

Integrated Historical Child Welfare Database

NDACAN Dataset Number 309 USER'S GUIDE



National Data Archive on Child Abuse and Neglect

Duke University / Cornell University

Phone: 607-255-7799

Email: ndacan@cornell.edu

Website: www.ndacan.acf.hhs.gov

Initial release: January 14th, 2026

Integrated Historical Child Welfare Database

Data Collected by

Alexander F. Roehrkasse, Ph.D., Butler University

Lauren Moreland, B.A., Butler University

Funded by

U.S. Administration for Children and Families

Processed and Distributed by

National Data Archive on Child Abuse and Neglect (NDACAN)

User's Guide Written by

Alexander F. Roehrkasse, Ph.D., Butler University

Lauren Moreland, B.A., Butler University

©2025 National Data Archive on Child Abuse and Neglect

Table of Contents

<i>Integrated Historical Child Welfare Database</i>	1
Preface.....	5
Acknowledgement of Source.....	5
Publication Submission Requirement	5
Abstract.....	6
Study Overview	7
Study Identification	7
Purpose of the Study	7
Study Design	7
Date(s) of Data Collection.....	7
Geographic Area	7
Unit of Observation.....	7
Sample.....	7
Data Collection Procedures.....	8
Response Rates.....	8
Sources of Information.....	8
Type of Data Collected.....	8
Related Publications and Final Reports.....	9
Analytic Considerations	9
Confidentiality Protection	11
Extent of Collection	11
Extent of Processing.....	11
Data File Information	12
File Specifications	12
Data File Notes.....	12
Acronyms and Abbreviations	12

Preface

The data for the *Integrated Historical Child Welfare Database (IHCWD)* have been given by the authors to the National Data Archive on Child Abuse and Neglect (NDACAN) for public distribution. The data were either collected by the authors from publicly available sources, or compiled by the authors from restricted data and reported in aggregate form, preventing disclosure risk. Funding for the project was provided by the U.S. Administration for Children and Families (Award IDs HHSP233201750093A and 140D0422P0227). The research was carried out in complete independence of the funding sources, and the authors accept responsibility for the data.

Acknowledgement of Source

Authors should acknowledge the National Data Archive on Child Abuse and Neglect (NDACAN) and the original collector(s) of the data when publishing manuscripts that use data provided by the Archive. Users of these data are urged to follow some adaptation of the statement below.

The data used in this publication were made available by the National Data Archive on Child Abuse and Neglect and have been used with permission. The data were either collected by the authors from publicly available sources, or compiled by the authors from restricted data and reported in aggregate form, preventing disclosure risk. Funding for the project was provided by the U.S. Administration for Children and Families. The collector(s) and distributors of the original data, the funder(s), NDACAN, Duke University, Cornell University, and their agents or employees bear no responsibility for the analyses or interpretations presented here.

The bibliographic citation for this data collection is:

Roehrkasse, A., & Moreland, L. (2025). *Integrated Historical Child Welfare Database* (Version 1) [Data set]. National Data Archive on Child Abuse and Neglect.
<https://doi.org/10.34681/33C6-RV69>

Publication Submission Requirement

In accordance with the terms of the *Data License* for this dataset, users of these data are required to notify the National Data Archive on Child Abuse and Neglect of any published work or report based wholly or in part on these data. A copy of any completed manuscript, thesis abstract, or reprint should be emailed to NDACANsupport@cornell.edu. Such copies will be used to provide our funding agency with essential information about the use of NDACAN resources and to facilitate the exchange of information about research activities among data users and contributors.

Abstract

The *Integrated Historical Child Welfare Database* compiles aggregate, historical administrative data from the United States between 1945 and 2021. Measures include children receiving child welfare services, expenditures on child welfare services, child welfare services personnel, and children entering, in, or exiting out-of-home care. By combining multiple sources of data and by standardizing measures, the IHCWD enables analysis of long-term trends, geographic variation, and population differences in the child welfare system and the children interacting with it. IHCWD data can be linked to other NDACAN data or to external data at the state-year level, enabling various research designs using demographic and causal inference techniques.

The IHCWD contains novel aggregate data from historical child welfare publications, namely the Children's Bureau Statistical Series (1945–1968) and the National Center for Social Statistics (1970, 1973–1975). These are combined with aggregate data from the Voluntary Cooperative Information System (1982–1995; NDACAN Dataset 247) and the Adoption and Foster Care Analysis and Reporting System (1995–2021; NDACAN Datasets 73, 80, 82, 86, 91, 97, 101, 105, 118, 124, 131, 137, 143, 149, 153, 163, 167, 176, 187, 192, 200, 215, 225, 235, 239, 255, and 274).

The IHCWD is a unique and powerful resource for conducting comparative and longitudinal research, but it contains many significant limitations of which users should be aware. Users should read the User's Guide fully and carefully in order to use the data responsibly.

Study Overview

Study Identification

Integrated Historical Child Welfare Database

Investigator(s):

Alexander F. Roehrkasse, Ph.D., Butler University

Lauren Moreland, B.A., Butler University

Funded By:

U.S. Administration for Children and Families

Award Number(s): 140D0422P0227

Purpose of the Study

Research objectives for the study:

- To understand long-term trends and subnational geographic variation in child welfare services provision and out-of-home care in the United States.
- To understand the demographic attributes of children experiencing out-of-home care.

Study Design

This was a compilation of administrative data collected by U.S. states.

Date(s) of Data Collection

1945–2021

Geographic Area

United States (including District of Columbia and Puerto Rico, excluding other U.S. territories)

Unit of Observation

State-year

Sample

Children receiving child welfare services; expenditures on child welfare services; child welfare services personnel; children entering, in, or exiting out-of-home care.

Data Collection Procedures

The authors compiled data from four sources.

Children's Bureau Statistical Series (CBSS)

Data for 1945–1968 come from the CBSS. Data from this source were published contemporaneously, digitized by Google, and accessed via HathiTrust. Operators manually entered information from digital images, with the authors resolving miscodes.

National Center for Social Statistics (NCSS)

Data for 1970 and 1973–1975 come from the NCSS. Data from this source were published contemporaneously, digitized by Google, and accessed via HathiTrust. Operators manually entered information from digital images, with the authors resolving miscodes.

Voluntary Cooperative Information System (VCIS)

Data for 1982–1995 come from the VCIS, administered by the American Public Welfare Association and accessed through the National Data Archive on Child Abuse and Neglect. These data were already aggregated to the state-year level.

Adoption and Foster Care Analysis System (AFCARS)

Data for 1995–2021 come from the AFCARS, operated by the Children's Bureau and accessed through the National Data Archive on Child Abuse and Neglect. The authors aggregated child-level data to the state-year level.

Response Rates

For 4,019 possible state-year-source observations, at least one measure was observed for 3,317 observations. For 4,004 possible state-year observations, at least one measure was observed for 3,302 observations. An average of 220 measures are observed across all state-year-source observations, and an average of 26.7 measures are observed across all state-year-source observations for which at least one measure is observed.

Sources of Information

Children's Bureau Statistical Series (CBSS), National Center for Social Statistics (NCSS), Voluntary Cooperative Information System (VCIS), Adoption and Foster Care Analysis and Reporting System (AFCARS)

Type of Data Collected

Administrative data

Related Publications and Final Reports

Roehrkasse, Alexander F. 2021. “Long-Term Trends and Ethnoracial Inequality in U.S. Foster Care: A Research Note.” *Demography* 58(5):2009–17. doi:[10.1215/00703370-9411316](https://doi.org/10.1215/00703370-9411316).

Users are strongly encouraged to review published works, based upon these data, before doing analyses. To view a complete list of publications for this dataset, please visit our online citations collection called “canDL” (Child Abuse and Neglect Digital Library): <https://www.ndacan.acf.hhs.gov/cndl/cndl.cfm>.

Analytic Considerations

The *Integrated Historical Child Welfare Database* provides a basis for both cross-sectional and longitudinal comparative analysis of foster care. However, the data have significant limitations that users are strongly urged to consider before conducting analyses. Users are strongly encouraged to consult directly the User Guides for the VCIS and AFCARS, which are distributed by NDACAN; these provide additional details specific to those sources.

Variation in administrative survey instrument and compliance:

- The data underlying the IHCWD derive from multiple sources. In many cases, survey instruments and measures are designed specifically to maximize reliability across sources. However, several known issues affect the general reliability of the data.
 - o Annual reporting periods varied across sources, so seasonal fluctuation in foster care will affect cross-source reliability.
 - o State’s compliance with survey instruments varied and changed in the VCIS. Poor documentation about compliance prevents systematic correction.
- The authors have made every effort to maximize the reliability and validity of measures, including by correction or suppression of unreliable or invalid measures. Nevertheless, users should exercise caution in making inferences from the data, especially when relying on a small number of measures or measures with small values.

Reliability of foster care measure:

- The VCIS and AFCARS define out-of-home care to include children under the supervision of public welfare agencies and living in relative and nonrelative foster homes, pre-adoptive homes, group homes, childcare and emergency facilities, or supervised independent living arrangements, as well as those who had run away from foster care.
- The CBSS and NCSS tabulate children receiving casework or social services from public welfare agencies, stratified by children’s residence. To measure out-of-home care, the authors include all such children living in foster family homes, institutions,

group homes, independent living arrangements, or “elsewhere,” excluding children living with their parents or with relatives not designated as foster family homes.

- Therefore, reliability of the foster care measure across these sources depends in part on the assumption that all children in out-of-home care receive casework or social services.

Reliability of age measure:

- Data from the VCIS (1982–1995) have been suppressed when reporting patterns or documentation make clear that states’ survey-instrument noncompliance regarding age makes reported figures unreliable or invalid. Users are encouraged to consult the stand-alone VCIS data product for additional data and documentation.

Reliability of race and ethnicity measure:

- Information on the race and ethnicity of children entering, exiting, and in out-of-home care is available for 1982–2021. However, the schemes used to categorize children by race and ethnicity were different in the VCIS (1982–1995) and the AFCARS (1995–2021): the former included an “Other non-Hispanic” category but not a “Multiracial non-Hispanic” category; the latter include information for constructing a “Multiracial non-Hispanic” category but not an “Other non-Hispanic” category. Counts for these and other ethnoracial groups are therefore comparable before and after 1995 only on the assumption that the “Other” and “Multiracial” response categories measure the same group. See below (“Overlapping data in 1995”) for suggestions about assessing the assumption and addressing this limitation.
- Data have been suppressed when reporting patterns or documentation make clear that states’ survey-instrument noncompliance regarding race and ethnicity makes reported figures unreliable or invalid. In other cases, such as the apparent uneven adoption of multiracial measures, users should exercise extreme caution.

Missing data:

- The IHCWD contains considerable missing data. The majority of missingness results from state non-reporting to the relevant source. Some data, however, is missing because it is suppressed by the authors to maximize the validity and reliability of included data. These missing data take the form of empty cells. Suppressed data is not distinguished from missing data arising from other sources, but users can consult the VCIS and AFCARS data products for additional data and documentation.
- For counts of children entering, in, or exiting out-of-home care that are stratified by children’s demographic attributes, many state-years include children for whom such demographic information is unobserved. These missing data are counted in variables corresponding to each such stratum (age, race/ethnicity, sex) that report the number of children with missing, or “unknown,” values for that stratum.
- Users should always evaluate the extent of missing data, and use strategies for addressing it that are appropriate to their analytic goals. Analysis only of complete-case data or demographic data without reference to counts of children with unknown demographic attributes will lead to invalid inferences.

Overlapping data in 1995:

- Both the VCIS and AFCARS report data for 1995, and both sources of data are included in the IHCWD. This inclusion presents both opportunities and challenges to data users.
- Users analyzing the data for descriptive or statistical analyses should be aware that duplicate observations exist for many states in 1995. Based on their analytic goals, users may want to exclude one or another source, or average the two sources.
- Different states reported to each source in each year. Under certain assumptions, this provides users an opportunity to address data missing in one source by incorporating data from the other source.
- Overlapping data in 1995 provide an opportunity for users to assess the reliability of measures of out-of-home care across sources. The administrative survey instruments differ between the VCIS and AFCARS, including but not limited to the categorization of race and ethnicity (see above). By comparing 1995 values of key variables derived from the VCIS and AFCARS, users can make more informed decisions about the reliability of time series and the best approaches for addressing unreliability.

Confidentiality Protection

The dataset underwent a confidentiality review by NDACAN and it was determined that no recodes were necessary since data are sufficiently aggregated to prevent disclosure risk. There are no primary or secondary identifiers in the dataset.

Extent of Collection

This dataset package includes a User's Guide, a Codebook, and the dataset DS309_IHCWD in multiple formats: Excel (.xlsx), text-format (.dat), SPSS (.sav), Stata (.dta), and SAS (.sas7bdat).

In addition, import syntax files are provided for SAS (.sas), SPSS (.sps), and Stata (.do) to read the dataset from the accompanying text-format (.dat) file. A comma-delimited (.csv) version of the dataset is also included for use with spreadsheet programs.

Extent of Processing

The authors processed the data in the following ways:

1. Aggregation. Record-level data from the AFCARS was aggregated to counts of annual entrances into out-of-home care, annual exits from out-of-home care, and point-in-time populations in out of home care, stratified by state-year and furthermore by age, race/ethnicity, or sex. Aggregation was formatted to correspond to count data in the CBSS, NCSS, and VCIS.
2. Cleaning. Data were recoded to make measures consistent within and across sources, such as with the recoding of erroneous zero counts as missing values. Where consistency could not be achieved on the basis of the data or data documentation, data was suppressed, such as with the suppression of ethnoracial data submitted out of compliance with the survey

instrument. Data from multiple sources (see above) were assigned consistent variable values and variable names.

3. Combination. Data from multiple sources (see above) were appended using consistent state identifiers.

Data File Information

File Specifications

This dataset contains one data file containing 4,019 observations (rows) and 74 variables (columns).

Data File Notes

Data are organized in a long format by state-year

Acronyms and Abbreviations

The following abbreviations and acronyms are used in this document:

AFCARS - Adoption and Foster Care Reporting and Analysis System

CBSS - Children's Bureau Statistical Series

IHCWD - Integrated Historical Child Welfare Database

NCSS - National Center for Social Statistics

NDACAN - National Data Archive on Child Abuse and Neglect

VCIS - Voluntary Cooperative Information System

U.S. - United States

Technical support for this dataset is provided by NDACAN.

Please send your inquiries to NDACANsupport@cornell.edu

Visit the User Support page of the NDACAN website for help documents and videos (<https://www.ndacan.acf.hhs.gov/user-support/user-support.cfm>).