testing), or because care to meet an emergent health need requires multiple disciplines (e.g., an encounter in Radiology that coincides with an emergency department visit for concern of a broken bone). While these encounters may be associated with each other in the EHR, they each generate data that is independent, and that can lead to what appears to be multiple encounters that occur on the same day. To account for this and avoid over-inflating the occurrences of healthcare encounters, we chose to code healthcare encounters by day rather than by documented count, where any day over the study period where 1 or more encounter occurred was given a value of 1, and any day over the study period where no encounter occurred was given a value of 0. This was further delineated using department or clinic affiliated with the encounter and type of visit to distinguish among the following:

- Planned healthcare days. These are days where $\geq$1 visit(s) occurred that were scheduled in advance, and could be outpatient (e.g., well child exam, follow-up visit with a specialist) or pre-scheduled inpatient (e.g., same-day surgery) where a caregiver would have anticipated that the visit was going to occur and had control over when it was scheduled. A complete list of visit types and codes for planned healthcare days is provided in Table 2.
- Unplanned healthcare days. These are days where $\geq$1 visit(s) occurred that were not schedule in advance, and could be outpatient (e.g., emergency and urgent care visits) or inpatient (e.g., unplanned hospital admission) where a caregiver would not have anticipated that the visit was going to occur and/or may not have had control over when it was scheduled. A complete list of visit types and codes for unplanned healthcare days is provided in Table 3.
- Total healthcare days. These are days where $\geq$1 visit(s) occurred that were either planned or unplanned.
- Mandated healthcare days. These are visits that occurred at the foster care clinic and were associated with either a mandated entry into foster care exam or a mandated change of placement exam. These visits are required by the child protection services system and are therefore not included as either planned or unplanned healthcare days.
- Missed healthcare days. These are visits that were scheduled in advance (any department or visit type), but where the patient either did not attend the visit or cancelled the visit within 24 h.

Once healthcare days were coded, we were able to summarize counts of healthcare days per calendar month, which was used in the final analysis [2].

Chronic Medical Conditions. For youth who had received healthcare prior to the observation period, including those who did not have any healthcare encounters during the observation period, chronic medical conditions were determined using the child's problem list, which is intended to capture diagnoses that will persist across multiple healthcare encounters. Any diagnosis can be added to the problem list, including acute illnesses not intended for inclusion, such as an ear infection; for that reason, we applied the Chronic Condition Indicator criteria provided by the Agency for Healthcare Research and Quality [4] to only include diagnoses from the problem list that reflected chronic medical conditions. Further, as youth experienced additional healthcare encounters during the course

of the study, new diagnoses made at those encounters or added to the problem list were extracted, following the same Chronic Condition Indicator criteria to only capture chronic medical conditions. Once a complete list of all chronic medical conditions identified for a child over the course of the study was available, we removed duplicate diagnoses (e.g., the same diagnostic code could be used for more than one encounter, as often occurs for youth with chronic conditions). Further, because there can be multiple diagnostic codes used for the same underlying condition (e.g., G80.8 and G80.9 both reflect a cerebral palsy diagnosis). International Classification of Diseases 9th and 10th revision structures (ICD9/ICD10) [5] were used to conservatively estimate chronic diseases, where chronic conditions were grouped by system (e.g., diseases of the blood and blood-forming organs; endocrine, nutritional, and metabolic disease, diseases of the nervous system, etc.; see Table 4). Diagnoses with high frequency of occurrence in paediatrics (e.g., Asthma) were coded separately from the system in which they are classified (e.g., Diseases of the respiratory system). The count of unique chronic condition diagnostic categories was used in these analyses, to reflect a child's unique number of chronic medical conditions. Diagnoses captured in the DSM-5 ([6]; e.g., ADHD) were not included in counts for chronic medical conditions as they were included with mental health conditions.

Mental HealthConditions. Mental health diagnoses made by or reported to a Cincinnati Children's Hospital provider were included. Diagnoses made and services delivered by mental health providers unaffiliated with Cincinnati Children's Hospital (e.g., for community-based outpatient behavioral health services) are not represented. Cincinnati Children's Hospital commonly provides behavioral health services integrated with other medical care (e.g., primary care, the foster care clinic) and inpatient care (e.g., psychiatric admissions) but is not the primary behavioral health service provider for outpatient behavioral health services. Mental health diagnoses were coded according to DSM-5 diagnostic criteria into one of 18 categories (Table 5), with ADHD, adjustment disorder, mood disorders, and suicidal behavior coded separately from the DSM category they align with due to the high frequency with which ADHD, adjustment disorders, mood disorders and suicidal behavior were observed. The count of unique disorders across DSM categories was used to determine the number of mental health diagnoses a child had.

c) Demographic data. The child's date of birth and date of encounter was used to calculate age, which was included as a time-varying covariate in all models. The biological sex of participants (male = 0, female = 1) was also included as a covariate. While other assessments of gender are now available in the EHR, at the time of this data collection nonbinary and non-conforming gender fields were not available.

2. Child Welfare Administrative Data: Only structured data was extracted from the statewide administrative child welfare information system. This included fields related to the child's entry into CPS custody (e.g., indicated or substantiated maltreatment occurrences, reasons for placement into out-of-home care, dates of entry into and exit from CPS custody), or to a specific placement (e.g., dates of entry into and exit from that placement, placement type, level of care expected by the placement provider for that child (traditional, therapeutic, therapeutic – high needs, medically complex), number of other children placed in the home at the same time). Children may have experienced multiple custody episodes and multiple placement changes over the study period. Those variables were processed to create variables for analysis as described below.

a) Custody and placement history prior to the start of the study. Age of entry into CPS custody for the first time was derived using the child's date of birth and earliest date of entry into CPS custody stored in the child welfare information system.

Given that some children were already in CPS custody on the first day of the study period, we included several covariates in each model to adjust for disparate exposure to CPS custody before the study period, as well as varying amount of time youth were stable in a placement. These variables included:

- Number of days in custody prior to the start of the study,
- Lifetime number of custody episodes,
- Number of days in placement prior to the start of the study (for youth already in CPS custody on July 1, 2012),
- Lifetime number of placements prior to study start, and
- Length of stay in each placement prior to the study start.

b) Custody and placement during the study. Custody information (i.e., episodes where children were in CPS custody) during the study period included number of days in custody during the observation period. Placements during the study period had additional data associated with it, which was coded to include the following:
   - Number of placement changes during the study period
   - Number of placements up until the current observation (i.e., time-varying placement count)
   - Length of stay in each placement.
   - Placement types, provided by the child welfare agency, which were dummy-coded as 0 (*not in that placement type)* or 1 (i*n that placement type*). Placement types included Adoptive Placement, Certified Approved Non-Relative, Certified Approved Relative, Certified Children's Residential Center, Certified Emergency Shelter Care Facility, Certified Foster Home, Certified Group Home, Detention Facility, Independent Living, Licensed Medical/Educational Facility, and Trial Home Visit.
   - To understand differences among family-style placements (i.e., one or two adult caregivers residing continuously with children in a home-like environment) versus non-family-style placement settings (e.g., congregate care facilities with multiple children and staff that supervise and provide care in shifts; independent living where children reside alone in an apartment) an additional dichotomous variable was created that indicated whether the current placement was family-style (0) or non-family style (1). Family-style placements included Certified Foster Home, Certified Approved Relative, Certified Approved Non-Relative, Adoptive Placement, and Trial Home Visit. Non-family-style placement types included Independent Living, Detention Facility, Certified Group Home, Certified Children's Residential Center, Licensed Medical/Educational Facility, and Certified Emergency Shelter Care Facility.
   - A separate "kinship" variable was created to understand the effect of placement with relatives (1) compared to stranger care (0) within family settings. Placement types of Certified Approved Relative, Certified Approved Non-Relative, and Trial Home Visit were recoded as Kinship (1), while Certified Foster Home was coded as non-kinship (0) and all other placement types were missing.
   - To capture variation in resources available to support caregiver monitoring across various placements, a variable was created that ranged from 0 (low level of monitoring), to 5 (high level of monitoring). Specifically, youth placed in independent living or semi-independent living were coded 0 (lowest level of monitoring), youth with a traditional level of care in placement settings with more than 6 other children were coded as 1 (second-lowest level of monitoring), youth with therapeutic or high needs levels of care in placement settings with more than 6 other children were coded as 2 (moderate level of monitoring), youth with a traditional level of care placed in kinship or foster homes with 3–6 other children were coded as 3 (third-highest level of monitoring), youth with therapeutic or high needs levels of care placed in kinship or foster homes with 3–6 other children were coded as 4 (second-highest level of monitoring), and youth in kinship or foster homes with no more than 1 other child were coded as 5 (highest level of monitoring).

c) Child welfare data to inform time metrics in analytic models. For analyses examining healthcare use with month of observation as the time metric (ranging from 0, start of study, to 63, end of study), months where entry into foster care or a placement change occurred were coded as "1" and months where no placement change occurred was coded

as "0". To accont for a child experiencing more than one placement in a study month, a multiple placement variable (yes = 1, no = 0) was also created.

For analyses examining healthcare use following entry into foster care or a placement change, we narrowed the sample to only individuals who had their first placement during the study (i.e., no prior placements before study start, entry into foster care occurred after the study start) and modeled healthcare use. Data were subsetted to create one database representing all children who entered CPS custody for the first time after the start of the study period. All data relevant to the child's first placement episode were included in the first dataset. Number of months in first placement was the time metric, ranging from 0 (start of placement) to 60 (60 months in placement).

Children who experienced a second placement following first entry into CPS custody during the study period were included in a second dataset representing the second placement following entry into foster care over the study period. Number of months in second placement was the time metric, ranging from 0 (start of placement) to 60 (60 months in placement).

Children who experienced a third placement following first entry into CPS custody during the study period were included in a third dataset representing the third placement following entry into foster care over the study period. Number of months in third placement was the time metric, ranging from 0 (start of placement) to 60 (60 months in placement).

d) Demographic data from the child welfare information system included:
   - Date of birth (used to calculate age over the study period),
   - Sex assigned at birth (male = 0, female = 1)
   - Race (White = 0, Black or African American = 1, other = 2),
   - Ethnicity (non-Hispanic = 0, Hispanic = 1)

3. Variables from Both Data Sources: One question of interest was whether mandated healthcare encounters that occurred close in proximal time to placement changes or entry into foster care influenced other types or patterns of healthcare use. For that reason, timing of mandated medical screenings following placement changes were coded as the number of days between placement entry date (from the child welfare information system) and the date of the visit to the foster care clinic (from the electronic health record), classified based on best-practice guidelines for healthcare delivery [7] as $\leq$ 3 days (0), 4 to 7 days (1), 8 to 30 days (2), or $>$ 30 days (3). This was included as a predictor in models where time since a placement change was the time metric.

   Some data elements were available in both data sources (i.e., date of birth, sex assigned at birth, race, ethnicity), introducing opportunity for inconsistencies in the data. This occurred for 6% of the total sample. When inconsistencies occurred, we selected which data source to use based on the presumed level of accuracy across each system.
   - Date of birth: child welfare in our state is required to secure a child's birth certificate and update their comprehensive child welfare information system with those identifying records within 60 days of entry into custody, which we determined was the most accurate source for child date of birth.
   - Race and ethnicity: child welfare also collects child race and ethnicity from the family of origin and stores that data with high consistency in their information system; the electronic health record captures race and ethnicity, but it is often assessed by registration staff, who may rely on the child's physical appearance or the report of the caregiver accompanying the child, introducing more opportunity for error. When race and ethnicity data were discrepant, child welfare data was used.
   - Sex assigned at birth: we chose to include sex assigned at birth from the electronic health record because full physical exams (including of a child's genitals) are common practice at our primary care and foster care clinics, and our pediatricians are trained to ensure accuracy in coding of child's biological sex in the electronic health record.

Ethics Statements

This study was conducted with Institutional Review Board approval under protocol number 2017–4747. As this is secondary analysis of archival data collected for non-research purposes and analyses included limited data, a waiver of consent was granted. Hamilton County Job and Family Services, the legal custodian for children in CPS custody who were included in this study, provided approval and permission to use these data for the research purposes described above.

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.

Data Availability

CAREFul (Reference data) (https://fairsharing.org/4036).

CRediT Author Statement

Sarah J. Beal: Conceptualization, Methodology, Validation, Writing – original draft, Project administration, Funding acquisition; **Katie Nause:** Software, Formal analysis, Data curation, Writing – review & editing; **Robert T. Ammerman:** Conceptualization, Writing – review & editing; **Eric S. Hall:** Software, Data curation, Writing – review & editing; **Constance A. Mara:** Formal analysis, Writing – review & editing; **Mary V. Greiner:** Conceptualization, Writing – review & editing.

Acknowledgments

The project described was supported in part by the National Institute for Minority Health and Health Disparities under Award Number 1R03MD011419 and the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number 5UL1TR001425. Dr. Beal's time was supported through the National Institute of Drug Abuse under Award Number 1K01DA041620. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.dib.2022.108507.

References

[1] M.V. Greiner, S.J. Beal, Developing a health care system for children in foster care, Health Promot. Pract. (2018) 621–628, doi:10.1177/1524839917730045.
[2] S.J. Beal, R.T. Ammerman, C.A. Mara, K. Nause, M.V. Greiner, Patterns of healthcare utilization with placement changes for youth in foster care, Child Abuse Negl. 128 (2022) in press, doi:10.1016/j.chiabu.2022.105592.
[3] J.W. Dexheimer, S.J. Beal, P. Divekar, E.S. Hall, V. Patel, M.V. Greiner, Automated patient linking for electronic health record and child welfare databases. J. Technol. Hum. Serv., 37(4), 286-292.
[4] Agency for Healthcare Research and Quality, Chronic Condition Indicator (CCI) for CID-9-CM. Healthcare cost and utilization project (2016). https://www.hcup-us.ahrq.gov/toolssoftware/chronic/chronic.jsp. Last accessed January 5, 2022.

[5] World Health Organization. International statistical classification of diseases and related health problems (2021). https://www.who.int/classifications/classification-of-diseases. Last accessed January 5, 2022.
[6] American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 5th ed. (2013) 947, doi:10.1176/appi.books.9780890425596.
[7] American Academy of Pediatrics, Fostering Health: Health Care for Children and Adolescents in Foster Care, 170, 2nd ed., University of Michigan Press, Ann Arbor, MI, 2005.