# CALHHS EQUITY DASHBOARD

## DATA QUALITY REPORT

SPRING 2024

# Table of Contents

## Purpose and Overview (Executive Summary)

This Data Quality Report is an overview of the creation of the Equity Dashboard, emphasizing the importance of accurate and comprehensive demographic data while spotlighting the valuable lessons learned and enhancements implemented along the way.

Key Takeaways:
- Initial struggles with standardizing racial identifiers underscored the necessity for additional controls and instructions in data collection to accommodate variability in categories used by departments.
- Varying reporting periods across programs and departments highlighted the importance of establishing standardized timelines to ensure consistency and accuracy in data reporting.
- Challenges in automating and converting paper forms into electronic forms underscored the need for ongoing adaptation to evolving legislation and resource limitations.
- Masking data for small-numbered program participants (less than 11) was a strategic decision made to uphold privacy standards and protect individual identities.
- Evaluation of Client Linkage revealed the need to implement a Business Use Case Proposal (BUCP) to enhance data management processes.
- Challenges in sharing geographical data, particularly from smaller counties, emphasized the necessity for robust privacy measures to address concerns about the potential identification of individuals.

These lessons highlight the complexities in data collection, standardization, and privacy protection encountered during the development of the CalHHS Equity Dashboard, emphasizing the ongoing efforts to enhance methodologies and address challenges for future iterations. By embracing these challenges as opportunities for growth, the Equity Dashboard team enhanced the understanding and also paved the way for more effective strategies moving forward.

## Equity Dashboard Overview

The CalHHS Equity Dashboard is a Managed Analytics Project commissioned by CalHHS Undersecretary Marko Mijic, sponsored by CalHHS Chief Equity Officer Dan Torres, and developed in partnership with departments and programs across CalHHS. The Equity Dashboard (EDB) is a cross-departmental tool designed to help Agency, departmental leadership, and the public to better understand the Californians who use CalHHS programming, and to identify and address disparities in CalHHS services. The Equity Dashboard was released to CalHHS department and agency staff on July 13, 2023 and will be updated iteratively.

The current iteration of the Equity Dashboard includes race, ethnicity, sexual orientation, and gender identity demographic data from the following departments: the Department of Rehabilitation (DOR), the Department of Managed Health Care (DMHC), the Department of Health Care Services (DHCS), the Department of State Hospitals (DSH), Department of

Development Services (DDS), the California Department of Aging (CDA), and the California Department of Public Health (CDPH).

## Lessons Learned

### Data Collection

Demographic data collection varies greatly by data element collected, method of collection, level of detail, and data type across departments and even across programs within the same department. Most programs adhere to standards dictated by specific state or federal legislation, which are often directly tied to a program's funding source. Additionally, in many cases, department or program data systems are manual and/or outdated, causing difficulties in changing current data standards or data collection practices.

Through the development of the Equity Dashboard, a need for data collection standardization became apparent. A Data Standards Community was established to provide CalHHS with the guidelines for demographic data best practices and to identify and suggest department standards. In May 2023 the eight (8) departments[1] that participated in the Data Standards Community were sent a survey to pinpoint the 3-5 most common legislative mandates for demographic data (race, ethnicity, sexual orientation, and gender identity) collection and reporting. Responses were received from seven (7) of the eight (8) departments. Each department follows multiple legislation and mandates. The only common collection and reporting standards are Assembly Bill 1726, and the 1997 OMB Standard, and these are only common across a few CalHHS departments.

### Open Data Portal

The Equity Dashboard team researched existing CalHHS datasets and data sources to assess viable usage within the Equity Dashboard. The Open Data Portal was originally perceived as a possible resource for gathering program-specific demographic data. As an existing repository that houses public data collected by the state through its routine business activities, the Open Data Portal was proposed as a data source for CalHHS demographic data. The EDB technical team focused on two data sets that appeared to be useful: The Record Reconciliation and the COVID datasets. Both data sets presented a unique set of challenges that rendered them unusable for the specific use case of the CalHHS Equity Dashboard. The scope of the CalHHS Equity Dashboard initially aimed to determine discrepancies in programmatic demographic data collection; these data sets were not able to fulfill that scope.

---

[1] Departments participating in the CalHHS Data Standards Community: California Department of Social Services (CDSS), California Department of Public Health (CDPH), DDS,Department of Health Care Services (DHCS), Department of Managed Health Care (DMHC), Department of State Hospitals (DSH), Department of Rehabilitation (DOR), Department of Health Care Access and Information (HCAI).

### Record Reconciliation

Four of the twelve departments that collect demographic data are using the Record Reconciliation data file set. Each data set must be analyzed individually to determine if it applies to the EDB. An in-depth analysis[2] by the EDB technical team determined the specific issues with the Record Reconciliation data file. Overall, the Equity Dashboard team discovered that the demographic data included in the RR annual data file is secondary to the report's primary focus. However, interpreting the aggregations in relation to how individual departments collect demographic data poses uncertainty, particularly regarding racial and ethnic identifiers. The format required by the Equity Dashboard presents challenges in accurately categorizing these identifiers. Additionally, the RR annual data file does not represent the methodology or standard of data collection by individual department programs, and aggregated fields like "other_missing" and "Unknown_Other gender" combine actual and missing or unidentified responses, further complicating data analysis.

### COVID Data Sets

Several metrics of data are collected in the COVID data sets: cases, fatalities, testing based on race, ethnicity, and other social determinants of health, such as income, housing conditions, and access to healthcare coverage. This specific data set only represented individuals who reported COVID-19 symptoms and infections and is not able to be broken down on the programmatic level.

Ultimately the Open Data Portal datasets were unusable for the first iteration of the Equity Dashboard as it was concluded they did not align with the specific needs of the Equity Dashboard. As the scope of the Equity Dashboard evolves the usability of the Open Data Portal may change. To continue moving forward with the creation of the Equity Dashboard alternate approaches to collecting departmental and programmatic demographic data were explored.

## Race and Ethnicity

The technical engineers initially struggled with the racial identifiers data provided by the departments, as it could not be standardized to the Race, Ethnicity, Sexual Orientation, and Gender Identity (RE/SOGI) measures level. For example, certain departments combined Hispanic/Latino with other racial categories, while others treated Hispanic/Latino as a distinct ethnic category.

Recognizing the need for standardization, the engineers reassessed the data collection process. They identified the diverse methods by which departments collected information and concluded that it was imperative to provide comprehensive instructions and incorporate examples into the second workbook template. The team

---

[2] Analysis can be found in the appendix section

found during the second of the data collection template that providing specific examples of how to categorize the various data points and scenarios on how to fill out the workbook closed the gap in the majority of misunderstandings that departments' staff had during the first iteration of the collection template.

### Gender Identity vs Gender Assigned at Birth (Census)

The U.S. Census identifies individuals based not on gender identity, but on gender assigned at birth. This initially presented a problem to the technical team as it was unclear if the gender data presented was based on gender identity or assigned gender. The Equity Dashboard demographic data looks to identify gender identity not gender assigned at birth. To mitigate any possible confusion, a definition for "gender identity" was provided in the second iteration of the data collection template. This allowed programmatic gender identity data to be aligned across departments and programs.

### Reporting Periods

Due to varying data collection and reporting standards, each program (even within the same department) has a different reporting period. This presented difficulties for the technical team when creating visuals for the Equity Dashboard. The Department of Rehabilitation for example has a quarterly reporting period. However, the bar charts only visualize data annually (fiscal and gregorian). To rectify this, if departments provide quarterly data, the 4th quarter is represented for previous years and the most recent quarter is represented for the current year. The dashboard's graphs' axes reflect this.

### Data Collection Methods

Programmatic data collection methods presented a unique opportunity for the Equity Dashboard technical team. CalHHS departments indicated that automating and/or converting paper forms into electronic forms presents an obstacle due to evolving privacy legislation, resource limitations, and accessibility requirements.  In many cases, department or program databases are outdated, leading to difficulty in standardizing data collection. The Equity Dashboard team is working with the Data Standards Community and CalHHS leadership to create solutions and best practices to bring data collection into the 21st century.

### De-identified Data

While exploring the possibility of using the Record Reconciliation as a data source, a significant amount of de-identified data was presented. The Equity Dashboard team received de-identified data from each of the departments to prevent exposure of small-numbered data - data values, counts, and aggregations based on program participants numbering less than 11.

De-identification refers to the process of removing or obscuring any personally identifiable information to minimize the risk of unintended disclosure of individual identities. Masking and suppression are methods utilized to protect data privacy, which

involves removing or modifying data to prevent the identification of individuals in small groups or those with unique characteristics.

While there was a readily available solution for de-identified data, the technical team realized the impact program size has on creating pipelines for the Equity Dashboard. Because of varying program size, both at the state level and at the county level some counties have data values of zero that do not require masking, while others with a small number of participants need their data to be suppressed or hidden. As a result, the technical team incorporated explicit instructions on how to handle data suppression and true-0 counts in the second iteration of the Data Collection Template. These instructions have been sourced from the CalHHS Suppression Standards.

### Confidential Data and Business Use Case Proposal (BUCP)

The Equity Dashboard technical team evaluated the Client Linkage when reviewing the Record Reconciliation datasets. Client Linkage tracks program participants who are in multiple programs during a single reporting period. It was determined that a Business Use Case Proposal (BUCP) was needed to allow CDII to utilize confidential data in both the Data Hub and the Equity Dashboard.

CDII and the Equity Dashboard technical team are currently working with CDSS to create a BUCP to request the following entities for all CDSS programs:
1. Client details (PII file used for CDN Client Linkage)
2. Facilities/License details
3. Service Provider details (may be the same as the license/facility data)

This BUCP will be amended in the future to include additional departments as the Equity Dashboard evolves. While this information is already being provided to the Children Data Network, a BUCP is needed for the technical team to use the raw, client-level data to aggregate race, ethnicity, sexual orientation, and gender identity data from the necessary departments, specifically CDSS. This provides a unique opportunity to create a foundation for future BUCPs with other departments allowing CDII and the Equity Dashboard team to document lessons learned over the last year to streamline future efforts.

### Geographical Data

Sharing geographical data poses a significant challenge for the Equity Dashboard technical team, particularly regarding demographic data obtained from smaller counties and programs. The potential identification of individuals based on shared demographic information, which includes sensitive details such as race, ethnicity, sexual orientation, and gender identity poses a security risk. Unlike larger jurisdictions, smaller counties and programs often have fewer participants in programs, making it more likely for individuals to be uniquely identifiable through the disclosed data.

As the Equity Dashboard seeks to provide a comprehensive overview of demographic trends across different regions, the balance between data transparency and privacy protection serves a unique challenge. The concern is that releasing detailed geographical information may inadvertently lead to the identification of individuals, compromising their privacy and confidentiality. This challenge is further amplified by the varying reporting periods and data collection methods adopted by different counties, adding complexity to creating meaningful visualizations for the Equity Dashboard.

To address these privacy concerns, the technical and strategic teams are actively exploring strategies to enhance data anonymization and protect individual identities within smaller counties. This includes evaluating methods such as aggregation at different geographical levels to ensure that released data is sufficiently de-identified while still providing the dashboard user with value. Moreover, the team recognizes the need for explicit guidelines on handling geographical data within the Data Collection Template, drawing on the lessons learned from previous challenges, such as those encountered with de-identified data and reporting periods.

## Next Steps

After considering the lessons learned in this Data Quality Report and optimizing its resources, the Equity Dashboard team has decided to refocus on finalizing a BUCP with CDSS allowing the technical team to access and share disaggregated data. This will be the blueprint for other CalHHS departments. The Equity Dashboard technical team will be able to aggregate client-level data as needed to ensure client safety and adhere to agency standards.  As the dashboard evolves the Equity Dashboard team will continue to document lessons learned and share them with our data and equity partners.

## Conclusion

As the CalHHS Equity Dashboard evolves, so will the data methodology. The technical team will continue to mitigate and document data occurrences that impact data and could benefit future CalHHS data initiatives and product creation.

If you have any questions or would like further information about the CalHHS Equity Dashboard or this report's contents, please get in touch with the Equity Dashboard team at cdii@chhs.ca.gov.

# APPENDIX

This section provides an in-depth, detailed view of the process and approach that resulted in this Data Quality Report. As well as the documents mentioned in the body of the report.

# RECORD RECONCILIATION ANALYSIS

The Equity Dashboard data engineers attempted to use the Record Reconciliation Annual Data file to acquire demographic data for 4 departments across 9 of their respective programs. At first glance, this approach seemed to work albeit with caveats. However, due to the problem statements, this approach did not work for the first iteration of the Equity Dashboard.

| Problem 1 | Problem 2 | Problem 3 | Problem 4 | Problem 5 | Problem 6 |
|---|---|---|---|---|---|
| The demographic data included in the RR annual date file is not the primary focus of the report. | We cannot be sure of what the aggregations mean in the context of how individual departments' programs are collecting the demographic data that is reported in the RR annual data file. | The RR annual data file contains the aggregations of demographic data in columns such as racial/ethnic identifiers such as: Black, White, Hispanic, Asian_PI, Native American, Other_Missing, which makes it impossible for the ED data engineers to categorize these identifiers accurately in the format that the ED Dashboard requires: Race measure; Ethnicity measure. | The RR annual data file is by nature, a 'reporting' of multiple departments' programs' participants and does not represent the methodology or standard on how those individual department programs 'collected' the data from the participants. | The aggregated field value of other_missing as a racial/ethnic identifier column combines actual responses and missing responses.<br><br>The aggregated field value of Unknown_Other gender as a gender identifier column combines actual responses and unidentified responses | The client linkage data elements sent from CDSS to CDN do not match the EDB Program Inventory Compiled Survey. |
| **Whys** | | | | | |
| This RR data file is a report that is generated by at third-party organization: Children's Data Network (CDN). CDN utilizes an AI model to generate the report for criteria matching at the 'client/person' level to provide a cross- | Because Children's Data Network (CDN) produces the RR Annual Data File, we don't know the logic of how the aggregations are calculated. Nor do we have any visibility on the | We are unable to determine how to map the racial/ethnic identifiers in the RR Annual Data file to the ED Dashboard measures: Ethnicity and Race. Thus, the question we cannot answer is "what racial/ethnic | The measures that the ED team are trying to ascertain from the departments' programs pertain to demographic data 'collection' and not necessarily 'reporting' | We are unable to separate the race other/missing aggregated field to derive people who responded to other vs. people who did not respond at all e.g missing.<br><br>We are unable to separate the race Unknown_Other | The client linkage data that is sent from CDSS to CDN is sourced from DHCS's MEDS (Medi-cal) system. Thus, demographic data elements included for CDSS-specific programs are not included. Examples of this are |

| tabular matrix of aggregations of people who have participated in multiple programs across departments during specified time/frames and locations. | departmental program datasets that that were provided to CDN to generate the RR annual data file. | identifiers in RR annual data file constitutes Race vs Ethnicity per department programs?" Also, the ED Dashboard is concerned with program participant aggregations at the measure-level and not at the racial/ethnic identifier-level. | standards which is what RR Annual Data file seems to be. | gender aggregated field to derive people who responded to other vs. people who are unknown.<br><br>Because these are CDN aggregates, the departments do not know how these aggregations were calculated and thus we cannot conclusively group them into our measure aggregates | the CDSS administered program, In Home Support Services (IHSS). IHSS collects SOGI data and stores it in CMIPS DDL system. However, this SOGI data is not included in the MEDS data file that DHCS produces for CDSS to transform and sent to CDN. |
|---|---|---|---|---|---|
| | | | **Final Problem Statement** | | |
| Because the RR Annual Data file is produced by a third-party and is a general report/representation of multiple departments' programs, the ED data engineers are finding that there are too many unknowns and variables working with the RR Annual Data file to provide accurate aggregations in the ED data pipelines. | | | | | |

## Equity Dashboard Deep Dive

The Equity Dashboard technical team was tasked with using existing tools and creating new data pipelines to automate the ingestion of demographic data in the Equity Dashboard. As the Equity Dashboard was planned to be released iteratively, a similar approach was taken to gathering demographic data and creating data pipelines.

The EDB technical team worked with each department to answer four questions:
1. What demographic data is currently collected by departments?
2. How complete is the collection of demographic data?
3. How many people participate in each program?
4. What is the demographic composition of program participants?

The Equity Dashboard technical team originally cast a wide net sending a Quantitative Program Inventory to all 12 CalHHS departments. In tandem, the team researched existing data sets in the Open Data Portal. They then narrowed their request and began working directly with the Department of State Hospitals (DSH), Department of Managed Health Care (DMHC), Department of Rehabilitation (DOR), and Department of Health Care Services (DHCS). The first Data Collection Template allowed the EDB technical team to begin the conversations about how demographic data is collected. The second version of the Data Collection Template expanded and improved upon the first iteration and was customized to the specific needs of the Equity Dashboard.

# DATA TOOLS

This section provides an in-depth, detailed view of the process and approach that resulted in this Data Quality Report. As well as the documents mentioned in the body of the report.

# QUANTITATIVE PROGRAM INVENTORY

In 2022, in an effort to understand current demographic data collection practices and standards, the Equity Dashboard team received program inventories from the 12 CalHHS departments. The program inventory focused on the programmatic collection of race, ethnicity, sexual orientation, and gender identity demographic data. The Equity Dashboard team was subsequently able to identify demographic data collection challenges, opportunities for improvement, and what support would advance their demographic data collection efforts.

# CalHHS/CDII Equity Dashboard:  Program Inventory and Equity Data Survey

**Version 2**
**Last Updated as of 4/5/2022**

**Purpose:**
California's Center for Data Insights and Innovation (CDII) is developing an implementation plan for the 'Equity Dashboard' as funded and described in the 'Equity-Centered Programs' CalHHS BCP1.  To inform the planning, CDII is conducting interviews with department leaders across CalHHS to understand current equity priorities, their successes and challenges in collecting and utilizing data to understand the populations they serve and measure disparities, and hear from them what Agency-wide support might be most helpful to accelerate their progress. The Equity Dashboard will be implemented in phases, with the first phase focused specifically on assessing the current state of demographic data collection on race, ethnicity, sex (assigned at birth), sexual orientation, and gender identity across the Agency, a key first step to identifying disparities and advancing equity.

**Instructions:**
Following an initial interview with the Department's Director, Chief Equity Officer, and Chief Data Officer (CDO),  CDII will follow up with a survey to collect more detailed information on programs and program data sets/sources that contain beneficiariary information. This inventory of programs/data sets will collect information related to the demographic information collected within the program with a focus on understanding where programs have consistent, complete, and accurate beneficiary demographic information that can be used for equity measurement and tracking disparities reductions.

Sources:
1. Equity-Centered Programs: https://esd.dof.ca.gov/Documents/bcp/2122/FY2122_ORG0530_BCP4482.pdf

*For questions, please email Kameron Knab (kknab@manatt.com)*

# DATA COLLECTION TEMPLATE I

The Equity Dashboard technical team used the lessons learned while exploring the Open Data Portal datasets to create Data Collection Templates and pinpoint the data needs of the CalHHS Equity Dashboard. The metadata template provided with the Data Collection Template was fashioned after the template used to publish data to the Open Data Portal.

The first iteration of the Data Collection Template (DCT) was intended to collect the department's program data in a highly flexible way. This DCT was a generalized CDII template not specific to the EDB.

To improve the Data Collection Template and get the precise information needed to complete data pipelines:
The data engineering team encountered several issues while reviewing data collected using the first DCT from various departments, particularly when dealing with small, numbered data (less than 11), and defining data elements (race, ethnicity, sexual orientation, and gender identity). This prevented continuity between program data and hindered the ability to provide usable data to the dashboard.

Version Enhancement Findings

The Data Collection Template underwent several enhancements from version 1 to version 2. These enhancements were carefully thought out after departments has many misunderstandings of what data we were asking for and how we wanted the DCT to be filled out. Are findings are below:

- Because of the de-identified data received during version 1, the team had difficulty in displaying data at the county level and at the variable (detailed breakdown of race, ethnicity, sexual orientation, and gender identity) level due to program size and/or county population size.
    - o V2 Enhancement (and EDB Decision): The singular data table sheet was split into 4 separate data table tabs to collect total program participants by California and California county and to collect the high-level demographic data points within those programs by California and California county.
    - o V2 Enhancement (and EDB decision): Remove the 'variable' data column that allowed for drill down of detailed race, ethnicity, sexual orientation, and gender identity data values. This modified the Count columns to only represent the program participants who's race, ethnicity, sexual orientation, and gender identity data responses were collected and not any further detailed breakdown. This also placed the responsibility of categorizing detailed breakdown data points into the higher level categories (race, ethnicity, sexual orientation, and gender identity) on the responding department.
- When using version 1, respondents were unsure of how to fill out the template as it lacked enough instruction.
    - o V2 Enhancement: Added as the first tab, an Instruction sheet that outlined in sections the purpose of each data table sheet, how to fill out each data table sheet in an explicit order, how to (with examples and scenarios) aggregate their data for each tab's count column, and what was expected for de-identification.
- When using version 1, respondents were able to enter invalid data in columns. For example, inputting "three" in the Count column when "3" was actually expected. Similarly with date formats and measure inputs
    - o V2 Enhancement: Place data type and formatting controls on the worksheet columns to ensure that respondents had to follow specific data input rules.
- When using version 1, there was nowhere to collect the department's program's demographic data collection and reporting standards and/or methods.
    - o V2 Enhancement: Added a worksheet to the DCT asking for each the responding Department's Program's data collection and reporting standard in free-form textual format.

# DATA COLLECTION TEMPLATE II

The second DCT contained three elements used to capture demographic data in machine-readable format: a Data Dictionary tab, a Reference Data tab, and a Demographic Data tab. Together, with the instructions tab, they were utilized to format the demographic data used in the Equity Dashboard.

The second Data Collection Template made several assumptions about the data provided:

1. The data provided has been aggregated (summarized) and effectively de-identified
2. Each demographic data field/area is defined as a measure (e.g. sexual orientation), and all measures are summarized individually to avoid/reduce small cells
3. Data will be provided in a machine-readable 'narrow/tall' format to allow for standardized data pipeline ingestion

While there were several improvements The Equity Dashboard technical team failed to instruct departments on how to indicate the suppression type used, resulting in inconsistent representation across programs. Some used one asterisk, while others entered "<11" or "small value." This inconsistency exemplified the diverse ways the departments reported their data. The lesson learned here was the need to define what values must be entered to denote the type of suppression in future iterations of the DCT.

## Equity Dashboard Demographic Data Collection Template

This workbook contains the data collection templates (worksheets) used to define the format for data provided by CalHHS departments. CDII requests that departments use these templates.

Table of Contents (by row #):
-3: Data Collection Worksheet Information
-11: Count Aggregation Instructions
-22: De-Identification Instructions
-28: Additional Instructions

### Data Collection Worksheet Information (The worksheets are to be completed in numerical order 1-5)

| Order of Completion | Worksheet Name | Usage Guidance | Explanation | Examples (Related worksheet column letters in square brackets [ ]) |
|---|---|---|---|---|
| 1 | 1-Total State Participants | For a reporting time period (quarterly preferred), this sheet should be populated with a count of unique program participants for the entire state of California.<br><br>The counts in this sheet should be higher than the counts in worksheet 2. | Each unique individual who participated in the program, state-wide, should be included in this total count, regardless of whether demographic details were collected for this individual or not. **Each individual should be counted only once**, even if they participated in a specific program more than once during the reporting time period. | In the entire state of California[F] during the reporting period from 1/1/2021[C] to 4/1/2021[D], Program 1[B] served 25364[H] total deduplicated people |
| 2 | 2-Demographics by State | For a reporting time period (quarterly preferred), this sheet should be populated with a count of person-level responses for each demographic category for the whole state of California.<br><br>The counts in this sheet should be **lower** than the counts in worksheet 1. | Each unique individual who participated in the program, by state, should be included in the demographic measure counts that their response has been categorized as. **Each individual should be counted only once** per demographic measure, even if they responded with multiple selections of race/ethnic identifiers during the reporting time period specified. | For Program 1[B], in the entire state of California[F] during the reporting period from 1/1/2021 [C] to 4/1/2021[D],<br>8754[H] deduplicated people responded to an Ethnicity[G] question,<br>7584[H] deduplicated people responded to a Race[G] question,<br>7244[H] deduplicated people responded to a Gender Identity[G] question,<br>253[H] deduplicated people responded to a Sexual Orientation[G] question |
| 3 | 3-Total County Participants | For a reporting time period (quarterly preferred), this sheet should be populated with a count of unique program participants for each of the 58 California counties.<br><br>The counts in this sheet should be **higher** than the counts in worksheet 4. | Each unique individual who participated in the program, by county, should be included in this total count, regardless of whether demographic details were collected for this individual or not. **Each individual should be counted only once**, even if they participated in a specific program more than once during the reporting time period. | For Alameda[F] county during the reporting period from 1/1/2021[C] to 4/1/2021[D], Program 1 [B] served 25364[H] total deduplicated people |
| 4 | 4-Demographics by County | For a reporting time period (quarterly preferred), this sheet should be populated with a count of person-level responses for each of the 58 California counties.<br><br>The counts in this sheet should be **lower** than the counts in worksheet 3. | Each unique individual who participated in the program, by county, should be included in the demographic measure counts that their response has been categorized as. **Each individual should be counted only once** per demographic measure, even if they responded with multiple selections of race/ethnic identifiers during the reporting time period specified. | For Program 1[B], in Alameda[F] county during the reporting period from 1/1/2021[C] to 4/1/2021[D],<br>758[H] deduplicated people responded to an Ethnicity[G] question,<br>1025[H] deduplicated people responded to a Race[G] question,<br>1001[H] deduplicated people responded to a Gender Identity[G] question,<br>25[H] deduplicated people responded to a Sexual Orientation[G] question |
| 5 | 5-Data Standard Statements | Please explain the standard used to collect the demographic data as well as the standard(s) used to report out the demographic data.<br><br>Please also explain any nuance to the data counts in worksheets 2-4. | Given a demographic section in the program form/application, what standard has been used to develop that section?<br><br>For that data collected, what reporting rules/standards is the data transformed with to send to whoever requires the data reported? | "For Program A, demographic data is collected using the State standard referenced in legislation ABC. Race and Ethnicity are collected as a single measure. However, because Program A serves children 1-12 months old, gender identity and sexual orientation are not collected due to appropriateness."<br><br>"For Program A, demographic data is reported to Agency XYZ using the Federal standard referenced in legislation ABC." |

### Count Aggregation Instructions (count column [H] in each data worksheet)

| Question | Answer | Program Demographic Collection Method (forms, online application, etc...) Examples | Worksheet `Count` Column Action | Scenario Example |
|---|---|---|---|---|
| **For the [specified] program, are race and ethnicity collected as the same category or separate categories?** | Race and ethnicity are collected as one category | Program paper/electronic form asks the participant to select the Race and Ethnicity that the participant identifies with.<br><br>Example:<br>What race/ethnicity(s) do you identify with?<br>a. Black or African American<br>b. Asian/Pacific Islander<br>c. White/Caucasian<br>d. Hispanic/Latino<br>e. Native American<br>f. Other<br>g. Choose not to Identify / Declined to State | If there has been any response to the Race/Ethnicity question, including "Choose not to identify/respond," the response will be considered as one and should be included in the Race AND Ethnicity count aggregations.<br><br>Multi selection is considered as one response. Regardless of how many racial or ethnic identifiers a person has selected the response will be considered as one.<br><br>Enter the same number of deduplicated people in the Count column for both the Race and Ethnicity categories (measure column). | The program had 100 total deduplicated participants.<br>85 of the participants provided some answer to this question (including some multi-selections and choices not to identify).<br>The other 15 participants left the question blank and did not respond.<br><br>**85 should be recorded under BOTH Race and Ethnicity.** |
| | Race is collected but Ethnicity is NOT collected. | Program paper/electronic form asks the participant a single question to select the Racial identifier(s) that the participant identify(s) with.<br><br>Example 1:<br>What Race(s) do you identify with?<br>a. Black or African American<br>b. Asian/Pacific Islander<br>c. White/Caucasian<br>d. Native American<br>e. Other<br>f. Choose not to Identify / Declined to State<br>*This example collection form uses the Federal standard with Ethnicity defined as "Hispanic/Latino", but is not collected in the example above.<br><br>Example 2:<br>What Race(s) do you identify with?<br>a. Black or African American<br>b. Asian/Pacific Islander<br>c. White/Caucasian<br>d. Hispanic/Latino<br>e. Native American<br>f. Other<br>g. Choose not to Identify / Declined to State | If there has been any response to the Race question, including "Choose not to identify/respond," the response will be considered as one and should be included in only the Race count aggregations.<br><br>Multi selection is considered as one response. Regardless of how many racial or ethnic identifiers a person has selected the response will be considered as one.<br><br>Enter the number of deduplicated people in the Count column for Race category (measure column). Leave the Count column BLANK for the Ethnicity category (measure column). | The program had 200 total deduplicated participants.<br>150 of the participants provided some answer to this question (including some multi-selections and choices not to identify).<br>The other 50 participants left the question blank and did not respond.<br><br>**150 should be recorded under ONLY Race and Ethnicity should be left blank.** |
| | Ethnicity is collected but Race is NOT collected. | Program paper/electronic form asks the participant a single question to select the Ethnicity identifiers that the participant identifies with.<br><br>Example 1:<br>What Ethnicity(s) do you identify with?<br>a. Hispanic/Latino<br>b. Not Hispanic/Latino<br>c. Choose not to identify/respond<br>*This example program's collection form uses the Federal standard with Ethnicity defined as "Hispanic/Latino", but is not collected in the example above.<br><br>Example 2:<br>What Ethnicity(s) do you identify with?<br>a. Black or African American<br>b. Asian/Pacific Islander<br>c. White/Caucasian<br>d. Hispanic/Latino<br>e. Native American<br>f. Other<br>g. Choose not to identify/respond | If there has been any response to the Race question, including "Choose not to identify/respond," the response will be considered as one and should be included in the Race count aggregation.<br><br>Multi selection is considered as one response. Regardless of how many racial or ethnic identifiers a person has selected the response will be considered as one.<br><br>Enter the number of deduplicated people in the Count column for Ethnicity category (measure column). Leave the Count column BLANK for the Race category (measure column). | The program had 200 total deduplicated participants.<br>150 of the participants provided some answer to this question (including some multi-selections and choices not to identify).<br>The other 50 participants left the question blank and did not respond.<br><br>**150 should be recorded under ONLY Ethnicity and Race should be left blank.** |

| | | Program paper/electronic form asks the participant to select the Ethnicity identifier(s) that the participant identifies and a question to select the Race identifier(s) that the participant identifies with. | If there has been any response to the Ethnicity question, including "Choose not to identify/respond," the response will be considered as one and should be included in the Ethnicity count aggregation. | The program had 100 total deduplicated participants. |
|---|---|---|---|---|
| | Race and ethnicity are collected as separate categories | Example: What Ethnicity(s) do you identify with? a. Hispanic/Latino b. Not Hispanic/Latino c. Choose not to identify/respond  What Race(s) do you identify with? a. Black or African American b. Asian/Pacific Islander c. White/Caucasian d. Native American e. Other f. Choose not to identify/respond | If there has been any response to the Race question, including "Choose not to identify/respond," the response will be considered as one and should be included in the Race count aggregation.  Multi selection is considered as one response. Regardless of how many racial or ethnic identifiers a person has selected the response will be considered as one.  Enter the number of deduplicated people in the Count column for Race category (measure column). Enter the number of deduplicated people in the Count column for Ethnicity category (measure column). | 80 of the participants provided some answer to the Ethnicity question (including some multi-selections and choices not to identify). 20 participants left the question blank and did not respond. **80 should be recorded for Ethnicity**  90 of the participants provided some answer to the Race question (including some multi-selections and choices not to identify). 10 participants left the question blank and did not respond. **90 should be recorded for Race.** |
| **For the [specified] program, is Gender Identity collected? Gender Identity only counts for the people that specifically choose their gender, NOT their assigned gender at birth.** | Gender assigned at birth is collected BUT Gender Identity is not collected | Program paper/electronic form asks the participant to select the gender (assigned at birth)/sex choice(s) that the participant identifies with.  Example: What is your gender (assigned at birth)/sex? a. Male b. Female c. Choose not to Identify / Declined to State | If there has been any response to a Gender assigned at birth question or similar, including "Choose not to identify/respond", the response should not be included in the Gender Identity count aggregation.  Leave the Count column BLANK for the Gender Identity category (measure column). | The program had 50 total deduplicated participants. All 50 of the participants provided some answers to the Gender (assigned at birth) question.  **No number should be recorded for Gender Identity** |
| | Gender Identity is collected BUT Gender assigned at birth is not. | Program paper/electronic form asks the participant to select the gender identity that the participant identifies with.  Example: What gender(s) do you identify as? a. Male b. Female c. Transgender Male to Female d. Transgender Female to Male e. Another Gender Identity f. Non-Binary g. Other h. Choose not to Identify / Declined to State | If there has been any response to a Gender Identity question or similar, including "Choose not to identify/respond", the response will be considerd as one and should be included in the Gender Identity count aggregation.  Enter the number of deduplicated people in the Count column who responded to the gender identity question in the Gender Identity category (measure column). | The program had 50 total deduplicated participants. 50 of the participants provided some answers to the Gender Identity question (including some multi-selections and choices not to identify).  **50 should be recorded for the Gender Identity.** |
| | Both Gender Identity and Gender are collected | Program paper/electronic form asks the participant to select the gender identity identifier(s) that the participant identifies and a question to select the gender (assigned at birth)/sex identifier(s) that the participant identifies with.  Example: What is your gender (assigned at birth)/sex? a. Male b. Female c. Choose not to Identify / Declined to State  What gender(s) do you identify as? a. Male b. Female c. Transgender Male to Female d. Transgender Female to Male e. Another Gender Identity f. Non-Binary g. Other h. Choose not to Identify / Declined to State | If there has been any response to a Gender Identity question or similar, including "Choose not to identify/respond" and/or a Gender (At birth)/Sex question, ONLY the Gender Identity response SHOULD be included in the Gender Identity count aggregation and should be considered as one.  Enter the number of deduplicated people in the Count column who responded to the gender identity question in the Gender Identity category (measure column). **Do not include the gender (assigned at birth)/sex respondents in this number.** | The program had 50 total deduplicated participants. 30 of the participants provided some answer to the Gender Identity question (including some multi-selections and choices not to identify). 15 participants responded to the Gender (assigned at birth) question.  **30 should be recorded for Gender Identity** |
| **For the [specified] program, is Sexual Orientation collected?** | Sexual Orientation is collected | Program paper/electronic form asks the participant to select the Sexual Orientation that the participant identifies with.  Example: What Sexual Orientation(s) do you identify with? a. Gay, Lesbian or Homosexual b. Bisexual c. Pansexual d. Queer e. Asexual f. Two-spirit g. Questioning/not sure h. Not listed, please describe i. Choose not to Identify / Declined to State | If there has been any response to a Sexual Orientation question or similar, including "Choose not to Identify / Declined to State", the response will be considered as one and should be included in the Sexual Orientation count aggregation.  Enter the number of deduplicated people in the Count column who responded to the Sexual Orientation question in the Sexual Orientation category (measure column). | The program had 50 total deduplicated participants. 30 of the participants provided some answer to the Sexual Orientation question (including some multi-selections and choices not to identify). 20 participants left the question blank and did not respond.  **30 should be recorded for Sexual Orientation** |

| **De-Identification (Suppression) Instructions (Please use the following values for the de-identified counts and/or program not collected)** |
|---|
| 0 = no data, but is collected |
| *1 = cell suppressed for small numbers (n=1-11) |
| *2 = cell suppressed for complementary cell (if exists) |
| *3 = not applicable/collected for program |

| **Additional Instructions** |
|---|
| For the Total Unique Participants By County and Demographics by County worksheets, prefilled records are provided for one program in a single reporting date ranges (quarterly preferred, but annual acceptable) across all 58 California counties. |
| If there are multiple programs and/or reporting date ranges, please copy the base prefilled records and change the program name and/or covereage dates |
| Please view the notes in the column headers |
| If there are any explanations required for the count columns, please put them in column K |
| If participants in a specific reporting period resided in multiple counties, please use the same methodlogy across all programs' data to determine the county to 'place' that participant. |

# LEGISLATIVE SURVEY

In May 2023 the eight (8) departments that participate in the Data Standards Community were sent a survey to pinpoint the 3-5 most common legislative mandates for demographic data (race, ethnicity, sexual orientation, and gender identity) collection and reporting. Responses were received from seven (7) of the eight (8) departments. Each department follows multiple legislation and mandates. The only common collection and reporting standards are Assembly Bill 1726, and the 1997 OMB Standard, and these are only common across a few CalHHS departments.

1.  Name

    _____

2.  Title

    _____

3.  Email

    _____

4.  Department

    _____

5.  Are there currently any state/federal legislative mandates and/or industry standards that your department adheres to for demographic data **collection**?

    *Mark only one oval.*

    ◯ Yes

6.  Are you currently required to collect **race and ethnicity** demographic data?

    *Mark only one oval.*

    ◯ Yes, we are required to collect both race and ethnicity demographic data

    ◯ No

    ◯ We are only required to collet race demographic data

    ◯ We are only required to collet ethnicity demographic data

7.  Please list up to five (5) state/federal **race** legislation and mandates and/or industry standards your department adheres to.

    _____

    _____

    _____

    _____

8.  Please list up to five (5) state/federal **ethnicity** legislation and mandates and/or industry standards your department adheres to.

_____

_____

_____

_____

_____

9.  Are you currently required to collect **sexual orientation and gender identity** demographic data?

    _Mark only one oval._

    ◯ Yes, we are required to collect both sexual orientation and gender identity demographic data

    ◯ No

    ◯ We are only required to collect sexual orientation demographic data

    ◯ We are only required to collect gender identity demographic data

10. Please list up to five (5) state/federal **sexual orientation** legislation and mandates and/or industry standards your department adheres to.

_____

_____

_____

_____

11. Please list up to five (5) state/federal **gender identity** legislation and mandates and/or industry standards your department adheres to.

_____

_____

_____

_____

12. Within the next three years is your department anticipating any new legislative or regulatory guidelines regarding the **collection or reporting** of demographic data collection? If so, please list those guidelines and intended date of compliance below.

_____

13.   Do you follow the same state/federal legislation and mandates and/or industry standards mentioned above for **reporting** demographic data collection?

*Mark only one oval.*

◯ Mostly

◯ Some

◯ No, to All

14.   Please list up to five (5) state/federal legislation and mandates and/or industry standards your department adheres to for **reporting** demographic data.

_____

_____

_____

_____

_____

This content is neither created nor endorsed by Google.

Google Forms

# DATA STANDARDS SUMMARY REPORT

In June 2023 the Equity Dashboard team completed a Data Standards Report that summarizes the data standards learnings from an initial quantitative program inventory, qualitative department interviews, department presentations in the Data Standards Community, and a survey to determine state and federal legislative mandates. This report led to the aforementioned legislative survey and allowed the Equity Dashboard team to get a clear view of the current data standards being used to collect and report race, ethnicity, sexual orientation, and gender identity demographic data.

# DATA STANDARDS SUMMARY REPORT

Equity Dashboard

## Abstract

This report provides a summary on data standards learnings from program inventories, departmental interactions, and surveys on legislative mandates.

California Health and Human Services Agency

Center for Data Insights and Innovation

Friday, June 9, 2023

# Table of Contents

## Introduction

The California Health and Human Services (CalHHS) Equity Dashboard is a Managed Analytics Project commissioned by CalHHS Undersecretary Marko Mijic, sponsored by CalHHS Chief Equity Officer Dan Torres, and developed in partnership with departments and programs across CalHHS. The Equity Dashboard is a cross-departmental tool designed to help Agency, departmental leadership, and the public to better understand the Californians who use CalHHS programming, and to identify and address disparities in CalHHS services.

The Equity Dashboard team collected information, input, and feedback from departments through a variety of methods. This report summarizes the data standards learnings from an initial quantitative program inventory, qualitative department interviews, department presentations in the Data Standards Community, and a survey to determine state and federal legislative mandates.

### Initial Data Standards Recommendations

To improve the state of demographic data collection in CalHHS, the Agency must prioritize standardization and leverage the Data Standards Community to enforce proposed standards and best practices.

The Office of Data and Innovation (ODI) is in the process of providing recommendations to the Governor's Office for standards and best practices for race and ethnicity data collection. ODI held a deep dive session and presented their recommendations in a Data Standards Community meeting and provided an opportunity for department staff to give feedback on the proposed standards and best practices.

ODI recommendations include expanding race and ethnicity selections to include MENA categories, increasing the number of race and ethnicity selections to six, and providing the opportunity to select more than one race or ethnicity option. Addendum #1 is a copy of the presentation given by ODI to the Data Standards Community.

At a minimum CalHHS should adopt ODI recommendations for race and ethnicity data collection as a baseline Agency standard with the opportunity to expand those recommendations to be inclusive of more Californians. The Equity Dashboard team can leverage the Data Standards Community to shape departmental efforts to incorporate those recommendations in their demographic data collection.

## Lessons Learned

### Department Program Inventory

In 2022, to understand current demographic data collection practices and standards, the Equity Dashboard team received program inventories from the 12 CalHHS departments. The program inventory focused on the programmatic collection of race, ethnicity, sexual orientation, and gender identity demographic data. The program inventory allowed the Equity Dashboard team to explore the varying collection methods of each department and its programs. In addition, the

Equity Dashboard team was able to identify demographic data collection challenges, opportunities for improvement, and what support would advance their demographic data collection efforts.

During this research, it was found that demographic data collection varies greatly by method, level of detail, and data type across departments and even across programs within the same department. Most programs adhere to standards dictated by specific state or federal legislation, and often these standards are directly tied to a program's funding source.  Additionally, in many cases, department or program databases are outdated, causing difficulties in changing current data standards or data collection practices.  Creating best practices to align with IT requirements will certainly help advance technological efforts.

Addendum #2 is a departmental breakdown of the program inventory that details program demographic data collection based on the responses received from the program inventory.

## Data Standards Community

The Data Standards Community aims to provide California Health and Human Services (CalHHS) and its departments and offices with best practices, policies, standards, and guidance for effective demographic data collection. These priorities are critical to achieving unbiased collection, storage, and cross-Agency analysis of data. The Data Standards Community increases the ability to identify and better understand current state and federal data standards, and regulatory and privacy steps, and realize opportunities for cross-Agency data coordination.

To further the knowledge of CalHHS demographic data collection, each department is in the process of sharing data collection barriers. To date, presentations have been made by five (5) departments: Department of Managed Health Care (DMHC), Department of Rehabilitation (DOR), California Department of Social Services (CDSS), California Department of Public Health (CDPH), and Department of Health Care Access and Information (HCAI). The Department of Health Care Services (DHCS) and Department of State Hospitals (DSH) will present in future 2023 meetings. From these presentations, additional insight has been gained into current demographic data collection and reporting processes and obstacles.

Data Standards Community Collection Obstacles

To date, programmatic demographic data collection is challenged by the following, which prevent data analysis across programs and affect how data is used to drive decision-making or track equitable outcomes:

- Inconsistent race and ethnicity data collection requirements and methods
- General reluctance to self-identify by program participants
- Constraints with manual and antiquated data collection systems
- Current processes that require program providers to make race and ethnicity selections on behalf of program participants
- General lack of requirements for collecting SOGI data

- Inconsistencies between state and federal options for gender identity

Another challenge with demographic data collection is that the available demographic data selections do not necessarily reflect the actual racial and ethnic identity of all the Californian respondents; California is a diverse state with more than 39 million people and every demographic is represented. Despite this diversity, often, only the five most common races are provided with no options for reporting multiple options. According to the Office of Data and Innovation (ODI) "Most of the state's collection methods are out of sync with evolving nations' rise in multiple identified, growing consensus of gaps in current standards and recognition of the importance of disaggregated data." Despite selections not reflecting California's population, several programs are required to collect data and make a selection, whether it is accurate or not.

Finally, CalHHS departments have indicated that automating and/or converting paper forms into electronic forms presents an obstacle due to evolving legislation, resource limitations, and accessibility requirements.

## State and Federal Legislation and Data Standards

In May 2023 the eight (8) departments that participate in the Data Standards Community were sent a survey to pinpoint the 3-5 most common legislative mandates for demographic data (race, ethnicity, sexual orientation, and gender identity) collection and reporting. Responses were received from seven (7) of the eight (8) departments. Each department follows multiple legislation and mandates. The only common collection and reporting standards are Assembly Bill 1726, and the 1997 OMB Standard, and these are only common across a few CalHHS departments.

- Assembly Bill 1726 requires any state agency, board, or commission that directly or by contract collects demographic data as to the ancestry or ethnic origin of Californians to use separate collection categories and tabulations for specified Asian groups and Pacific Islander groups
- The 1997 OMB Standard provides standards for the classification of race and ethnicity

Addendum #4 details the responses received from the seven (7) departments.

# Conclusion

The CalHHS Equity Dashboard team and the Data Standards Community will be prioritizing race, ethnicity, sexual orientation, and gender identity collection standards and best practices. As the Equity Dashboard evolves, so will the demographic data categories and the accompanying standards. Additionally, this report will be updated regularly as additional information about and recommendations for department data standards are established.

# Addendums

The following pages offer an in-depth explanation of the summary report

## Addendum 1

The Office of Data Innovation (ODI) is in the process of making recommendations to the Governor's Office for standards and best practices surrounding race and ethnicity data collection.

*The ODI presentation and accompanying recommendations can be found attached.

## Addendum 2

After a review of the program inventories submitted in 2022 by 12 departments, the CalHHS Equity Dashboard team was able to track each department's programs and associated demographic data collection standards.

| Department | Total Reported # of Programs | # of Programs that Collect Demographic Data | Race/Ethnicity Data Collection | SOGI Data Collection | Notes |
|---|---|---|---|---|---|
| CDA | 1 | 1 | Both race and ethnicity demographic data collected | Both sexual orientation and gender identity demographic data collected | |
| CDSS | 24 | 20 | Race and ethnicity demographic data collection vary by program. | SOGI demographic data collection varies by program. | |
| CDPH | 45* | Cannot be determined with the information provided. | Cannot be determined with the information provided. | Cannot be determined with the information provided. | Program count includes internal and external services as well as offices within CDPH. |
| CSD | 2 | 2 | Both race and ethnicity demographic data collected | Sexual orientation demographic data not collected. Gender identity demographic data collected. | |
| DCSS | Cannot be determined with the information provided. | Cannot be determined with the information provided. | Cannot be determined with the information provided. | Cannot be determined with the information provided. | Child Support Services race/demographic info for the participants exists in other databases. The same is true for the hospital admissions. If DCSS had access to these databases, the demographic data could potentially import a department system. Information is input by child support caseworkers in each county. |
| DDS | 3 | 3 | Both race and ethnicity demographic data collected | Sexual orientation demographic data not collected. Gender identity demographic data collected. | |
| DHCS | 9 | 9 | Both race and ethnicity demographic data collected | Sexual orientation demographic data not collected. Gender identity demographic data collected. | |
| DMHC | 1 | 1 | Both race and ethnicity demographic data collected | Sexual orientation demographic data not collected. Gender identity demographic data collected. | |
| DSH | 5 | 5 | Both race and ethnicity demographic data collected | Sexual orientation demographic data not collected. Gender identity demographic data collected. | |
| DOR | 7 | 6 | Both race and ethnicity demographic data collected | Sexual orientation demographic data not collected. Gender identity demographic data collected. | |
| EMSA | 3 | 2 | Ethnicity is collected, but not for all programs | Sexual orientation demographic data not collected. Gender identity demographic data collected. | |
| HCAI | 15 | 14 | Race and ethnicity demographic data collection vary by program. | SOGI demographic data collection varies by program. | |

## Addendum 3

Version 1 of the CalHHS Equity Dashboard will include data from DOR, DMHC, DSH, and DHCS. During the building phase, Equity Dashboard Data Engineers met with each department's data resources to further discuss data collection. As a result, the Equity Dashboard team was able to further assess programmatic demographic data collection.

| Department | Program | Data Standards Type | Race and Ethnicity Collection Process | SOGI Data Included? | Data Standard Description from Department |
|---|---|---|---|---|---|
| Department of State Hospitals (DSH) | State Hospitals<br><br>Jail Based Competency Treatment (JBCT)<br><br>Community Based Restoration<br><br>Conditional Release Program (CONREP)<br><br>Institutions for Medical Disease/Sub-Acute | Departmental Data Standards | Race/Ethnicity collected as one data element | SOGI Data is collected | DSH's standard data collection practice for the race category is through a multi-system approach due to the complexities of the programs and the unique data management capabilities at each facility. The race and ethnicity data are aggregated into the following racial groupings: Asian, Black, Hispanic, Other/Unknown, and White. Additional steps are taken to categorize the smaller racial categories into the larger groupings referenced.<br><br>DSH's standard data collection practice for gender identity is through documentation in patients' pre-admission documents and/or self-identification through the course of treatment and is subject to change. The gender identity data is aggregated into the following: male, female, transgender female, transgender male, nonbinary, and unknown.<br><br>DSH's reporting standard is based on internal annual metrics that are produced as part of year end caseload and governor's budget estimate related items. |

| Department | Program | Standard | Race/Ethnicity | Sexual Orientation/Gender Identity | Notes |
|---|---|---|---|---|---|
| Department of Managed Healthcare (DMHC) | Independent Medical Review (IMR) | Federal Office of Management and Budget | Race/Ethnicity collected as one data element | Sexual Orientation not collected | **Collection Standard**<br>For the IMR program, race and ethnicity data is collected following:<br><br>1. OMB's Statistical Policy Directive No. 15, Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity; however, race and ethnicity are collected as one category,<br>2. California Government Codes 8310.5 and 8310.7 to break out various Asian and Pacific Islander races (note that the DMHC is an optional participant for 8310.7),<br>3. The 2015 National Content Test Race and Ethnicity Analysis Report to break out Middle Eastern/North African from White/Caucasian, and<br>4. Other values (Multi-Ethnicity, Other Asian/Pacific Islander, and Other) to capture remaining write-in values.<br><br>Gender identity data is collected as Male, Female, Something Else, or Declined to State, where Something Else was added at the consultation of consumer advocates. Consumers who choose Something Else can then write in their gender identity.<br><br>Demographic data is self-reported on our complaint forms and is not reconciled with an external or administrative file.<br><br>Demographic data counts do not include "Declined to State," which "no response" is also assigned to. They cannot disaggregate these in their system.<br><br>**Reporting Standard**<br>Demographic profile data, including age and gender, is reported publicly on the DMHC's public IMR database in accordance with Health and Safety Code 1374.33(h). |
| Department of Rehabilitation (DOR) | Vocational Rehabilitation (VR) | Rehabilitation Administration Federal Standard | Race and Ethnicity collected as separate data elements | Sexual Orientation not collected | The Department of Rehabilitation follows an expanded version of the federal OMB standards with race and ethnicity separated.<br><br>Hispanic/Latino is considered Ethnicity.<br><br>American Indian or Alaska Native Asian Indian Black or African American Cambodian Chinese Filipino Guamanian or Chamorro Hawaiian Japanese Korean Laotian Not Available Other Asian Other Pacific Islander Samoan Vietnamese White are considered Race.<br><br>This is collected by case workers and enforced via the DOR AWARE case management system. |
| | Potentially Eligible (PE) | | | | |
| Department of Health Care Services (DHCS) | Family Planning, Access, Care, and Treatment (FPACT) | Centers for Medicare and Medicaid Services (CMS) Standards | Race and Ethnicity collected as separate data elements | Both are collected, but are re-coded for reporting and internal uses | Race and ethnicity counts: race and ethnicity are optional data fields on DHCS applications and do not have an option of "Choose not to Identify / Declined to State"; DHCS counts NULL/blank responses as "Choose not to Identify / Declined to State".<br><br>Gender Identity counts: DHCS applications include a required "sex" question. As an example, on the Medi-Cal application, the question is worded as: "Are you:" [Male, Female] (paper version of the application) or "What is [applicant's name] sex?" [Female, Male, Transgender: male to female, Transgender: female to male] (on-line version). DHCS recodes the transgender response options to M or F for program use and for public reporting. |
| | Medi-Cal (MediCal) | | Race and Ethnicity collected as separate data elements | Sexual Orientation and Gender Identity are both collected | |

| | | | | |
|---|---|---|---|---|
| | Every Woman Counts (EWC) - Clinical Coordination and Health Education for EWC Region (CHEER) System | | Race and Ethnicity collected as separate data elements | Both are collected, but are re-coded for reporting and internal uses |
| | Every Woman Counts (EWC) - Detecting Early Cancer (DETEC) System | | Race and Ethnicity collected as separate data elements | Both are collected, but are re-coded for reporting and internal uses |
| | Genetically Handicapped Persons Program (GHPP) | | Race and Ethnicity collected as a single data element | Both are collected as a single data element, but are re-coded for reporting and internal uses |
| | California Children's Services (CCS) | | Race and Ethnicity collected as a single data element | Both are collected as a separate data elements, but are re-coded for reporting and internal uses |

## Addendum 4

Based on the review of information pertaining to the priority demographic data categories, a disconnect was found in the collection and reporting of race, ethnicity, sexual orientation, and gender identity demographic data.

| Demographic Category | Findings | Example | Data Elements |
|---|---|---|---|
| Race | Race and ethnicity can be collected as the same category or separate categories depending on program. Race and ethnicity is also defined differently between departments and programs. | For Program A, ethnicity is categorized using the Office of Management Budgets (OMB) standard which assigns Hispanic or Latino as 'ethnicity' while Program B categorizes Hispanic or Latino as 'race' | Example 1) Race and Ethnicity collected separately.<br>What Ethnicity(s) do you identify with?<br>a. Hispanic/Latino<br>b. Not Hispanic/Latino<br>c. Choose not to identify/respond<br><br>What Race(s) do you identify with?<br>a. Black or African American<br>b. Asian/Pacific Islander<br>c. White/Caucasian<br>d. Native American<br>e. Other<br>f. Choose not to identify/respond<br><br>Example 2) Race and Ethnicity collected as one.<br>Select the race/ethnicity(s) you identify with?<br>a. Black or African American<br>b. Asian/Pacific Islander<br>c. White/Caucasian<br>d. Hispanic/Latino<br>e. Native American<br>f. Other<br>g. Choose not to Identify / Declined to State |
| Ethnicity | | 23 of programs collected / 68 total programs | |
| Sexual Orientation | Based on the inventory survey, most departments' programs do not yet collect or report Sexual Orientation, defined as "A person's sexual and emotional attraction to another person and the behavior and social affiliation that may result from this attraction." | For Program A, the participant is asked to select their Sex which is assumed to mean 'assigned at birth', while in Program B the participant is asked to select the 'gender they identify'.<br>54 of programs collected / 68 total programs | Example)<br>What Sexual Orientation(s) do you identify with?<br>a. Gay, Lesbian or Homosexual<br>b. Bisexual<br>c. Pansexual<br>d. Queer<br>e. Asexual<br>f. Two-spirit<br>g. Questioning/not sure<br>h. Not listed, please describe<br>i. Choose not to Identify / Declined to State |
| Gender | While most programs collect Sex (assigned at birth) many do not collected Gender Identity as defined as people who specifically choose their gender, NOT their assigned gender at birth. Another definition) "An individual's sense of self as man, woman, transgender, or other."<br><u>superscript 1</u> | For Program A, the participant is asked to select their Sex which is assumed to mean 'assigned at birth', while in Program B the participant is asked to select the 'gender they identify'.<br>54 of programs collected / 68 total programs | Example 1) Gender (sex) assigned as birth asked<br>What is your gender (assigned at birth)/sex?<br>a. Male<br>b. Female<br>c. Choose not to Identify / Declined to State<br><br>Example 2) Gender Identity asked |

| | | | What gender(s) do you identify as?<br>a. Male<br>b. Female<br>c. Transgender Male to Female<br>d. Transgender Female to Male<br>e. Another Gender Identity<br>f. Non-Binary<br>g. Other<br>h. Choose not to Identify / Declined to State |
| --- | --- | --- | --- |
| | | | |

## Addendum 5

A survey was issued to the 12 CalHHS departments that collect demographic data. From this survey, it was determined the top state and federal legislative mandates used to collect and report race, ethnicity, sexual orientation, and gender identity.

| Department | Industry Standards | State Legislation/Mandates | Federal Legislation/Mandates | Notes |
|---|---|---|---|---|
| CDA | N/A | N/A | N/A | CDA is not currently represented in the Data Standards Community. CDA Representatives will be present beginning June 19th. CDA will have the opportunity to complete the survey by the end of June 2023 |
| CDSS | | • CA Executive Order N-16-22<br>• AB 1726<br>• CA EO B-10-11<br>• SB-435 | • AFCARS regulations for child welfare<br>• Federal Executive Order 13985 | |
| CDPH | | • AB 1726 - Asian and Pacific Islander disaggregation,<br>• AB 532 Multi-race and Multi-ethnic individuals<br>• AB 959: Collect and display SOGI data<br>• AB 435: Latino disaggregation | | |
| CSD | N/A | N/A | N/A | CSD is not currently represented in the Data Standards Community. |
| DCSS | N/A | N/A | N/A | DCSS is not currently represented in the Data Standards Community. |
| DDS | | • CA Government Code section<br>• 11135 CA Labor Code section<br>• 1197.5 CA Welfare and Institutions Code section 4666<br>• The California Fair Employment and Housing Act (Gov. Code, § 12900 et seq.)<br>• CA Health and Safety Code section 1522.41<br>• CA Welfare and Institutions Code section 4502 | • Section 1557 of the Patient Protection and Affordable Care Act<br>• Title VII of the Civil Rights Act of 1964 (see Bostock v. Clayton County, Georgia (2020) 207 L.Ed.2d 218 [140 S.Ct. 1731, 1734].) | |
| DHCS | | • California Government Code (GC) Section 8310.5<br>• Assembly Bill 959 | • 42 Code of Federal Regulations Section 435.907<br>• Federal regulations and requirements provided by Centers for Medicare and Medicaid Services | |

| | | | | |
|---|---|---|---|---|
| | | | • U.S. Office of Management and Budget (OMB) with Statistical Policy Directive No. 15<br>• 1997 OMB standards<br>• 2011 HHS Data Collection Standards for Race, Ethnicity, Sex, Primary Language and Disability Status<br>• National Committee for Quality Assurance (NCQA) Race and Ethnicity Stratifications | |
| DMHC | | • HSC 1374.33(h)(1)(A) | • 1997 OMB Testing | |
| DSH | • Centers for Medicare and Medicaid Services<br>• The Joint Commission | • CA Government Code section 19792 (h) | • Section 4302 of the Affordable Care Act<br>• HR 1370<br>• 1997 OMB standards | |
| DOR | Unknown | Unknown | Unknown | DOR did not respond to the survey |
| EMSA | N/A | N/A | N/A | EMSA is not currently represented in the Data Standards Community. |
| HCAI | • CDC Race and Ethnicity File<br>• National Uniform Billing Committee Uniform Billing format (NUBC UB-04)<br>• US Office of the National Coordinator (ONC) Interoperability Standards Advisory USCDI V2 data elements | • HSC 128735 (g) (3),<br>• HSC 128736 (a) (3) & (4)<br>• HSC 128737 (a)(3) & (4)<br>• HSC 128738<br>• OMB 1997 Standard<br>• HSC 127345<br>• HSC 1216 (a) (1)<br>• HSC 128735 (f), 1<br>• HSC 28736 (d)<br>• HSC 128737 (d)<br>• HSC 127673 (b) (4)<br>• HSC 127673(a) (1) | • 1997 OMB standards | |

Addendum 5

# DEMOGRAPHIC DATA CATALOG

In the journey to gain a comprehensive view of demographic data practices within CalHHS, the Equity Dashboard team created a data catalog focusing on Race, Ethnicity, Sexual Orientation, and Gender Identity (RE/SOGI). The catalog compiled the respective departmental, Federal, State, and Industry data collection standards and practices for RE/SOGI data with the future goal of building out data crosswalks for comparison. Through meticulous analysis, the Equity Dashboard team sought to identify discrepancies, commonalities, and areas for alignment across programs and departments. The intricate process involved aligning and contrasting data standards related to RE/SOGI, recognizing the diversity in data elements, collection methods, and levels of detail adopted by different departments. The insights gained from the catalog will contribute to the ongoing efforts to enhance the Equity Dashboard's data pipelines and provide a foundation for establishing unified best practices in collecting and reporting RE/SOGI demographic data. This database is an iterative and due to the format cannot be shared on a large scale.

CalHHS | California Health & Human Services Agency — CENTER FOR DATA INSIGHTS AND INNOVATION