CIVHC
CENTER FOR IMPROVING
VALUE IN HEALTH CARE

# Data Release Application

## Limited and Identifiable Extracts

## Navigation

# Data Release Application

Limited and Identifiable Extracts

CIVHC
CENTER FOR IMPROVING VALUE IN HEALTH CARE

## Client Application Revision History

The following reflects the history of changes made to this document during the application process prior to project production. Once in production, any further changes to the application may result in additional cost and production delays.

| To be completed by CIVHC staff | | | |
|---|---|---|---|
| Date | New Version Number | Description of Change(s) | CIVHC Change Author (full name, complete title) |
| 1/12/2024 | V.01 | Initial version drafted with client. | Mason Thaxton, Health Data Consultant |
| 10/8/2024 | V.02 | Updates From Client | Mason Thaxton, Health Data Consultant |
| 10/29/2024 | V.03 | Version Update | Mason Thaxton, Health Data Consultant |
| 2/8/2025 | V.04 | Updates to general project details and member/patient filter. | Lucía Sanders, Key Account Manager |
| 3/14/2025 | V.05 | Updates to research aims, methodology, linkage, PHI justification, financial data elements. | Lucía Sanders, Key Account Manager |
| 3/24/2025 | V.06 | corrected the date range from January 1, 2007-December 31, 2030 on page 11 to January 1, 2007-December 31, 2024. | Lucía Sanders, Key Account Manager |
| 3/24/2025 | V.07 | Document formatting corrections | Kelsey Foland, Compliance Process Manager |
| | V.08 | | |
| | V.09 | | |
| | V.10 | | |

# Data Release Application

Limited and Identifiable Extracts

CIVHC
CENTER FOR IMPROVING VALUE IN HEALTH CARE

## Data Requestor Details

### General Project Details

| | |
|---|---|
| Project Title: | Health Outcomes and Expenditures for Children using All Payer Claims Data |
| Application Start Date: | 6/5/2023 |
| Requested Project Delivery Date: | 6/30/2025 |
| Client Organization (legal name): | Icahn School of Medicine at Mount Sinai |
| Client Organization Address: | 1 Gustave L. Levy Place New York, NY 10029-5674 |
| CIVHC can publicly share the Client Organization's name in its Change Agent Index. | ☒ Yes ☐ No |
| **To be completed by CIVHC staff** | |
| CIVHC Contact (full name, complete title): | Lucía Sanders |
| Project Number: | 25.08 |
| Condensed Project Title: | Health Outcomes Expenditures |

### Project Contacts

| | |
|---|---|
| **Project Contact Name:** | **Brett Anderson, MD MBA MS** |
| Title: | Director, Center for Child Health Services Research |
| Email: | Brett.Anderson@mssm.edu |
| Phone Number: | 203-901-6923 (cell) |
| **Analytic Contact Name:** | **Sarah Crook, PhD** |
| Title: | Director of Analytics |
| Email: | Sarah.Crook@mssm.edu |
| Phone Number: | |

| | |
|---|---|
| **Invoice Contact Name:** | **Yohaira Guzman** |
| Title: | Administrative Director |
| Email: | Yohaira.Guzman@mssm.edu |
| Phone Number: | |
| **Data Release Fee Signatory:** | **Matthew Rosamond** |
| Title: | Chief Financial Officer |
| Email: | Matthew.Rosamond@mountsinai.org |
| Phone Number: | |
| **Data Use Agreement Signatory:** | **Hadijah Vactor** |
| Title: | Director, Grants and Contracts / Authorized Official |
| Email: | Hadijah.Vactor@mssm.edu |
| Phone Number: | |

## Project Schedule and Purpose

| | |
|---|---|
| Proposed Project Start Date[1]: | 3/31/2025 |
| Anticipated Project End Date: | 3/31/2030 |
| Proposed Publication or Release Date: | 3/31/2030 |

1. Detail the specific research question(s) you are trying to answer or problem(s) you are trying to solve with this data request. Please list and number the individual questions.

*Individual research questions:*

It is estimated that 40% of children seen at tertiary children's hospitals and 25% of children nationally, have complex chronic conditions.[1] These conditions are responsible for nearly a quarter of deaths in the U.S. among people 0 to 24 years old.[2] Among those who survive, these conditions are often associated with significant and life-long disease burdens that impact the quality of life and productivity of both children and their families and costs families and the health care system dearly. Further, it is known that patients from some demographic, economic, or rural/urban backgrounds are more likely to die or have chronic complications. What is not known is why and what can be done about it. Congenital heart defects, for example, effect ~1% of live births[3] and result in annual acute care costs of >$10 billion (US $ 2024).[4] Treatment typically involves at least one—if not multiple—open heart surgeries, with half of children undergoing initial operation in their first year of life. Infant mortality for these children exceeds 10%, and survivors often describe significant healthcare burdens and challenging, lifelong, journeys.[5-11] Our team and others have described 15-20% higher mortality for non-Hispanic Black and Hispanic children, even after adjusting for cardiac anatomy, clinical risk factors, family income (payer), and neighborhood sociodemographics.

This project aims to use the CO All Payer Claims Data (CO APCD), linked to the National Death Index, neighborhood characteristics from the U.S. Census Bureau, and clinical nuance for risk stratification from clinical registry data to identify modifiable drivers (mediators) of health outcomes and healthcare utilization / expenditures, to inform targeted interventions (policies, programs, and system organization) to improve outcomes and healthcare value for children and young adults.

[1] After all required documents have been signed, typical production time is 30-60 days for a Limited or Identifiable Extract. Anticipate a longer production period for projects including a Finder File or creation of a Member Match File.

In prior NIH-funded research, we focused on NY State Medicaid patients with congenital heart disease, as it is the most prevalent and resource intensive birth defect managed in the United States.[3]

1. We quantified—for the first time—longitudinal healthcare utilization and health expenditures for congenital heart patients and validated novel longitudinal outcomes for this congenital heart patients.[12] We found that the average child on NY Medicaid spent >90 days in hospitals and doctors' offices in their first 5 postoperative years ($139K in Medicaid expenditures in first 5 years, not including initial surgery).[13]
2. We developed the first longitudinal congenital heart risk models,[12,13] the first congenital heart risk models that consider neighborhood characteristics,[14-16] and the first congenital heart risk model for ICD-10data.[17]
3. We validated novel measures of **access to high-quality / high-experience congenital heart providers** as predictors of outcomes among NY Medicaid patients.[15] Drawing from Penchansky and Thomas, access is often operationalized as patient-to-provider geographic distance.[18] We expanded on this, evaluating both distance to and a range of potential descriptors of the quality / experience of the birth centers, primary care physicians, prenatal / pediatric cardiologists, congenital heart surgeons, surgical centers, and emergency rooms that patients used. Candidate measures were selected for consideration by expert consensus and include nationallyendorsed measures, as available. We tested associations with risk-adjusted, acute and longitudinal outcomes, retaining measures reaching statistical significance.[15]
4. We established the **scientific premise** for evaluation of **provider characteristics as mediators** of differences in congenital heart outcomes across patient demographics and neighborhood characteristics. Among NY Medicaid patients, for example, we found not only that Non-Hispanic Black and Hispanic children saw less experienced healthcare provider (defined across multiple measures) than their Non-Hispanic White, Medicaid-insured peers, but also that some of these **differences explained 10-20% of observed differences in risk-adjusted outcomes**—even when examining differences in providers within surgical centers.[15]
5. In preliminary work, we have also described strong mediating effects of potentially modifiable neighborhood-level characteristics (e.g. exposure to environmental toxins / air particulate matter levels) on outcomes and healthcare utilization. We are working currently to disentangle these effects from the impacts of poverty at large and neighborhood / geographic effects on access.

We propose to use the CO APCD to expand upon this work, examining aspects of healthcare system as mediators of differences in outcomes and healthcare resource utilization / expenditures for children both with congenital heart disease and with other chronic conditions. We will bring together APCDs from New York, Colorado, Texas, and Massachusetts, allowing us to assess the generalizability of our prior findings across four states in four distinct regions of the country, and including publicly and commercially insured patients to allow us to explore the impacts of poverty. In particular, we are interested in answering the following specific research questions:

**Aim 1: We will first quantify the degree to which access to greater experience / quality multidisciplinary providers explains differences in risk-adjusted, longitudinal congenital heart patient outcomes across the life course.**

**Aim 1.1.** We will describe outcomes and healthcare utilization for congenital heart patients across the life course by disease severity, state, payer, patient and neighborhood demographics / economics / rurality. To assist in interpretation of our results, we will compare outcomes and healthcare utilization patients with congenital heart disease to the general, healthy, pediatric population. Heretofore, the phrase outcomes and healthcare utilization will be used to describe: our primary outcomes (longitudinal mortality and days alive and out of healthcare) and our secondary outcomes (Number / duration of inpatient admissions, number of emergency department visits, ambulatory / observation stays, outpatient primary care and subspecialty visits, and prescriptions**;** total payer expenditures; and disease-specific healthcare encounters).

**Aim 1.2.** We will apply mediation analyses to quantify the degree to which measures of distance to or the experience / quality of obstetricians, birth centers, primary care providers, prenatal / pediatric subspecialists, surgeons, or surgical centers explain differences in outcomes and healthcare utilization across the lifespan, considering mediating and modifying effects of state, payer, patient and neighborhood demographics / economics / rurality / environmental toxins.

**Aim 2: Quantify how much congenital heart multidisciplinary provider team (network) and payer characteristics explain differences in risk-adjusted, longitudinal patient outcomes and healthcare utilization across the lifespan.**

**Aim 2.1.** We will leverage innovative network analyses to describe congenital heart provider team characteristics -- the connections between multidisciplinary providers who share congenital heart patients across time.

**Aim 2.2** We will apply mediation analyses to quantify whether and by how much provider network and payer characteristics explain observed differences in outcomes and healthcare utilization across the lifespan, considering mediating and modifying effects of state, payer, and neighborhood demographics / economics / rurality / environmental toxins.

**Aim 3: We will leverage the lessons learned in Aims 1 and 2, to quantify the degree to which access to greater experience / quality multidisciplinary providers, and provider team / payer characteristics explains differences in risk-adjusted, outcomes and healthcare utilization for all children across the life span.**

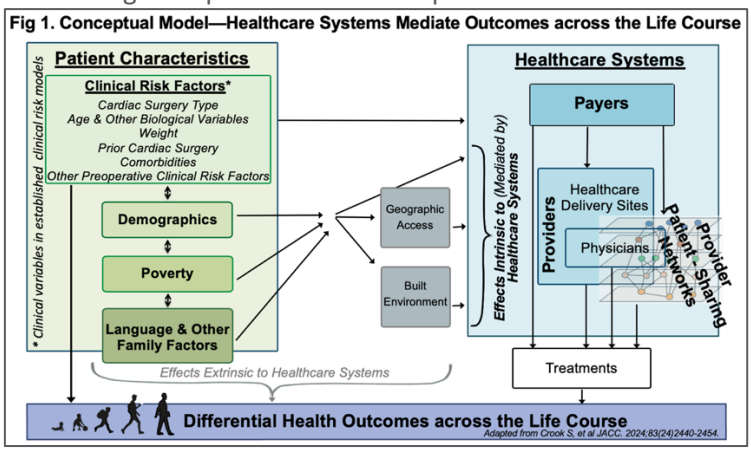
**Aim 3.1** We will define longitudinal, disease and risk-stratified outcomes and healthcare utilization for children and young adults with chronic disease, adjusting for clinical characteristics, demographics, and family and neighborhood characteristics, comparing to the general population, and identifying populations with the highest healthcare burdens / the greatest variation in outcomes. Children with chronic conditions will initially be identified using the Complex Chronic Conditions v 3, developed by Colorado researcher and colleague, James Feinstein. [1] We will consider the use of other pediatric,

administrative data specific algorithms, such as the Pediatric Medical Complexity Algorithm v 3.[19]

**Aim 3.2.** We will apply mediation analyses, as above, to quantify the degree to which measures of distance to or measures of the experience / quality of multidisciplinary providers explain observed explains differences in risk-adjusted, outcomes and healthcare utilization for non-cardiac children with chronic disease across the lifespan.

**Aim 3.3.** We will apply mediation analyses to quantify whether and by how much provider network and payer characteristics explain observed differences in outcomes and healthcare utilization across the lifespan, considering mediating and modifying effects of state, payer, and neighborhood demographics / economics / rurality / environmental toxins.

Figure 1 presents our conceptual model:



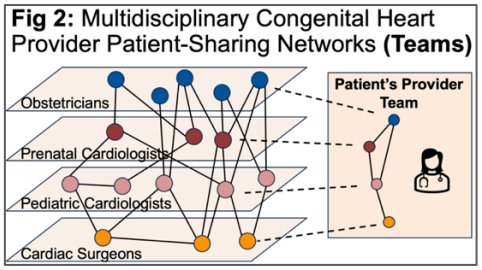




Figure 2 illustrates an interdisciplinary provider team/network, in the context of congenital heart patient care:

2. Describe your methodology or how you will be using data from the Colorado All Payer Claims Database (CO APCD) to answer your research questions.

**In Aim 1**, we validate across states and payers (income) the predictive validity and mediating effects of **provider** characteristics assessed in the parent R01. **In Aim 2**, we combine network analysis, causal mediation, and advanced statistical modeling, to characterize multidisciplinary congenital heart **provider networks** (patient-sharing teams) and to understand how patients'

access to provider networks impacts which downstream providers they see and, ultimately, their outcomes. Obstetricians, prenatal / pediatric cardiologists, congenital heart surgeons, and others work formally and informally as teams, providing complementary care for patients across the life course **(Fig 2)**. We hypothesize that these teams are more important than any one provider alone. Further, we hypothesize that both upstream providers and health insurers (payers) influence team compositions. Understanding these relationships has potential to influence how healthcare system leaders' approach, for example, future patient referrals, teamwork, mentorship, and states' managed care contracts. **In Aim 3**, we expand beyond congenital heart disease, using the lessons learned from congenital heart disease to identify modifiable mediators of outcomes and healthcare utilization across childhood chronic disease. We hypothesize that upstream healthcare factors (e.g. insurance, prenatal care, birth center) impact downstream care access (e.g. quality or experience of congenital heart team) and outcomes, and that analyses will reveal modifiable healthcare system features as targets for care improvement.[2]

**Data Sources for Quantitative Analyses:**

We will expand our prior work, including Medicaid and commercial claims (APCD) data from four states in four regions of the country, and clinical registry data from 25 children's hospitals.

*1) All Payer Claims Databases (APCDs)*: APCDs are state-administered / state-mandated data repositories that include direct identifiers and longitudinal data on all billed services associated with—but not limited to— inpatient, outpatient, emergency room, urgent care, pharmacy, rehabilitation, and home healthcare services, across institutions and across time. APCDs capture both 100% of Medicaid and 60-80% of non-Medicaid (commercial and other governmental) insurance claims. Unique identifiers follow patients even when they change insurers. We will use APCDs from **NY, Massachusetts, Colorado**, and **Texas**. While the NY and Texas ACPDs are not broadly available, we have worked with the NY and Texas APCDs and have been given permission to access these data; **these APCDs are ready for our use.**
*2) National Death Index:* As in our prior work, longitudinal mortality will be obtained from the *Centers for Disease Control and Prevention's National Death Index (NDI).*
*3) Patient Residential Neighborhood Characteristics:* Measures of neighborhood-level demographics / economics / rurality / environmental toxins will be obtained via the *US Census American Community Survey* 5-year estimates. As in our prior work, key measures will be identified via principal components analysis.[3] Composite measures will again be considered (e.g. Yost, Childhood Opportunity Index).[4, 5] *R*urality will again be assessed as commuting time (patients' home addresses to both the nearest and used (accessed) providers), as well as via Rural-Urban Commuting Area / Combined Statistical Areas[6, 7] or other publicly available measure as appropriate for the geography. We already have these data on our local server.
*6) Provider Characteristics:* Provider years of experience, board certifications, and addresses come via the American Medical Association Physician Masterfile and Departments of Health's rosters. Other measures will be derived via claims.
*7) Clinical Registry:* For Aims 1 and 2, cases will be identifies / clinical detail enhanced via linkage to The Society of Thoracic Surgeons-Congenital Heart Surgery Database (**STS-CHSD**). The STS-CHSD is the world's largest congenital heart surgery registry. It captures >90% of US congenital heart operations, with great clinical granularity. Data include direct identifiers, key demographics, 110 anatomically specific cardiac diagnoses, 187 procedures, 22 comorbidities / preoperative risk-factors, bypass, cross- clamp, and circulatory arrest times, 11 major postoperative morbidities, and operative mortality. Missingness for key demographics in the last decadeis <1%.[8] Colorado

Children's Hospital and HCA HealthONE Rocky Mountain Children's own their own locally-held STS-CHSD and are partners in this application. These institutions have agreed to share their locally held clinically registry data with us. These data will be stored on our local server. For Aim 3, the cohort will initially be defined the CCC and PMCA algorithms. To the extent to which we find key chronic diseases to lack specificity / sensitivity in these algorithms, Colorado Children's Hospital and HCA HealthONE Rocky Mountain Children's will work with us to identify the most appropriate, locally-held clinical registry necessary to enhance scientific rigor.

Patient Cohorts:

Chronic Disease Cohorts: For Aim 1 and 2, the STS-CHSD clinical registry data from each contributing center will be used for congenital heart patient case capture, with analyses conducted at the patient-level. The full cohort will include all patients <30 years of age at initial operation, from any contributing centers, Jan 1, 2007 (or first available data in the state's APCD) to December 31, 2024 (or last available date in the APCD). For Aim 3, the cohort will initially be defined the CCC and PMCA algorithms. To the extent to which we find key chronic diseases to lack specificity / sensitivity in these algorithms, Colorado Children's Hospital and HCA HealthONE Rocky Mountain Children's will work with us to identify the most appropriate, locally-held clinical registry necessary to enhance scientific rigor.

Comparison Cohorts: Comparators—children without chronic disease—will be matched 2-to-1 in APCDs on dates of birth (±6 months), age at insurance-enrollment (±4 months), other key demographics (race/ethnicity if sufficient data quality and sex), health plan ID, plan enrollment duration, and county of residence.[9]

Parental Cohorts: For assessment of prenatal care, we will include all pregnant parents of subjects in the chronic disease or comparison cohorts, using our previously established algorithm. For patients whose births are observed in the APCD, this will include linking to women 14-45 with the same family ID with billing codes for delivery that match the date of the baby's delivery (+/- 4 days) (to avoid matching to a pregnant sibling instead of parent).

DATA LINKAGES: Linkages will be performed using methods established in our prior research and in compliance appropriate privacy protections and with state-specific and local laws, regulations, and policies. Clinical registry and claims data will be linked via multi-pass, iterative, deterministic matching on direct and indirect identifiers, including first and last names, mothers' and fathers' last names, alternate or prior names, dates of birth, admissions, surgeries, and discharges, and National Provider Identifiers (NPIs). In our prior work, matching **sensitivity and specificity were >99.9%**.[10] NDI will be linked to registry data via multi-pass, iterative, deterministic matching on direct and indirect identifiers. With prior CO APCD permission, NDI would be linked directly to the APCD. Neighborhood characteristics will be linked at the block group-level by residential addresses as available. Provider characteristics will be linked byNPI.[10]

Analytic Approach:

Patient characteristics, mortality, healthcare utilization, and healthcare expenditures will be summarized by year of initial surgery, disease severity, patient age, key patient and neighborhood demographics, state, and payer. To provide context to assist in interpretation, clinical outcomes, healthcare utilization, and health expenditures will be compared for patients with cardiac and

other chronic diseases and for matched strata of otherwise healthy children from the comparison cohort. Univariable and standard multivariable analyses will test associations between predictor variables and outcomes. We will consider fractional logit models for percent days alive and out of healthcare; Cox proportional hazards models for mortality; Poisson models that allow for over dispersion for visit counts, number of inpatient days, and prescriptions; and log linear models for expenditures. Models will adjust for clinical characteristics, as well as key patient and neighborhood demographics, operative year and an offset term for time insurance enrolled. We will consider center fixed vs. random effects and test for interactions or effect modification by state, payer, rurality, surgical era, and neighborhood characteristics, stratifying as appropriate and quantifying effect attenuation. Fractional logit models use a generalized linear model framework with a logit link function to model outcomes as proportions on a 0-1 interval (including 0 and 1); this allows for retention of patients who die before discharge (zero day alive and out of healthcare), while ensuring predicted probabilities remain in this interval.[11] We will conduct sensitivity analyses to understand effects of death and insurance disenrollment / change, comparing patients who: **a)** survived vs. died, **b)** enrolled continuously in the same health plans vs. changed plans; **c)** enrolled continuously in a Medicaid vs. in a non-Medicaid plan vs. switched (and timing of switch); and **d)** enrolled continuously in any plan vs. churned (defined as ≤11 months per year enrolled) vs. moved or lost insurance without re-enrolling—standardized for months enrolled. We will also consider Tobit models that directly address truncation of healthcare utilization among patients who die, move, or lose insurance,[12] or other models as appropriate.

Provider Characteristics: We will calculate commuting time from patients' homes to primary providers via *OpenTripPlanner*. We will calculate measures of experience / quality of primary providers established in our prior research via APCDs, clinical registry, and AMA. Primary providers will be defined as the providers who submit claims for the plurality of patients' visits within each subspecialty each year.[13] For obstetrician, prenatal cardiologist, or other prenatal provider, we will also consider first provider, as some providers routinely refer complex patients. Primary surgeons and surgical centers will be identified in registry data at first index operation. We will describe provider characteristics across disciplines (obstetricians, prenatal cardiologists, birth centers, pediatric cardiologists, surgeons, surgical centers). Univariable and multivariable analyses will compare provider characteristics on patient clinical characteristics, demographics, states, payers, and residential neighborhood characteristics, adjusting for multiple comparisons. We will use linear and fractional logit models (or other models as indicated) to regress key provider characteristics on patient demographics *(Model 1)* and outcomes on provider characteristics adjusting for patient demographics *(Model 2)*. We will plan to adjust models for clinical characteristics and operative year and will consider including an offset term for time insurance enrolled. We will consider center fixed vs. random effects. We will test for interactions, mediation, or effect modification by state, payer, surgical era, and key neighborhood characteristics—including neighborhood demographics, economics, rurality, or environmental toxins—stratifying as appropriate and quantifying effect attenuation. We will retain provider characteristics that reach statistical significance and apply formal mediation analyses to estimate the proportion of the total effects of key patient or neighborhood characteristics on outcomes mediated by provider characteristics, estimating each

characteristic in a separate model (**Fig 3**).[14] Sensitivity analyses will consider insurance disenrollment/change and other statistical challenges.

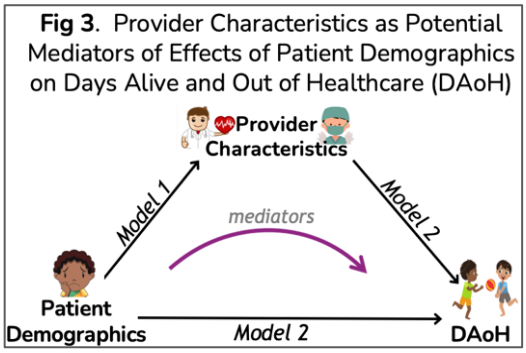



Table 1 lists sample provider charactersitics:

| Table 1: Candidate Measures of Congenital Heart Provider Quality / Experience / Distance* | OB | Prenatal Cardiologist | Birth Center | Pediatric Cardiologist | Surgeon | Surgical Center | Primary Care |
|---|---|---|---|---|---|---|---|
| % Prenatal visits timely completed | x | | | | | | |
| % Level II ultrasound | x | | | | | | |
| % CHD patients referred to fetal cardiologist | x | | | | | | |
| % CHD patients referred to cardiac center | x | | | | | | |
| NICU level of birth hospital | | | x | | | | |
| Cardiac surgical capability of birth hospital | | | x | | | | |
| Time to initial postnatal cardiac echocardiogram, for critical CHD | | | x | | | | |
| Observed to expected operative mortality ratio | | | | x | x | x | |
| Volume cardiac surgical patients | | x | | x | x | x | |
| Volume of critical neonatal surgical patients | | x | | x | x | x | |
| Volume of single ventricle surgical patients | | x | | x | x | x | |
| Median surgical timing, for index operations | | x | x | x | x | x | |
| % Medicaid patients Q1 surgical timing | | x | | x | x | x | |
| % Patients on Medicaid | x | x | | x | x | x | x |
| % Patients from key demographics | x | x | | x | x | x | x |
| Age / years of experience | x | x | | x | x | | x |
| % Patients w 30 months well-child visits, *NCQA* | | | | | | | x |
| Volume patients w Complex Chronic Conditions or Pediatric Medical Complexity | x | | x | | | | x |
| Geographic distance / commute time to provider | x | x | x | x | x | x | x |
| **Candidate measures derived via APCD, STS-CHSD, or AMA Physician Masterfile* | | | | | | | |

For network analyes, we will follow **Co-Is Moen** & **O'Malley's** prior analytic workflow, innovative measure development, and statistical modeling to assemble and analyze multidisciplinary congenital heart **provider networks** (patient-sharing teams), using the igraph R package. Providers will serve as nodes, stratified by provider discipline. Patients shared between providers will serve as ties, with tie weights calculated as the number of shared patients. To assess within- / between-hospital physician ties, physicians will be attributed to hospitals at which they submit the plurality of claims or to which the plurality of their patients is admitted.[13, 15] We will restrict analyses to providers who share ≥2 patients. Networks will be estimated by year; we will consider annualizing over two years. We will use community detection to identify naturally occurring, patient-sharing sub-networks in the care trajectories visualized in **Fig 2**.[16] We will use an extension of the constrained community detection approach developed by **Co-I O'Malley** to identify sub-networks of providers.[17] We will allow providers to serve as members of multiple

sub-network.[18] We will include network characteristics that describe sub-network structures, as well as provider and care team-level measures within sub-networks, to capture referral behaviors. **Table 2** lists sample network measures. Network characteristics will be compared across patient/neighborhood characteristics, years, states, and payers.

**Table 2. Sample Characteristics of Provider Patient-Sharing Networks—Measures of Interdisciplinary Provider Teams**

| Network Measure | Measure Concept | Measure Definition |
|---|---|---|
| **Care Density** | Familiarity of providers with the other, cross- disciplinary providers in patients' care *teams* | Calculated as the sum of ties within each patient's care *team*, adjusted for *team* size[19, 20] |
| **Clustering Coefficient** | How tightly knit provider networks are; degree to which sets of three providers share patients | Sum of closed triangles in shared-patient network divided by total possible triplets--open &closed[20, 21] |
| **Degree Centrality / Physician Node Strength**<br>*-Within-hospital*<br>*-Between-hospital* | Local vs. regional provider prominence within networks. For surgeons, we will compare to the effects of total case volumes from clinical registry (national prominence) | Number of providers with whom providers shares patients / sum of weighted ties, calculated within and across hospitals per calendar year. We will consider adjusting metrics for volume[13, 22] |
| **Assortativity,** by:<br>*-Node strength*<br>*-Experience / Quality* | Degree to which similar providers share patients; extent to which access to one high- experience, high-quality, connected provider is associated with access to future providers | Calculated as correlation coefficients for each sub-network. We will consider assortativity on node strength, as well as on provider characteristics outlined in **Table 1**[23] |
| **Referral Bias**<br>-Patient demographics<br>-Payer | Degree to which providers systematically share / refer patients to different colleagues across patients' demographics (or payers) | Patient sharing networks are calculated separately and compared for patients of different race and ethnicities (or payers)[24] |
| **Network Segregation / Dissimilarity Index**<br>-Demographics<br>-Payer | Similar to referral bias, but across network sub-networks. Extent to which patient-sharing patterns across sub-networks differ by patient demographics (or payers) | Assesses differences in provider-provider ties, measuring whether referrals differ by patients' demographics (or payers) / compares to distribution by geography[25, 26] |

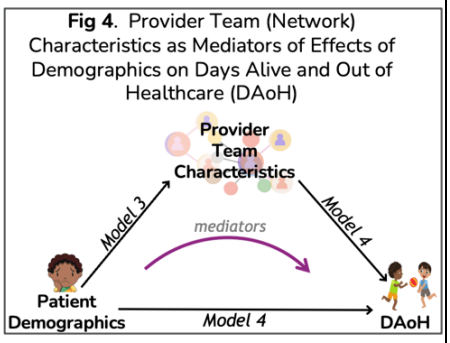



**Fig 4.** Provider Team (Network) Characteristics as Mediators of Effects of Demographics on Days Alive and Out of Healthcare (DAoH)

To estimate the proportion of total effects of patient demographics on outcomes mediated by **provider team characteristics**, we will repeat the mediation process described above. Linear and fractional logit models will regress team characteristics on patient demographics *(Model 3)* and percent outcomes on team characteristics, adjusting for key patient demographics *(Model 4).* We will test team characteristics individually in separate models. We will plan to adjust for patient clinical characteristics, year, center random vs. fixed effects, and an offset term for time insurance enrolled. We will test for interactions or effect modification by state, payer, rurality, era, and key neighborhood characteristics—including neighborhood demographics,

economics, rurality, and environmental toxins, stratifying as appropriate, quantifying effect attenuation, and assessing insurance disenrollment / change.

To further understand how patients get to the providers they see, we will examine roles of upstream providers and payers on teams. We will repeat mediation methods, estimating the proportions of the total effects of patient demographics on provider characteristics, mediated by team characteristics, and the proportions of total effects of patient demographics on team characteristics that are mediated by payer characteristics, as available or as derived via the APCDs (**Table 3**; **Fig 5**; *Models 5-7*). We will test for interactions or effect modification and assess insurance disenrollment / change.

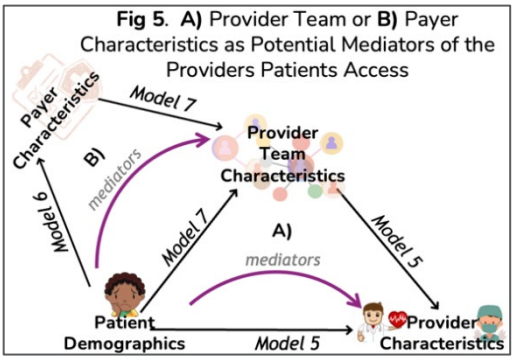


**Fig 5. A)** Provider Team or **B)** Payer Characteristics as Potential Mediators of the Providers Patients Access

| |
|---|
| Payer Type (Medicaid vs. non-) |
| Business Structure. (e.g. for-profit, managed care) |
| Health Plan |
| Geographic Regions Served |
| Rurality Served |

Understanding that the characteristics of multidisciplinary providers, payers, and teams might have non-linear relationships with each other and / or with outcomes, we will also consider nonparametric series regression to flexibly model associations between provider, payer, and team characteristics and outcomes and interactions. One might imagine, for example, how experienced surgeons might benefit from engaging with more experienced subspecialist or pediatricians, but how the benefits might not be the same as they are for less experienced surgeons—or how experienced pediatricians, obstetricians, or birth centers might benefit from engaging with experienced subspecialists, but how the benefits might not be the same as they are for less experienced pediatricians. Nonparametric series regression uses fully non-parametric models, agnostic to functional form.[27] Finally, in network science, networks are characterized by observed patient sharing (used / "accessed" providers). This might differ from the "provider networks" health plans report for network adequacy compliance (the "ghost provider" effects).[28] We will compare characteristics of providers in observed networks vs. plan rosters—using coded plan identifiers. Cox proportional hazard and fractional logit regression and simulation modeling will quantify the degree to which changing real-world access to providers to match plan rosters might impact patient outcomes. We will consider other models as appropriate.

3. Explain how this project will benefit Colorado and its residents.[2]

It is estimated that 40% of children seen at tertiary children's hospitals and 25% of children nationally, have complex chronic conditions.[1] These conditions are responsible for nearly a quarter of deaths in the U.S. among people 0 to 24 years old.[2] Among those who survive, these conditions are often associated with significant and life-long disease burdens that impact the quality of life and productivity of both children and their families and costs families and the health care system dearly. Further, it is known that patients from some demographic, economic, or rural/urban backgrounds are more likely to die or have chronic complications. What is not known is why and what can be done about it.

By linking the Colorado APCD to nuanced clinical, provider, and neighborhood charactersitics and analyzing it along side other states' data, we can identify modifiable dimensions of the healthcare system, health care, and communities causally associated with worse outcomes. In future work, these results could be used to design interventions to improve outcomes and reducing health disparities experienced by children across the state of Colorado. Further, as part of our research collaborative, we have partnered with physician investigators / institution leaders providing care in the State of Colorado to Colorado residents (Children's Hospital Colorado and HCA HealthONE Rocky Mountain). These institutions share their clinical registry data and clinical perspectives with us such that we can, together, use these data to directly improve the quality and value of care delivery for their patients, which will directly improve health outcomes for Colorado children.

4. Describe how your project will improve health care quality, increase health care value, or improve health outcomes for Colorado residents.[2]

This work focuses on children who are high healthcare utilizers and for whom there is known to be large variation in outcomes and healthcare utilization. We then center this application on identifying modifiable mediators of outcomes for these children--things that might be changeable by providers or policy makers and that directly contribute to worse outcomes for select groups of patients. Because our fundamental goal is to have direct and immediate impacts on improving the quality and value of care delivery and outcomes, we partnered with physician investigators providing care in the State of Colorado to Colorado residents. These investigators share their clinical registry data and clinical perspectives with us such that we can, together, use these data to directly identify and implement strategies to improve the quality and value of care delivery for their patients, thereby directly leading to improvments in health outcomes for Colorado children.

[2] It is a statutory requirement for all non-public releases of CO APCD data to benefit Colorado or its residents. Contributions to generalizable knowledge alone are not sufficient to satisfy this requirement.

5. Health equity is defined as the state in which everyone has a fair and just opportunity to attain their highest level of health. Explain how your project addresses health equity.

| |
|---|
| Significant disparities are known to exist in short-term outcomes and resource utilization. Neighborhood economics, education, environment, and interpersonal bias are thought to contribute to these inequities. However, the specific mechanisms underlying these inequities remain unclear. We aim to use these linked claims data on health outcomes and healthcare resource utilization to not only describe differences in outcomes across groups of patients in the State of Colorado, but also to identify potential drivers (mediators) of unequal health outcomes and to quantify the degree to which these measures mediate (drive) health inequities. Using causal mediation analyses, we focus primarily on modifiable dimensions of health care and the healthcare system over which we, our collaborators, and state departments of health have agency, such that the knowledge gained from these analyses can be used directly by clinicians within the State of Colorado to improve equitable outcomes for their patients. To a more limited extent, we also consider slightly harder to modify neighborhood characteristics, such as environmental toxins (both as mediators themselves and as confounders). |

6. Describe any publication you plan to develop based on your use of CO APCD data, its intended audience, and whether it will be made publicly available.

| |
|---|
| We intend to disseminate our findings through publications in academic journals, presentation at national academic conferences, and by providing actionable feedback to Colorado healthcare providers and policymakers. |

# Data Matching and Linkage

## Finder File

A Finder File is a file you submit to CIVHC with information about a pre-selected cohort for matching to CO APCD data. Ask your CIVHC Contact for more information about this process and requirements for Finder File submission.

Will you provide CIVHC with a Finder File as part of this project?

- ☒ No
- ☐ Yes

## Member Match File

A Member Match File is a file that CIVHC creates on your behalf to send to a registry or other outside entity to create a crosswalk connecting data from the CO APCD to the other entity's data.

Does this project require the creation of a Member Match File?

- ☐ No
- ☒ Yes. Consult with your CIVHC Contact about completing a separate Data Element Selection Form specifying the data elements that should be used to create the Member Match File.

  Answer the following:

| Who will receive the Member Match File? |
| --- |
| Vital Statistics at CDPHE |

## Control Group

A Control Group is a group of individuals who can be used to compare against the cohort identified in the Finder File.

Will you need CIVHC to create a Control Group as part of this project?

- ☒ No
- ☐ Yes. Consult with your CIVHC Contact about completing a separate Control Group Data Element Selection Form specifying the data elements that should be used to define the Control Group.

## Linkage

Data Linkage is a method of joining data from different sources together to create a new data set.

Will the CO APCD data be linked to another data source?

☐ No
☒ Yes. Answer the following:

| What is/are the other data source/s? |
| --- |
| As detailed above, we will include the following data sources: 1) All Payer Claims Databases (APCDs): APCDs are state-administered / state-mandated data repositories that include direct identifiers and longitudinal data on all billed services associated with—but not limited to—inpatient, outpatient, emergency room, urgent care, pharmacy, rehabilitation, and home healthcare services, across institutions and across time. APCDs capture both 100% of Medicaid and 60-80% of non-Medicaid (commercial and other governmental) insurance claims. Unique identifiers follow patients even when they change insurers. We will use APCDs from NY, Massachusetts, Colorado, and Texas. While the NY and Texas ACPDs are not broadly available, we have worked with the NY and Texas APCDs and have been given permission to access these data; these APCDs are ready for our use.<br><br>2) National Death Index: As in our prior work, longitudinal mortality will be obtained from the Centers for Disease Control and Prevention's National Death Index (NDI).<br><br>3) Patient Residential Neighborhood Characteristics: Measures of neighborhood-level demographics / economics / rurality / environmental toxins will be obtained via the US Census American Community Survey 5-year estimates. As in our prior work, key measures will be identified via principal components analysis.14 Composite measures will again be considered (e.g. Yost, Childhood Opportunity Index).21, 22 Rurality will again be assessed as commuting time (patients' home addresses to both the nearest and used (accessed) providers), as well as via Rural-Urban Commuting Area / Combined Statistical Areas23, 24 or other publicly available measure as appropriate for the geography. We already have these data on our local server.<br><br>4) Provider Characteristics: Provider years of experience, board certifications, and addresses come via the American Medical Association Physician Masterfile and Departments of Health's rosters. Other measures will be derived via claims.<br><br>5) Clinical Registry: For Aims 1 and 2, cases will be identifies / clinical detail enhanced via linkage to The Society of Thoracic Surgeons-Congenital Heart Surgery Database (STS-CHSD). The STS-CHSD is the world's largest congenital heart surgery registry. It captures >90% of US congenital heart operations, with great clinical granularity. Data include direct identifiers, key demographics, 110 anatomically specific cardiac diagnoses, 187 procedures, 22 comorbidities / preoperative risk-factors, bypass, cross- clamp, and circulatory arrest times, 11 major postoperative morbidities, and operative mortality. Missingness for key demographics in the last decade is <1%.25 Colorado Children's Hospital and HCA HealthONE Rocky Mountain Children's own their own locally-held STS-CHSD and are partners in this application. These institutions have agreed to share their locally held clinically registry data with us. These data |

will be stored on our local server. For Aim 3, the cohort will initially be defined the CCC and PMCA algorithms. To the extent to which we find key chronic diseases to lack specificity / sensitivity in these algorithms, Colorado Children's Hospital and HCA HealthONE Rocky Mountain Children's will work with us to identify the most appropriate, locally-held clinical registry necessary to enhance scientific rigor.

6) Vital Statistics at CDPHE: Crosswalk of mother-infant pairs for children aged 0-18 years in CO APCD, which is required to assess impacts of prenatal care.

**Who will perform the data linkage?**

Data analytic team at the Center for Child Health Services Research at the Icahn School of Medicine performs all linkages.

**What identifying data elements will be used to perform the data linkage?**

Names, parental names, DOB, geolocation in the smallest allowable unit, Date of Service, National Provider Index number (NPI), family ID on insurance plan, payer ID

**What non-CO APCD data elements will appear in the new linked file?**

Date of Death, cause of death, Mortality Status (linkage to the National Death Index will allow us to collect death information even among patients who move out of state), Distance to healthcare providers, drive time to healthcare providers, public transit time to healthcare providers, more specific information on clinical diagnoses and procedures from clinical registry data, provider charactersitics (including provider training, age, referral network, etc.), and neighborhood characteristics (poverty rates, crime rates, percent of population who owns a car, etc.), as well as calculated variables, such as provider, provider-team, and health plan characteristics.

Admit/birth time will be used in assessment of time to services delivered (example, a baby born/admitted at 10pm on a Friday, might wait longer for initial diagnosis and treatment, then a baby born/admitted at 9am on a Monday.

## Data Inclusion Criteria

Make selections in the following sections based on what data you want to have included in this extract. If you will be creating a Control Group, complete this section for your study population and not the Control Group.

### Protected Health Information (PHI)

Indicate which Protected Health Information data elements you require for your project purpose:

| Available for Limited and Identifiable extracts: | | |
|---|---|---|
| ☐ Member 5-Digit Zip Code | ☐ Member County | ☐ Member City |
| ☒ Member Dates of Service | ☒ Member Eligibility Dates | ☐ Claim Paid Dates |
| ☐ Employer Name | ☒ Member Census Tract | ☒ Member Census Block |
| ☒ Member Census Block Group | | |
| **Available for Identifiable extracts only (see also Identifiable Data Use Approval):** | | |
| ☒ Member Name | ☒ Member Date of Birth (if requesting more than year only) | |
| ☐ Member Street Address | ☒ Member Latitude and Longitude | |
| **Provide detailed justification for the inclusion of all PHI data selected above, and explain how its inclusion meets the Minimum Necessary Requirement.[3]** | | |
| Geolocations will be used to link publicly available information about neighborhood-level social determinants of health / their built environment to assess how neighborhood factors impact health and access to healthcare providers (including distance, drive times, and public transit times). We use longitude and latitude to calculate the distances / time that patients must travel for care. Given how rural many parts of the State are, there might be significant differences in travel time if we used Census Tract/Block/Group. Ages and dates are needed to link to future datasets and will also be used to asses the impacts of age and timing of birth on healthcare utilization and outcomes / to examine provider differences in timing of care delivery, as a potential measure of provider quality. Coded health plan identifiers are needed to assess impacts of health plan / plan characteristics on outcomes and to children with | | |

[3] Limited and Identifiable extracts must adhere to the Minimum Necessary Requirement under the HIPAA Privacy Rule; only that data required to answer the project purpose can be included in the request.

chronic disease to otherwise similar healthy controls. Names are needed to perform the necessary linkages

## Line(s) of Business

- ☒ Commercial Payers
- ☒ Health First Colorado (Colorado's Medicaid and CHP+ programs)[4]
- ☐ Medicare Advantage
- ☐ Medicare Fee for Service (FFS)[5]

## Year(s) of Data

| | | | | | |
|---|---|---|---|---|---|
| ☒ 2012 | ☒ 2013 | ☒ 2014 | ☒ 2015 | ☒ 2016 | ☒ 2017 |
| ☒ 2018 | ☒ 2019 | ☒ 2020 | ☒ 2021 | ☒ 2022 | ☒ 2023 |
| ☒ 2024[6] | | | | | |

## Claim Type(s)

| | | |
|---|---|---|
| ☒ Inpatient Facility | ☒ Outpatient Facility | ☒ Professional |
| ☒ Pharmacy | ☐ Dental | |

## Financial Detail by Line Item

| | | |
|---|---|---|
| ☐ Charged Amount | ☒ Allowed Amount | ☒ Plan Paid Amount |
| ☐ Plan Pre-Paid Amount | ☐ Member Copay | ☐ Member Deductible |
| ☐ Member Coinsurance | ☒ Total Member Liability | |

[4] Medicaid-only data requests must be approved by the Colorado Department of Health Care Policy and Financing.
[5] Medicare FFS data are not available for all requests and must go through a separate approval process.
[6] This year's data is incomplete and not fully adjudicated. Consult with your CIVHC Contact to find out what data is available at the time of your request.

## Filter Criteria – Services, Providers, Facilities

If you need data for specific services, providers and/or facilities, specify that filter criteria below (ask your CIVHC Contact about including an additional file with this application for large code lists):

| ICD Diagnosis Code(s): |
| --- |
| |
| Procedure(s) (list CPT, HCPCS, DRG, ICD, and/or CDT codes): |
| |
| Drug(s) (list pharmacy NDC and/or HCPCS codes): |
| |
| Facility Type(s): |
| |
| Facilities (list NPIs and/or Pharmacy IDs): |
| |
| Facilities within these geographical areas (list county, zip code, Census Tract, etc.): |
| |
| Provider Type(s): |
| |
| Provider(s) (list NPIs): |
| |
| Providers within these geographical areas (list county, zip code, Census Tract, etc.): |
| |
| Specific payers (minimum of five): |
| |

| Other claim specification: |
| --- |
| |

## Filter Criteria – Members/Patients

If you need data for specific member/patient groups, specify that filter criteria below (ask your CIVHC Contact about including an additional file with this application for large code lists):

| Ages: | | |
| --- | --- | --- |
| DOB 01/01/1990-12/31/2024 to capture all patients 0-18 years of age from 2012-2024.<br>Additionally, all women in crosswalk file provided by CDPHE (crosswalk represents mothers matched to children born 01/01/1990-12/31/2024.) | | |
| ☐ At the time of service | ☒ At year end | ☐ By another anchor date:<br>*Specify here* |
| With these ICD Diagnosis Code(s): | | |
| | | |
| Who have had the following procedure(s) (list CPT, HCPCS, DRG, ICD, and/or CDT codes): | | |
| | | |
| Within these geographical areas (list county, zip code, Census Tract, etc.): | | |
| | | |

## Value-Add Data Elements

- ☒ Medicare Severity Diagnosis Related Group Codes (MS-DRGs)
- ☒ 3M All Patient Refined Diagnosis Related Group Codes (3M APR DRGs)
- ☐ Medicare Repricer (available at the claim line level)
- ☐ Fields from the American Community Survey (available at the Census Tract level):

| *Specify here* |
| --- |

# Additional Documentation

## Data Element Selection Form (DESF)

The Data Release Application must be accompanied by a completed Data Element Selection Form. Ask your CIVHC Contact for more information about completing this form.

☒ By checking this box, the Client Organization confirms that the Data Element Selection Form has been completed.

☐ If applicable, by checking this box the Client Organization confirms that a separate Member Match File Data Element Selection Form has been completed.

☐ If applicable, by checking this box the Client Organization confirms that a separate Control Group Data Element Selection Form has been completed.

## Identifiable Data Use Approval

If you are requesting Identifiable information, approval from an Institutional Review Board (IRB) or a Privacy Board is required before such data can be released.

☐ Not applicable; the Client Organization is requesting a Limited Extract.

### Approval Type

☒ IRB Approval

☐ Privacy Board Approval

### Approval Type

☐ Approval request not yet submitted.
Anticipated submission date:

☐ Approval request submitted and under review.
Anticipated project approval date: 12/13/2024

☒ Approval already received.

### Approval Documentation

☐ By checking this box, the Client Organization confirms that the IRB or Privacy Board **application and approval documents** have been provided to CIVHC.

### Data Management Plan

An organization requesting CO APCD data must submit an organizational Data Management Plan to CIVHC outlining the organization's data security and data management policies and procedures to safeguard the data. This Data Management Plan must be approved by CIVHC prior to any data release.

| | |
|---|---|
| Date Submitted to CIVHC: | 2/24/2025 |
| Date Approved by CIVHC: | 3/10/2025 |

## Client Acknowledgements and Signatures

### Report or Product Distribution

If your project results in the production of a report for public distribution in any format (print, electronic, lecture, slides, etc.), including peer-reviewed publication, it must be submitted to CIVHC for review prior to public release. CIVHC will assess compliance with CMS Cell Size Suppression Policy, risk of inferential identification, CIVHC and CO APCD citations, and consistency with the purpose and methodology described in this Data Release Application. CIVHC will not assess the accuracy of the study results or attempt to recreate results.

This requirement is further defined in the Data Use Agreement. Failure to pursue and obtain CIVHC approval prior to publication will be a violation of the Data Use Agreement and may put the organization's future access to data from the CO APCD at risk.

☒ By checking this box, the Client Organization acknowledges this requirement.

### Data Destruction Period

All data must be destroyed within 30 days of the project end date. If your project end date changes from this application, please reach out to your CIVHC Contact for a project extension request form.

☒ By checking this box, the Client Organization acknowledges that CIVHC's Data Destruction Certificate[7] must be completed and returned to DataCompliance@CIVHC.org by 4/30/2030 based on the Anticipated Project End Date.

[7] Available on the Data Release Application and Documents page of CIVHC's website under *Privacy, Security, and Regulatory Information*.

## Data Users

List any individuals that will be working with the data. The Data Use Agreement must be updated through your CIVHC Contact every time individuals are granted access to the data during the course of the project.

| Full Name | Title/Role | Organization |
|---|---|---|
| **Brett R. Anderson** | Principal Investigator | Icahn School of Medicine at Mount Sinai |
| **Sarah Crook** | Director of Analytics | Icahn School of Medicine at Mount Sinai |
| **Pengfei Jiang** | Senior Data Analyst | Icahn School of Medicine at Mount Sinai |
| **Eric Zhou** | Biostatistician | Icahn School of Medicine at Mount Sinai |
| **Anna Chorniy** | Assistant Professor/Investigator | Icahn School of Medicine at Mount Sinai |
| **Michael Cassidy** | Assistant Professor/Investigator | Icahn School of Medicine at Mount Sinai |
| **Ellerie Webber** | Assistant Professor/Investigator | Icahn School of Medicine at Mount Sinai |
| **Son Duong** | Assistant Professor/Investigator | Icahn School of Medicine at Mount Sinai |
| **Emily Bucholz** | Assistant Professor/Investigator | Children's Hospital Colorado |
| | | |

## Data Release Application Version Approvals

The Client Organization has reviewed and confirms that the final version number of the Data Release Application reflected below correctly represents the project objectives.

| Version | Checkpoint |
|---|---|
| V.05 | Presented at CIVHC Application Review |
| V.07 | Presented to the Data Release Review Committee (DRRC) |
| V.00 | Final version approved for production |

| CIVHC Sign-Off | | Receiving Organization Sign-Off | |
|---|---|---|---|
| Signature: | | Signature: | |
| Name: | Lucía Sanders | Name: | Brett Anderson |
| Title: | Key Account Manager | Title: | Director, Center for Child Health Services Research |
| Date: | | Date: | |

## Data Element Selection Form Version Approvals

The Client Organization has reviewed and confirms that the final version number of the Data Element Selection Form reflected below correctly represents the data specifications needed to meet the project objectives.

| Version | Checkpoint |
|---|---|
| V.04 | Presented at CIVHC Application Review |
| V.05 | Presented to the Data Release Review Committee (DRRC) |
| V.00 | Final version approved for production |

| CIVHC Sign-Off | | Receiving Organization Sign-Off | |
|---|---|---|---|
| Signature: | | Signature: | |
| Name: | Lucía Sanders | Name: | Brett Anderson |
| Title: | Key Account Manager | Title: | Director, Center for Child Health Services Research |
| Date: | | Date: | |