EVALUATION DESIGN PLAN FOR COLORADO'S ADULT PRENATAL COVERAGE IN CHILD HEALTH PLAN PLUS (CHP+) SECTION 1115 DEMONSTRATION WAIVER

FINAL DRAFT
SEPTEMBER 10, 2021

HEALTH MANAGEMENT ASSOCIATES

EVALUATION TEAM MEMBERS:

MARK PODRAZIK, PRINCIPAL INVESTIGATOR

AKHILESH PASUPULATI
RYAN SANDHAUS
DEBBIE SAXE
BARRY SMITH

TABLE OF CONTENTS

Listing of Exhibits

Abbreviations List

Abbreviation	Meaning
ACA	Affordable Care Act
HMA-Burns	Burns & Associates, a Division of Health Management Associates
BIDM	Business Intelligence Data Management
CDC	Centers for Disease Control and Prevention
CHIP	Children's Health Insurance Program
CHIPRA	Children's Health Insurance Program Reauthorization Act of 2009
CHP+	Child Health Plus
CMS	Centers for Medicare and Medicaid Services
CoHID	Colorado Health Information Dataset
CPT	Current Procedural Terminology
CY	Calendar Year
DOS	Date of Service
DR	Desk Review
DS	Descriptive Statistics
DXC	DXC Technologies (now Gainwell)
E&M	Evaluation & Management
ED	Emergency Department
EDW	Enterprise Data Warehouse
FFS	Fee-For-Service
FG	Focus Groups
FI	Facilitated Interviews

Abbreviation	Meaning
FPL	Federal Poverty Level
HCPCS	Healthcare Common Procedure Coding System
HCPF	Department of Health Care Policy & Financing
IBM	IBM Corporation (formerly Truven Health Analytics)
ITS	Single Segment Interrupted Time Series
LBW	Low Birth Weight
MAGI	Modified Adjusted Gross Income
MCO	Managed Care Organization
MMIS	Medicaid Management Information System
NCQA	National Committee for Quality Assurance
OR	Onsite Reviews
PCP	Primary Care Provider
PRAMS	Pregnancy Risk Assessment and Monitoring System
PS	Provider Surveys
RAE	Regional Accountable Entity
RCT	Randomized Control Trials
SFY	State Fiscal Year
STC	Special Terms and Conditions
SUD	Substance Use Disorder
TJC	The Joint Commission

SECTION I: GENERAL BACKGROUND INFORMATION

I.A Waiver Demonstration Information[1]

Colorado has had a long-standing Section 1115(a) demonstration which was originally approved in 2002 and most recently extended from December 18, 2020 through July 31, 2025. The demonstration waiver was selected as a mechanism to allow Colorado to continue to provide coverage to uninsured pregnant women with family income using Modified Adjusted Gross Income (MAGI) equivalent between 141 and 195 percent of the federal poverty level (FPL). Colorado continues to use the Child Health Plus (CHP+) 1115 Demonstration to improve the health status of low-income pregnant women and their newborns by using the goals as described in Section I.B to guide the administration and implementation of the demonstration.

Name: Colorado Adult Prenatal Coverage in Child Health Plus (CHP+)
Project Number: 21-W-00014/8
Approval Date: December 21, 2020
Time Period Covered by Evaluation: December 18, 2020 through July 31, 2025

I.B Waiver Demonstration Goals[2]

Colorado's goals in operating the demonstration are to improve the health status of low-income Coloradoans by enabling a:

1. Decrease in the uninsurance rate for pregnant women;
2. Increase in prenatal and postpartum care for pregnant women enrolled in the demonstration; and
3. Increase in the number of healthy babies born to pregnant women enrolled in the demonstration.

I.C Brief Description and History of Implementation[3]

The Colorado Adult Prenatal Coverage in the CHP+ demonstration was initially approved on September 27, 2002 to provide coverage to uninsured pregnant women with family income above the CHP+ state plan level, from 133 to 185 percent of the FPL. At the time of initial approval, states only had the option to cover pregnant women above the CHP+ state plan level under title XXI, i.e., the Children's Health Insurance Program (CHIP) through a section 1115 demonstration.

[1] Colorado Adult Prenatal Coverage in Child Health Plan Plus (CHP+) Section 1115(a) Demonstration Special Terms and Conditions, accessed at https://www.CHP+.gov/CHP+-CHIP-Program-Information/By-Topics/Waivers/1115/downloads/co/co-adult-prenatal-coverage-ca.pdf
[2] Ibid, page 5 of 31
[3] Ibid, page 4 of 31

The Children's Health Insurance Program Reauthorization Act of 2009 (CHIPRA) added section 2112 to the Act which created the option for states to cover pregnant women in the CHIP state plan, but only if the state covered pregnant women in CHP+ up to at least 185 percent of the FPL. Consistent with CHIPRA, Colorado extended coverage in the CHIP state plan to pregnant women with family income up to 250 percent of the FPL but had to amend its CHP+ state plan to move pregnant women from 133 to 185 percent of the FPL from coverage under the CHIP section 1115 demonstration to the CHP+ state plan (effective January 1, 2013).

To support Colorado with continuing its pre-CHIPRA coverage of pregnant women from 133 to 185 percent of the FPL, the Centers for Medicare and Medicaid Services (CMS) grandfathered title XXI coverage for this population of uninsured pregnant women (at the MAGI-equivalent eligibility level of above 141 percent through 195 percent of the FPL) with the July 30, 2012 extension of the demonstration. Grandfathering title XXI coverage for these pregnant women is consistent with section 2112(f) of the Act (enacted by CHIPRA) that authorizes the continuation of other state options for providing medical assistance to pregnant women, including *pregnancy-related services through the application of any waiver authority (as in effect on June 1, 2008)*.

Colorado continues to operate the Adult Prenatal Coverage in CHP+ demonstration within the program authorities and implementation parameters in existence on June 1, 2008. In accordance with section 2112(f) of the Act, CMS approved a five-year extension of Colorado's grandfathered title XXI coverage in September 2015 (through July 31, 2020; temporarily extended through December 31, 2020) and is approving another five-year extension through July 31, 2025 with these STCs and associated expenditure and non-applicable authorities. The program authorities granted with this approval are solely limited to, and contingent upon, Colorado's continued implementation of its pre-CHIPRA coverage of pregnant women from 133 to 185 percent of the FPL (at the MAGI-equivalent of 141-195% of the FPL) in accordance with section 2112(f) of the Act.

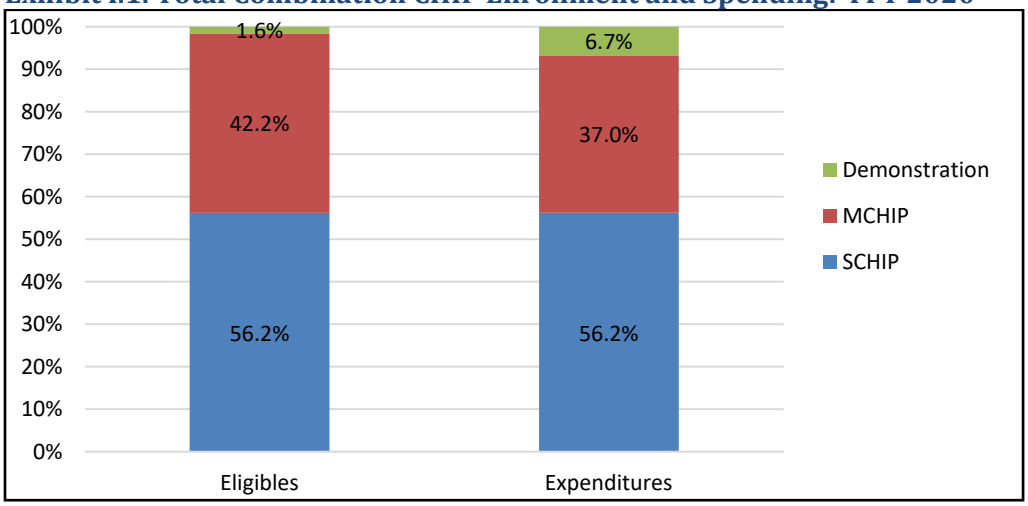
This demonstration furthers the objectives of title XXI by improving access to high-quality prenatal, delivery, and postpartum care services to low-income pregnant women that is producing positive health outcomes for beneficiaries. For example, the state's interim evaluation report for the 2015 – 2020 demonstration period shows that the state realized an 8.6 percent increase in the proportion of eligible beneficiaries accessing postpartum care from the state's baseline to demonstration year one. After the first demonstration year, this proportion remained relatively stable across the remaining demonstration years. Another positive outcome is the proportion of beneficiaries who gave birth to a low birth weight (LBW) baby decreased each year of the demonstration.

I.D Population Groups Impacted

Overview of Colorado's CHP+ Program

The Department of Health Care Policy & Financing (HCPF) has responsibility for the administration and oversight of Colorado's CHIP as well as the CHP+ program under the waiver and state plan authorities. As seen in Exhibit I.1, during federal fiscal year (FFY) 2020, CHP+ comprised 1.6% of the total enrollment of 135,265 and 6.7% of the total of $330 million in expenditures for Colorado's total combined CHIP program.

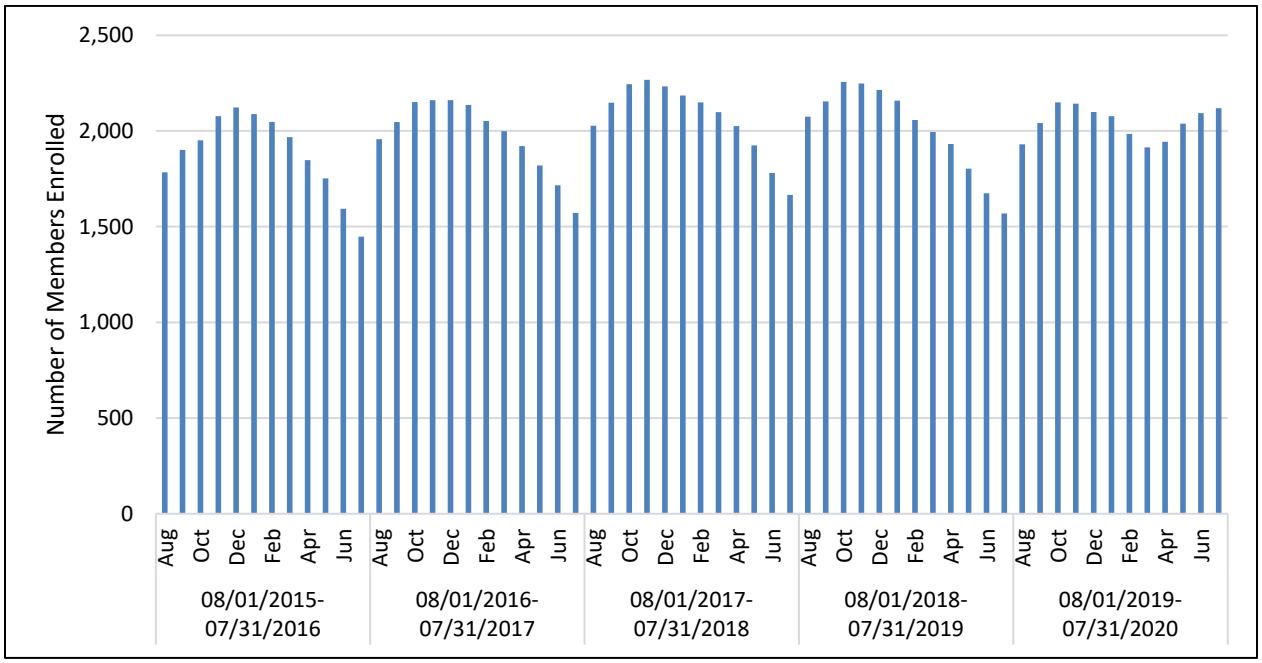
Exhibit I.1. Total Combination CHIP Enrollment and Spending: FFY 2020



Source: CHP+ Demonstration Extension Application and FFY 2020 Allotment Neutrality Report

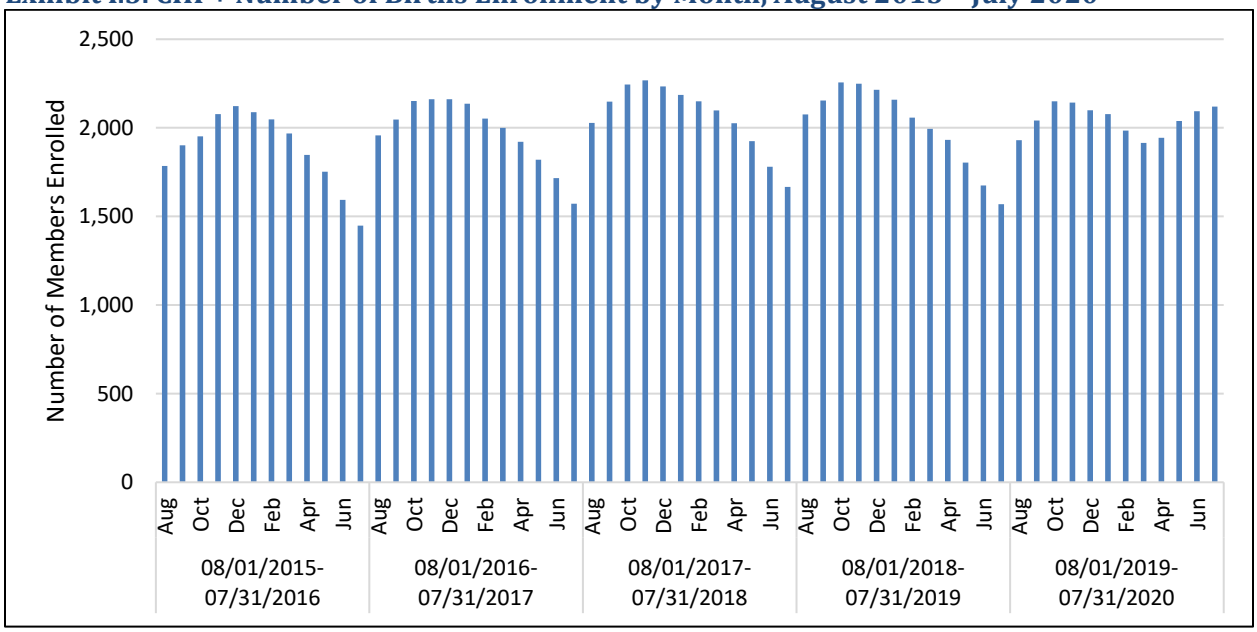
In the most recent demonstration year, there were 2,938 unduplicated pregnant women enrolled. Since 2015, monthly enrollment of pregnant women and births has trended upward as found in Exhibits I.2 and I.3.

Exhibit I.2. CHP+ Number of Women Enrolled in Prenatal Demonstration, August 2015 – July 2020



Source: CHP+ Client Data

Exhibit I.3. CHP+ Number of Births Enrollment by Month, August 2015 – July 2020

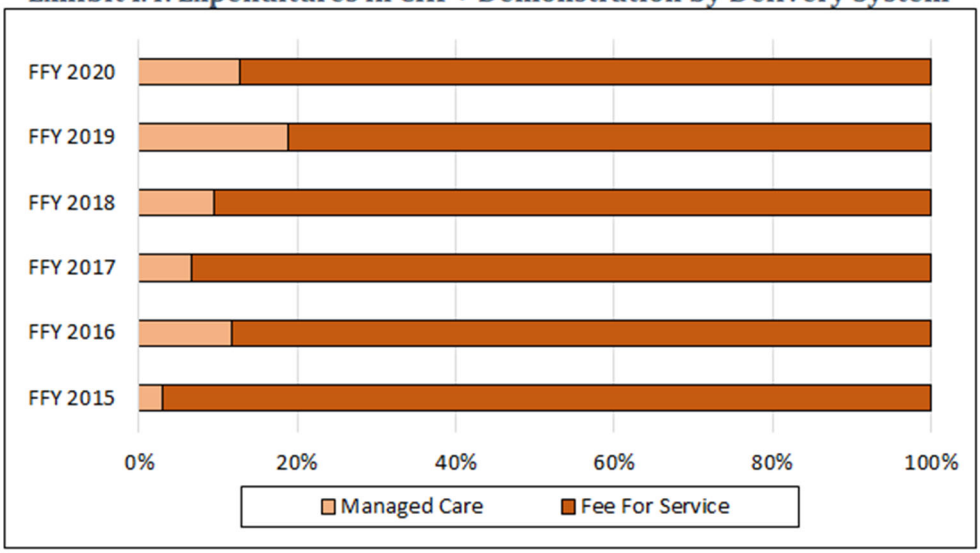


Source: CHP+ Client Data

CHP+ enrollees are entitled to receive all mandatory and optional state plans services approved under the Medicaid state plan. Services are provided through a combination of fee-for-service (FFS) and managed care delivery systems that vary geographically.

During this same time, the majority of Colorado's CHP+ demonstration expenditures were for care provided through the FFS delivery system, although the proportion of payments to managed care plans is increasing over time (refer to Exhibit I.4).

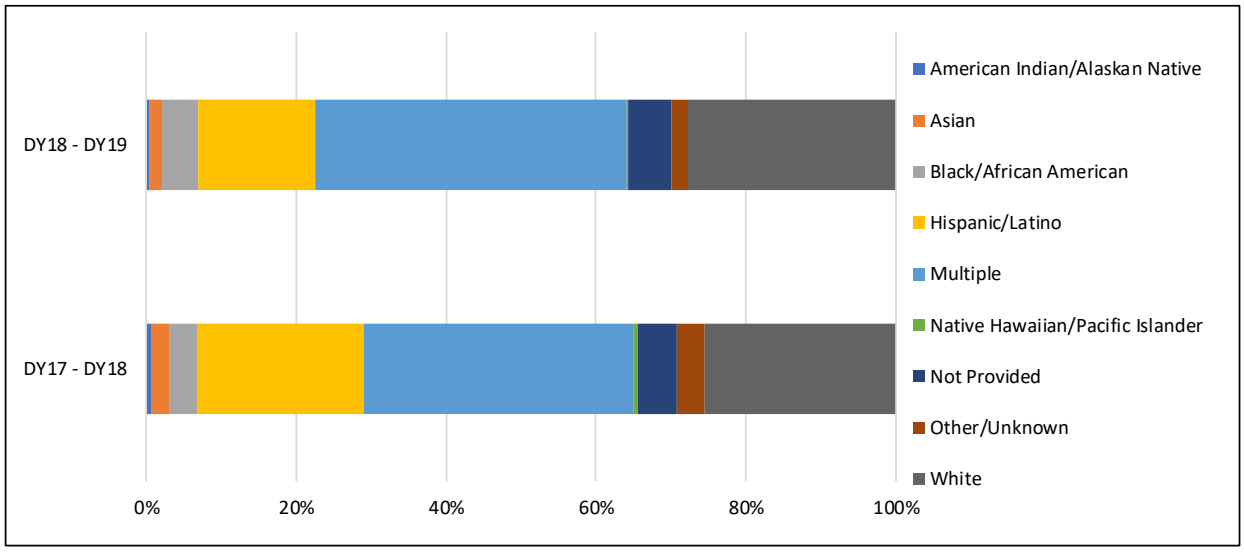
Exhibit I.4. Expenditures in CHP+ Demonstration by Delivery System



Source: CO CHP+ Allotment Neutrality Report

Of those members enrolled in the demonstration from 2018 to 2019, the most predominant race/ethnicity reported was multiple (41.6% of the total), followed by White (27.7%), Hispanic/Latino (15.5%), Black/African American (4.8%), Asian (1.7%), American Indian/Alaskan Native (0.4%) and Native Hawaiian/Pacific Islander (0.2%), and other/unknown or not provided (8.1%) (refer to Exhibit I.5 on the following page).

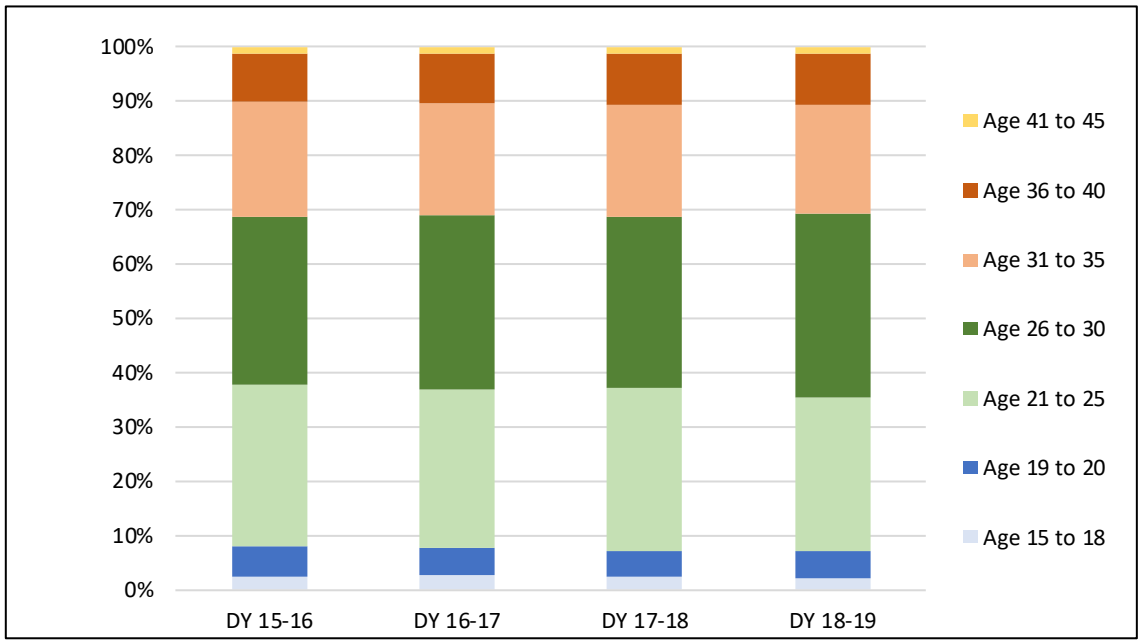
Exhibit I.5. Demonstration Population by Race/Ethnicity



Source: CO CHP+ Client Data

Exhibit I.6 distributes enrollment in the demonstration by the age of the members. Just over 60 percent of the women enrolled are between the ages of 21 and 30 (green portions of exhibit).

Exhibit I.6. Demonstration Population by Age Group



Source: CO CHP+ Client Data

SECTION II: EVALUATION QUESTIONS AND HYPOTHESES

II.A Defining Relationships: Waiver Policy, Short-term and Longer-term Outcomes

As part of the examination of the relationships between demonstration goals and the maturity of evaluating a long-term demonstration, the evaluation team at Burns & Associates, a Division of Health Management Associates (HMA-Burns) constructed logic models delineating short-term and longer-term outcomes associated with the three principle policy objectives of the demonstration.

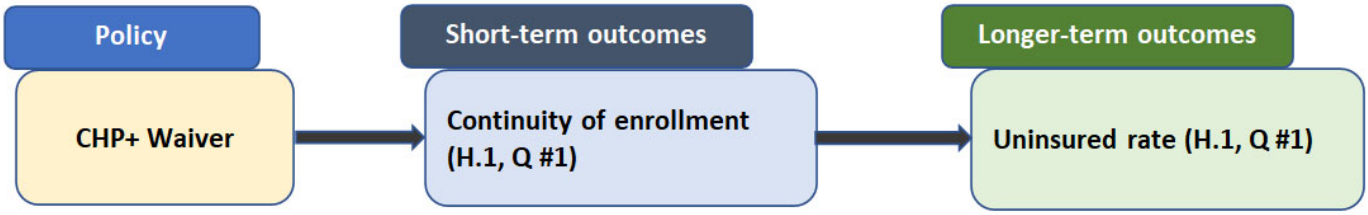
1. Maintain Continuity of Enrollment,
2. Maintain Access to Care, and
3. Maintain or Improve Health Outcomes

The determination of whether an outcome is short-term or longer-term is dependent on the measure specifications, including measurement period, and the data needed to adequately assess trends with the waiver policy. For example, because national outcome measures tend to have annual measurement periods, they are considered in this evaluation to be longer-term indicators of policy outcomes. Each of the three principle policy objectives are described in detail below and include logic models to illustrate both short-term and longer-term outcomes. Each logic model also provides a reference to specific hypotheses and research questions that will be described in Section II.B.

Maintain Continuity of Enrollment

HMA-Burns chose Maintain Continuity of Enrollment as the first policy objective as it is responsive to Waiver Goal #1, decreasing the rate of pregnant women who do not have insurance. Exhibit II.1 illustrates the baseline assumption that continuing the demonstration will not have an adverse impact on trends in the continuity of CHP+ enrollment in the short term. On a longer-term basis, the assumption is that trends in prenatal care paid by some type of insurance will not worsen over the course of the demonstration. Both process and outcome measures are proposed to assess impact.

Exhibit II.1. Logic Model 1: Maintain Continuity of Enrollment

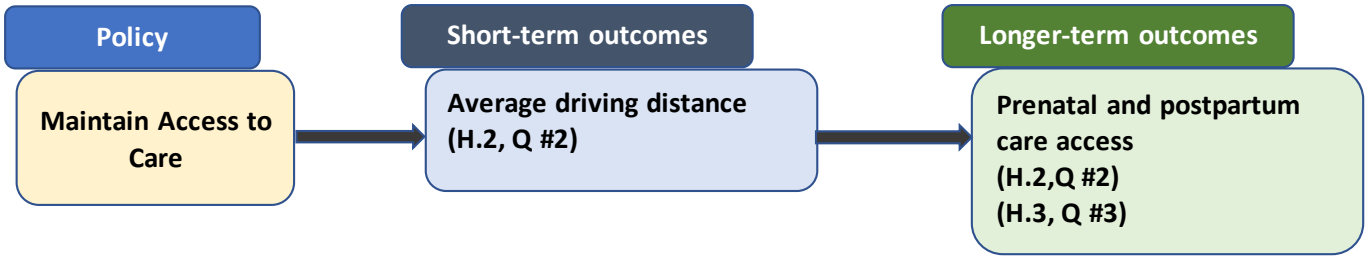


Maintain Access to Care

Maintain Access to Care is the second policy objective and it is based on Waiver Goal #2, increase in prenatal and postpartum care during the demonstration. Exhibit II.2 on the following page illustrates the assumption that trends in access to care continue or do not worsen. HMA-Burns is proposing to use outcome measures to assess trends in access to care. In the short term, trends in average driving

distance to prenatal care services will be assessed. To evaluate access to care on a longer-term basis, HMA-Burns is proposing to use established outcome measures of access and utilization.

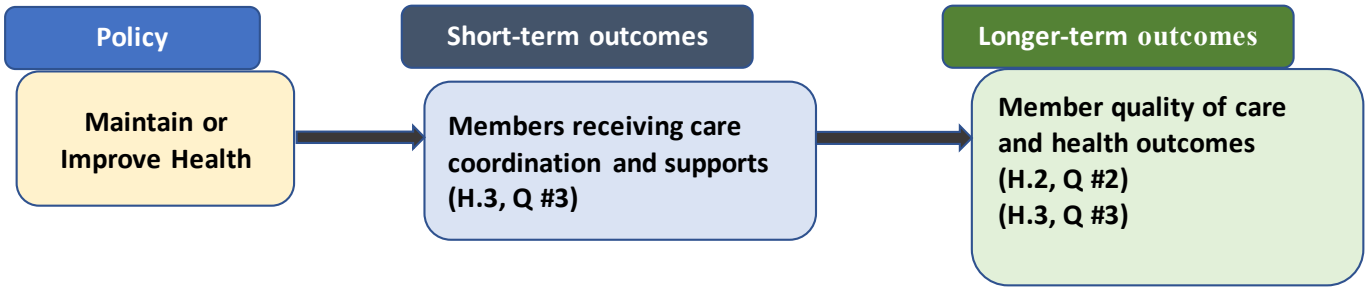
Exhibit II.2. Logic Model 2: Maintain or Improve Access



Maintain or Improve Health Outcomes

The third policy objective is Maintain or Improve Health Outcomes and it encompasses Waiver Goal #3, increase in the number of healthy babies born to pregnant women enrolled in the demonstration. Exhibit II.3 illustrates the assumption that CHP+ beneficiaries enrolled in the demonstration will maintain or improve health outcomes. In the short term, a process measure will measure access to care coordination and supports. On a longer-term basis, national health outcome metrics and HMA-Burns customized process measures focusing on care coordination will complete the assessment of the third principle policy objective.

Exhibit II.3. Logic Model 3: Maintain or Improve Health Outcomes



HMA-Burns found that there are existing, nationally-recognized outcome measures associated with principle policy objectives two and three. The specifications and data sources for many of these were already described as part of Colorado CHP+'s Quality Strategy. In addition to using nationally recognized outcome measures, HMA-Burns will fill gaps with custom measures developed by us where needed.

A more detailed description of the data, measures, and analyses to be used are described in Section III of the Evaluation Design document.

II.B Hypotheses and Research Questions

The three principle policy areas depicted in the logic models in Section II.A were converted into four hypotheses (H) and four research questions (Q). Each research question has assigned measures and a targeted analytic methodology which is described in detail in Section III. Methodology. Exhibit II.4 provides a high-level overview of each hypothesis and the associated research question. In most cases,

the research question assesses impact on both a short- and longer-term basis, except for Q #4 which has measures that only assess longer-term impact.

Exhibit II.4. Hypotheses and Research Questions

		Outcomes	
Hypothesis	**Research Question**	**Short-term**	**Longer-term**
H.1: Trends in continuity of enrollment in the demonstration continue (or do not worsen) for pregnant women in the current waiver period.			
	Q #1: *Does the waiver improve or maintain the uninsured rate of pregnant women in Colorado during the demonstration period?*	X	X
H.2: Trends observed in access to health care for pregnant women continues (or do not worsen) in the current waiver period.			
	Q #2: *Do CHP+ members achieve similar (or improved) access and health outcomes in the current waiver period?*	X	X
H.3: Trends observed in the health of the mother continues (or do not worsen) in the current waiver period.			
	Q #3: *Do CHP+ members achieve similar (or improved) pregnancy and postpartum outcomes in the current waiver period?*	X	X
H.4: Trends observed in the number of healthy babies (i.e., over 2500 grams) continues (or do not worsen) in the current waiver period.			
	Q #4: *Do CHP+ members achieve similar (or improved) birth outcomes in the current waiver period?*		X

II.C Alignment with Demonstration Goals

Building upon the matrix shown in Section II.B, each hypothesis was cross-referenced to demonstration goals. This was to ensure that the evaluation hypotheses and research questions are responsive to the CMS guidance in the approved waiver STCs. As demonstrated in Exhibit II.5 on the next page, each hypothesis addresses at least one demonstration goal and, in one case crosses two goals.

Exhibit II.5. Alignment of Hypotheses with Demonstration Goals

		Hypotheses			
		H.1	H.2	H.3	H.4
		Continuity of Enrollment	Acces to Health Care	Outcomes for Mother	Outcomes for Baby
Waiver Goals					
G.1	Decrease the uninsurance rate for pregnant women	X			
G.2	Increase prenatal and postpartum care for pregnant women enrolled in the demonstration		X		
G.3	Increase the number of healthy babies born to pregnant women enrolled in the demonstration			X	X

II.D How Hypotheses and Research Questions Promote Objectives of Titles XIX and XXI

The Evaluation Design Plan hypotheses were also cross referenced with the objectives of the CHP+ program[4] to ensure that the plan promotes the objectives of Titles XIX and XXI of the Social Security Act as required in Attachment A of the approved waiver STCs. Each hypothesis supports the principle objective to improve access to services that promote positive health outcomes. In the case of CHP+, the demonstration provides access to health care services for pregnant women and their newborns who otherwise would not qualify for these services.

[4]Accessed at: https://www.medicaid.gov/medicaid/section-1115-demonstrations/about-section-1115-demonstrations/index.html

SECTION III: METHODOLOGY

III.A Evaluation Design

The evaluation design is a mixed-methods approach, drawing from a range of data sources, measures and analytics to best produce relevant and actionable study findings. HMA-Burns tailored the approach for each of the research questions described in Section II, Evaluation Questions and Hypotheses. The evaluation plan reflects a range of data sources, measures and perspectives. It also defines the most appropriate study population and sub-populations, as well as describes the analytic methods included in the evaluation design.

The analytic methods proposed for use across the four hypotheses and four research questions include the following:

1. Descriptive statistics (DS),
2. Statistical tests (ST),
3. Desk reviews (DR), and
4. Facilitated interviews (FI).

Exhibit III.1 below presents a chart displaying which method(s) are used for each hypothesis. It also includes a brief description of the indicated methods as well as the sources of data on which they rely.

Exhibit III.1. Summary of Four Analytic Methods by Hypothesis

		Method				Analytic Method Examples
	Hypothesis Description	DS	ST	DR	FI	
1	Trends in continuity of enrollment in the demonstration continue (or does not worsen) for pregnant women in the current waiver period.	X		X		**DS**: trends in frequencies and percentages of enrollment duration and insurance status stratified by subpopulations of interest. Data sources: enrollment and CO PRAMS data.
2	Trends observed in access to health care for pregnant women continues (or does not worsen) in the current waiver period.	X	X	X	X	**DS**: trends in frequencies and percentages. **ST**: chi square or t-tests of significance; ITS. **DR/FI**: Prenatal Care focus study (2 rounds). Data sources: claims data and enrollment data, beneficiary interviews.
3	Trends observed in the health of the mother continues (or does not worsen) in the current waiver period.	X	X	X	X	**DS**: trends in frequencies and percentages. **ST**: chi square or t-tests of significance; interrupted time series. **DR/FI**: Prenatal Care focus study (2 rounds). Data sources: claims and enrollment data, reports submitted by MCOs/RAEs validated by HMA-Burns.
4	Trends observed in the number of healthy babies (over 2500 grams) continues (or does not worsen) in the current waiver period.	X	X	X		**DS**: trends in frequencies and percentages. **ST**: chi square or t-tests of significance; interrupted time series. Data sources: claims and enrollment data, state vital records, and CoHID.

DS = Descriptive Statistics; ST = Statistical Tests; DR = Desk Reviews; FI = Facilitated Interviews

As described in Section II.A, the majority of the hypotheses and associated research questions focus on whether the 1115 Demonstration made an impact on key CHP+ waiver goals (i.e., short-term and longer-term outcomes). In order to facilitate evaluation on whether a statistically significant difference between the pre-waiver and current waiver period can be detected, the data, measures and methods for these research questions will be tested using healthcare claims, member enrollment data, managed care organization (MCO) or regional accountable entity (RAE) report submissions, and provider enrollment data. The proposed metrics blend nationally-recognized measure specifications with custom metrics developed by HMA-Burns (where national metrics are unavailable). Analytic methods include interrupted time series (ITS) and descriptive statistics using chi-square tests or t-tests as applicable.

The focus shifts to assessing member perception of access to insurance, and quality. Given that these require information beyond what is available in claims or other public data sets, this section draws upon a set of mixed methods to evaluate progress. Where possible, measures will be incorporated into a reporting dashboard that tracks results from the pre-waiver period and the waiver-to-date period. Wherever possible, data will be tracked and reported on a quarterly basis.

III.B Target and Comparison Populations

Target Population

The target population is any Colorado CHP+ beneficiary enrolled in the demonstration in the study period. HMA-Burns will use Section III in the approved waiver STCs as the basis for identification of beneficiaries enrolled in the demonstration. HMA-Burns will create flags to identify CHP+ members and providers that will be part of the analytics. Flags will be assigned to attribute individuals to each sub-population group which includes, but is not limited to:

- MCO or RAE enrolled with
- Member race and ethnicity
- New member enrollment due to COVID
- Birthweight of newborn
- Member age (for specified age groups)
- Member home location (e.g., city/county/region)
- Substance Use Disorder

There will also be flags assigned to providers. The provider type and specialty will be tracked. HMA-Burns will use these indicators and create other flags that may require the joining of existing variables to assign providers by:

- Regional location
- Level of care
- Newly-enrolled and long-standing enrolled providers

The matrices included in Section III.G identify the target population and stratification proposed for each hypothesis and research question.

Comparison Groups

Two ideal comparison groups described in the CMS technical advisory guidance on selection of comparison groups include another state CHIP population and/or prospectively collected information

prior to the start of the intervention.[5] Specifically, a CHP+ population with similar demographics but in another state without those waiver flexibilities described in Colorado would be an ideal comparator. However, identifying whether such a state exists or the ability to obtain data from another state given the sensitivity of privacy concerns as it relates to data sharing is not feasible; therefore, it is outside the scope of this evaluation.

The other example of a control group described in the design guide is to collect prospective data. To our knowledge, there is no known prospective data collection on which to build baselines. Given the lack of an available and appropriate comparison group, HMA-Burns will use an analytic method which creates a pre-waiver and current waiver (intervention) group upon which to compare outcomes. See Section III.F for more details on the analytic methods.

Available results from CMS's Core Set of Health Care Quality Measures for Children in Medicaid and CHIP and the Initial Core Set of Health Care Quality Measures for Medicaid-Eligible Adults will be used as a benchmark comparator for those nationally-recognized metrics included in the evaluation design. Results of these measures are reported at a statewide level by CHIP program, as well as national values. In this case, comparator states will be identified and included, along with national values, within the Summative Evaluation. Comparator states will be chosen in consultation with the State, CMS and other stakeholders. For non-Core measures that align with Colorado Medicaid goals and initiatives for pregnant women, HMA-Burns will compute a benchmark using Colorado Medicaid as the comparator population. For average driving distance, HMA-Burns will use Colorado Medicaid and CHIP managed care organization, and Accountable Care Collaborative RAE distance standards to benchmark access.

III.C Evaluation Period

A pre-waiver and current wavier period will be defined as three calendar years before and five calendar years after waiver implementation. The pre-waiver period is defined as enrollment or dates of service from August 1, 2017 through December 17, 2020. The current waiver period is defined as enrollment or dates of service from December 18, 2020 through July 31, 2025. In support of the analytic methods described in Section III.F, the calendar year data will be further defined into both monthly and quarterly segments such that both the pre-periods will include 12 quarters or 36 months from the pre-waiver period, and 20 quarters or 60 months from the current waiver period.

To simplify the analytic plan, HMA-Burns is making an assumption about the first six months of 2020 prior to the current waiver being approved. For annual measures in which a national steward has defined measure specifications, HMA-Burns will consider August 1, 2019 to July 31, 2020 in the period prior to the current approved demonstration that became effective December 18, 2020. Although CMS approved Colorado's 1115 waiver in December 2020, waiver-related activities were moving forward in anticipation of approval of the extension throughout 2020. For ease of conducting and describing the analysis, the evaluation period will be defined as follows:

- For monthly and quarterly metrics, the six months in the 2020 calendar year prior to December 18, 2020 approval will be defined as the current waiver period (not the pre-waiver period).

[5] Comparison Group Evaluation Design. https://www.medicaid.gov/medicaid/section-1115-demo/downloads/ evaluation-reports/comparison-grp-eval-dsgn.pdf.

- For annual metrics, August 1, 2020 through July 31, 2025 will be considered the demonstration period.

It should be noted that, while this is the expected current evaluation period, modifications may be warranted to better reflect differences in the time period upon which one would expect to see a change in outcomes resulting from waiver activities. At this time, there was little data or similar studies available on which to base specific alternatives to the proposed current evaluation period. HMA-Burns, therefore, will examine time series data in order to identify whether the current evaluation period should be delayed. For example, if review of the data shows a distinctive change in the first and second quarter of 2021, then the current period would be adjusted such that the third and fourth quarter data would not be considered in the interrupted time series analysis described in Section III.F.

III.D Evaluation Measures

The measures included in the Evaluation Design Plan directly relate to the three principle policy objectives and short-term and longer-term outcomes described in Section II.

The measures fall into two primary domains: quality and access. Exhibit III.2 summarizes the list of measures included in the evaluation plan. A comprehensive summary of measures, which includes measure stewards as well as a description of numerators and denominators, can be found in the detailed matrices in Section III.G. Where possible, measure results will be stratified by race, ethnicity and region.

Exhibit III.2. Evaluation Measures by Domain

Quality
• Timeliness of Prenatal Care (PPC)
• Postpartum Care (PPC)
• Utilization of emergency department among PPC population
• At risk of poor maternal and/or infant health outcome
• Percentage of women who follow ACOG guidelines
• Proportion of at-risk deliveries
• Live births weighing less than 2,500 grams
• Well-child visits in the first 15 months of life

Access
• Utilization of prenatal care services per 1000 members
• Average driving distance to prenatal care services
• Proportion of enrollees continuously enrolled in CHP+
• Enrollment duration during pregnancy
• Prenatal care paid by type of insurance
• Proportion of PPC women, prenatal, using emergency department
• Proportion of PPC women, postpartum, using emergency department

III.E Data Sources

As described in Section III.A, Evaluation Design, HMA-Burns will use existing secondary data sources as well as collect primary data. The evaluation design relies most heavily on the use of Colorado CHP+ administrative data, i.e., enrollment, claims and encounter data. Supplemental administrative data, such as survey data, will also be incorporated. Primary data will be limited and will include data created by desk review and facilitated interview instruments. A brief description of these data and their strengths and weaknesses follow.

Colorado CHP+ Administrative Data

Claims and encounters with dates of service (DOS) from August 1, 2017 and ongoing will be collected from the Colorado Medicaid Management Information System (MMIS) Data Warehouse (EDW), facilitated by HCPF's MMIS vendor, Gainwell (formerly DXC) Technologies and IBM Corporation (formerly Truven Health Analytics) Business Intelligence Data Management (BIDM). A data request specific to the 1115 Evaluation Design Plan will be given to HCPF and the data will be delivered to the evaluators in an agreed-upon format. The initial EDW data set will include historical data up to the point of the delivery date. Subsequent data will be sent to HMA-Burns on a periodic basis. The last query of the EDW will occur on August 1, 2026 for claims with DOS in the study period. All data delivered to HMA-Burns from the HCPF will come directly from the EDW, including Vital Statistics data matched to

CHP+ enrollees. HMA-Burns will leverage all data validation techniques used by Gainwell before the data is submitted to the EDW. HMA-Burns will also conduct its own validations upon receipt of each monthly file from the HCPF to ensure accuracy and completeness when creating our multi-year historical database.

When additional data is deemed necessary for the evaluation, HMA-Burns will outreach directly to the MCOs and/or RAEs when they are determined to be the primary source. HMA-Burns will build data validation techniques specific to the ad hoc requests from the MCOs and/or RAEs.

Additional data from the MCOs and/or RAEs and the State will be collected on care coordination activities. There could be some data validity or quality issues with these sources as they are not as rigorously collected as claims and encounters data. That being said, we will use a standard quality review and data cleaning protocol in order to validate these data, as well as provide detailed specifications and reporting tools to the MCOs, RAEs and the State to minimize potential for differences in reporting of the requested ad-hoc data.

Survey and Facilitated Interview Data

Colorado Pregnancy Risk Assessment Monitoring System (PRAMS)[6]

The Colorado Pregnancy Risk Assessment and Monitoring System (PRAMS) is a survey of women to assess their experiences before, during and after pregnancy and includes CHP+ beneficiaries. Data is reported for women and infants at a granular level including, but not limited to, demographics and insurance level. The data will be used to review for descriptive trends over time of the percent of Colorado women who report being uninsured prior to, during, and after their pregnancy.

Facilitated Interview Guides

The evaluation team will construct facilitated interview guide instruments as a means to collect primary data for the prenatal care focus study. The types of respondents that the evaluators propose to interview are identified at the metric level in Section III. G. Respondents will include beneficiaries, the MCOs and RAEs. Beneficiary perspectives will be gathered using Colorado's Maternity Advisory Council, which leverages the lived experiences of maternity care to inform existing and emerging policy and is comprised primarily of Black, Indigenous and People of Color.[7] Where focused interviews are used to collect data, B&A will use semi-structured interview protocols that are intended to be standardized within the population being interviewed. Although semi-structured in nature, each stakeholder will have the opportunity to convey additional information that he/she would like to convey to the evaluators in an open-ended format at the conclusion of each interview.

Whereas the Colorado CHP+ administrative data will be collected and used on a monthly basis throughout the waiver period and after the waiver concludes to produce the Summative Evaluation, HMA-Burns anticipates that data from our sources will be collected in CY 2023 and CY 2025 for use in evaluation activities. Exhibit III.3 that appears on the next page contains the proposed primary data

[6] Accessed at https://cdphe.colorado.gov/center-for-health-and-environmental-data/survey-research/pregnancy-risk-assessment-monitoring

[7] Accessed at https://hcpf.colorado.gov/maternity-advisory-committee

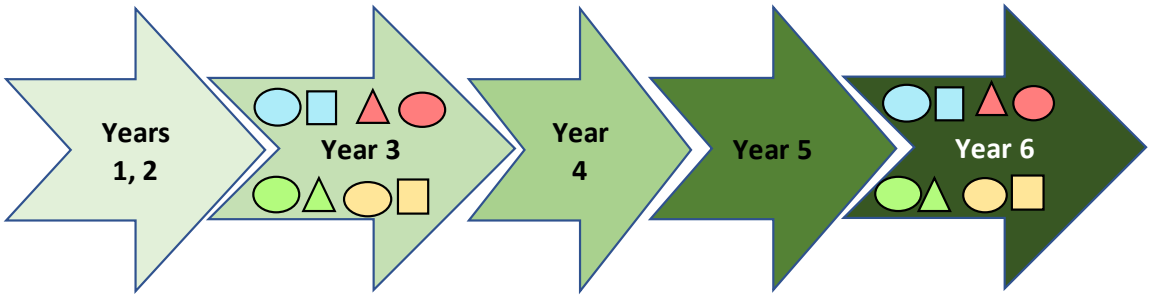
collection activities by source, year, and hypotheses. Exhibit III.4 that appears on page III-7 demonstrates the proposed primary data collection timeline by type, year, and hypotheses.

Exhibit III.3. Proposed Primary Data Collection Activities, by Source, Year and Hypotheses

Hypotheses	Source	Desk Review: MCOs RAEs	Desk Review: Other State Partners	Desk Review: State Agencies	Facilitated Interviews / Focus Groups: Members	Facilitated Interviews / Focus Groups: Other State Partners	Facilitated Interviews / Focus Groups: State Agencies	Facilitated Interviews / Focus Groups: MCOs RAEs
	Contract Year 1&2, CYs 2021-2022							
	All Hypotheses			X				
	Contract Year 3, CY 2023							
	1 Continuity of Enrollment		X	X	X			
	2 Trends in Access to Care			X	X			
	3 Trends in Outcomes for Mother	X		X			X	X
	4 Trends in Outcomes for Baby	X	X	X		X	X	X
	Contract Year 4, CY 2024							
	All Hypotheses			X				
	Contract Year 5, CY 2025							
	All Hypotheses			X				
	Contract Year 6, CY 2026							
	1 Continuity of Enrollment		X	X	X			
	2 Trends in Access to Care			X	X			
	3 Trends in Outcomes for Mother	X		X			X	X
	4 Trends in Outcomes for Baby	X	X	X		X	X	X

* Years shown correspond to Independent Evaluator contract years. Note: Presently, the State only has the authority to contract with HMA-Burns through December 31, 2022. There are deliverables due to CMS after this period reflected above.

Exhibit III.4. Proposed Primary Data Collection Timeline, by Type, Year and Hypotheses



Hypotheses

1 Continuity of Enrollment
2 Trends in Access to Care
3 Trends in in Outcomes for Mother
4 Trends in Trends in Outcomes for Baby

Methods

Desk Review
Member Survey
Facilitated Interview/Focus Group

Evaluator contract years. Note: Presently, the State only has the authority to contract with HMA-Burns through 12/31/22. There are deliverables due to CMS after this period which are reflected in this timeline.

III.F Analytic Methods

Exhibit III.1 depicted the analytic methods to be used in the analysis. A detailed discussion of each method is described below. This includes, where applicable, HMA-Burns' approach to address the impact of the COVID-19 pandemic within each method.

Method #1: Descriptive Statistics

In order to facilitate ongoing monitoring, all measures will be summarized on an ongoing basis over the course of the waiver. The descriptive statistics will be stratified by MCO, RAE and FFS delivery systems, and/or by region where possible. For reporting purposes, the descriptive studies will be subject to determination of a minimum number of beneficiaries in an individual reported cell (i.e., minimum cell size) and subject to blinding if the number falls below this threshold. While a conventional threshold is 10 or fewer observations, given the sensitivity of the small population size and the public dissemination of report findings, a higher threshold may be established by the evaluators upon review of the final data.

Results will primarily be reported in terms of longitudinal descriptive statistics of defined groups of beneficiaries and using regional maps where possible.

COVID-19 Considerations

For metrics where descriptive trends is the appropriate methodology, the evaluators propose to include a marker of pre- and post- COVID overlaid onto any graphs so one can visually inspect if there is an obvious change in the particular outcome starting mid-2020 and adding a comparator group.

In both cases, newly eligible members who became CHP+ eligible as a result of COVID will be identified and treated as a subpopulation in the analysis. This will allow the evaluators to continue to include those newly eligible members for which enrollment is unrelated to the pandemic.

Method 2: Statistical Tests

T-test or Chi-square test

Tests will be used to determine whether the observed differences in the mean value or rate differs for the most recent evaluation two-year period compared to the two-year period prior to waiver implementation. To assess if results for each metric compared to the pre-waiver timeframe are not due to chance alone, the evaluators will use chi-square tests for categorical data and t-tests for continuous data. Testing of the assumptions of normality and adjustments will be made before performing the final statistics and discussed below.

COVID-19 Considerations

For those metrics where simple statistics (chi square or t-test) is the appropriate quantitative methodology, the evaluators propose testing two separate post years to baseline to estimate the treatment effects before, during and after the pandemic. In both cases, members who became newly-

eligible for CHP+ as a result of COVID will be identified and treated as a subpopulation in the analysis. By doing this, HMA-Burns will be able to continue to include other newly-eligible members for which enrollment in CHP+ is unrelated to the pandemic.

T-test

The t-test is a type of inferential statistics. It is used to determine whether there is a significant difference between the means of two groups. Conceptually, it represents how many standardized units of the means of the pre- and post- populations differ. There are generally five factors to contribute to whether a statistically significant difference between the pre- and post-periods will be considered significant:[8]

1. How large is the difference? The larger the difference, the greater the likelihood that a statistically significant mean difference exists, and confidence increased.
2. How much overlap is there between the groups? The smaller the variances between the two groups, the greater probability a difference exists, hence increasing confidence in results.
3. How many subjects are in the two samples? The larger the sample size, the more stable and hence, confidence in results.
4. What alpha level is being used to test the mean difference? It is much harder to find differences between groups when you are only willing to have your results occur by chance 1 out of 100 times ($p < .01$) as compared to 5 out of 100 times ($p < .05$) but confidence in results is less.
5. Is a directional (one-tailed) or non-directional (two-tailed) hypothesis being tested? Other factors being equal, smaller mean differences result in statistical significance with a directional hypothesis so less confidence can be assigned to the results.

The assumptions underlying the t-test include:

- The samples have been randomly drawn from their respective population.
- The scores in the population are normally distributed.
- The scores in the populations have the same variance (s1=s2). A different calculation for the standard error may be used if they are not.

There are two types of errors associated with the t-test:

- Type I error —whereby the evaluator would detect a difference between the groups when there really was not a difference. The probability of making a Type I error is the chosen alpha level; therefore, an alpha level at $p < .05$, results in a 5% chance that you will make a Type I error.
- Type II error —whereby the evaluator detects no difference between the groups when there really was one.

The evaluators will consider results significant at a level of probability of $p < .05$. A test statistic will be generated in the SAS© statistical program. Assumptions will be tested and addressed if detected,

[8] T-test. https://researchbasics.education.uconn.edu/t-test/#. Accessed May 14, 2020.

including tests of normality and variance in the pre- and post- data. Metrics which are continuous will be tested using a t-test. The lowest level of reliable granularity available and reliability will be used for conducting tests (i.e., monthly or quarterly observations instead of annual).

Chi-square test

A chi-square test may be used in lieu of the t-test for some categorical variables. Chi-square may be preferable to t-test for comparing rates. All χ^2 tests are two sided.

The chi-square test for goodness of fit determines how well the frequency distribution from that sample fits the model distribution. For each categorical outcome tested, the frequency of patients in the pre- and post-period would be tested. The chi-square test for goodness of fit would determine if the observed frequencies were different than expected; in other words, whether the difference in the pre- and post-outcomes were significantly different statistically than what would have been expected given the pre-period. The null hypothesis, therefore, is that the expected frequency distribution of all wards is the same. Rejecting the null would indicate the differences were statistically significant (i.e., exceeded difference more than would be expected at a given confidence level).

The chi-square formula is: $\chi 2=\sum i=1k(O^i-E^i)2/E^i$

The assumptions of the chi-square are:

- Simple random sample
- Sample size. Small samples subject to Type II error.
- Expected cell count. Recommended 5-10 expected counts.
- Independence. Evaluation of the appropriateness of a McNemar's test may be warranted.

The evaluators will consider results significant at a level of probability of $p < .05$. A test statistic will be generated in the SAS© statistical program. Annually-reported categorical metrics for chi-square testing will either be derived from pooled population data (i.e., create one rate in pooled years of pre- and post-data) or two calendar year time periods (i.e., compare last year pre-waiver to last year post-waiver). Final approach will be determined upon examination of the data.

Interrupted Time Series (ITS)

Interrupted time series (ITS) is a quasi-experimental method used to evaluate health interventions and policy changes when randomized control trials (RTC) are not feasible or appropriate.[9,10,11] As it would not be ethical or consistent with CHP+ policy to withhold services resulting from waiver changes from a

[9] Bonell CP, Hargreaves J, Cousens S et al. Alternatives to randomisation in the evaluation of public health interventions: Design challenges and solutions. J Epidemiol Community Health 2009;65:582-87.
[10] Victora CG , Habicht J-P, Bryce J. Evidence-based public health: moving beyond randomized trials. Am J Public Health 2004;94:400–05.
[11] Campbell M , Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. . Framework for design and evaluation of complex interventions to improve health. BMJ 2000;321:694.

sub-set of beneficiaries for purposes of evaluation, an RTC is therefore, not possible. Per CMS technical guidance, the ITS is the preferred alternative approach to RTC in the absence of an available, adequate comparison group. The ITS method is particularly suited for interventions introduced at the population level which have a clearly defined time period and targeted health outcomes.[12,13,14]

An ITS analysis relies on a continuous sequence of observations on a population taken at equal intervals over time in which an underlying trend is "interrupted" by an intervention. In this evaluation, the waiver is the intervention, and it occurs at a known point in time. The trend in the post-waiver is compared against the expected trend in the absence of the intervention.

While there are no fixed limits regarding the number of data points because statistical power depends on a number of factors like variability of the data and seasonality, it is likely that a small number of observations paired with small expected effects may be underpowered.[15] The expected change in many outcomes included in the evaluation are likely to be small; therefore, the evaluators will use 72 monthly observations where possible and 24 quarterly observations where monthly data are not deemed reliable.

In order to determine whether monthly or quarterly observations will be created, a reliability threshold of having a denominator of a minimum number of 100 observations at the monthly or quarterly level will be used. If quarterly reporting is not deemed reliable under this threshold, the measure and/or stratification will not be tested using ITS. Instead, these measures will be computed using calendar year data in the pre- and post- period and reported descriptively.

ITS Descriptive Statistics

All demographic, population flags, and measures will be computed, and basic descriptive statistics will be created: mean, median, minimum, maximum, standard deviation. These data will be inspected for identification of anomalies and trends.

To identify underlying trends, seasonal patterns and outliers, scatter plots of each measure will be created and examined. Moreover, each outcome will undergo bivariate comparisons; a Pearson correlation coefficient will be produced for each measure compared to the others as well as each measure in the pre- and post- periods.

---

[12] Soumerai SB. How do you know which health care effectiveness research you can trust? A guide to study design for the perplexed. Prev Chronic Dis 2015;12:E101.

[13] Wagner AK , Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. J Clin Pharm Ther 2002;27:299-309.

[14] James Lopez Bernal, Steven Cummins, Antonio Gasparrini; Interrupted time series regression for the evaluation of public health interventions: a tutorial, International Journal of Epidemiology, Volume 46, Issue 1, 1 February 2017, Pages 348–355, https://doi.org/10.1093/ije/dyw098

[15] James Lopez Bernal, Steven Cummins, Antonio Gasparrini; Interrupted time series regression for the evaluation of public health interventions: a tutorial, International Journal of Epidemiology, Volume 46, Issue 1, 1 February 2017, Pages 348–355, https://doi.org/10.1093/ije/dyw09 8

Regression Analysis

Wagner et al. described the single segmented regression equation as[16]:

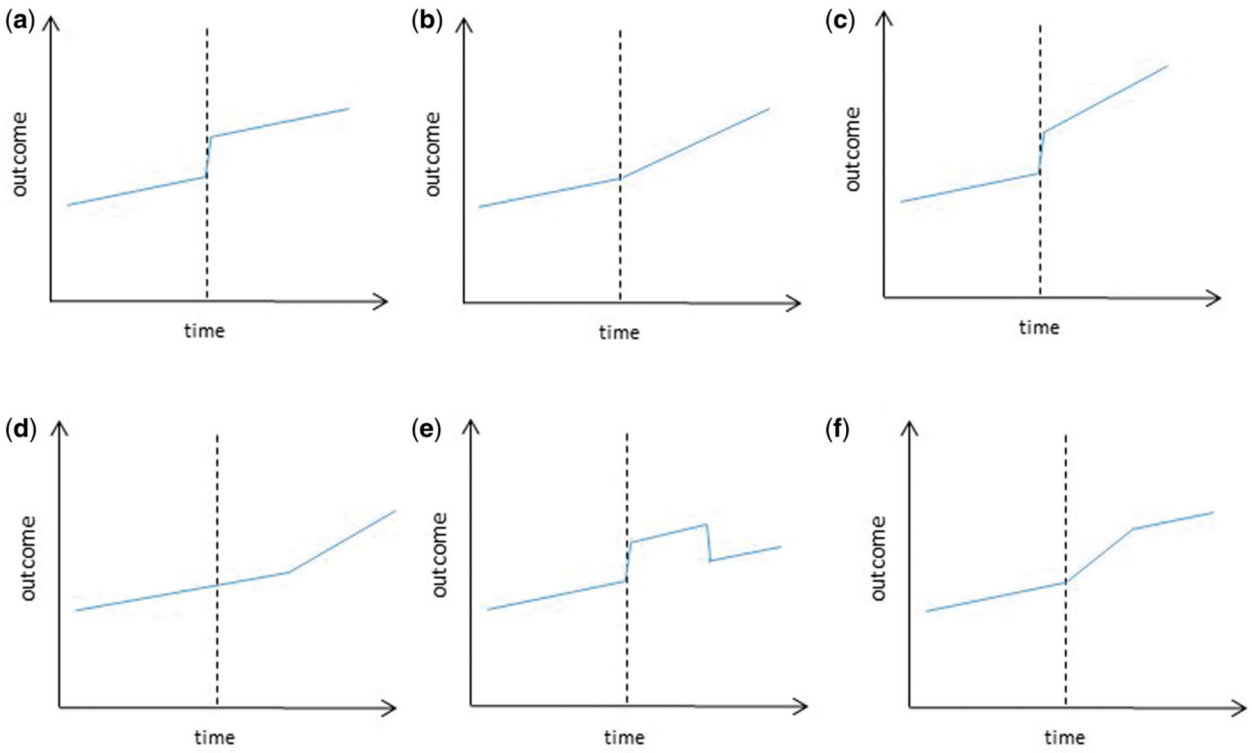
$$\hat{Y}_t = \beta_0 + \beta_1 * time_t + \beta_2 * intervention_t + \beta_3 * time\_after\_intervention_t + e_t$$

Where: Y_t is the outcome *time* indicates the number of months or quarters from the start of the series *intervention* is a dummy variable taking the values 0 in the pre-intervention segment and 1 in the post-intervention segment *time_after_intervention* is 0 in the pre-intervention segment and counts the quarters in the post-intervention segment at time t	β_0 estimates the base level of the outcome at the beginning of the series β_1 estimates the base trend, i.e., the change in outcome in the pre-intervention segment β_2 estimates the change in level from the pre- to post-intervention segment β_3 estimates the change in trend in the post-intervention segment e_t estimates the error

Visualization and interpretation will be done as depicted in the Exhibit III.5. Each outcome will be assessed for one of the following types of relationships in the pre- and post-waiver period: (a) Level change; (b) Slope change; (c) Level and slope change; (d) Slope change following a lag; (e) Temporary level change; (f) Temporary slope change leading to a level change.

[16] Wagner AK , Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. J Clin Pharm Ther 2002;27:299-309.

Exhibit III.5. Illustration of Potential ITS Relationships [17]



Seasonality and Autocorrelation

One strength of the ITS approach is that it is less sensitive to typical confounding variables which remain fairly constant, such as population age or socio-economic status, as these change relatively slowly over time. However, ITS may be sensitive to seasonality. To account for seasonality in the data, the same time period, measured in months or quarters, will be used in the pre- and post-waiver period. Should it be necessary, a dummy variable can be added to the model to account for the month or quarter of each observation to control for the seasonal impact.

An assumption of linear regression is that errors are independent. When errors are not independent, as is often the case for time series data, alternative methods may be warranted. To test for the independence, the evaluators will review a residual time series plot and/or autocorrelation plots of the residuals. In addition, a Durbin-Watson test will be constructed to detect the presence of autocorrelation. If the Durbin-Watson test statistic value is well below 1.0 or well above 3.0, there is an indication of serial correlation. If autocorrelation is detected, an autoregressive regression model, like the Cochrane-Orcutt model, will be used in lieu of simple linear regression.

[17] From: Interrupted time series regression for the evaluation of public health interventions: a tutorial Int J Epidemiol. 2016;46(1):348-355. doi:10.1093/ije/dyw098. Int J Epidemiol.

Other assumptions of linear regression are that data are linear and that there is constant variance in the errors versus time. Heteroscedasticity will be diagnosed by examining a plot of residuals verses predicted values. If the points are not symmetrically distributed around a horizontal line, with roughly constant variance, then the data may be nonlinear, and transformation of the dependent variable may be warranted. Heteroscedasticity often arises in time series models due to the effects of inflation and/or real compound growth. Some combination of logging and/or deflating may be necessary to stabilize the variance in this case.

Controls and Stratification

As described in Section III.B, the regression analysis will be run both on the entire target population and stratified by relevant sub-populations. The sub-population level analysis may reveal waiver effects that would otherwise be masked if only run on the entire population. Similarly, common demographic covariates such as age, gender, and race will be included in these models to the extent they improve the explanatory power of the ITS models.

COVID-19 Considerations

For those metrics where multivariate analysis is the appropriate quantitative methodology, the evaluators propose to construct a 0/1 dummy variable that indicates if the observations are post-March 2020 until a defined "post" COVID period for use as a control in the regression model. Members who became newly-eligible for CHP+ as a result of COVID will be identified by aid category and benefit plan and treated as a subpopulation in the analysis. This will allow the evaluators to continue to include those newly-eligible members for which enrollment is unrelated to the pandemic (e.g., aged, blind and disabled, pregnant women, newborns).

Method #3: Onsite Reviews

A limited number of desk reviews will supplement the other study methods included in the evaluation. These reviews will focus on hypotheses which are directed at assessment of process outcomes like avoidance of implementation delays, system changes according to schedules, transparency of policy and rates, and utility of stakeholder tools and analytics. Each desk review will use a questionnaire that asks for the information sought, the documentation reviewed, and the finding. Any gaps in information will also be noted as findings. The evaluators will review publicly-available information and/or documentation specifically requested from the HCPF and/or the MCOs and RAEs.

Method #4 Facilitated and/or Focus Group Interviews

As needed, the evaluators will construct facilitated interview guide instruments as a means to collect primary data for the focus studies. Intended respondents will include the MCOs, the RAEs, and beneficiaries eligible under this waiver demonstration. Where focused interviews are used to collect data, the evaluators will use semi-structured interview protocols that are intended to be standardized within the population being interviewed. The interview protocols will vary, however, for each population interviewed due to the type of information that is intended to be collected. Although semi-

structured in nature, each stakeholder will have the opportunity to convey additional information that he/she would like to convey to the evaluators in an open-ended format at the conclusion of each interview.

HMA-Burns will ensure that, for each population that interviews are conducted, there is sufficient representation within the population among those being surveyed. Sampling may be completed by using geographic location, provider size (large and small), and beneficiary age, to name a few.

III.G Other Additions

Beginning on the next page, Exhibit III.16 provides information on each measure selected for use in the evaluation, by research question and hypothesis.

Exhibit III.6. Summary of Evaluation Questions, Evaluation Hypotheses, Data Sources, and Analytic Approaches

Outcome	Measure description	Measure steward, endorsement	Numerator	Denominator	Data source	Analytic approach
Evaluation Question #1: ***Does the waiver improve or maintain the uninsured rate of pregnant women in Colorado during the demonstration period?***						
Demonstration Goal: G.1 Decrease the uninsurance rate for pregnant women.						
Evaluation Hypothesis #1: Trends in continuity of enrollment in the demonstration continue (or does not worsen) for pregnant women in the current waiver period.						
Short Term (Continuity of Enrollment)	Proportion of enrollees continuously enrolled in CHP+	HMA-Burns	Frequency distribution of enrollees continuously enrolled for the 9 months prior to delivery in the measurement period, stratified subpopulations of interest	Total number of enrollees during the measurement period.	Enrollment data	Descriptive statistics (trends in the proportion of enrollees continuously enrolled by subpopulations of interest)
	Enrollment duration during pregnancy	HMA-Burns	Frequency distribution of CHP+ enrollees by the number of months of eligibility in the measurement period, stratified by aid category and assignment plan.		Enrollment data	Descriptive statistics (trends in enrollment duration by subpopulations of interest)
Long Term (Continuity of Enrollment)	Prenatal care paid by type of insurance	Colorado PRAMS	Weighted percentage of respondents who reported the type of insurance coverage for prenatal care		Colorado PRAMS	Descriptive statistics (trends in Colorado reported percentages over the demonstration period); comparison to baseline period and available national and regional values

Outcome	Measure description	Measure steward, endorsement	Numerator	Denominator	Data source	Analytic approach
Evaluation Question #2: ***Do CHP+ members achieve similar (or improved) access and health outcomes in the current waiver period?***						
Demonstration Goal: G.2 Increase prenatal and postpartum care for pregnant women enrolled in the demonstration.						
Evaluation Hypothesis #2: Trends observed in access to health care for pregnant women continues (or does not worsen) in the current waiver period.						
Short Term (Access to Care)	Average driving distance to prenatal care services	HMA-Burns	Sum of the driving distances traveled from member home to their prenatal care provider	Sum of the unique trips to the member's prenatal care provider in the year	Claims data	Descriptive statistics (trends in average driving distance stratified by MCO/RAE and region)
Long Term (Access to Care)	Utilization of prenatal care services per 1000	HMA-Burns	Count of prenatal care services in the measurement period for CHP+ enrollees, and overall by sub-populations of interest	Total CHP+ enrollee member months for a 12-month study period (result of this formula expressed as per 1,000 member months)	Claims data	Descriptive statistics (frequencies and percentages) stratified by populations of interest; chi square or t-tests of significance comparing target population to baseline for Interim Evaluation; ITS for Summative Evaluation
	Proportion of PPC women using the emergency department	HMA-Burns	Number of PPC Timeliness of prenatal care women who had an emergency department visit during the pregnancy	Number of PPC Timeliness of Prenatal Care members	Claims data	Descriptive statistics (frequencies and percentages) stratified by populations of interest; chi square or t-tests of significance comparing target population to baseline for Interim Evaluation; ITS for Summative Evaluation
Long Term (Improved Outcomes)	Prenatal care for pregnant women (PPC): Timeliness of Prenatal Care	NCQA	1. Timeliness of Prenatal Care. Number of women having a prenatal care visit as a member of the organization in the first trimester, on the enrollment start date or w/in 42 days of enrollment in the organization.	1. Timeliness of Prenatal Care. Number of deliveries of live births.	Claims data	Descriptive statistics (frequencies and percentages) stratified by populations of interest; chi square or t-tests of significance comparing target population to baseline for Interim Evaluation; ITS for Summative Evaluation

Outcome	Measure description	Measure steward, endorsement	Numerator	Denominator	Data source	Analytic approach
Evaluation Question #3: ***Do CHP+ members achieve similar (or improved) pregnancy outcomes in the current waiver period?***						
Demonstration Goal: G.3 Increase the number of healthy babies born to pregnant women enrolled in the demonstration						
Evaluation Hypothesis #3: Trends observed in the health of the mother continues (or does not worsen) in the current waiver period.						
Short Term (Improved Outcomes)	Percentage of women determined to be at risk of poor maternal and/or infant health outcome	HMA-Burns	Count of women determined to be at risk of poor maternal and/or infant health outcome	Count of women screened	MCO/RAE specific report	Descriptive statistics (trends in the proportion of members determined to be at risk by subpopulations of interest)
Long Term (Improved Outcomes)	Percentage of women who follow ACOG guidelines overall and by subpopulation of interest	HMA-Burns	Count of pregnant women who followed ACOG guidelines overall	Number of CHP+ members	Claims data	Descriptive statistics (frequencies and percentages) stratified by populations of interest; chi square or t-tests of significance comparing target population to baseline for Interim Evaluation; ITS for Summative Evaluation
	Proportion of at-risk deliveries	HMA-Burns	Number of at-risk deliveries	Number of deliveries of live births	Claims data	Descriptive statistics (frequencies and percentages) stratified by populations of interest; chi square or t-tests of significance comparing target population to baseline for Interim Evaluation; ITS for Summative Evaluation
	Prenatal care for pregnant women (PPC): Postpartum Care	NCQA	2. Postpartum Care. Number of women having a postpartum visit on or between 21 and 56 days after delivery.	2. Postpartum Care. Number of deliveries of live births.	Claims data	Descriptive statistics (frequencies and percentages) stratified by populations of interest; chi square or t-tests of significance comparing target population to baseline for Interim Evaluation; ITS for Summative Evaluation
Long Term (Access to Care)	Proportion of PPC women using the emergency department	HMA-Burns	Number of PPC Postpartum Care women who had an emergency department visit during the pregnancy	Number of PPC Postpartum Care members	Claims data	Descriptive statistics (frequencies and percentages) stratified by populations of interest; chi square or t-tests of significance comparing target population to baseline for Interim Evaluation; ITS for Summative Evaluation

Outcome	Measure description	Measure steward, endorsement	Numerator	Denominator	Data source	Analytic approach
Evaluation Question #4: ***Do CHP+ members achieve similar (or improved) birth outcomes in the current waiver period?***						
Demonstration Goal: G.3 Increase the number of healthy babies born to pregnant women enrolled in the demonstration						
Evaluation Hypothesis #4: Trends observed in the number of healthy babies (over 2500 grams) continues (or does not worsen) in the current waiver period.						
Long Term (Improved Outcomes)	Live Births Weighing Less Than 2,500 Grams (LBW-CH)	CDC	Number of babies born low birthweight (less than 2500 grams).		State vital records, CoHID	Descriptive statistics (frequencies and percentages); chi square or t-tests of significance comparing target population to baseline for Interim Evaluation; ITS for Summative Evaluation
	Well-Child Visits in the First 15 Months of Life (W15)	NCQA	Number of children who turned 15 months old during the measurement year who had 6 or more well-child visits with a PCP	Number of children who turned 15 months old during the measurement year.	Claims data	

SECTION IV: METHODOLOGICAL LIMITATIONS

There are inherent limitations to both the study design and its specific application to the 1115 waiver evaluation. That being said, the proposed design is feasible and is a rational explanatory framework for evaluating the impact of the 1115 waiver on the demonstration population. Moreover, to fill gaps left by the limitations of this study design, a limited number of desk reviews and facilitated interviews/focus groups are proposed to provide a more holistic and comprehensive evaluation. Some known limitations are addressed below.

Since Colorado's population will be small compared to other states, some metrics and/or sub-populations may not be meaningful for reporting and there will be a concern about insufficient statistical power to detect a difference. For any observational studies, it may be difficult to find statistically significant results, particularly if the population size is low. We will recommend a threshold for a minimum number of observations. For any measures below this threshold, the expectation of statistical testing would be waived.

While CMS prefers a true comparator group from another state, this would require significantly more resources and cooperation with another state on sharing data. Therefore, HMA-Burns is recommending the use of ITS and descriptive statistics including the use of chi square or t-tests as the starting point in development of the evaluation design. One exception to this would be to use available results from CMS's Core Set of Health Care Quality Measures for Children in Medicaid and CHIP and the Initial Core Set of Health Care Quality Measures for Medicaid-Eligible Adults as a benchmark comparator for nationally recognized metrics included in the evaluation design. While the populations and benefit packages may be similar, there will still be differences from Colorado's demonstration population. In this scenario, HMA-Burns would compare these trends to two other states and national values if desired and if the data is available. The determination of the states to compare to would be done in consultation with the State, CMS and other stakeholders, and will note the limitations associated with the selected benchmarks.

For non-Core measures that align with Colorado Medicaid goals and initiatives for pregnant women, HMA-Burns will compute a benchmark using Colorado Medicaid as the comparator population. Using Medicaid as the comparator has its limitations as the benefit package is identical, with the only difference being the demonstration population has income that is more than 141% to 195% FPL.

For average driving distance, HMA-Burns will use Colorado Medicaid and CHIP managed care organization, and Accountable Care Collaborative RAE distance standards to benchmark access. Using Medicaid, RAE and CHIP distance standards are comparators is limitations as they include a broader population than the demonstration.

Use of Colorado's Maternity Advisory Council to obtain beneficiary perspectives on lived experiences of maternity care offers a unique opportunity to collect qualitative information. However, the council is not specific to the demonstration population and will also include Medicaid beneficiary input. Therefore, it may not be possible to attribute qualitative observations solely to the demonstration population.

The use of Colorado PRAMS data as the source for insurance status was proposed because it is obtained using a standard survey instrument collecting data from pregnant women and includes CHP+ breakouts

as well as commercial insurance and Medicaid breakouts. While it can provide broad context, there is no ability to link the survey results to demonstration enrollees.

The fact that the 1115 waiver components have been in place during what would be considered the pre-waiver period for evaluation purposes will make identifying any changes in outcomes directly attributable to waiver implementation difficult. Therefore, it is expected that not all outcomes or process measures included in the study will show a demonstrable change descriptively, and in fact may show no change in trends from the prior demonstration period. Where possible, the use of national or benchmark trends may provide context in this instance.

Equally, observed changes in outcome metrics in the current waiver period will be difficult, if not impossible, to attribute to one specific demonstration component given the interrelationship of the components themselves and the longstanding nature of the demonstration. Therefore, it will be important to use statistical tests of significance so that findings are properly put into context.

Related to the issues mentioned above, many of the outcome measures are multi-dimensional and influenced by social determinants of health. In addition, the State has multiple efforts underway to address prenatal care and birth outcomes that may influence the results of the demonstration. While changes under the waiver related to access to care may be one dimension of various outcomes of interest, and may contribute to improvements, it may be difficult to achieve statistically significant findings in the absence of data on other contributing dimensions, such as housing and employment.

Lastly, the evaluators recognize that the utilization patterns that will occur relatively early in this demonstration period will be severely disrupted due to the COVID-19 pandemic. The predictability of future utilization patterns remains uncertain as of the date of this document. The evaluators are prepared to work with CMS in the event that guidance is provided to states for all waiver evaluations as to options that CMS will offer with respect to how to account for the acute period of the pandemic. The initial plan for handling COVID-19 effects are addressed in Section III. Methodology.

ATTACHMENT A: INDEPENDENT EVALUATOR

Process

Burns & Associates, a division of HMA, (HMA-Burns) submitted a proposal through a competitive bid process to be retained for professional services to facilitate the research and design of the Colorado Adult Prenatal Coverage in CHP+ Section 1115 demonstration evaluation with the Colorado Department of Health Care Policy & Financing (HCPF). The current contract was entered into effective March 1, 2021 with an end date of December 31, 2022.

Vendor Qualifications

Burns & Associates (B&A) was founded in 2006 and was in continual operations until September 1, 2020 when it was acquired by Health Management Associates. The staff at Burns & Associates all migrated to Health Management Associates with this change. The B&A team, now a division of HMA, works almost exclusively with state Medicaid agencies or related social services agencies in state government. The B&A team has worked with 33 state agencies in 26 states. Current team members are also completing Section 1115 waiver evaluations in Delaware and Indiana. For Delaware, the evaluation of its 1115 Diamond State Health Plan Waiver Demonstration Project and its Substance Use Disorder waiver; for Indiana, the evaluation of its 1115 Substance Use Disorder waiver. For all three projects, the B&A team has developed the approved Evaluation Design Plan and completed CMS-approved Interim Evaluation and Mid-Point Assessment reports (in Indiana). B&A has also conduced independent assessments of Indiana's 1915(b) waiver for Hoosier Care Connect and served as the External Quality Review Organization (EQRO) for Indiana from 2007 to 2020.

Assuring Independence

In accordance with standard term and condition Section IX Evaluation of the Demonstration and Attachment A– Developing the Evaluation Design, HMA-Burns attests to having no conflicts to perform the tasks needed to serve as an independent evaluator on this engagement. The HMA-Burns Principal Investigator is prepared to deliver a signed attestation to this effect upon request.

ATTACHMENT B: EVALUATION BUDGET

As part of the procurement process, Burns & Associates, a Division of HMA (HMA-Burns) was required to submit a work plan that presents the level of effort to complete all deliverables associated with the independent evaluation of Colorado's Adult Prenatal Coverage in CHP+ Section 1115 demonstration evaluation. Presently, the State only has the authority to contract with HMA-Burns through December 31, 2022, and there are deliverables due to CMS after December 31, 2022 which are reflected in the Attachment C Timelines and Milestones.

In recognition of the contracting limitations, Exhibit B.1 Proposed Costs for 1115 Waiver Evaluation summarizes the amount to complete deliverables due to CMS through December 31, 2022.

Exhibit B.1 Proposed Costs for 1115 Waiver Evaluation through December 31, 2022

	Deliverable	Proposed Cost
SFY 21	2019-2020 Annual Monitoring Report	$30,000
	2020-2022 Project Work Plan	$3,875
	2020-2025 Evaluation Design	$15,200
	2020-2025 Final Evaluation Design	$4,950
SFY 22	2015-2020 Draft Summative Report	$42,325
	2015-2020 Final Summative Report	$3,300
	2020-2025 Project Charter	$2,750
	2020-2021 Annual Monitoring Report	$28,000
SFY 23	2021-2022 Annual Monitoring Report	$28,000
	Total Year 1 (SFY 2021)	$54,025
	Total Year 2 (SFY 2022)	$76,375
	Total Year 3 (SFY 2023)	$28,000
	TOTAL	$158,400

ATTACHMENT C: TIMELINE AND MILESTONES

As part of the procurement process, Burns & Associates, a Division of HMA (HMA-Burns) was required to submit a work plan, including major tasks and milestones to complete the scope of work. Presently, the State only has the authority to contract with HMA-Burns through December 31, 2022. There are deliverables due to CMS after December 31, 2022.

HMA-Burns has built a work plan for the independent evaluation of Colorado's Adult Prenatal Coverage in CHP+ Section 1115 demonstration that is constructed around the development of each deliverable identified as part of CMS required deliverables and the State's obligations related to monitoring and evaluation (M&E) activities.

The main sections of the work plan are as follows:

- Section A, ***Project Management***, includes Tasks 1, 2 and 3. The tasks in the section will be conducted across the entire engagement.
    - Tasks in this section:
        - Kickoff meeting
        - Project management and project plan
        - Project charter
    - Deliverables in this section:
        - Monthly status and other project management reports
        - Project charter

- Section B, ***Annual Monitoring Activities and Ongoing Assistance***, includes Tasks 4 through 6. It is anticipated that the work in this section will start immediately upon contract execution and continue until January 31, 2027.
    - Tasks in this section:
        - Obtain and read in data for project
        - Create Annual Monitoring Reports
        - Ongoing consultation and technical assistance
    - Deliverables in this section:
        - Creation and maintenance of the analytic data warehouse specific to the Evaluation Design Plan and associated focus study
        - Compute and validate metrics specific to the Evaluation Design Plan on an annual basis
        - Annual Monitoring Reports (6 total)

- Section C, ***Summative Evaluation and Evaluation Design Plan Activities***, includes Tasks 7 through 8. It is expected that the work in this section will start immediately upon contract execution and continue until September 30, 2021.
    - Tasks in this section:
        - Prepare Summative Evaluation for 2015 to 2020 Demonstration
        - Develop Evaluation Design Plan for 2020 to 2025 Demonstration
    - Deliverables in this section:
        - Draft Evaluation Design for 2020 to 2025 Demonstration to CMS (May 15, 2021)
        - Final Evaluation Design for 2020 to 2025 Demonstration to CMS (July 14, 2021)
        - Draft Summative Evaluation for 2015 to 2020 to CMS (July 14, 2021)

        - Final Summative Evaluation for 2015 to 2020 to CMS (September 13, 2021)

- Section D, ***Interim Evaluation Activities,*** includes Task 9. It is expected that the work in this section will start in Q4 of CY 2023 and continue until July 31, 2024. Task 9 includes a pregnancy services focus study with an internal report to HCPF along with work to produce the Interim Evaluation itself. Results from the focus study will be included in the Interim Evaluation to CMS.
    - Tasks in this section:
        - Conduct one focus study (September 2023 – January 2024)
        - Prepare Interim Evaluation
    - Deliverables in this section:
        - Intermittent reports for the focus study during the 4-month period study period
        - Detailed outline of the Interim Evaluation (January 2024)
        - Draft Version of Interim Evaluation to CMS (June 31, 2024)
        - Final Version of Interim Evaluation to CMS (July 2024)

- Section E, ***Summative Evaluation Deliverables***, includes Task 10 and is expected to repeat the pregnancy services focus study as a follow-up to what was reported on in the Interim Evaluation It is expected that the work in this section will start in Q1 of CY 2026 and continue until January 31, 2027.
    - Tasks in this section:
        - Conduct one focus study (March 2026 – June 2026)
        - Prepare Summative Evaluation
    - Deliverables in this section:
        - Intermittent reports for the focus study during this 4-month study period
        - Detailed outline of the Summative Evaluation (July 2026)
        - Draft Version of Summative Evaluation to CMS (December 2026)
        - Final Version of Summative Evaluation to CMS (January 2027)