Child and Caregiver Outcomes Using Linked Data (CCOULD) Dataset

NDACAN Dataset Number 272
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: Feb 3, 2023

Last Revision: Feb 2, 2023

CHILD AND CAREGIVER OUTCOMES USING LINKED DATA (CCOULD) DATASET

Data Contributors

Tami L. Mark, PhD
RTI International

Melissa Dolan, PhD
RTI International

Benjamin Allaire, MS
RTI International

Keith Smith, BS
RTI International

William Parish, PhD
RTI International

Christina Bradley, MA
RTI International

Claire Strack, BS
RTI International

Project Officers

Emily Madden
Office of the Assistant Secretary for Planning and Evaluation (ASPE)

Valeria Butler
Administration for Children and Families (ACF)

Funded by

Distributed by

National Data Archive on Child Abuse and Neglect

User's Guide and Codebook Written by

Sarah Sernaker, MS
National Data Archive on Child Abuse and Neglect

in collaboration with

Keith Smith
CCOULD Data Archive Lead, Project Director
RTI International

Table of Contents

Preface

The data for Child and Caregiver Outcomes Using Linked Data (CCOULD) have been given to the National Data Archive on Child Abuse and Neglect (NDACAN) for public distribution by RTI International. Funding for the project was a partnership provided by the Office of the Assistant Secretary for Planning and Evaluation (ASPE) and the Administration for Children and Families (ACF). (Award Number: HHSP233201500039I).

Acknowledgement of Source

Authors should acknowledge the National Data Archive on Child Abuse and Neglect (NDACAN) and the original collector of the data when they publish manuscripts that use data provided by the Archive. Users of these data are urged to use 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. Data from CCOULD were originally provided by the project team at RTI International, including Tami Mark, Melissa Dolan, Benjamin Allaire, Keith Smith, Will Parish, Christina Bradley, and Claire Strack, as well as Emily Madden at ASPE and Valeria Butler at ACF. Funding for the project was provided by the Patient-Centered Outcomes Research Trust Fund (PCORTF) From SSF1] (Award Number: HHSP233201500039I. The collector of the original data, the funder, NDACAN, Duke University, and Cornell University and their agents or employees bear no responsibility for the analyses or interpretations presented here. The opinions and conclusions expressed herein belong solely to the author(s) and should not be construed as representing the opinions or policy of the U.S. Department of Health and Human Services.

The bibliographic citation for this data collection is:

> Mark, TL., Dolan, M., Allaire, B., Smith, K., Parish, W., Bradley, C., Madden, E., & Butler, V. (2022). Child and Caregiver Outcomes Using Linked Data [CCOULD]. National Data Archive on Child Abuse and Neglect. https://doi.org/10.34681/33ge-9z29

The state of Kentucky has also requested that the following acknowledgment be included in publications that use the CCOULD data:

> The analyses presented in this publication were based on data from [the name of the CCOULD dataset]. These data were provided by the National Data Archive on Child Abuse and have been used with permission. The data were originally collected under the auspices of the Kentucky Cabinet for Health and Family Services, Department for Community Based Services and the Kentucky Cabinet for Health and Family Services, Department for Medicaid Services. Funding was provided by the Office of the Assistant Secretary for Planning and Evaluation (ASPE) and the Patient-Centered Outcomes Research Trust Fund. The collector of the original data, the funder, NDACAN, Duke University, Cornell University, and the agents or employees of these institutions bear no responsibility for the analyses or interpretations presented here. The information and opinions expressed reflect solely the opinions of the authors.

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 CCOULD project worked with Kentucky and Florida (specifically the Florida Department of Children and Families, the Florida Agency for Health Care Administration, and the Kentucky Cabinet for Health and Family Services) to produce linked, state-level data to support research on the relationships among Medicaid utilization, behavioral health services, patient-centered outcomes, and child welfare outcomes. State-level linked data were combined into a de-identified, standardized research use dataset containing data from the two participating states.

Documentation for the dataset includes:

(1) codebooks that include a listing of the variables in the dataset, brief descriptions of the variables, variable names, variable labels, response categories, value ranges for continuous variables, and frequency distributions for categorical variables;

(2) a users' guide that will explain the background, purpose, and design of the study; the data de-identification procedures; the data limitations (e.g., missing data and tracking children who have moved to different states); and analytical considerations.

Study Overview

Study Identification

Child and Caregiver Outcomes Using Linked Data (CCOULD)

Investigator(s):

Tami Mark, RTI, Washington, DC

Melissa Dolan, RTI, Chicago, IL

Ben Allaire, RTI, Raleigh, NC

Keith Smith, RTI, Chicago, IL

Will Parish, RTI, Raleigh, NC

Christina Bradley, RTI, Washington, DC

Claire Strack, RTI, Raleigh, NC

Project Officers:

Emily Madden, ASPE, Washington, DC

Valeria Butler, ACF, Washington, DC

Funding Agencies:

Patient-Centered Outcomes Research Trust Fund (PCORTF).

Award Number: HHSP233201500039I

Purpose of the Study

The Child and Caregiver Outcomes Using Linked Data (CCOULD) project developed consistent state datasets from Florida and Kentucky containing linked child and caregiver welfare records and Medicaid claims and enrollment records. These datasets contain information from both child welfare and Medicaid information systems on case demographics, medical diagnoses, services, outcomes, and other relevant information. The purpose of the data is to support research on the relationships between Medicaid utilization, behavioral health services, patient-centered outcomes, and child welfare outcomes. Of particular interest are outcomes for families that may have substance use disorders, like opioid use disorder.

Study Design

The data are from Kentucky and Florida's child welfare and Medicaid agencies (the Florida Department of Children and Families, Florida Agency for Health Care Administration, and Kentucky Cabinet for

Health and Family Services). The project developed a common data model to ensure that both states provided data and variables in a consistent manner. States linked Medicaid and child welfare data using deterministic linkage methodologies. The data contain no personally identifiable information (PII).

Date(s) of Data Collection

The following are the date ranges of the key types of records that are in the CCOULD datasets.

Maltreatment report disposition dates
Dataset: ccould_ch_welfare_report,
Variables: RPT_DISPOSITION_DT_MO and RPT_DISPOSITION_DT_YR
- Florida: 10/2016 – 9/2020
- Kentucky: 1/2016 – 9/2021

Foster care discharge dates
Dataset: ccould_ch_fc_episodes
Variables: FC_DISCHRG_DT_MO and FC_DISCHRG_DT_YR
- Florida: 10/2016 – 8/2021
- Kentucky: 1/2016 – 9/2021

Title IV-E services dates
Dataset: ccould_title_iv_e_services
Variables: CH_SRVC_DT_MO and CH_SRVC_DT_YR
- Florida: 3/1995 - 12/2020
- Kentucky: 10/2013 - 10/2021

Medicaid enrollment dates
Dataset: ccould_mdcd_enrollmt
Variables: MONTH and CYEAR
- Florida: 1/2017 – 6/2021
- Kentucky: 1/2017 – 12/2020

Medicaid enrollment comparison group dates
Dataset: ccould_mdcd_enrollmt_compgrp
Variables: MONTH and CYEAR
- Florida: 1/2017 - 6/2021
- Kentucky: 1/2017 - 12/2020

Medicaid IP header admission dates
Dataset: ccould_mdcd_ip_hedr
Variables: ADMT_DT_MO and ADMT_DT_YR
- Florida: 1/2012 – 6/2021
- Kentucky: 6/2015 – 12/2020

Medicaid IP header comparison group admission dates
Dataset: ccould_mdcd_ip_hedr_compgrp
Variables: ADMT_DT_MO and ADMT_DT_YR
- Florida: 1/2000 - 6/2021
- Kentucky: 10/2016 - 12/2020

Medicaid IP line admission dates
Dataset: ccould_mdcd_ip_line
Variables: ADMT_DT_MO and ADMT_DT_YR
- Florida: 3/2001 - 6/2021
- Kentucky: 6/2015 - 12/2020

Medicaid IP line comparison group admission dates
Dataset: ccould_mdcd_ip_line_compgrp
Variables: ADMT_DT_MO and ADMT_DT_YR
- Florida: 1/2000 - 6/2021
- Kentucky: 10/2016 - 12/2020

Medicaid OT header service dates
Dataset: ccould_mdcd_ot_hedr
Variables: SRVC_FROM_DT_MO/SRVC_FROM_DT_YR and
SRCV_THRU_DT_MO/SRVC_THRU_DT_YR
- Florida: 4/2015 – 9/2021
- Kentucky: 1/2017 – 6/2021

Medicaid OT header comparison group service dates
Dataset: ccould_mdcd_ot_hedr_compgrp
Variables: SRVC_FROM_DT_MO/SRVC_FROM_DT_YR and
SRCV_THRU_DT_MO/SRVC_THRU_DT_YR
- Florida: 6/2010 - 9/2021
- Kentucky: 1/2017 - 7/2021

Medicaid OT line service dates
Dataset: ccould_mdcd_ot_line
Variables: SRVC_FROM_DT_MO/SRVC_FROM_DT_YR and
SRCV_THRU_DT_MO/SRVC_THRU_DT_YR
- Florida: 4/2015 - 9/2021
- Kentucky: 1/2017 - 6/2021

Medicaid OT line comparison group service dates
Dataset: ccould_mdcd_ot_line_compgrp
Variables: SRVC_FROM_DT_MO/SRVC_FROM_DT_YR and
SRCV_THRU_DT_MO/SRVC_THRU_DT_YR
- Florida: 6/2010 - 9/2021
- Kentucky: 1/2017 - 7/2021

Medicaid pharmacy prescription dates
Dataset: ccould_mdcd_rx_line
Variables: RX_DT_MO and RX_DT_YR
- Florida: 1/2017 – 6/2021
- Kentucky: 1/2017 – 12/2020

Medicaid pharmacy comparison group prescription dates
Dataset: ccould_mdcd_rx_line_compgrp
Variables: RX_DT_MO and RX_DT_YR
- Florida: 1/2017 - 6/2021
- Kentucky: 1/2017 - 12/2020

Geographic Area

The data are from all counties in Florida and Kentucky.

Unit of Observation

The data are organized into 16 files described in Table 1.

Table 1. Data File Names and Descriptions

File Name	Description of content	Unit of analysis
1. ccould_ch_welfare_report	This file contains data on child welfare events, e.g., maltreatment reports.	Child-report
2. ccould_title_iv_e_services	This file contains data on Title IV-E services provided to families.	Child-report
3. ccould_ch_fc_episodes	This file contains data on child placement episodes.	Child-removal date
4. ccould_mdcd_enrollmt 5. ccould_mdcd_enrollmt_compgrp	Medicaid enrollment and eligibility of the person, with corresponding data for the comparison group.	Person (child/caregiver)-month-year
6. ccould_mdcd_ip hedr 7. ccould_mdcd_ip hedr_compgrp	Medicaid claims headers: Inpatient diagnostic and procedure codes, with corresponding data for the comparison group.	Person (child/caregiver)-claim

File Name	Description of content	Unit of analysis
8. ccould_mdcd_ip line 9. ccould_mdcd_ip line_compgrp	Medicaid claim lines: Inpatient diagnostic and procedure codes, with corresponding data for the comparison group.	Person (child/caregiver)-claim ID-claim line
10. ccould_mdcd ot_hedr 11. ccould_mdcd ot_hedr_compgrp	Medicaid claims headers: Outpatient diagnostic and procedure codes, with corresponding data for the comparison group.	Person (child/caregiver)-claim ID
12. ccould_mdcd_ot_line 13. ccould_mdcd_ot_line_compgrp	Medicaid claim lines: Outpatient diagnostic and procedure codes, with corresponding data for the comparison group.	Person (child/caregiver)-claim ID-claim line
14. ccould_mdcd_rx_line 15. ccould_mdcd_rx_line_compgrp	Medicaid claims: Prescription drugs filled, with corresponding data for the comparison group.	Person (child/caregiver)-claim ID
16. ccould_ch_cg_crosswalk	Crosswalk between caregivers and children, as well as demographic varibles for the children and caregivers.	Person (child/caregiver)

Two files (the child welfare report and Title IV-E services files) provide data on children who had a report of maltreatment in Florida or Kentucky with a disposition date between 2016-2021 and capture information on the maltreatment report and services provided in response to the maltreatment report (if any). The unit of analysis for these files is child-report, with each record providing information from a unique child and report of maltreatment. A third child welfare system file (the foster care episodes file) provides data on children who were removed from their home. The unit of analysis for this file is the child-removal episode.

Medicaid data on the children and their caregivers are provided in six files (enrollment, inpatient [IP] header, IP line, other therapy [OT] header, OT line, and pharmacy claims). The unit of analysis for the Medicaid enrollment file is the person (child or caregiver) and month. The unit of analysis for the header files (IP and OT headers) as well as the pharmacy claims is the person and claim level. The unit of analysis for the line files (IP and OT lines) is the person, claim, and claim-line level.

A crosswalk file is provided to link children to their caregivers. The unit of analysis in the crosswalk file is the person (child or caregiver). A variable called CH_ID can be used to identify a child's caregivers.

This variable is repeated within all members of a "family" unit. See Table 3 in the Merging Across Data Files section for more details.

Sample

Florida Population

The Florida Child Welfare dataset includes maltreatment reports from 2016 to 2021. The child welfare dataset includes 824,677 unique children, identified from 645,401 reports. All 824,677 children have demographic records. The Florida Medicaid claims data represents Medicaid claims filed January 2017 – June 2021 that could be linked to the child welfare data.

Florida also provided a Medicaid comparison sample of individuals not associated with the child welfare system.

Kentucky Population

The Kentucky Child Welfare dataset includes maltreatment reports from 2016 to 2021. The child welfare dataset includes 263,086 unique children, identified from 461,742 child welfare reports or investigations. The Kentucky Medicaid claims data represents Medicaid claims filed 2017 – 2020 that could be linked to the child welfare data.

Kentucky also provided a Medicaid comparison sample of individuals not associated with the child welfare system.

Data Collection Procedures

Data was collected from NCANDS, AFCARS, the Florida Department of Children and Families, Florida Agency for Health Care Administration, and Kentucky Cabinet for Health and Family Services.

Response Rates

Not applicable.

Sources of Information

For Florida, the source of the child welfare data was the Florida Department of Children and Families and the Florida Agency for Health Care Administration was the source of the Medicaid enrollment and claims data.

For Kentucky, the source was the Cabinet for Health and Family Services. Departments within the cabinet collaborated to generate the data files, including the Department for Community Based Services, the Department for Medicaid Services, and the Office of Health Data and Analytics. The source of the Medicaid enrollment and claims data was their Department for Medicaid Services.

Type of Data Collected

The following is a description of the data files and the types of data included:

- Child welfare report files: Administrative data on child welfare events, e.g., maltreatment reports.
- Child foster care episode files: Administrative data on child placement episodes.
- Title IV-E Services Provided: Administrative data on Title IV-E services provided to families.
- Caregiver-Child Crosswalk and Demographic (Derived) Files: Crosswalk between caregivers and children. Children and their caregivers were linked to Medicaid using SSN or via linkage on demographic information, and the linkage type indicators allow the researcher to identify which type of linkage was used to produce a specific match in this file. Also demographic variables for the children and caregivers, which were derived from the child welfare databases if available or from Medicaid if not available in the child welfare databases.
- Medicaid enrollment files: Data on Medicaid enrollment and eligibility of the person.
- Medicaid core claims files: Data on inpatient, outpatient, long term care, prescription drugs, and diagnostic and procedure codes.

Measures

Not applicable.

Related Publications and Final Reports

Users are strongly encouraged to review these references before doing analyses. To view a complete list of publications visit our online citations collection called "canDL" at: http://www.ndacan.acf.hhs.gov/publications/publications.cfm. Once on the webpage, navigate to the DS #272 folder to view all publication citations relevant to this dataset.

A brief description of the CCOULD project is located on ASPE's website:
https://www.acf.hhs.gov/sites/default/files/documents/opre/OPRE-CCOULD_ProjectBrief-Jan22.pdf

Analytic Considerations

Helpful links

In the Medicaid data files there are several standardized coding fields. These include diagnosis and procedure codes, national drug codes, revenue center codes, type of service codes, bill type codes, and taxonomy codes. There are some resources that users can avail themselves to understand what these codes represent.

Most diagnosis codes are ICD-10 diagnosis codes, and information on ICD codes can be retrieved from the Centers for Medicare & Medicaid Services (CMS) website: https://www.cms.gov/Medicare/Coding/ICD10.

Other procedure codes are described here: https://www.cms.gov/Medicare/Fraud-and-Abuse/PhysicianSelfReferral.

The Food & Drug Administration (FDA) maintains a NDC directory, which can be accessed here: https://www.fda.gov/drugs/drug-approvals-and-databases/national-drug-code-directory.

Additional details on the values of these codes can be found here: https://resdac.org/cms-data/variables/revenue-center-code-taf.

Additional details on the values of these codes can be found here: https://resdac.org/cms-data/variables/bill-type-code.

Type of service codes are described here: https://resdac.org/cms-data/variables/type-service-code.

Additional details on the values of these codes can be found here: https://nucc.org/index.php/code-sets-mainmenu-41/provider-taxonomy-mainmenu-40.

Outliers

Some variables have a few very large values that could be characterized as outliers. For example, a few individuals have very high medical care costs and a few individuals have many of placements with different care settings. There is no way to discern whether the outliers are true values or data entry errors therefore they were left unchanged. The most extreme values were top-coded for disclosure risk reduction (see Confidentiality Protection section below).

There were outlier admission dates in the Medicaid IP header, Medicaid IP line, and Medicaid IP line comparison group data files, with years as early as 1916. Outlier admission dates (year before 2000) were replaced with missing values. This affected a small number of rows (7, 4, and 9 in the Medicaid IP header, Medicaid IP line, and Medicaid IP line comparison group files, respectively).

Merging Across Data Files

The structure of the CCOULD datasets is displayed in Figure 1 below, including the source, description, and unique identifier for each file. Table 2 below shows the various identifiers, descriptions of each, and which files contain each ID.
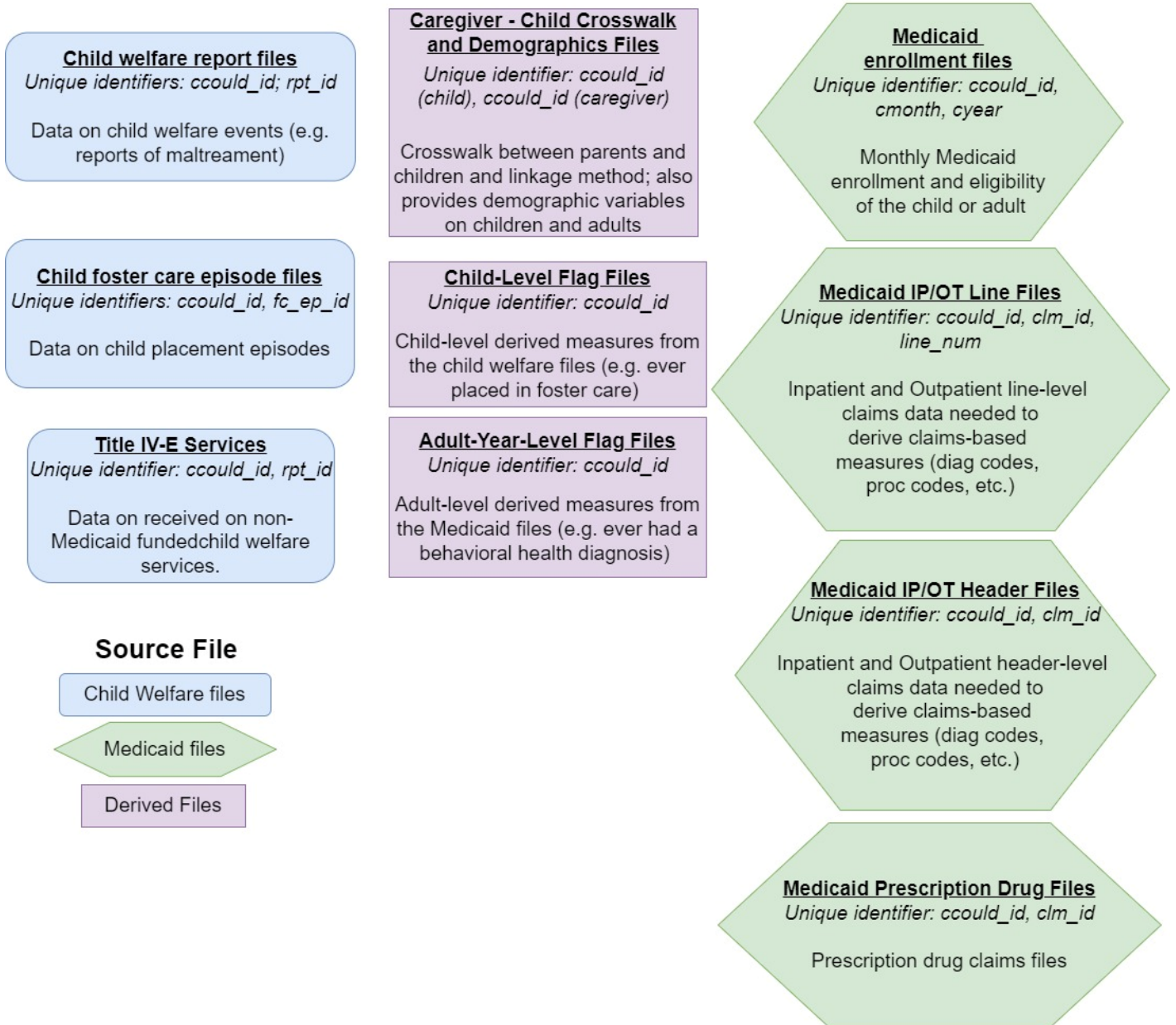
Child welfare report files
Unique identifiers: ccould_id; rpt_id

Data on child welfare events (e.g. reports of maltreament)

Caregiver - Child Crosswalk and Demographics Files
Unique identifier: ccould_id (child), ccould_id (caregiver)

Crosswalk between parents and children and linkage method; also provides demographic variables on children and adults

Medicaid enrollment files
Unique identifier: ccould_id, cmonth, cyear

Monthly Medicaid enrollment and eligibility of the child or adult

Child foster care episode files
Unique identifiers: ccould_id, fc_ep_id

Data on child placement episodes

Child-Level Flag Files
Unique identifier: ccould_id

Child-level derived measures from the child welfare files (e.g. ever placed in foster care)

Medicaid IP/OT Line Files
Unique identifier: ccould_id, clm_id, line_num

Inpatient and Outpatient line-level claims data needed to derive claims-based measures (diag codes, proc codes, etc.)

Title IV-E Services
Unique identifier: ccould_id, rpt_id

Data on received on non-Medicaid fundedchild welfare services.

Adult-Year-Level Flag Files
Unique identifier: ccould_id

Adult-level derived measures from the Medicaid files (e.g. ever had a behavioral health diagnosis)

Medicaid IP/OT Header Files
Unique identifier: ccould_id, clm_id

Inpatient and Outpatient header-level claims data needed to derive claims-based measures (diag codes, proc codes, etc.)

Source File

Child Welfare files

Medicaid files

Derived Files

Medicaid Prescription Drug Files
Unique identifier: ccould_id, clm_id

Prescription drug claims files

Figure 1. CCOULD Dataset Structure

Table 2. Identifiers used in the CCOULD data files

ID	Description	Files Containing ID
CCOULD_ID	Person-level identifier.	All files
CH_ID	Child identifier that can be used to link children to their caregivers. In the crosswalk file, the child and all linked caregivers are assigned the same value for CH_ID, and it is equal to the CCOULD_ID for that child.	CH-CG Crosswalk and Demographics file
RPT_ID	Report identifier.	Child welfare and Title IV-E services files
CLM_ID	Claim identifier.	Medicaid IP/OT Header and Line files; Medicaid Pharmacy Claims

Child Welfare Files. The child welfare report file is organized by child and unique report submitted for each child. The unique key variables are the child identifier (CCOULD_ID) and the report identifier (RPT_ID). The child foster care episode file is organized by child and removal date (FC_LST_RMV_DT) for the foster care episode. The Title IV-E Services file is organized by child and unique report submitted for each child. Data from the child welfare report and Title IV-E services files can be merged by matching on CCOULD_ID and RPT_ID. Data from the child foster care episode and Title IV-E services files can be merged by matching on CCOULD_ID and removal dates (FC_LAST_RMV_DT in the child foster care episode file and CH_SRVC_FOSTER_CR_DT in the Title IV-E services file).

Derived Files. The caregiver-child crosswalk and demographic file is organized at the person-level (CCOULD_ID). For both groups the person-level identifier is the CCOULD_ID. Data from these derived files can be merged by matching on CCOULD_ID.

Medicaid Files. The Medicaid files include an enrollment file, line files, header files, and a prescription drug file. The enrollment file is organized at the person-month-year level (CCOULD_ID-CMONTH-CYEAR). The Medicaid IP/OT Line Files are organized at a claim line level. Each record is a unique line within a claim submitted for a Medicaid beneficiary. To uniquely identify a record, use CCOULD_ID to identify the person on the claim, CLM_ID to identify the unique claim, and LINE_NUM to identify the specific line within a claim. The Medicaid IP/OT Header Files are organized

at the claim level and there is only one row of data for each unique claim. The purpose of the header files is to provide summary information about the claim, with more detail provided on the line file for each claim. The Medicaid Prescription Drug file is organized at a claim level only and can be uniquely identified by CCOULD_ID and CLM_ID. Data from these files can be merged by matching CCOULD_ID. To match lines to headers, also match on CLM_ID. To match enrollment data to the claim files (headers/lines or prescription drugs), also match the month and year of the enrollment record to the END_DT associated with each claim.

Identifying Caregivers. One of the main purposes of the caregiver-child crosswalk and demographics file is to allow the identification of all caregivers associated with a particular child. The child ID field (CH_ID) is unique to the combination of children and their parent(s)/caregiver(s), while CCOULD_ID is unique to the individual (see example in Table 3). Therefore, to identify all caregivers associated with a child, identify all of the CCOULD_IDs associated with records that share a CH_ID value.

Table 3. Example of how the data in the crosswalk file are organized around unique child and parent/caregiver combinations

CH_ID	CCOULD_ID	FAM_MBR_TYPE
1234	1234	1 (child)
1234	5678	2 (primary parent/caregiver)
1234	9999	3 (secondary parent/caregiver)

Confidentiality Protection

The CCOULD data contain no personal identifying information (PII) and are considered to be de-identified under the Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy Rule, following the Expert Determination Method. The steps taken to protect the data are as follows. First, all 18 types of identifiers under the Safe Harbor rule of the HIPAA Privacy Rule were removed from the files except for the month (and year) contained on the insurance claims service date records, the month indicated on the enrollment files, and the month indicated on the child welfare records.

18 Types of Safe Harbor Identifiers

1. Names
2. All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP code, and their equivalent geocodes, except for the initial three digits of the ZIP code if, according to the current publicly available data from the Bureau of the Census: (1) The geographic unit formed by combining all ZIP codes with the same three initial digits contains more than 20,000 people; and (2) The initial three digits of a ZIP code for all such geographic units containing 20,000 or fewer people is changed to 000
3. All elements of dates (except year) for dates that are directly related to an individual, including birth date, admission date, discharge date, death date, and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older.
4. Telephone numbers
5. Vehicle identifiers and serial numbers, including license plate numbers
6. Fax numbers
7. Device identifiers and serial numbers
8. Email addresses
9. Web Universal Resource Locators (URLs)
10. Social security numbers
11. Internet Protocol addresses
12. Medical record numbers
13. Biometric identifiers, including finger and voice prints
14. Health plan beneficiary numbers
15. Full-face photographs and any comparable images
16. Account numbers
17. Any other unique identifying number, characteristic, or code, (except as permitted by #18 below), and
18. Certificate/license numbers

Second, the data were screened and data reduction methods were applied. Data reduction methods protect against re-identification of respondents by making respondents more similar to each other on identifying variables. The data reduction methods applied were:

- Removal of ID variables for individuals and providers that are not necessary for analysis of children and caregivers. This includes IDs for case workers, medical providers, perpetrators of harm, etc. The remaining ID variables are encrypted and are those necessary to link and analyze the data files.
- Removal of binary variables with between 1 and 10 "Yes" cases in a state. If removal was only indicated in one state, the values were suppressed.
- Screening of categorical variables with small categories and coarsening where appropriate. In some cases small categories were allowed where they were not deemed to be identifying (e.g. taxonomies, drug codes) and there was not a natural way to apply coarsening. Rare diagnostic codes were coarsened by removing the last digit.
- Top- and bottom-coding of numeric variables, where outliers are present. For dollar amounts, the

values of the largest cases in each state were replaced with the mean of the outlying values, such that the mean values by state are preserved. Year of birth is bottom-coded at 1933 and perpetrator age is bottom-coded at 18.

Lastly, the data are protected by the security protocols specified in the data license and data user agreement. It is important that all researchers adhere to the terms of the license agreement.

The data use agreement specifies that the investigator will not release or disclose information where the number of observations (i.e., individual discharge records) in any given cell of tabulated data is <10. The publication of values of 1-10 is prohibited in text and tables.

<u>**Extent of Collection**</u>

Table 4 below displays the list of accompanying documentation for the data files (each of which come in SAS, SPSS, Stata, and text file formats).

Table 4. List of documents included in dataset, in addition to the User's Guide.

Document File Name	Brief description	Data File(s) to which the document is relevant
User Guide	This document contains the study-level metadata which describe integral facets of the study.	All data files
ccould_ch_welfare_report.pdf	This document contains the variable level metadata including the variable name, variable label, variable type, values, value labels, mean, range, and frequencies for each variable in the Child Welfare Report data file.	ccould_ch_welfare_report
ccould_title_iv_e_services.pdf	This document contains the variable level metadata including the variable name, variable label, variable type, values, value labels, mean, range, and frequencies for each variable in the Title IV-E Services data file.	ccould_title_iv_e_services
ccould_ch_fc_episodes.pdf	This document contains the variable level metadata including the variable name, variable label, variable type, values, value labels, mean, range, and frequencies for each variable in the Child Foster Care (FC) Episodes data file.	ccould_ch_fc_episodes
ccould_mdcd_enrollmt.pdf	This document contains the variable level metadata including the variable name, variable label, variable type, values, value labels, mean, range, and frequencies for each variable in the Medicaid enrollment data file.	ccould_mdcd_enrollmt

Document File Name	Brief description	Data File(s) to which the document is relevant
ccould_mdcd_enrollmt_compgrp.pdf	This document contains the variable level metadata including the variable name, variable label, variable type, values, value labels, mean, range, and frequencies for each variable in the Medicaid enrollment comparison group data file.	ccould_mdcd_enrollmt_compgrp
ccould_mdcd_ip_hedr.pdf	This document contains the variable level metadata including the variable name, variable label, variable type, values, value labels, mean, range, and frequencies for each variable in the Medicaid Inpatient (IP) Header data file.	ccould_mdcd_ip_hedr
ccould_mdcd_ip_hedr_compgrp.pdf	This document contains the variable level metadata including the variable name, variable label, variable type, values, value labels, mean, range, and frequencies for each variable in the Medicaid Inpatient (IP) Header comparison group data file.	ccould_mdcd_ip_hedr_compgrp
ccould_mdcd_ot_hedr.pdf	This document contains the variable level metadata including the variable name, variable label, variable type, values, value labels, mean, range, and frequencies for each variable in the Medicaid Other Therapy (OT) Header data file.	ccould_mdcd_ot_hedr
ccould_mdcd_ot_hedr_compgrp.pdf	This document contains the variable level metadata including the variable name, variable label, variable type, values, value labels, mean, range, and frequencies for each variable in the Medicaid Other Therapy (OT) Header comparison group data file.	ccould_mdcd_ot_hedr_compgrp
ccould_mdcd_ip_line.pdf	This document contains the variable level metadata including the variable name, variable label, variable type, values, value labels, mean, range, and frequencies for each variable in the Medicaid Inpatient (IP) Line data file.	ccould_mdcd_ip_line
ccould_mdcd_ip_line_compgrp.pdf	This document contains the variable level metadata including the variable name, variable label, variable type, values, value labels, mean, range, and frequencies for each variable in the Medicaid Inpatient (IP) Line comparison group data file.	ccould_mdcd_ip_line_compgrp

Document File Name	Brief description	Data File(s) to which the document is relevant
ccould_mdcd_ot_line.pdf	This document contains the variable level metadata including the variable name, variable label, variable type, values, value labels, mean, range, and frequencies for each variable in the Medicaid OT Lines data file.	ccould_mdcd_ot_line
ccould_mdcd_ot_line_compgrp.pdf	This document contains the variable level metadata including the variable name, variable label, variable type, values, value labels, mean, range, and frequencies for each variable in the Medicaid OT Lines comparison group data file.	ccould_mdcd_ot_line_compgrp
ccould_mdcd_rx_line.pdf	This document contains the variable level metadata including the variable name, variable label, variable type, values, value labels, mean, range, and frequencies for each variable in the Medicaid Pharmacy data file.	ccould_mdcd_rx_line
ccould_mdcd_rx_line_compgrp.pdf	This document contains the variable level metadata including the variable name, variable label, variable type, values, value labels, mean, range, and frequencies for each variable in the Medicaid Pharmacy comparison group data file.	ccould_mdcd_rx_line_compgrp

Extent of Processing

The data contributor prepared the Users Guide, Codebook, data file(s) and supporting documentation which were reviewed/revised by NDACAN staff prior to the dataset's release.

Data File Information

File Specifications

There are 16 data file(s) containing 489 variables in total. A full list of the included data files is detailed in Table 5 below. The data files are all available in SAS, Stata, SPSS, and text file formats.

Table 5. Data File Names and Descriptions

Data File Name	Description	Number of variables	Number of observations
ccould_ch_welfare_report	This file contains data on child welfare events, e.g., maltreatment reports.	104	1,841,908
ccould_title_iv_e_services	This file contains data on Title IV-E services provided to families.	40	1,841,908
ccould_ch_fc_episodes	This file contains data on child placement episodes.	116	105,104
ccould_ch_cg_crosswalk	Crosswalk between caregivers and children, as well as demographic varibles for the children and caregivers.	10	1,178,341
ccould_mdcd_enrollmt	Medicaid enrollment and eligibility of the person.	10	53,395,248
ccould_mdcd_enrollmt_compgrp	Medicaid enrollment and eligibility of the person (comparison group).	10	62,267,071
ccould_mdcd_ip_hedr	Medicaid claims headers: Inpatient diagnostic and procedure codes.	43	687,100
ccould_mdcd_ip_hedr_compgrp	Medicaid claims headers: Inpatient diagnostic and procedure codes (comparison group).	43	728,707
ccould_mdcd_ip_line	Medicaid claim lines: Inpatient diagnostic and procedure codes.	56	4,364,497
ccould_mdcd_ip_line_compgrp	Medicaid claim lines: Inpatient diagnostic and procedure codes (comparison group).	56	5,203,789
ccould_mdcd_ot_hedr	Medicaid claims headers: Outpatient diagnostic and procedure codes.	42	83,964,959
ccould_mdcd_ot_hedr_compgrp	Medicaid claims headers: Outpatient diagnostic and procedure codes (comparison group).	42	83,048,726
ccould_mdcd_ot_line	Medicaid claim lines: Outpatient diagnostic and procedure codes.	54	153,378,745

Data File Name	Description	Number of variables	Number of observations
ccould_mdcd_ot_line_compgrp	Medicaid claim lines: Outpatient diagnostic and procedure codes (comparison group).	54	162,327,174
ccould_mdcd_rx_line	Medicaid claims: Prescription drugs filled.	14	44,732,243
ccould_mdcd_rx_line_compgrp	Medicaid claims: Prescription drugs filled (comparison group).	14	47,123,190

Data File Notes

Missing value codes vary by variable. Generally, unknown or missing values are coded as 9 or 99, and NULL values indicate cases where the variable was not collected or not applicable. Users should refer to the accompanying data file codebooks.

Medicaid data files with the suffix "_compgrp" relate to the comparison sample of individuals that are not associated with the child welfare system (see more details in the Sample section). Data without the "_compgrp" suffix relate to individuals who have interacted with the child welfare system.

Data users interested in using R, can go to our online User Support page to review help documents that provide instructions for pulling in the SPSS native data file or the tab-delimited data file into R (https://www.ndacan.acf.hhs.gov/user-support/user-support.cfm).

Acronyms and Abbreviations

Commonly used abbreviations in the study documentation and data files:

ACF – Administration for Children and Families

ASPE – Assistant Secretary for Planning and Evaluation

CCOULD – Child and Caregiver Outcomes Using Linked Data

CMS – Centers for Medicare and Medicaid Services

DUA – Data Use Agreement

FY – Fiscal Year

IP – In Patient

OT – Other Therapy

NDACAN – National Data Archive on Child Abuse and Neglect

OPRE – Office of Planning, Research & Evaluation

PCORTF – Patient-Centered Outcomes Research Trust Fund

PHI – Protected Health Information

PII – Personally Identifiable Information

RTI International – Research Triangle Institute International

SDA – Statistical Disclosure Analysis

SSN – Social Security Number

URL – Universal Resource Locators

Technical support for this dataset is provided by NDACAN.

Please send your inquiries to

NDACANsupport@cornell.edu

Check the User Support page of the NDACAN website for help documents related this dataset (https://www.ndacan.cornell.edu/user-support/user-support.cfm).