COLORADO
Lieutenant Governor Dianne Primavera

January 17, 2020

Greetings:

The Colorado State Innovation Model (SIM) changed the way health care is delivered and paid for in Colorado. And while SIM's overarching objectives were directed at systemic reforms across the state's health care landscape, at its core, the initiative was about people- and improving the health of Coloradans by increasing access to "whole person" care.

The final evaluation reports now available on the SIM website – including the SIM Final Report, SIM Final Evaluation Outcomes Report, SIM Final Evaluation Process Report, and SIM Return on Investment (ROI) Analysis – offer a detailed analysis on SIM's many successes, as well as the challenges and lessons learned.

While SIM officially came to an end on July 31, 2019, its impact will be felt for years to come. As Governor Polis and I continue to work with all of you to help implement our "Roadmap to Saving Coloradans Money on Health Care," the following SIM lessons and findings will be at the top of our minds:

- **Integrated physical and behavioral health results in cost savings**. Results from the analyses of SIM's ROI are extremely encouraging, showing an estimated cost savings of $178.6 million through January 1, 2018. In addition, several cost and utilization measures analyzed in the SIM Final Evaluation Outcomes report also showed positive impacts- such as a reduction in emergency department utilization, and lower rates of 30-day hospital readmissions for mental health conditions. Evaluators used different methodologies to calculate cost savings (or avoided costs), and the results of their analyses raise questions that merit future investigation and research.
- **Integrated physical and behavioral health also improves care delivery.** SIM's success in improving access to the right care, at the right time, in the right place is most powerfully expressed through the stories of the patients and the providers who were involved in the initiative, which can be found on the SIM website. The Evaluation Reports offer further evidence of improved care quality, resulting in improved outcomes. This information will be critical in directing future state efforts to strengthen and improve primary care delivery- work that is currently being pursued by the Colorado Primary Care Payment Reform Collaborative.
- **Systems change requires strong relationships and cross-sector partnerships**. Colorado SIM was unique, among other states who received SIM awards, in its level of stakeholder engagement. The relationship and trust building that occurred over the course of the initiatives - between payers and providers, care team members working in integrated setting, state agencies and public partners - were instrumental to SIM's success.

I encourage you not only to read the wealth of information contained in the reports, but to find new ways to engage in care delivery and payment reform efforts currently underway in Colorado. SIM shows that true reform takes sustained engagement, motivation, and cooperation- it is now up to all of us to take up the reigns and work together to advance the health of all Coloradans.

Sincerely,

Dianne Primavera

Dianne Primavera
Lieutenant Governor

1876

SIM
State Innovation Model

COLORADO
Governor Jared Polis

Final Evaluation Reports Submission Cover Letter

July 31, 2019

Background: In our initial proposal, Colorado SIM proposed to include Final and Summative Evaluation Reports as part of the final evaluation. Over the course of time it became apparent that these reports made more sense titled the Process and Outcomes Evaluation Reports. In 2015 TriWest was selected through a competitive selection process to be the State Led Evaluator (SLE) for Colorado SIM. The following is a review of challenges encountered in our work on the evaluation that should be kept in mind when reviewing this report.

Data Lag and Quality Challenges:

- Substantial portions of the analyses are based off claims data from the All-Payer Claims Database (APCD), managed by the Center for Improving Value in Health Care (CIVHC). Due to the lag in reporting of claims data, this analysis includes data from 2015 – 2017. This limits our ability to measure the impact of the initiative since we have limited data for the implementation periods of the cohorts: practice transformation support was provided to cohort 1 from February 2016 through March 2018; to cohort 2 from September 2017 through June 2019; and to cohort 3 practices from June 2018 through June 2019. This means that only eleven months of cohort 1 and four months of cohort 2 are included in these analyses and cohort 3 is excluded. Our logic model posits that the initiative will impact cost and utilization first by increasing utilization of certain upstream services when patients are able to access the physical and behavioral care that they need and that this improvement in care will lower the utilization of more costly downstream acute services. Since it may take years to see these effects, future analyses may be able to more accurately measure the impact that SIM had on cost and utilization.
- Medicaid and CIVHC both underwent data processing vendor changes during the 2016-2017 period. There was a significant delay in available data and inconsistencies in the data across the partners. The SIM Office worked extensively with Medicaid and CIVHC to identify the time period and extent of variation and agree to a process moving forward.
- Payer data was regularly asked for but was extremely difficult to collect. As a result, the evaluation was unable to address several payment reform questions.
- The operational activities of the initiative ended so close to the initiative closeout date of July 31st, which created a considerable backlog of work to finalize the data and assessments that was collected from practices occurred. The initiative had essentially six weeks to finalize much of this data. Just checking the data quality and finalizing the data in and of itself would have been taxing in that short timeframe. To additionally analyze that data, create a report, review, and finalize said report was extremely challenging.

The project described was supported by Funding Opportunity Number CMS-1G1-14-001 from the U.S Department of Health and Human Services (HHS), Centers for Medicare & Medicaid Services (CMS). The Colorado State Innovation Model (SIM), a four-year initiative, is funded by up to $65 million from CMS. The content provided is solely the responsibility of the authors and does not necessarily represent the official views of HHS or any of its agencies.

- In late 2017 it became apparent that some of the evaluation questions that initially selected in 2015 were too extensive or did not have an available data source. To address this issue the SIM office instructed TriWest to lead an effort to narrow the scope of the evaluation. While this effort occurred, it was not a thorough as it could have been. Evaluation questions were maintained without available data sources. This did not become apparent until final evaluation plans were presented midway into AY 4 with the outline of the methodology.
- The SIM Office worked to review and fact-check the evaluation report but are not responsible for the results, analysis or interpretation included in this evaluation.

The project described was supported by Funding Opportunity Number CMS-1G1-14-001 from the U.S Department of Health and Human Services (HHS), Centers for Medicare & Medicaid Services (CMS). The Colorado State Innovation Model (SIM), a four-year initiative, is funded by up to $65 million from CMS. The content provided is solely the responsibility of the authors and does not necessarily represent the official views of HHS or any of its agencies.

Colorado State Innovation Model Evaluation

Outcome Evaluation Report

Final: July 31, 2019 (revised October 9, 2019)

Colorado SIM Office
Denver, CO 80203

The project described was supported by Funding Opportunity Number CMS-1G1-14-001 from the U.S. Department of Health & Human Services, Centers for Medicare & Medicaid Services. The contents provided are solely the responsibility of the authors and do not necessarily represent the official views of the U.S. Department of Health and Human Services, or any of its agencies. The research presented here was conducted by the awardee. Findings might or might not be consistent with or confirmed by the findings of the independent federal evaluation contractor.

TriWest
Health + Human Service Evaluation + Consulting
4450 Arapahoe Avenue,
Suite 100, Boulder, CO 80303

Suggested Citation

TriWest Group. (2019). *Colorado State Innovation Model (SIM) evaluation: Outcome evaluation report*. Denver, CO. Colorado State Innovation Model Office.

For more information, contact Tonya Aultman-Bettridge of TriWest Group.
taultman-bettridge@triwestgroup.net.

Table of Contents

Executive Summary

This report is one of two final evaluation reports we (TriWest Group, the Colorado SIM statewide evaluator) prepared, addressing the implementation of and outcomes achieved by the Colorado State Innovation Model (SIM), a federally funded, Governor's Office health reform initiative. It addresses outcomes resulting from the SIM initiative. A companion report contains findings from the SIM Final Process Evaluation Report.

In this Executive Summary, we present brief descriptions of the SIM initiative, our evaluation, key findings, and select recommendations. The remaining chapters, as listed below with chapter numbers in parentheses, provide detailed expansions of these items:

1. **Introduction and Report Organization (1)**. Outcomes methodology, attribution methodology, comparison group development, and analysis methodology.
2. **Practice Transformation: Integration Efforts (2)**. Data sources, movement on the integration continuum, behavioral integration, integration by practice characteristics, provider satisfaction and burnout, burnout and satisfaction by practice characteristics, workplace satisfaction and suggested areas of improvement, patient engagement.
3. **Practice Transformation: Access to Care (3)**. Improving access to care through practice transformation and public health drivers, measures and methods, comparison groups, access to physical healthcare in Colorado, stigma and access to behavioral health services, effectiveness of the Colorado healthcare system.
4. **Clinical Outcomes (4)**. Proxy clinical quality measures, changes in clinical quality, the relationship between practice transformation and changes in clinical quality. The effect of integration, staff satisfaction and burnout, and HIT improvements on CQM changes. Differences in CQMs by practice site characteristics.
5. **Payment Reform (5)**. Moving towards value-based payment, value-based payments effects on integration and quality of care, cost of integration and sustainability for SIM primary care practice sites.
6. **Cost and Utilization (6)**. Milliman cost and utilization metrics, transformation factors influencing cost and utilization outcomes, detailed analysis of individual measures, effect of practice transformation factors on cost and utilization outcomes, changes in cost and utilization by practice characteristics.
7. **Population Health (7)**. Population health findings.
8. **Conclusions (8)**. Results and findings.

Evaluation Overview

In 2015, Colorado was awarded up to $65 million in federal funding and support from the Center for Medicare and Medicaid Services (CMS) in the form of a cooperative agreement to implement SIM. The proposal and planning process included large-scale stakeholder engagement and contributions to ensure the statewide model would be comprehensive and sustainable. This model was developed to address the Quadruple Aim[1] to improve patient experience (both the quality of and satisfaction with care), improve population health, reduce/avoid healthcare costs, and improve the work life of providers. There were four key elements to this model:

- **Practice Transformation:** Over the course of implementation, SIM selected and provided support to three cohorts of primary care practices in their efforts to integrate behavioral and physical healthcare. Additionally, SIM selected four community mental health centers (CMHCs) to implement bi-directional integration efforts. These practice sites received practice transformation support, specifically focused on integrating physical care and behavioral healthcare. In addition, sites received value-based payments from health plans to support their work to provide patient-centered, team-based, integrated care. SIM helped practice sites create infrastructure and new processes to prepare them for greater success with value-based payment models. During their participation, sites also completed a variety of assessments designed to help guide quality improvement efforts. Further description of these sites can be found in the Practice Transformation chapter of this report.
- **Payment Reform.** SIM engaged seven public and private payers that worked together prior to SIM implementation and committed to work with SIM to support behavioral and physical healthcare integration. As a requirement of SIM participation, primary care practice sites received compensation through at least one alternative payment model. In some cases, this support was a new or enhanced payment model started as part of SIM participation. But the support often represented a continuation of value-based payment arrangements that were in place prior to a practice site applying for SIM participation. The Multi-Payer Collaborative had six health plan members at the end of SIM.
- **Health Information Technology (HIT).** The SIM strategy for improving HIT quality and utilization focused on support at the individual practice-site level and at the state level. At the individual practice-site level, efforts focused on optimization of electronic health records (EHRs) to support practice transformation efforts, quality improvement, and reporting of clinical quality measures. Wider-ranging statewide efforts included SIM

---

[1] SIM began with a focus on the “Triple Aim” of lower costs, better care, and better patient experience, then elected to add a focus on workforce during its initial planning year.

contributions to a statewide HIT roadmap, support for increased broadband access and telehealth capacity in the state, support for eConsult initiatives, development of an electronic clinical quality measure (eCQM) solution, development and support of a Clinical Health Information Technology Advisor (CHITA) workforce, and health information exchange (HIE) connectivity.

- **Population Health**. The SIM strategy for improving health at the state level included local support for community efforts to reduce stigma, promote coordination of health systems, and reduce barriers to accessing care. This strategy was developed through two major efforts. The first was funding for local public health agencies (LPHAs) and behavioral health transformation collaboratives (BHTCs) for projects targeting stigma reduction, community education, and coordination. The second was partial funding of the Regional Health Connector (RHC) program to connect the systems that keep people healthy, including primary care, public health, social services, and other community resources.

The SIM model recognized variances across practices, and it was designed to produce different types of outcomes for different groups. For example, patients experienced changes in access to care and utilization. Practice sites, similarly, experienced changes in levels of integration and practice operations. Furthermore, changes in approaches to value-based payments affected some payers. Because of these different components, a single methodological approach was insufficient for evaluating the many aspects of the various SIM efforts. Therefore, our evaluation utilized a mixed-methods approach that used qualitative data, quantitative data, and multiple analytical methods (e.g., descriptive, time-series, within and between-group comparisons).

Outcome Evaluation Findings

Complete findings are presented throughout the major portions of this evaluation. Below, we present abbreviated findings from each of the four primary drivers (i.e., Payment Reform [PR], Practice Transformation [PT], Population Health [PH] and Health Information Technology [HT]).

Practice Transformation (PT1). **To what extent did practice sites and bi-directional programs move along the continuum of integration?** High percentages of SIM cohort practice sites reported they were moving towards or had “completed” most if not all milestone activities associated with physical-behavioral health integration. All cohorts and the CMHCs demonstrated significant improvement on behavioral integration over the course of SIM participation.

Practice Transformation (PT3). **Was access to integrated care improved for 80% of Coloradans? (The original source of this question is the goal of 80% access to integrated care,**

supported by value-based payments, in coordinated community systems.) We were unable to answer this question because we did not have integrated care measures. However, we were able to answer questions about access to care measures in APCD and the Colorado Health Access Survey (CHAS). (See below.)

Practice Transformation (PT4). **Do patients attributed to SIM participating practices have better access to primary care relative to patients attributed to comparison practices? Better access to behavioral health care relative to patients attributed to comparison practices?** Analysis of the SIM and comparison observed rates in time series graphs lead us to conclude that the SIM intervention did not result in improvements to the adult and pediatric indicators. The adult composite indicators showed some improvements after the start of the implementation, but improvements for the comparison groups were equally large or larger. The pediatric indicators showed increased admissions for cohort 1 after the start of SIM, but the same increase was observed in the comparison group. For cohort 2, the pediatric indicators showed almost no change during the first year of the intervention.

Practice Transformation (PT6). **What specific transformation factors (level of integration, milestone targets, data quality, clinician and staff experience, etc.) most influence outcomes (CQMs, costs, population health measures)?** Based on our analyses, level of integration and clinician and staff experience did not have statistically significant relationships to outcomes. Select CQMs had statistically significant relationships with improved HIT data quality, and these results show some positive effects in the increase in clinical quality over the course of SIM implementation. While the lack of a comparison group means that we cannot definitively state that SIM alone contributed to these improvements, the results are encouraging.

Practice Transformation (PT8). **To what extent are primary care and behavioral health providers satisfied with the experience of integrating primary and behavioral health care? Report burden? Does satisfaction increase and burden decrease over time?** Overall, the average workplace satisfaction across all respondents was generally high, with large majorities (85%) of respondents agreeing or strongly agreeing that they are satisfied with the work they do at their practices. Most respondents reported either no burnout or occasional stress in the workplace, but 7% reported high levels of burnout. The remaining 17% reported that they were gradually burning out. In terms of burnout, results of chi-square tests indicated that time period effects are negligible and do not reach statistical significance for cohorts 2 and 3, with very little noticeable change taking place between assessment periods. However, statistically significant effects of time period did appear at cohort 1, which was longer than the other two cohorts.

Payment Reform (PR1). **To what extent were value-based payment models implemented? What were the barriers to this transition? Did implementation result in improved integration and quality of care?** Implementation was varied and not well understood among stakeholders and practice staff. Many practices and providers are frustrated over the perceived lack of results from their efforts and from a perceived disinterest among payers to support those efforts.

Payment Reform (PR3). What is the cost of integration transformation efforts to SIM practice sites and CMHCs? (Reporting will be separate for primary care and CMHC sites.) The original SIM evaluation planning called for assessing the degree to which the implementation of VBPs influenced improvements in practice-level outcomes (e.g., clinical quality and access to care) and for assessing the costs of implementing specific VBP models. However, available data were not sufficient to create a quantitative measure that could be used as a discrete variable to assess the degree to which specific APMs might be associated with improvements in outcomes.

Payment Reform (PR3.1). Is this cost sustainable through revenue generated by the APMs? While fee-for-service remains the most prevalent payment model, many practices reported that they will seek to continue offering integrated care, and SPLIT assessments indicated that practice sites from all cohorts have improved their data literacy and data usage to support these transformations. This data literacy is a foundational step in practices' abilities to present their work and care and qualify for reimbursements and compensation by payers.

Payment Reform (PR3.2). Are practice sites willing to absorb some unreimbursed costs as a result of increased satisfaction? Responses were mixed as to whether practices could estimate revenue from their integrated BHP to inform their budgeting. Even though revenue uncertainty might deter some practices from expanding integrated care efforts, some practice sites did indicate a belief in the value of integrated care, even if right now it cannot be fully supported by existing payment models. A subset of SIM practice sites already offering integrated care and who reported at closeout that they were planning to continue largely indicated that they were willing to offer it regardless of revenue.

Payment Reform (PR4). To what extent did the utilization of services and total cost of care (including out-of-pocket costs to consumers) differ over time for consumers attributed to SIM participating practices? Was this different compared to consumers in comparison practices? In comparison to patients attributed to non-SIM providers, we did not observe statistically significant declines in the total cost of care for any SIM cohort. In that sense, it does not appear that practices participating in SIM experienced savings in the total cost of care. A general conclusion for patients attributed to cohort 1 practice sites is that relative to a matched comparison group, the total cost of care did not change. Primary care costs increased slightly, but the costs of other services decreased. Out-of-pocket expenses were relatively stable during the period covered by the SIM program.

Payment Reform (PR5). What alternative payment models result in the best outcomes for different populations served (children, adults, type of payer, urban vs. rural vs. frontier areas)? We are unable to answer this question because we do not have access to the necessary data.

Health Information Technology (HT1). Are primary care practice sites and CMHCs using valid, reliable data (in the form of Clinical Quality Measures—CQMs and others) to drive change? Assessment data support other evidence that practice sites are continuing to incorporate increased comfort with data into their practice routines and operations. Cohorts reported considerable progress in data reporting and subsequently reviewing data quarterly to better guide their improvement efforts.

Health Information Technology (HT1.3). To what extent is data quality improving (data capture and CQM reporting)? Data quality considerably improved in aggregate, but changes were uneven across individual data elements and CQMs. Data related to physical conditions improved more than data related to behavioral conditions, and the subset of substance use data improved the least.

Health Information Technology (HT4). To what extent did practices increase or improve use of data to coordinate care? Milestone activity assessments support that practice sites are more comfortable incorporating data into their practice routines and operations. Practices are expanding use of registries, HIE and EHRs, and care compacts with outside providers.

Population Health (PH1). To what extent did the 14 behavioral and physical health related population health measures change over time? Did more resources and improved coordination/alignment result in improved population health measures?. As expected, there were no dramatic changes in state or HRS-level population health measures over the four years of SIM implementation. However, monitoring and reporting on these measures may provide a foundation for future population health policy makers to explore longer-term impacts of SIM.

Evaluation Summary, Lessons Learned, and Recommendations

The SIM initiative was an ambitious and comprehensive effort, touching virtually every aspect of healthcare in Colorado. Major activities took place in four areas: practice transformation, payment reform, health information technology, and population health. Each of these was supported by a series of stakeholder engagement workgroups, which were guided by a SIM Steering Committee, with an Advisory Board providing oversight.

Outcome evaluation questions, like those in the process evaluation, were developed jointly with stakeholders. Because so many diverse perspectives and priorities were involved, much evaluation time was dedicated to trying to find data sources for questions rather than just focusing on questions for which there were definitive data sources. This, combined with significant data source challenges, meant that not all evaluation questions could be answered.

Performing a large-scale program evaluation within this context lead to these five evaluation lessons learned:

Data availability for addressing evaluation questions was a challenge.

As mentioned in the process evaluation report, in the early planning stages of the initiative, stakeholders helped TriWest and SIM office staff to develop evaluation questions for which there were no readily available data sources. As the evaluator, we attempted to accommodate the diverse and broad interests of the stakeholders across the workgroups. This approach may have been too ambitious. Over the course of the evaluation, the SIM office experienced significant staff changes. With these changes came new considerations of the value and viability of some of the original evaluation questions.

As a result, the evaluation, in partnership between the SIM office and TriWest, adjusted emphases. Through this process, we reconsidered data sources and whether, despite our best intentions, some of the original questions were answerable without new data sources. In some cases, we attempted to adapt existing sources to respond to questions to honor the interests and direction of the initial evaluation stakeholders. Ultimately, we recognized that the evaluation could have been improved if we, as evaluators, had worked with the SIM office to better refine initial evaluation questions to the limitations of existing data while acknowledging the diverse and significant interests of stakeholders (many of whom remained throughout the full evaluation, often retaining their initial interests even as data limitations to address them became more apparent).

The UCDFM SPLIT set of assessments provided data helpful in examining the implementation of SIM drivers at the individual primary care practice and CMHC levels. For the Process Evaluation report, these data provided useful information on practice successes and challenges that were used during the implementation for mid-course corrections and for annual/final process evaluation reports.

Although it was very helpful to have access to data gathered specifically for the SIM initiative and related to SIM drivers, the SPLIT data also had significant limitations. One is related to the timing of reporting. Initially SPLIT measures were intended to be reported quarterly, and initial evaluation designs were predicated on having two years of quarterly observations for each practice site. Timing of these assessments changed over time and made evaluation adjustments necessary. Cohort 1 practice sites found that the frequency of assessments was a burden, and, therefore, assessments were done less frequently and sites had greater reporting flexibility (e.g., in choosing which CQMs to report). This flexibility was important for the implementation effort but resulted in the ability to report only one change over time for most practice sites: from the last calendar quarter of the initial participation year to the last calendar quarter of the second participation year. This data structure makes detection of change driven by SIM participation less reliable.

A second limitation to the SPLIT data is that they were practice-reported. TriWest did not have any independent method of assessing data standardization and quality, and based recommendations of CHITAs we interviewed, we limited our analysis to the two periods mentioned and did not use other quarters of reported data.

A third limitation is that key data elements, especially measures of integration, were only available for SIM-participating practices. This prevented comparisons to non-SIM practice sites for any evaluation question involving the degree of integration and its impact on outcomes.

Claims data offer both benefits and challenges.

The Colorado SIM project was unique because it involved payers beyond just Medicaid and Medicare. The All Payers Claims Database (APCD) provides an enormous opportunity to include data claims from private payers in order to include a broader patient population in evaluation efforts. The APCD was only source of data on true baseline (pre-SIM participation) outcomes and for outcomes for patients not attributed to SIM providers.

However, the complexity of these data leads to challenges in merging disparate sources. Both CIVHC, the manager of the APCD, and the state Medicaid office changed data processing vendors during the course of SIM initiative. Reconciliation of Medicaid claims in the APCD versus in Medicaid’s own data systems was not entirely completed, leading to concerns about Medicaid APCD data quality for this important payer.

In addition, claims data have limitations in their utility for measuring outcomes because they are tied to billing for services rather than provision of services. Some key services for this evaluation, such as depression screening, are not normally billed and therefore included in the APCD. This and other limitations are discussed extensively throughout this report.

Small grants offer concrete examples of progress.

Throughout this report and the Final SIM Process Evaluation report, examples provided from practice sites receiving small grants provide concrete examples of practice transformation activities that furthered integration and improved patient care. These examples could be very helpful for other practices interested in greater physical-behavioral healthcare integration. We were unable to directly measure outcomes from these grants. Doing so would be a useful next evaluation step.

With a short follow-up period to complete the evaluation, we are more likely to identify process improvements than large changes in health, utilization, or cost outcomes.

All cohorts saw gains in the level of integration and overall progress in reaching the milestones designed to indicate practice improvement. In addition, practice sites saw some significant increases in indicators of quality integrated practice, such as increased depression screening and better diabetes control. Further, as discussed in the process evaluation report, many of the foundations laid by SIM, in the efforts made to facilitate better communication and collaboration around VPBs, the population health Call-to-Action document, and many others, are likely to serve as conduits for ongoing healthcare reform in the state.

We had much less success in identifying resulting changes in cost and utilization outcomes of patients. This was driven in part by the reporting lag in the APCD and potentially by the need for practices to further refine integration procedures through experience.

The timing of the assessments for the final process evaluation was also contributor to the limited number of outcomes. We recommend for any similar future efforts that the evaluation timeframe have at least a six-month lag between end of all project activities and data gathering and the final evaluation report completion.

1 Introduction and Report Organization

Introduction

This report is one of two final evaluation reports prepared by TriWest Group (TriWest), the Colorado SIM Statewide evaluator, addressing the implementation of and outcomes achieved by the Colorado State Innovation Model (SIM), a federally funded, Governor's Office health reform initiative. This report addresses the **outcomes** of the initiative, focusing on a summary of the observed and potential future impacts of SIM across the state. A companion report contains TriWest findings from the SIM process evaluation.[2]

In 2015, Colorado was awarded up to $65 million in federal funding and support from the Center for Medicare and Medicaid Services (CMS) in the form of a cooperative agreement to implement SIM. The proposal and planning process included large-scale stakeholder engagement and contributions to ensure the statewide model would be comprehensive and sustainable. This model was developed to address the Triple Aim and was expanded to the "Quadruple Aim" to improve patient experience (both the quality of and satisfaction with care), improve population health, reduce/avoid healthcare costs, and improve the work life of providers.

Colorado SIM's overarching goal was to improve the health of Coloradans by increasing access to integrated primary care and behavioral health services in coordinated community systems, with value-based payment structures, for 80% of state residents by 2019.

To achieve its goals, the SIM office implemented multiple strategies, including the following:

- Help 319[3] practice sites integrate behavioral and physical health in primary care settings and test alternative payment models,
- Assist four community mental health centers (CMHCs) in their bi-directional efforts to integrate physical and behavioral healthcare,
- Facilitate communication between providers and payers and support practice sites as they navigated multiple aspects of value-based payments,
- Support population health improvement efforts through funding projects within local public health agencies (LHPAs) and two Behavioral Health Transformation Collaboratives (BHTCs),

---

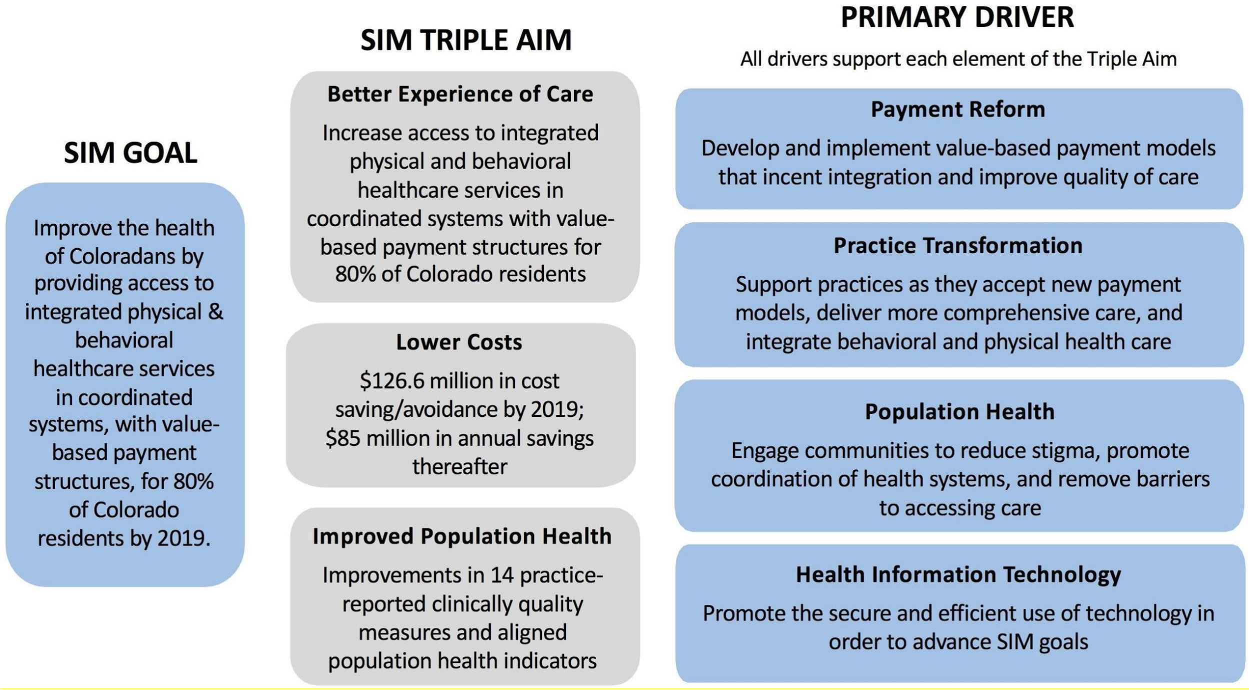
[2] The Final SIM Process Evaluation Report can be found at https://www.colorado.gov/pacific/healthinnovation/sim-data-hub.

[3] While more sites initially completed participation agreements, 319 practice sites completed SIM.

- Improve community coordination and support population health efforts by providing partial funding for a Regional Health Connector program,[4]
- Facilitate consumer engagement and workforce development needed to support overall transformation efforts through structured workgroups and targeted small projects,
- Participate in statewide HIT infrastructure improvement efforts in close collaboration with the Governor's Office of eHealth Innovation (OeHI).

These strategies were organized into four primary drivers, as illustrated in Figure 1.

Figure 1. SIM Updated Driver Diagram: Year 4 Operational Plan



SIM GOAL

Improve the health of Coloradans by providing access to integrated physical & behavioral healthcare services in coordinated systems, with value-based payment structures, for 80% of Colorado residents by 2019.

SIM TRIPLE AIM

Better Experience of Care

Increase access to integrated physical and behavioral healthcare services in coordinated systems with value-based payment structures for 80% of Colorado residents

Lower Costs

$126.6 million in cost saving/avoidance by 2019; $85 million in annual savings thereafter

Improved Population Health

Improvements in 14 practice-reported clinically quality measures and aligned population health indicators

PRIMARY DRIVER

All drivers support each element of the Triple Aim

Payment Reform

Develop and implement value-based payment models that incent integration and improve quality of care

Practice Transformation

Support practices as they accept new payment models, deliver more comprehensive care, and integrate behavioral and physical health care

Population Health

Engage communities to reduce stigma, promote coordination of health systems, and remove barriers to accessing care

Health Information Technology

Promote the secure and efficient use of technology in order to advance SIM goals

This summary of SIM outcomes is organized based on specific areas of outcomes, which align with the four primary drivers. We present the following list in the order in which chapters appear in this report, starting with the more narrowly focused efforts (e.g., those efforts in individual SIM practice sites) and moving to broader, state-focused goals.

---

[4] Additional funding was provided by the EvidenceNOW Southwest program: http://www.practiceinnovationco.org/ensw/

- **Practice Transformation: Integration Efforts:** Over the course of implementation, SIM selected and provided support to 100 physical health practice sites[5] in cohort 1 (February 2016–April 2018), 156 sites in cohort 2[6] (August 2017–June 2019), and 88 practice sites in cohort 3[7] (June 2018–June 2019) in their efforts to integrate behavioral and physical healthcare. Additionally, SIM selected four community mental health centers (CMHCs) to implement bi-directional integration efforts. The Practice Transformation: Integration Efforts chapter of this report describes the degree to which practice sites moved along the integration continuum. Specifically, the chapter presents outcomes on overall integration progress, staff and clinician experiences in SIM as well as efforts to increase patient engagement in SIM practice sites.

- **Practice Transformation: Access to Care:** Access to care is part of both a better patient experience of care and is also linked to overall improvements in population health. The SIM driver diagram mapped a path whereby SIM activities supported "access to integrated physical and behavioral healthcare services…" and "remove[d] barriers to accessing care." These barriers to care can include cost, stigma, and workforce issues, including provider burnout. The Agency for Healthcare Research and Quality (AHRQ) defines "access to care" as "the timely use of personal health services to achieve the best health outcomes."[8] In explaining this definition, AHRQ lists four components:

    - **Coverage**: Facilitates entry into the healthcare system. Uninsured people are less likely to receive medical care of any kind and more likely to have poor health status.
    - **Services**: Having a usual source of care is associated with adults receiving recommended screening and prevention services.
    - **Timeliness**: ability to provide healthcare when the need is recognized.
    - **Workforce**: capable, qualified, culturally competent providers.

---

[5] Of the 100 sites that initially signed PPAs or MOUs, eight withdrew over the course of SIM participation. The final count of cohort 1 practices was 92 sites.

[6] Of the 156 sites that initially signed PPAs or MOUs, 12 withdrew over the course of SIM participation. The final count of cohort 2 practices was 144 sites.

[7] Of the 88 sites that initially signed PPAs or MOUs, five withdrew over the course of SIM participation. The final count of cohort 3 practice sites was 83.

[8] https://www.ahrq.gov/research/findings/nhqrdr/chartbooks/access/elements.html

- **Clinical Outcomes**. An additional important aspect of practice transformation was practice site reporting of clinical quality measures (CQMs). CQMs offered one way to measure the degree to which practice transformation efforts in SIM-participating sites lead to better care for patients. SIM-participating primary care practice sites and CMHCs reported these measures quarterly in order to track processes and outcomes for their populations. These measures focused primarily on increased screening of patients for important physical and behavioral health conditions. They also included two measures related to positive health outcomes.
- **Payment Reform.** SIM engaged seven public and private payers that were participants of the Multi-Payer Collaborative that worked together prior to SIM implementation and that committed to work with SIM to support behavioral and physical healthcare integration. As a requirement of SIM participation, the primary care practice sites received compensation through at least one alternative payment model. In some cases, this support was a new or enhanced payment model started as part of SIM participation. But the support often represented a continuation of value-based payment arrangements that were in place prior to a practice site applying for SIM participation. The Multi-Payer Collaborative had six health plan participants at the end of SIM.
- **Cost and Utilization.** This chapter examines changes in healthcare costs and the types of healthcare utilization associated with higher spending. The SIM model test put forth in the Colorado SIM proposal sought to reduce these costs in both the short and long term and demonstrate the association between integration, better care, and lower or avoided costs.
- **Health Information Technology (HIT).** The SIM strategy for improving HIT quality and utilization focused on support at the individual practice site level and at the state level. At the individual practice site level, efforts focused on optimization of electronic health records (EHRs) to support practice transformation efforts, quality improvement, and reporting of CQMs. Wider-ranging statewide efforts included SIM contributions to a statewide HIT roadmap, support for increased broadband access and telehealth capacity in the state, support for e-Consult initiatives, development of an eCQM solution, development and support of a Clinical Health Information Technology Advisor (CHITA) workforce, and health information exchange (HIE) connectivity.
- **Population Health**. The SIM strategy for improving health at the state level included local support for community efforts to reduce stigma, promote coordination of health systems, and reduce barriers to accessing care. This strategy was developed through two major efforts. The first was funding for local public health agencies (LPHAs) and behavioral health transformation collaboratives (BHTCs) for projects

targeting stigma reduction, community education, and coordination. The second was partial funding (along with EvidenceNOW Southwest) of the Regional Health Connector (RHC) program to connect the systems that keep people healthy, including primary care, public health, social services, and other community resources.

Organization of This Report

This report contains a chapter for each one of the outcome areas in the bulleted list above. In addition, the Methodology chapter of this document details the data sources and overall methodologies used to conduct the analysis of SIM outcomes. Each chapter has additional methodological details unique to the element that it addresses, to provide some context for readers who may read some chapters independently, rather than this document in its entirety.

The last chapter of this report includes a summary of findings and notes recommendations that may be applicable to similar initiatives going forward in Colorado.

Outcomes Methodology

Primary Data Sources for Outcomes Analysis

TriWest used multiple data sources to measure the SIM outcomes presented in this report. This section lists our most prominent and used data sources in the general order they appear in the report and describes the nature of the data as well as important caveats and limitations associated with each source. We discuss other data sources used to provide context and information for specific chapters within each individual chapter. This approach allows chapters to be read independently for those who do not wish to read the entire report.

Practice Monitor – Measure of Integration and Other Building Blocks

The Medical Home Practice Monitor (Practice Monitor) is a practice self-assessment of the current level of implementation of core aspects of advanced primary care, also called the "building blocks of primary care." The assessment is completed by practice site staff with assistance from Practice Facilitators (PF).

Practice sites rate their implementation of the core components on a 5-point scale:

- 1-not implemented or routine across the entire practice
- 2-occurs sometimes
- 3-occurs partially
- 4-occurs inconsistently
- 5-completely implemented and routine across the entire practice site

We calculate practice-site-level averages of domains. These are aggregated into cohort-level averages and are further broken out by practice site characteristics (i.e., type, location, size, volume of underserved patients) for additional analyses. We consider practice site averages, cohort averages, and progress over time with data from other SPLIT assessments to analyze characteristics of integration and analyze correlation. We also consider averages and change over time with small grant data to evaluate the effect of competitive funding on integration.

Practice Site Characteristics

We have taken practice site characteristics from the information provided in the SIM participation application. In some cases, the University of Colorado Department of Family Medicine (UCDFM) may have corrected some of this information based on information learned through working with practice sites. The practice site characteristics used for the outcome analyses are presented below (Table 1).

Table 1. Colorado SIM Practice Site Subgroupings

Colorado SIM Practice Site Subgroupings for Data Analysis		
Subgroup Category	**Groupings**	**Definitions**
Practice Type		
	Adult Primary Care	Entirely serves adult patients
	Pediatric Primary Care	Entirely serves patients < 18 years
	Mixed Primary Care	Serves both adults and children
Urban/Rural Practice		
	Urban	Urban areas defined by RUCA codes 1–3
	Rural	Rural areas defined by RUCA codes 4–10
Practice Size		
	Small	0–5,999 annual patient visits
	Medium	6,000–14,999 annual patient visits
	Large	15,000+ annual patient visits
Proportion of Patients Underserved		Defined by percentage of patients insured by Medicaid or uninsured
	Low	0%–10% of all patients
	Medium	11%–30% of all patients
	High	31%–50% of all patients
	Very High	>50% of all patients

Colorado Health Access Survey

The Colorado Health Access Survey (CHAS) is a self-administered questionnaire administered to a random sample of individuals every other year to measure the health of Coloradans, their interactions with healthcare systems, and the impact of major healthcare policy developments. More than 10,000 households are surveyed anonymously, online, and by phone to gather a representative sample of Colorado's regions and population. The Colorado Health Institute (CHI) administers the CHAS through survey vendor SSRS. The survey is funded by The Colorado Trust and the Colorado Health Foundation. SIM sponsored additional questions regarding behavioral healthcare included on the 2017 and 2019 CHAS.[9]

Clinical Outcome Measures

Clinical quality measures (CQMs) are metrics collected internally by primary care practice sites and community mental health centers (CMHCs) using electronic health records or other internal

---

[9] For more information, see https://www.coloradohealthinstitute.org/research/colorado-health-access-survey.

data systems. Each measure comprises a numerator and a denominator and is expressed as a “rate per 100” (percentage). The numerator represents the total count of individual patients receiving a specific procedure (e.g., depression screening) or having a specific health outcome (e.g., diabetes patients with an A1c level <9%). The denominator represents the total population eligible for either the procedure or the outcome (e.g., all adult patients, all patients with diabetes).

Practice sites reported each numerator or denominator on a quarterly basis. Practice sites were required to report on a specific set of measures using standardized definitions. In some cases, sites reported measures using slightly different measure definitions. These were noted when data were submitted.

Data were collected in the SPLIT reporting tool and submitted to the SIM office. The SIM office reviewed and cleaned files then sent those files to TriWest via the Egnyte system.

All Payer Claims Database

The All Payer Claims Database (APCD) is managed by the Center for Improving Value in Health Care (CIVHC). These data comprise patient-level encounter and claims records detailing services received, service costs, and patient and provider characteristics. The data cover all patients of submitting payers in both SIM and non-SIM practice sites. We received and worked with multiple extracts of the APCD. Submitting payers include Medicaid, Medicare, and some private providers throughout the state. But not all payers submit data to the APCD and not all claims are submitted. Some self-insured (e.g., Administrative Only, ASO) plans do submit claims on a voluntary basis, but not all do. In other words, although we provide results based on analysis of APCD data, these data do not comprise the full population or activity of Coloradans.

Additionally, a separate file containing Behavioral Health Organization (BHO) Medicaid encounter data is submitted separately.

Data from the APCD contributes to the SIM Evaluation in four ways: (1) attribution of patients to participating SIM primary care practice sites and CMHCs and comparison groups, (2) cost and utilization analyses of claims data, (3) calculations of claim- based “access to care” measures, and (4) calculations of claims-based “proxy” clinical quality measures.

Patient Attribution (Primary Use #1)

CIVHC’s attribution of patients to practice sites. Attribution covers all patients and providers in the APCD data, including both SIM and non-SIM-site providers. CIVHC and stakeholders have jointly developed this attribution methodology, which provides us the specific patients attributed to each primary care provider (by National Provider ID). Attribution is annual, with separate attribution runs for each calendar year 2015–2018.

In the case of primary care providers, attribution is based on the provider receiving the plurality of the outpatient primary care claims for a specific patient. In the case of CMHCs, the SIM office facilitates the transfer of panel lists from the CMHCs to CIVHC. CIVHC then converts identified patients into de-identified patient composite IDs grouped by CMHC. CMHCs have worked with one another and with technical assistance providers to develop consistent attributions. Please see the Attribution Methodology Section below for additional information.

Cost and Utilization (Primary Use #2)

Cost and utilization analysis were conducted primarily by Milliman, SIM's actuarial partner. Milliman calculated per patient per month (PMPM) costs attributable to patients at SIM practice sites. In addition, Milliman used APCD data to calculate utilization of certain key services (e.g., emergency department visits) and to attribute rates for these utilization variables to SIM practice sites. The PMPM costs are key components to Milliman's Return on Investment (ROI) analysis. TriWest also requested that Milliman calculate PMPM costs and utilization for a matched comparison of non-SIM-attributed patients in addition to the calculations made for SIM practice sites. These data were used to assess **cost and utilization** outcomes presented in this report.

Access to Care Measures (Primary Use #3)

We calculated claims-based access to care indicators. These indicators reflect utilization of certain types of inpatient hospitalizations related to insufficient access to quality integrated primary care, such as hospitalization for complications arising from poorly managed diabetes. These indicators include four specific Agency for Health Care Research and Quality (AHRQ) access to care measures that we calculate for patients attributed to SIM-participating providers and a matched comparison group of patients attributed to providers not participating in SIM. (See http://www.qualityindicators.ahrq.gov/Modules/PQI_TechSpec.aspx for more details.)

Access to Care Measures (Primary Use #4)

CIVHC uses claims data to calculate proxy versions of clinical quality measures. Measures were then reported to the SIM office and shared with TriWest. We describe these measures in detail in the Clinical Outcomes chapter of this report.

Attribution Methodology

To measure patient- and practice-level outcomes using APCD claims and encounter data, we needed to determine which patients received came from each SIM primary care and CMHC practice site. For purposes of developing a comparison group, we also needed to determine which patients *did not* receive care from SIM practice sites.[10]

Before TriWest was hired as the evaluator, the SIM office worked with CIVHC, Milliman, and the University of Colorado Department of Family Medicine (UCDFM) to develop a process to attribute patients found in the APCD to practice sites. This alignment was based on a mix of information available in the APCD and provider information given by each practice site. Although we were not involved in the initial development and implementation of this attribution methodology, we used the resulting attributions in our analysis, and we have summarized the process below.

However, before providing methodological details, we hope to clarify what information was available from each data source. The APCD, for example, contains claims and encounter, enrollment, and pharmacy data for the majority of insured Coloradans. It also has extensive data related to billing providers. These data were received from payers (commercial, Medicare, and Medicaid) and processed by CIVHC into standardized data sets. The claims data include information on patients receiving medical services, what services they received, the costs of those services, and the National Provider Identifier (NPI) of the service provider. Several NPIs may be on any individual claim, including the NPI of a specific person providing medical services (providers) or the NPI of the site billing for the service.

Most claims lack reliable information on the exact place of service (e.g., a specific clinic address). Most claims also do not link the provider NPI to a specific organization in a manner that we are able to use. The APCD is therefore very useful in determining which patients received care from which providers but not useful in linking patients directly to SIM or non-SIM practice sites.

As we describe below, UCDFM requested the NPIs of providers working at each clinic from the applying SIM practice sites in order to address this limitation. As such, UCDFM was our source for which NPIs associated with SIM practice sites, whereas CIVHC's APCD was the source for which patients received care from the NPIs of providers.

In the following section, we discuss each half of this attribution process: (1) the assignment of patients to NPIs and (2) the assignment of NPIs to specific SIM practice sites. Note that this

---

[10] Although some payers worked with practices sites to develop patient panels, we do not have access to this information, and cannot use it to align patients with practice sites.

discussion applies only to the primary care practice sites. The attribution process for the CMHCs is entirely different, and we summarize that process separately at the end of this section.

Using the APCD to Attribute Primary Care Patients to NPIs

Patients receive care from many different healthcare providers, and the claims found in the APCD reflect these complex relationships. For purposes of evaluation, we needed to simplify the many possible relationships and to identify a single primary care provider for each primary care patient in each year.

The method developed by the SIM office and its evaluation partners was to first identify all individual patients in the APCD who qualified for attribution. In addition to meeting the technical requirements related to CIVHC uniquely identifying each patient, the patient must have also had insurance eligibility for at least one month during the activity period. Attribution is for an annual period (e.g., calendar year 2015), and eligibility and claims used to make attribution for that period are the same calendar year (e.g., 2015) and the previous calendar year (e.g., 2014). The activity period includes the year prior to the year of attribution because some primary care patients did not receive primary care services every year.

Once eligible patients are identified, all primary care related claims for each patient are identified. A claim is categorized as primary care related based on the following criteria:

- The claim type was “professional.”
- The service provider must have had a valid NPI.
- The service provider must have had a taxonomy code corresponding to a primary care provider (e.g., claims involving services provided by a psychologist were excluded in the attribution methodology).

Using these claims, CIVHC attributed each patient to the provider NPI with the largest number of claims using a series of “runs.”

Run 1: Only claims from the attribution year (e.g., 2015) were used, and provider types that were more tangentially related to primary care (e.g., women’s health providers, education-related health clinics) were excluded. Any patient who had claims meeting these criteria was attributed to the corresponding NPI and removed from further runs.
Run 2: Same as Run 1 but for the previous 12-month period (e.g., 2014).
Run 3: Using claims from the year of attribution (e.g., 2015), claims from the previously excluded provider types (e.g., women’s health, education-related health clinic providers) were used.
Run 4: Same as Run 3 but for the previous 12-month period.

For any remaining unattributed patients, Runs 5 and 6 used claims in which the NPI was for an organization (practice) instead of a person (provider). These NPIs include Federally Qualified Health Centers (FQHCs), rural health clinics, community health clinics, Public Health clinics (federal, state, and local), primary care clinics, and student health clinics. Run 5 used claims from the attribution year; Run 6 used claims from the previous year.

The resulting attribution assigned every qualifying patient in the APCD to an NPI, either of a single provider in runs 1–4, or to a practice in runs 5–6. This attribution covered a single year, and we requested a separate attribution by year for 2015, 2016, and 2017 for every patient in the APCD meeting the qualification requirements. For additional details on the methodology used by CIVHC to attribute patients to primary care NPIs, please see the model programming documentation produced by CIVHC.

Assigning Provider and Practice NPIs to SIM Primary Care Practice Sites

Attribution of patients to NPIs was only half of what was needed to attribute APCD patients to SIM primary care practice sites. The remainder of the attribution process required us to group provider and practice NPIs with specific SIM sites. TriWest and the SIM office attempted two different processes to accomplish this. First, we attempted to use a provider directory developed by CDPHE to link NPIs to the physical address of SIM primary care practice sites. As compared with our second method described below, we found poor alignment of NPIs to SIM sites and therefore followed up with three SIM sites—identified by UCDFM as good case studies for comparing provider and patient attributions—for a detailed review of each site's NPIs identified through the provider directory. The three sites confirmed that the selected NPIs were often not currently associated with the site. In some cases, the provider had practiced at the site many years ago. In other cases, the provider was part of the site's organization but practicing at another location. After discovering that the SIM attributions did not match well with the CDPHE provider directory, we abandoned its use for aligning provider NPIs with SIM or potential comparison primary care practice sites.

The second method, which was used for the final data analysis, was to use NPIs provided directly by the SIM primary care sites. UCDFM was the lead organizer of this process. On the initial application to participate in SIM, practice sites provider both provider and practice NPIs. UCDFM repeatedly surveyed practice sites, requesting updates to the NPI rosters. UCDFM faced several challenges in assembling a comprehensive list of NPIs for each practice site. These included:

- Providers moved between practices during each site's participation in SIM. This created additions and deletions of NPIs between UCDFM's surveys.

- The same provider sometimes worked at multiple clinics. In some cases, the same provider worked at more than one SIM clinic; in other cases, the provider worked at both SIM and non-SIM clinics.
- Practice NPIs were in some cases claimed by multiple SIM practice sites. As with the provide NPIs, a practice site NPI was sometimes claimed by two SIM sites in the same cohort. In other cases, the NPI was claimed by practice sites in differing cohorts.
- Some practice NPIs were also known to include both SIM and non-SIM sites.

Another complication in assigning NPIs to unique SIM practice sites related to the differing types of evaluation, each with unique needs. In the case of TriWest's outcome analysis, we attempted to use an intent-to-treat design whenever data allowed. In other words, we preferred to retain all practice sites that started SIM participation, including those that eventually withdrew. Consistent with this, we selected a single group of patients attributed to a practice site at baseline (2015 for cohort 1, 2016 for cohort 2) and follow their outcomes over the course of the evaluation.

For other purposes, such as Milliman's reporting of practice site cost and utilization to the participating practices, an intent to treat design was not relevant. Practices were more interested in cost and utilization for patients served in each year of SIM participation, and for this purpose holding the patient population constant was not optimal. There was also not any need to report to practice sites that withdrew from SIM.

In order to best address these complexities, CIVHC, UCFM, and TriWest performed the following procedures in assigning NPIs to SIM practice sites. First, CIVHC attempted to resolve the issue that some practice NPIs might be used for both SIM and non-SIM primary care sits. CIVHC's attribution methodology, described previously, already prioritized providers over practices in assignment of a patient. Additionally, CIVHC used information found in The NPI Registry Public Search[11] and practice web pages to exclude practice NPIs that covered combinations of SIM and non-SIM sites. Patients attributed to such practice NPIs were marked as unattributed.

In order to link provider NPIs to SIM practice sites, separate methods were developed for TriWest's and Milliman's use. Because our (TriWest's) focus is on program-wide evaluation, it is important to not double count patients, and we therefore assigned each NPIs to a single practice site. Although it is important that each NPI and its attributed patient is included only once, it is less important which site the NPI is attributed to since, in general, our analysis compares all patients attributed to all SIM practice sites to comparison pools of patients

---

[11] https://npiregistry.cms.hhs.gov/

attributed to non-SIM NPIs. Also, we used an intent-to-treat design and did not drop practice sites from our analysis even if they withdrew from SIM.

Milliman's work on cost and utilization had several aspects, one of which was reporting the cost and utilization of all patients served to each practice separately. Double counting of patients between practice sites was less important for this work, and the analysis also was only for practice sites that did not withdraw from SIM.

For our work, UCDFM developed a protocol to assign *provider* NPIs to practice sites based on the frequency in which the provider NPI was included in the survey's sent to SIM sites. Sites claiming a provider NPI most frequently were assigned the NPI. SIM sites that remained in the SIM program were also prioritized over sites that withdrew.

In order to uniquely link *practice* NPIs to a single SIM practice site, UCDFM developed a protocol, implemented by TriWest, to make a unique assignment. For multiple practice sites claiming a single practice NPI, we prioritized assignment first by cohort, with the ranking of cohort 1, then 2, then 3. In cases in which multiple practice sites within a single cohort claimed a practice NPI, we assigned the NPI to the practice site with the lowest SPLIT ID, reflecting earlier application to the SIM program.

To summarize TriWest's final attribution, we used the annual attribution of APCD patients to NPIs generated by CIVHC. For cohort 1, we used the 2015 attribution year, cohort 1's baseline year. For cohort 2 we used the 2016 baseline year. We used UCDFM's assignment of provider NPIs to SIM sites based on the most frequent claim of the NPI in SIM site surveys. We used UCDFM's assignment of practice NPIs to SIM sites based on a prioritization of cohort 1 over cohort 2 and of cohort 2 over cohort 3. In cases in which a practice NPI was claimed by two sites within a cohort, we assigned the NPI to the site with the lowest SPLIT ID.

For purposes of Milliman's practice site cost and utilization reporting, NPIs were assigned to practice sites using a somewhat different methodology. For provider NPIs, only those included in the final survey were included on a practice site's individual report. Milliman's reporting therefore reflects the cost and utilization of patients attributed at the end of SIM participation. The provider and practice NPIs were also not uniquely assigned to a single practice site. This approach resulted in double counting of some patients when comparing between sites. For more aggregate reporting, Milliman deduplicated all NPIs. Because Milliman's analysis excluded practice sites that withdrew from SIM, and only used NPIs reported in the last of the site surveys, Milliman uses a smaller number of attributed patients than TriWest uses in its analysis.

Many of these choices are arbitrary responses that prevent double counting of patients. For most of the evaluation analysis performed by TriWest, the arbitrary assignment of an NPI to one SIM practice over another is irrelevant since analysis is between all patients served by SIM practice sites and comparison patients attributed to non-SIM NPIs. In cases in which an NPI is assigned to cohort 1 practice site over a cohort 2 site, TriWest's evaluation results could be affected since we do not pool the two cohorts. Given the large number of SIM-attributed patients who are attributed without these complications, we view this issue as unlikely to change any conclusions based on the use of APCD data.

Community Mental Health Center (CMHC) Attribution Methodology

The CMHCs specialize in providing care for behavioral health conditions, and although their participation in SIM helped them provide integrated care, the APCD attribution methodology is unlikely to correctly identify their patients. One reason for this is that the CMHCs' provision of primary care services generally started with participation in SIM. So, attribution based on primary care claims in the baseline period will not correctly identify the patients treated as part of SIM.

The SIM office hired another contractor to work with the four CMHCs to determine an attribution methodology. Because the nature of the intervention varies between the four centers, a uniform method was not selected. Each practice site identified the patients receiving bi-directional care, and provided a list of Medicaid identification numbers, which CIVHC converted to the patient composite IDs used in the APCD. We received two separate lists, one in December 2018 and another in June 2019. Because information was not included on the dates of service of CMHC-attributed patients, we are using the June 2019 list to determine the CMHC treatment group.

Comparison Group Development

Primary Care Patients

In order to estimate the impact of SIM on the cost and utilization and access to care of SIM-attributed patients, we developed a comparison group of APCD patients who were not attributed to NPIs associated with SIM. This effort included cohorts 1 and 2 since data is not available for outcomes for cohort 3, due to the start later in SIM and the short amount of SIM involvement. The theoretical underpinnings of comparison group analysis is that while we might be able to measure changing outcomes for patients attributed to SIM participating practice sites, we do not know whether the observed changes were caused by involvement with SIM practice sites or factors that would have caused changes regardless of the primary care practice site these patients received care from. For example, the cost of medical services generally rises each year. If the cost of services provided to SIM-attributed patients rose by 1% over the two years of participation in SIM, we would need to consider whether this increase was more or less than what would have occurred absent SIM participation. If we had determined, for example, that without SIM, costs for these patients would have increased by

5%, then SIM involvement resulted in a net cost savings—even though costs were higher after SIM. To estimate changes in cost and utilization arising from SIM, we need to estimate the counterfactual situation of SIM not having occurred.

Milliman's return on investment (ROI) analysis made the counterfactual estimate using projected trends (see the Milliman ROI report for details). At TriWest, we estimated this counterfactual situation by examining outcomes of a set of patients who were as similar as possible to patients attributed to SIM practice sites and subject to the same Colorado-specific factors that cause change in outcomes of SIM patients. Our process of developing this comparison group is as follows.

We began with the multiyear attribution developed by CIVHC. We excluded all unattributed patients in each year. Using the full list of provider and practice NPIs created by UCDFM, we excluded any patient in 2015, 2016, or 2017 who was attributed to any SIM primary care practice site or provider NPI or who did not have eligibility for at least six months during each of these years. We then excluded any patient attributed to one of the four SIM CMHCs. This sequence left us with a very large pool of APCD patients with separate attribution in 2015, 2016, or 2017.

This pool of APCD patients were all attributed to Colorado primary care providers or practice NPIs but potentially had different baseline (pre-SIM) characteristics, making them inappropriate to use for measuring the effect of SIM involvement. We therefore needed to select from them patients with the same baseline characteristics as the SIM practice attributed patients through a matching process. The SIM evaluation committee suggested we use the following patient characteristics in performing matching:

- Gender
- Age
- Residence by Health Statistics Region (HSR)
- Taxonomy code of the primary care provider or practice
- Payer type
- Baseline Milliman Advanced Risk Adjuster (MARA) score

Each of these patient characteristics is an important predictor of the amount of utilization and cost of medical services. Therefore, the comparison group needed as similar levels as possible to the treatment group in baseline values. An examination of claims volume showed substantial variation by patient age and gender. Furthermore, there are 21 HSRs in Colorado, and ability to access medical services—and many other factors determining utilization—vary by HSR.

Regarding the many taxonomy code categories associated with primary care providers, and based on the evaluation committee's advice, we collapsed these categories into Pediatric, Family, Adult, Other, and Clinic/Center.

Payer type also strongly influenced cost and utilization. We initially used three payer types (commercial, Medicaid, Medicare), but at the request of the SIM office, we reran matching after dividing Medicare into Medicare Fee-for-Service and Medicare Advantage. Patients often had claims from multiple payers, and we assigned a single payer type to each patient using this prioritized ranking:

- Medicare Advantage
- Medicare Fee-for-Service
- Commercial
- Medicaid

Patients with any Medicare Advantage claims in the baseline year were assigned a payer type of Medicare Advantage, even if that patient had more claims in another payer category. In the same way, we selected Medicare fee-for-service over commercial or Medicaid.

The MARA risk score proved to be one of the most influential variables on cost and utilization. Milliman calculated, for each attributed patient in the APCD, a score that was helpful in predicting current and future medical costs.[12] To oversimplify a complex actuarial product, Milliman used large claims data sets to estimate the correlation of groups of patient diagnostic and procedure codes on utilization and costs. This predictive model was reduced to a single value closely tied to morbidity. By matching based on each patient's risk score, we formed a comparison group that had similar medical needs.

Gender, HSR, taxonomy code of the primary care provider, and payer type are all categorical variables, and the evaluation committee recommended we match exactly on these characteristics. Age and MARA risk scores are continuous variables, preventing exact matching. We initially attempted to use propensity score matching to select a comparison group with similar age and risks scores from the larger pool of non-SIM patients. This technique is commonly used for program evaluation, and we have used it in other statewide evaluations, usually as a requirement of the evaluation contract. In this case we were not satisfied with the quality of the matched comparison pool. We initially selected a single comparison patient for each SIM-attributed patient (one-to-one matching). However, examination of the matched pairs revealed many poorly matched pairs. This problem is not unique to our project, and

---

[12] For more information on risk scores, see http://www.millimanriskadjustment.com/risk-scores/

recent academic articles have emphasized the same limitations we encountered with the technique.[13]

In reaction to problems encountered with propensity score matching, we used Coarsened Exact Matching (CEM), a newer technique designed specifically to result in directly comparable matched groups. With CEM, all continuous variables are divided into categories based on the criteria that all values within a category are sufficiently similar to allow a comparison patient from that category to be used with any treatment patient in that category. For example, age is a determining factor for volume of medical claims: typically a large number of claims at birth, followed by a rapidly declining volume in a child's early life, then a rapidly increasing volume as a person enters middle age. CEM stratifies age into different categories so that a person at an age with very low levels of medical utilization (such as age 25) is in a different category as a person with very high levels of utilization (such as age 80). But people with very similar ages, such as age 25 and age 26, are in the same age category, since there are not systematic differences in medical use between these two ages.

For patient age, we created 12 categories with the following cut points: 0, 2, 15, 17, 22, 25, 42, 52, 63, 66, 74, and 117. We based these cut points by calculating volumes of claims in the APCD by year of age and adding a cut point at each noticeable change in claims volume. We performed a similar analysis on the MARA risk scores and selected 14 categories with the following cut points: 0, 0.25, 0.5, 0.75, 1, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100. Any patients with age or MARA risk scores outside of these categories were dropped from the analysis.

Using the age, MARA risk score, gender, HSR, payer, and provider taxonomy categories, we arrived at 107,520 unique combinations. CEM matching refers to each combination as a "stratum," and we sorted every SIM- and non-SIM-attributed patient into one of these unique combinations. We dropped any SIM-attributed patient in a stratum that does not contain at least one non-SIM comparison patient from the analysis. The logic behind this omission is that if the strata does not contain a comparison patient, there are not any available comparison patients with sufficiently comparable characteristics. Any comparison patients in a strata without at least one SIM patient were also dropped from the analysis, since they are unneeded for matching purposes (seeTable 2 and Table 4 below).

Matching requires complete data for each SIM and non-SIM patient, and any patient without complete data is dropped from the analysis. For cohort 1 we started with 338,003 SIM practice attributed patients and 2,062,177 patients attributed to non-SIM patient NPIs, and dropped patients based on missing data:

---

[13] See https://gking.harvard.edu/files/gking/files/psnot.pdf for an informative discussion on the limits of propensity score matching, and further information on coarsened exact matching.

Table 2. Patients Dropped Because of Missing Data (Cohort 1)

	Attributed Patients	Attributed Patients with 2015 Claims for Line of Business	Attributed Patients with Eligibility Minimum	Attributed Patients with Eligible HSR	Attributed Patients with MARA
Comparison Patients	2,062,177	1,893,933	1,591,759	1,562,065	1,400,615
Cohort 1 Patients	338,003	310,749	276,151	274,407	250,076
Dropped Comparison Patients		168,244	302,174	29,694	161,450
Dropped Cohort 1 Patients		27,254	34,598	1,744	24,331

We used the remaining SIM and comparison patients to perform CEM matching. Because any unmatched SIM patient or comparison patient was also dropped, we ended up with in the following final SIM and comparison groups:

Table 3. SIM and Comparison Group Patients (Cohort 1)

	Pre-Match Counts	Post-Match Counts	Difference
Comparison Patients	1,400,615	1,028,874	371,741
Cohort 1 Patients	250,076	243,478	6,598

For cohort 2, we dropped patients based on missing data following this pattern:

Table 4. Patients Dropped Because of Missing Data (Cohort 2)

	Attributed Patients	Attributed Patients with 2015 Claims for Line of Business	Attributed Patients with Eligibility Minimum	Attributed Patients with Eligible HSR	Attributed Patients with MARA
Comparison Patients	1,717,467	1,336,633	1,307,785	1,281,785	1,152,408
Cohort 2 Patients	209,468	165,687	162,545	161,614	146,043
Dropped Comparison Patients		380,834	28,848	26,000	129,377
Dropped Cohort 2 Patients		43,781	3,142	931	15,571

And the final SIM and comparison patient group counts were:

Table 5. SIM and Comparison Group Patients (Cohort 2)

	Pre-Match Counts	Post-Match Counts	Difference
Comparison Patients	1,152,408	1,023,238	129,170
Cohort 2	146,043	143,884	2,159

Because the pool of comparison patients was many times larger than the pool of SIM patients, on average there will be more comparison patients in each strata. In order to maintain comparability, each comparison patient needed to be "weighted" so that the weighted average of comparison patients had the same characteristics as the average of the SIM patients. As an example, if a particular stratum has 10 SIM patients and 100 comparison patients, each comparison patient must be weighted in such as way so the 100 comparison patients are counted in the outcome analysis as 10 patients in total.[14]

We performed CEM matching for patients attributed to cohort 1 and 2 separately. Because the claims data needed for outcome analysis were not available for cohort 3,[15] we did not develop a comparison group for this cohort. The output from the CEM matching process is a separate file for cohorts 1 and 2, with three variables in each file: (1) the patient composite ID of each matched SIM and comparison patient, (2) a variable indicating whether the patient was SIM or comparison, and (3) a weight for each patient. Every SIM patient had a weight of 1 and was therefore used equally in analysis of outcomes. Weights for all comparison patients averaged to 1 but varied by strata.

We used weights in analysis of outcomes in one of two ways. In cases in which we reported the average outcome for SIM versus comparison patients, we report weighted average using the CEM weights. For all regression-based analysis, we performed weighted least squares, which minimizes each observation's squared error times the observation's weight.

For calculation of the AHRQ Access to Care indicators, the software we used required us to input all inpatient claims of a population (i.e., in this case, the population made up of our treatment and comparison groups). The software does not facilitate use of "weighted" claims,

---

[14] Although each SIM patient has a weight of 1, the weights for comparison patients are not only based on the ratio of SIM to comparison patients in each stratum, but also take into account the total number of SIM and comparison patients in all strata. See https://gking.harvard.edu/cem for the specific weighting formula and its explanation.

[15] Cohort 3 started very late in the SIM implementation and only had about nine months of total participation. More information on the timing of the three cohorts can be found in the practice transformation chapters of this and the Final SIM Process Evaluation reports.

so in order to make the SIM and comparison groups equal size, we revised our comparison group methodology. Instead of using all comparison patients in a strata, we randomly selected an equivalent number of comparison patients to the number of SIM patients in that same strata. If the strata included more SIM patients than comparison patients, we randomly selected the comparison patients with replacement, allowing multiple versions of the same comparison patient. This matching process gave us the same number of SIM patients as our previously described comparison group but only used a randomly selected subset of matched comparison patients.

CMHC Patient Matching

We attempted to make the CMHC patient matching process as similar as possible to the primary care comparison group development but made significant adjustments based on the different medical and behavioral health characteristics of the people served by each type of provider. The four CMHCs each had somewhat unique interventions, and we therefore tailored a comparison group to each.

One of the CMHCs, Southeast Health Group, requested that we draw a comparison group from patients serviced by three similar CMHCs that did not participate in SIM. These were West Central Mental Health Center, Inc; San Luis Valley Community Mental Health Center (Alamosa); and Centennial Mental Health Center.

Two of the SIM CMHCs—Community Reach Center and Mental Health Partners—requested we draw comparison patients from each CMHC's non-integrated care patients. Jefferson Center for Mental Health, which had a complex intervention that includes patients with and without serious mental illness and with a significant pediatric focus, asked that we draw multiple comparison pools. We selected one based on all patients receiving care from non-SIM Colorado CMHCs and another from patients attributed to pediatric providers using the CIVHC attribution method described previously. This pediatric patient matching followed the same procedure primary care matching followed.

For the CMHC comparison group analysis, with the exception of the pediatric comparison group we used with Jefferson Center, we used a data file to supplement the claims in the APCD. This file of encounter data was generated from Colorado's Behavioral Health Organizations that were the primary payers for CMHCs. CIVHC worked with the Colorado Medicaid office to transform the Medicaid IDs associated with each encounter record into a patient composite ID as used in the APCD. CIVHC then transferred the file of encounter data to us as a supplement to the APCD. This file had a place of service field that allowed us to identify encounters by specific CMHCs, both SIM and non-SIM. And the relatively small number of CMHCs allowed us to resolve name and address variations in a way that would not be possible with the much larger number of primary care providers.

The matching criteria used for the CMHCs (except Jefferson Center's pediatric matching) were based on these criteria:

- Primary diagnosis
- Age
- Gender
- MARA risk score in 2016
- Months of BHO claims in 2017

Because Colorado Medicaid was the payer for the CMHCs, we did not include it as a matching criteria. And because the integrated service was provided by the same type of organization, a community mental health center, we did not match on this variable. Furthermore, because the CMHCs primarily serve patients with serious mental illness, and because different mental health conditions result in different patterns of cost and utilization of services, we added to our matching criteria a behavioral health diagnosis category. We drew each patient diagnosis from encounter records and developed a prioritized or ranked list, given that each patient had potentially many diagnoses. For matching purposes, the patient received a single diagnosis based on the first of the following diagnoses (Table 6) encountered on any patient encounter records:

Table 6. Diagnoses on Patient Encounter Records

Diagnoses on Patient Encounter Records
Schizophrenia related: F20xx
Schizoaffective disorder: F25xx
All other psychosis: F2xxx
Bipolar disorder: F31xx
Major depressive disorder, recurrent: F33xx
Major depressive disorder, single episode: F32xx
Trauma/Stress related disorders: F43
Anxiety disorders: F41
Personality disorders: F60xx
Impulse disorders: F63xx
Various disorders mostly associated with childhood (e.g., attention-deficit hyperactivity disorder): F90–F94 as a single category.
Mental disorder, not otherwise specified: F99
Opioid related: F11

Diagnoses on Patient Encounter Records
Cocaine related: F14
Alcohol related: F10
Cannabis related: F12
Sedative related: F13
Other stimulant related: F15

We excluded patients with other diagnoses from the analysis. For the age cut points, we used the following range:

- < 3, drop
- 3–11
- 12–17
- 18–25
- 26–35
- 36–50
- 51–64
- 65–75
- >75, drop

We based risk score stratification on these MARA risk score cut points:

- (0,0.25)
- (0.25,0.5)
- (0.5,0.75)
- (0.75,1)
- (1,2)
- (2,3)
- (3,4)
- (4,5)
- (5,10)
- (10,20)

- (20,30)
- (30,40)
- (40,50)
- (50,60)

We also matched based on months of service during Year 1 (2017) of the CMHC participation in SIM. Unlike primary care services, patients of CMHCs generally require ongoing outpatient therapy and medication management. We wanted to ensure we were including only patients who received similar months of service, and the cut points for months of service were 1–3, 4–6, 7–9 and 10–12.

In the case of the primary care attribution, all attributed patients had claims or encounter records in the APCD, since that was the only source for initially identifying each patient. But in the case of the CMHCs, patients were identified directly from the centers by Medicaid ID, and many patients were not in the BHO data file or APCD. The requirement for complete data, and for both a SIM and comparison patient to exist in each stratum, resulted in the following loss of attributed patients:

Table 7. Comparison Group and Southeast Health Group

	CMHC Identified Patients	Identified Patients BHO file Baseline Year	Identified Patients with Eligible Diagnoses	Identified Patients with MARA	Identified Patients with Eligible Age	Matched Identified Patients
Comparison Group		5,408	5,306	4,876	4,843	879
Southeast Health Group	1,962	583	583	542	540	323

Table 8. Comparison Group and Mental Health Partners

	CMHC Identified Patients	Identified Patients BHO file Baseline Year	Identified Patients with Eligible Diagnoses	Identified Patients with MARA	Identified Patients with Eligible Age	Matched Identified Patients
Comparison Group		3,602	3,340	3,340	3,264	625
Mental Health Partners	961	589	585	585	585	338

Table 9. Comparison Group and Community Reach Center

	CMHC Identified Patients	Identified Patients BHO file Baseline Year	Identified Patients with Eligible Diagnoses	Identified Patients with MARA	Identified Patients with Eligible Age	Matched Identified Patients
Comparison Group		4,923	4,389	4,389	4,370	328
Community Reach Center	384	230	228	228	227	137

Table 10. Comparison Group from BHOs and Jefferson Center for Mental Health

	CMHC Identified Patients	Identified Patients BHO file Baseline Year	Identified Patients with Eligible Diagnoses	Identified Patients with MARA	Identified Patients with Eligible Age	Matched Identified Patients
Comparison Group from Behavioral Health Organizations		8,858	6,237	5,305	5,305	1,221
Jefferson Center for Mental Health	28,415	3,462	25,554	2,329	2,330	445

Table 11. Comparison Group from APCD and Jefferson Center for Mental Health

	CMHC Identified Patients	Identified Patients BHO file Baseline Year	Identified Patients with Eligible Diagnoses	Identified Patients with MARA	Identified Patients with Eligible Age and HSR	Matched Identified Patients
Comparison Group From APCD		176,151	7,619	6,705	6,698	5,614
Jefferson Center for Mental Health	28,415	3,462	2,554	2,329	1,860	875

Analysis Methodology

Each outcome chapter provided non-technical details on the methodology used to develop the presented results. This section provides additional technical details. We divide the approaches into analysis of (1) practice site-level data in which a single value is reported for each SIM practice site and (2) patient-level data in which a cohort of patients attribute to SIM and non-SIM NPIs has a single value for each patient.

Practice Site Level Analysis

For measures of integration, CQMs, and cost and utilization, we were provided data sets with outcomes by SIM practice site and no data for non-SIM sites. We used the same general approach for each type of outcome. First, we analyzed change over time broken out separately by cohort. Steps included:

- Plotting each practice site's outcome value on a graph, with separate points for each period in which the outcome was measured. Plotting individual practice site values provides the reader with a sense of the distribution of practice site values, and whether the distribution has a wide variance and is skewed.
- Added a line between time periods showing the change over time of the average practice site value. We used separate lines for each cohort, and only calculated averages for all practice sites reporting in both periods connected by the line. For integration and cost and utilization data, data were available for all periods for any site that had data available in at least one period. This was not the case for CQMs, which had much more inconsistent reporting.
- Conduct a statistical test to determine if the average value changed between periods. Because we limited the data to practice sites reporting in both periods, we were able to use paired T-tests, which, in comparison to two sample T-tests, have greater power to detect statistically significant change.

After analysis of change over time in the outcome metrics, we analyzed the effect of three transformation factors on the outcome measures. The three factors were degree of integration, improvement in HIT capacity, and staff satisfaction. Each of these factors potentially drove any change over time in the outcome measures, and we used regression analysis to identify any impact. We measured the relationship between outcomes and transformation factors through both cross sectional and change over time (two period longitudinal) analysis.

To measure cross-sectional relationships, we used each outcome measure as the dependent variable in a regression, and the transformation factor as the independent variable. This approach is equivalent to performing correlation analysis. We performed this with contemporaneous variables (e.g., first assessment period depression screening rates as the dependent variable, and first assessment period integration scores as the independent variable). These correlations were informative on addressing the question of whether practices with higher independent variable scores (integration) also had higher dependent variable scores (depression screening). But this analysis did not inform us whether improvements in the independent variable were correlated with improvements in the dependent variable.

To analyze the effect of improvements in the independent variable, we calculated change over time in both the dependent and independent variable (also known as first differences) and determined if the change in the independent variable was correlated with a change in the dependent variable.

In the case of depression screening and integration, we performed further analysis to determine if the relationship between these variables was different for different practice characteristics. These include practice type (pediatric, mixed primary care, adult), practice location (urban versus rural), practice size, and practice population as measured by the proportion of patients who had payer types associated with lower levels of service (indigent and Medicaid).

We performed this analysis using cross-sectional data by adding the practice characteristic variable as an independent variable and by allowing it to interact with the independent variable, which was integration. We used the same approach in our change over time analysis. Since practice variables do not change over time, they only entered the regressions via interacting with the independent variable.

In the case of depression screening and integration, we found that the effect of integration on depression screening did not vary by practice site characteristics. This was true for both cohorts 1 and 2 and for all four types of practice characteristics. See the Clinical Outcomes chapter for these results. Because depression screening and integration were our most complete CQM and practice transformation data sets, we thought it unlikely that practice characteristics would be important in the analysis of other practice transformation factors and stopped further analysis of this type.

Patient-Level Analysis

Using the patient-level comparison groups described previously, we analyzed if improvement in outcome variables generated from the APCD was larger for patients of SIM-attributed practice sites than for their matched patients from non-SIM-attributed practices. We performed this analysis for the AHRQ access to care indicators and select patient-level cost and utilization measures generated by Milliman.

In the case of the AHRQ access to care indicators, because the SIM group had outcomes indicating poorer access to care than the comparison patients, and trends over time were very similar, we did not perform any statistical tests. See the Access to Care chapter for more details.

In the case of the Milliman-generated cost and utilization measures, we had patient-level outcomes measured at pre-SIM baseline, Year 1, and for cohort 1, Year 2. Because we did not include all patients in our comparison group analysis in the cost and utilization data set from

Milliman, we re-ran comparison group weighting for the included patients. With data from each period for each patient, we calculated difference-in-difference tests, which compared whether the change over time for the SIM group was different than the change over time for the comparison group. The model form was a two panel first difference test.

With this difference-in-difference approach, combined with a matched comparison group, we are decomposing any change over time in cost and utilization for patients attributed to SIM practice sites into two components: (1) changes caused by factors that would also affect the comparison group, such as inflation in medical costs, and (2) changes that only affected the patients attributed to SIM sites. We assume that participation in SIM is the source of the second type of changes. And by measuring change over time for both the SIM and comparison groups, we can subtract the first type of change from the total change over time experienced by the SIM group.

We used a regression approach, first creating differences over time for each patient in each outcome variable. Because our final data set of annual outcomes had missing quarters for many patients, we counted the number of quarters of data in both baseline and the final years and used the change over time in this count as a covariate. And because we used CEM to create the comparison group, we used weighted least squares with weights provided by the CEM output.

With this analytic approach, we included a binary variable indicating whether a patient was attributed to a SIM practice site. The regression coefficient on this “treatment” indicator corresponds to the average change over time for the SIM practice-site-attributed patients minus the average change over time for the comparison patients. A positive coefficient means that the patients attributed to SIM practices had a larger increase or a smaller decrease in the dependent variable.

For three outcome measures, we took a slightly different approach. The variables *30-Day Follow Up After Psychiatric Hospitalization, Readmission After Psychiatric Hospitalization,* and *Readmission After non-Psychiatric Hospitalization* are each limited to patients with an initial hospitalization and do not apply to patients without a hospitalization. We therefore limited our analysis to patients with qualifying hospitalizations, which made the matched comparison group weights incorrect. Because there was less than 50% overlap in patients between the two compared years (e.g., 2015 and 2017 for cohort 1), we could not use a two-panel first difference model. Instead we performed a two-sample difference-in-difference test, creating a time by treatment interaction term.[16] For these three measures, we did not use weighted analysis and instead included as control variables of the variables used in matching.

---

[16] Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data* (2nd ed.). Cambridge, Mass: MIT Press.

2 Practice Transformation: Integration Efforts

Introduction

The SIM Operational Plan provides an overview of the structured approach to achieving the overall SIM goal to improve the health of Coloradans by improving access to integrated primary care and behavioral health services in coordinated community systems, with value-based payment structures, for 80% of state residents by 2019. Integration of physical and behavioral healthcare within one clinical setting enhances whole-patient care by ensuring that fewer people are lost in the process of referral to external services, difficulties are identified earlier, interventions are initiated sooner, and overall care is better coordinated. Additionally, there is potential to reduce total cost of care as many routine behavioral health issues can be addressed from within primary care, without the need for referral to external subspecialists and because of the high degree of co-morbidity between physical and behavioral health conditions. Chronic diseases and overall better physical health can be influenced by behavior change. For patients, integrated behavioral health services often enhance their experience because of the convenience of receiving comprehensive care in one clinical setting and as the result of improved communication among treating providers. The SIM initiative aims to assist primary care practices and CMHC programs in their efforts to move along the continuum toward greater integration of medical and behavioral care within the medical home setting.

SIM Practice Transformation Milestones reflect common attributes of high-performing primary care practices and CMHC programs. They are organized based on a well-recognized framework, Bodenheimer's "10 Building Blocks of High-Performing Primary Care."[17] As depicted in the SIM Framework graphic (Figure 2), the building block concepts are consistent with themes articulated in other published frameworks. For example, the *Colorado SIM Practice Transformation Toolkit*[18] references the *Lexicon for Behavioral Health and Primary Care Integration*[19] for definitions and information on the concepts of integrated behavioral healthcare.

SIM cohort 1 practice sites and CMHC programs were encouraged to prioritize building blocks and milestones using the Milestone Activity Inventory (MAI) to meet their self-identified needs and plans for becoming more integrated. The practice site and CMHC program experiences and

---

[17] Bodenheimer, T., Ghorob, A., Willard-Grace, R. & Grumbach, K. 2014. The 10 building blocks of high-performing primary care. Annals of Family Medicine 12(2), 166-71.

[18] Colorado State Innovation Model (SIM) Practice Transformation Toolkit. http://resourcehub.practiceinnovationco.org/tools/

[19] Peek, C.J. and the National Integration Academy Council. 2013. Lexicon for Behavioral Health and Primary Care Integration: Concepts and Definitions Developed by Expert Consensus. AHRQ Publication No.13-IP001-EF. Rockville, MD: Agency for Healthcare Research and Quality. Available at http://integrationacademy.ahrq.gov/sites/default/files/Lexicon.pdf.

lessons learned, along with input from the Colorado Multi-Payer Collaborative regarding the milestones they wanted practices to achieve, led to the evolution of the MAI to become the Milestone Attestation Checklist (MAC) for the CMHC programs and cohorts 2 and 3. With the implementation of the MAC, the SIM office also implemented a more structured approach or model for prioritizing building block and milestone efforts. This approach retained the focus on integrated care and brought more focus on moving towards implementing strategies to be successful in alternative payment models and other areas of integration. The MAC is based on the Bodenheimer Building Blocks of Advanced Primary Care with milestones developed in partnership with the multi-payer collaborative and the practice transformation workgroup.

The SIM Implementation Guide clarifies the general philosophy behind the integration of these building blocks with the SIM initiative:

> *"it is important for each practice to become familiar with the models of integrated behavioral healthcare, be able to identify their current stage along the continuum of integrated care and develop plans to move along the continuum toward truly integrated care. The SIM practice transformation building blocks outline key activities and skills needed to develop integrated care in the context of healthcare reform. The intent of SIM is to support practices to develop the skills of an advanced primary care practice (as formulated through the 10 building blocks) and to use those skills to improve patient outcomes, practice productivity, and the integration of behavioral health."*

The following graphic presents the 10 building blocks as categorized by Bodenheimer and associates. As we detail in the following section, SIM adopted and adapted these building blocks by adding an eleventh: the Behavioral Health Integration Subscale of the Practice Monitor (referred to throughout this chapter as "Building Block 11").

Figure 2. Building Blocks[20]

These building blocks overlapped with an integrated care framework used extensively throughout the *Colorado SIM Milestone Implementation Guide,* and they acted as a model for conceptualizing the categories of collaboration between medical providers and behavioral health providers. That integrated care framework includes three general forms or levels of integration: coordinated care, co-located care, and integrated care. SIM practice sites and CMHC programs use the Integrated Practice Assessment Tool (IPAT)[21] to assess their levels of integration, specifying them into one of six IPAT levels (two each for the three general levels of integration). However, the IPAT worked better as an internal tool to facilitate conversations about integration and did not do well showing change over time. Therefore, we did not use it in our final evaluation analyses.

To further assess how completely each building block was implemented, we also used the *SIM Milestone Implementation Guide* Comprehensive Primary Care Practice Monitor[22] (Practice Monitor). We use data from the Practice Monitor in this chapter, depending on the evaluation question. We also use Milestone data and data from key informant interviews as appropriate.

The integrated care framework in the lower part of the graphic above (Figure 2) is used extensively throughout the *Colorado SIM Milestone Implementation Guide* and act as a model

---

[20] Bodenheimer, T., Ghorob, A., Willard-Grace, R., & Grumbach, K. (2014). The 10 building blocks of high-performing primary care. *Annals of Family Medicine 12*(2), 166–71.

[21] Waxmonsky, J.A., Auxier, A., Wise-Romero, P., Heath, B. Integrated Practice Assessment Tool (IPAT). Available at: https://www.integration.samhsa.gov/operations-administration/IPAT_v_2.0_FINAL.pdf

[22] Comprehensive Primary Care Practice Monitor. http://resourcehub.practiceinnovationco.org/filters/?tags[]=107

for conceptualizing the categories of collaboration between medical providers and behavioral health providers. That framework includes three general forms or levels of integration: coordinated care, co-located care, and integrated care, as shown in Figure 2 above.

This chapter examines the progress made in practice transformation achieved by SIM primary care practice sites and the CMHCs. There are three outcome-related evaluation questions related to practice transformation efforts.

- To what extent did practice sites and bi-directional programs move along the continuum of integration? How do they change over time?
- To what extent are primary care and behavioral health providers satisfied with the experience of integrating primary and behavioral healthcare? Report burden? Does satisfaction increase and burden decrease over time?
- To what extent are consumers in SIM practice sites and bi-directional programs satisfied with the experience of primary and behavioral healthcare?

Data Sources

Measuring Integration

SIM practice sites and CMHCs used the Comprehensive Medical Home Practice Monitor (Practice Monitor) to assess how completely each building block had been implemented. Milestone data and data from key informant interviews were also used as appropriate.

It is important to note that the SPLIT assessments were all self-reported, completed by practice sites and CMHCs with assistance from their PFs or CHITAs. There is a possibility of bias in the instruments: sites may not be completely objective when assessing the degree to which they have implemented specific activities. Additionally, the assessments may have been completed by different respondents within the practice sites across assessment periods, which raised limitations of rater reliability. Assistance from the PF or CHITA should help to minimize this bias, but these assessments remain primarily the views and reports of individuals making change within the practice, not entirely those of an objective outside observer.

Measuring Clinical and Staff Experience

During each assessment period, the University of Colorado Department of Family Medicine (UCDFM) conducted an online survey of clinicians and staff working in each SIM primary care practice site and CMHC. Respondents used their own definitions/conceptualizations of "burnout" to indicate which best described their work situation on a 5-point scale: no burnout and enjoy work (1), occasionally stressed out (2), definitely burning out (3), symptoms not going away (4), and completely burned out (5). Clinicians and staff also used a 5-point scale (1-strongly disagree to 5-strongly agree) to rate their job experience, including areas of work-life balance, staff-patient time, safety, care quality, and overall satisfaction. Respondents identified themselves as clinicians, nursing staff, front office staff, or “other” from cohort 1. The instrument was modified so respondents in cohorts 2 and 3 additionally identified whether they were primary care or behavioral health providers.

Measuring Patient Experience

The evaluation had two distinct data sources by which to describe the experiences of patients in SIM practice sites: (1) survey results provided directly to TriWest by SIM primary care practice sites that were already conducting patient surveys and were willing to share aggregate results and (2) the Department of Health Care Policy and Finance’s (HCPF) annual CAHPS® survey of Medicaid patients. In 2018, HCPF sampled only SIM practice sites for the annual survey; however, HCPF did not sample all SIM practice sites.

The SIM office, the University of Colorado Department of Family Medicine (UCDFM), and TriWest included questions in the closeout survey of cohort 1 practice sites to request patient experience of care data summaries from those sites willing to share that information. We present other non-survey-related efforts to improve patient engagement using milestone data and descriptions of relevant small-grant-funded activities.

In addition to these patient perspectives, and to get a broader perspective on what SIM practice sites were doing to better engage their patients, we also present a summary of results of the Practice Monitor Building Block 5, Patient-Team Partnership, which gave us practice sites’ self-assessments of their work to better engage patients.

Movement on the Integration Continuum

PT1. To what extent did practice sites and bi-directional programs move along the continuum of integration? How do they change over time? Do practices report an ability to sustain any changes made during SIM?

The Comprehensive Primary Care Practice Monitor (Practice Monitor) survey instrument included a series of questions designed to address various internal processes within practice sites. The survey is divided into 11 key domains that reflect the building blocks and was administered across assessment periods for primary care practice site cohorts and CMHCs. This section provides an overview of a specific domain within the survey, namely behavioral health integration. We will first provide a brief summation of key findings and then move to a more comprehensive presentation of empirical results.

Summary of Key Findings

The measure of behavioral health integration (BHI) used in this analysis is the completeness of BHI implementation as measured by items in Building Block 11 in the Practice Monitor assessment. To more accurately account for the distinct integration efforts of primary care practices versus CMHCs, Building Block 11 included 14 items for primary care practices and thirteen items for CMHCs. For both types of practices, we calculated composite scores by summing the points recorded for the practice and dividing by the maximum possible sum (56 for primary care practices and 52 for CMHCs) to get a score from 0–100% (see Table 12).

We explored the effects of cohort (1,2, 3) and changes over time by analyzing cohort baseline, midpoint, and final assessment data. Separate tables are presented for the primary care cohorts and for the CMHCs. Because of a shorter involvement in SIM, cohort 3 had only two assessments periods, with baseline assessments completed during fall 2018, followed by final assessments completed in spring 2019. Because of the small number of CMHCs (4), statistical analysis involving CMHCs was not completed.

Statistical analysis of baseline composite scores indicated a statistically significant overall difference between the cohorts (F=4.0, df=2/329, p<.05). Post hoc comparisons indicated that cohort 2 was statistically significantly lower at the .05 level than was cohort 1 in the completeness of integration composite score at baseline. Cohort 3 was not statistically significantly different from either cohort 1 or cohort 2.

Analysis of the final assessment for each cohort was also statistically significant (F=10.7, df=2, 317, p<.01) with post hoc comparisons revealing that cohort 2 was higher in completeness of integration than were cohorts 1 and 3. Because the assessments were completed approximately every 12 months, cohort 3's final reporting period was roughly equivalent to the time that passed between baseline and midpoint assessments for cohorts 1 and 2. To better control for cohort 3's abbreviated SIM participation, we completed an analysis to compare cohort 1 and 2's midpoint composite scores to cohort 3's final composite score. That analysis showed the differences between cohorts was not statistically significant. Analysis of change scores from baseline to the midpoint shows that cohorts 2 and 3's levels of integration increased statistically more than cohort 1 (F=11.1, df=2, 321, p<.01) although the improvement for all cohorts was evident.

The effect of assessment period on behavioral health integration was sizeable and highly statistically significant. Specifically, results of one-way ANOVA regression models indicated that average levels of behavioral health integration increased substantially over time (F=163.45, df=2, 890, p<.001).

Table 12. Primary Care BHI Composite Scores

Practice Monitor Building Block 11 – Primary Care Practices			
Percentage of Maximum Possible	**Baseline**	**Midpoint**	**Final**
Cohort 1	61.0%	71.6%	78.7%
Cohort 2	52.6%	74.2%	85.7%
Cohort 3	55.3%	-	77.4%

Table 13. CMHC BHI Composite Scores

Practice Monitor Building Block 11. Baseline, Midpoint and Final			
Percentage of Maximum Possible	**Baseline**	**Midpoint**	**Final**
CMHCs	71.2%	84.6%	90.4%

Analysis of Practice Monitor data yielded several additional, specific findings. First, item-level measures of integration indicated that practice sites were reporting measurable success integrating behavioral health into their sites. For example, 88%, 97%, and 89% of practice sites in cohorts 1, 2, and 3 (respectively) reported that the goal of having their practice site "actively working to improve [their] care of behavioral health conditions" had been "completely" implemented (see Table 14). As we will detail, high percentages of SIM cohort practice sites reported they were moving towards or had achieved "complete" behavioral health integration. Our results showed high integration scores to be most related to more general processes of BH integration and identification of patients with potential BH needs. However, it is equally important to note elements of workplace integration that were less completely implemented. These items were more directly related to the strategic implementation of integration systems that may include evidence-based tracking of patient progress, procedures for targeted follow-up care, and data collection for monitoring and assessing the impacts of integrated care.

Results

This review of the Practice Monitor will focus only on those practice sites that participated in all assessments for their respective cohorts across time periods. Please note that one cohort 2 site withdrew after completing all its final assessments. Although we did not include that practice site in the final count of cohort 2 practice sites that completed SIM, we did include the site's assessment data in this report, which accounts for the 145 practice sites referenced throughout this report and analyses. The final count of sites completing SIM participation were 92 in cohort 1, 144 in cohort 2, and 83 in cohort 3.

Our analysis of behavioral health integration began by reviewing the specific questionnaire items that comprised the Building Block 11 subsection, which included 14 items for primary care practice sites and 13 items for CMHCs. The response scale for each item ranged from 0 for "not at all" implemented to 4 for "completely" implemented. The individual items may be used as individual indicators or aggregated to form a composite index of integration. Our analysis of these items showed that the 14 items scored high in internal consistency (Cronbach's α=.92), implying that the composite index is a reliable measure of overall behavioral health integration. Because of the extremely small number of CMHCs, we did not conduct this analysis for those sites.

The table below shows the specific items and the percentage of practice sites in each cohort that responded "completely" implemented for each of the 14 items. According to the first item, high percentages of SIM cohort practice sites indicated movement towards or have achieved "complete" implementation of activities to improve their care of behavioral health conditions.

Table 14. BHI Item Scores by Cohort

Building Block 11 - Behavioral Health Integration Item Responses by Cohort Percent "Completely Implemented" at Final Assessment			
Survey Question	**Cohort 1 (n = 92)**	**Cohort 2 (n = 145)**	**Cohort 3 (n = 83)**
a. Our practice is actively working to improve our care of behavioral health conditions.	88.0%	97.2%	89.2%
b. We have defined the types of patients who we will particularly target for behavioral health services.	65.2%	88.3%	91.6%
c. We educate all patients and their family members on the benefits of integrated behavioral health and primary care.	46.7%	51.7%	44.6%
d. A system has been implemented to screen or otherwise identify patients with behavioral health conditions, concerns, or contributing factors	81.5%	93.8%	83.1%
e. Reliable registry data are used in our practice to identify and manage specific populations of patients with behavioral health concerns	25.0%	49.0%	42.2%
f. We have an effective system for identifying and assisting patients with mental health issues who are not improving with treatment	25.0%	34.5%	39.8%
g. We have an effective system for identifying and following up with patients with behavioral health issues who do not follow through with planned visits	16.3%	35.9%	34.5%
h. A referral system is in place for those identified as needing specialty behavioral health services	75.0%	88.3%	90.4%
i. Protocols and workflows have been implemented for coordination between primary care and behavioral health clinicians	64.1%	78.6%	63.9%
j. Our staff and primary care clinicians work closely as a team with the behavioral health clinicians in our practice to provide integrated care	76.1%	71.7%	61.4%
k. Behavioral health clinicians in our practice are readily available for warm handoffs and collaboration	62.0%	53.8%	44.6%

Building Block 11 - Behavioral Health Integration Item Responses by Cohort Percent “Completely Implemented” at Final Assessment			
l. Training on behavioral healthcare is provided to all clinicians and staff joining our practice	48.9%	60.0%	43.4%
m. We have developed collaborative agreements such as care compacts with specialty behavioral health clinicians, covering timely access, communication, handoffs, and coordination of services	27.2%	68.3%	48.2%
n. We systematically collect data to track the reach and outcomes of our behavioral healthcare services	23.9%	29.0%	26.5%

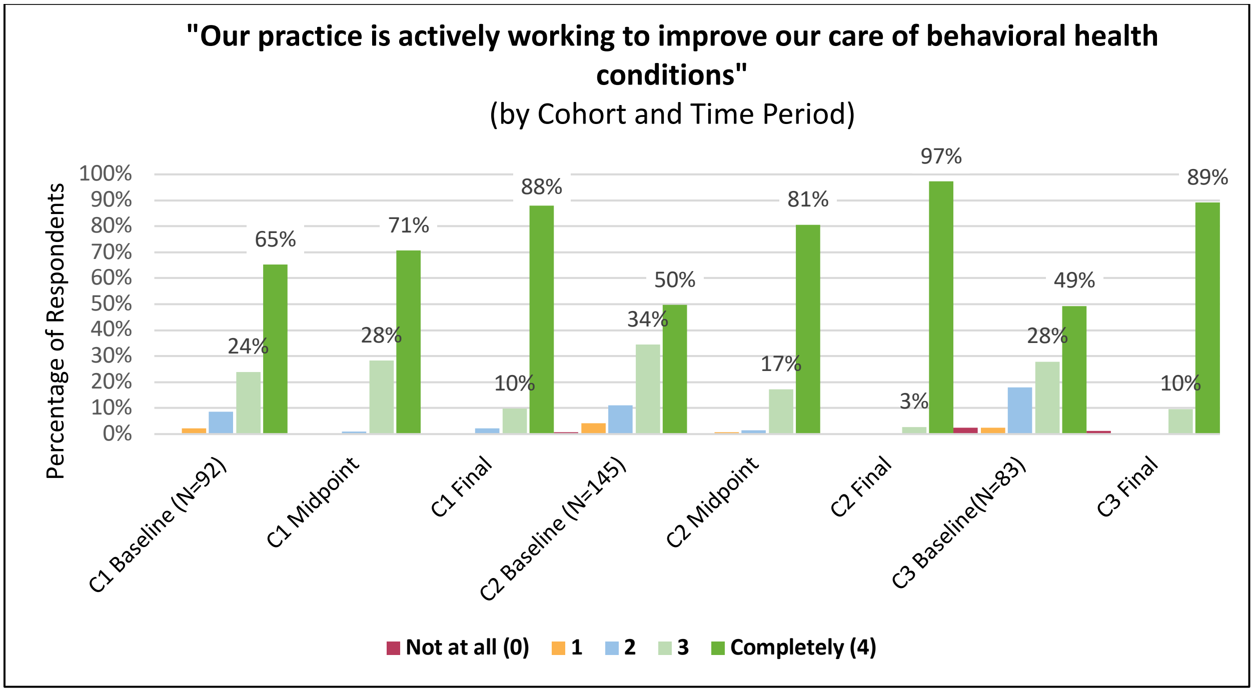
Behavioral Integration: Effects of Cohort and Assessment Period

To assess behavioral health integration we used the overall composite measure of BHI, and we also looked at changes in item A (Table 14 above): implementation of activities to improve their care of behavioral health conditions. We have based the remaining BHI analysis on both the composite integration measure as well as this single-item indicator.

The first graph below (Figure 3) provides a graphical display of percentages responding to the single-item integration question by cohort and assessment periods.

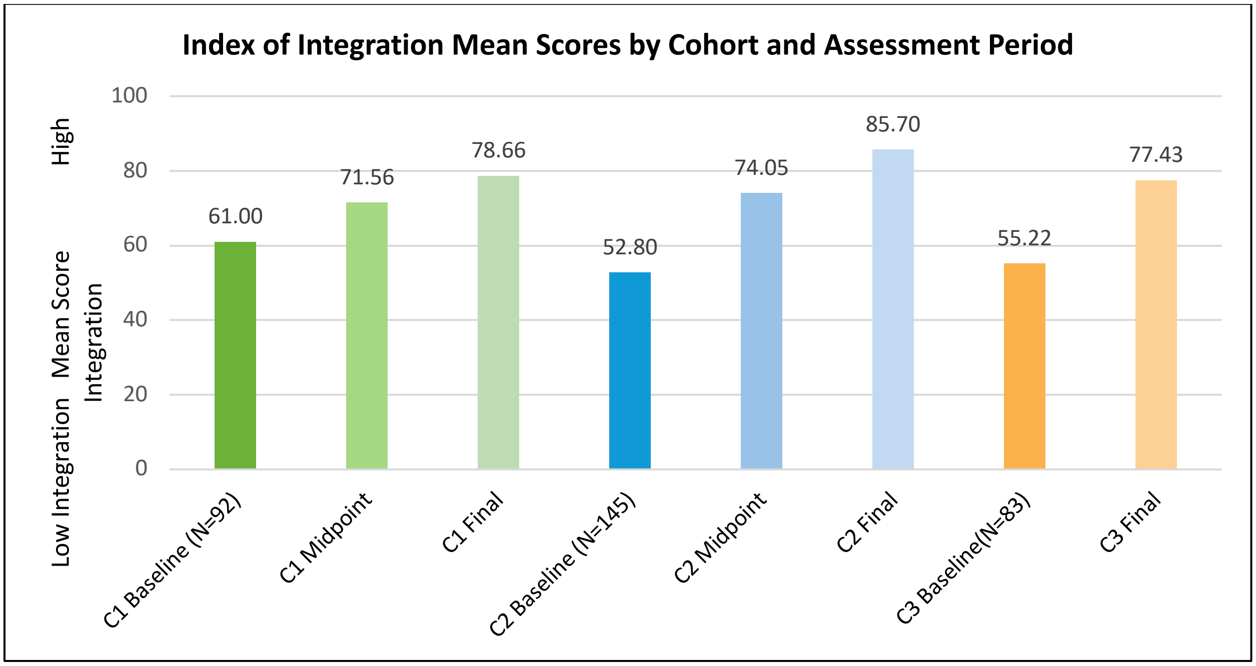
As the graph shows, practice sites reported that the percentage of practice sites who indicated they have “completely” implemented activities to improve their care of behavioral health conditions increased considerably across time periods for all cohorts. We assessed these changes by carrying out chi-square tests to assess levels of statistical significance across assessment points. The large magnitude of percentage change and the associated level of statistical significance for the chi-square test provided evidence of an increase in the degree practice sites report they have completely implemented these activities over assessment periods for each cohort. Results showed that chi-square tests were statistically significant for all cohorts (cohort 1: χ^2=23.63, df=6, p<.01; cohort 2: χ^2=98.77, df=8, p<.001; cohort 3: χ^2=34.06, df=4, p<.001). In short, the results of this analysis suggested that behavioral integration as measured by the first Practice Monitor item improved considerably across time periods for all cohorts.

Figure 3. All Cohorts – BHI Item A



Next, we assessed the effect of cohort and assessment period using the composite measure of integration, which we calculated as a percentage of the maximum possible. As evidenced in the graph below, the general pattern was similar to that shown above for the single item, giving us confidence in the reliability of both the individual assessment question and the composite index measures of BHI. Very noticeable was the sharp, almost linear increase in the composite measure of behavioral health integration from baseline to the final assessment period. We assessed statistical significance using one-way ANOVA regression models. These results indicated that the increases in the composite BHI scores were highly statistically significant for all cohorts (cohort 1 F=21.32, df=2, 273, p<.001; cohort 2: F=123.93, df=2, 432, p<.001; cohort 3: F=51.57, df=1, 164, p<.001). Taken in their totality, our analysis of integration provided compelling evidence of increased behavioral health integration over time and across cohorts.

Figure 4. All Cohorts – Composite BHI Score Over Time



Results from CMHCs

The table below contains responses from the four CMHC sites that participated in SIM. As noted earlier, Building Block 11 on the Practice Monitor contained 13 items for CMHCs. We calculated the scores below using the same methodology we applied to the primary care practice sites: we calculated the number of sites reporting each item as "completed" and presented that as a percentage of total CMHCs indicating completion.

Table 15. CMHC BHI Item Scores

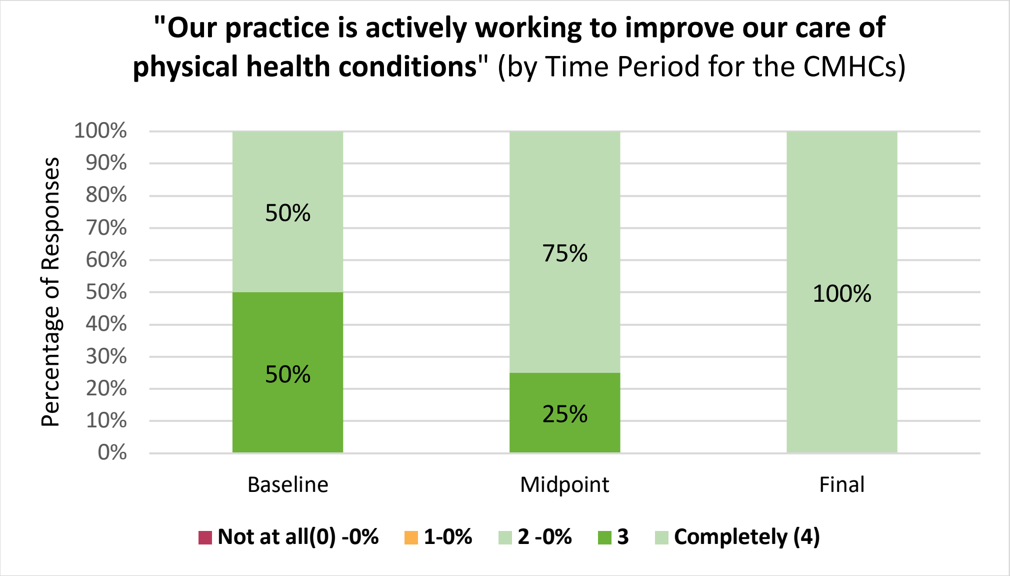
Building Block 11 - Behavioral Health Integration Item Responses (CMHCs) Percentage "Completely Implemented" at Final Assessment	
Survey Question	**CMHCs**
a. Our practice is actively working to improve our care of physical health conditions.	100%
b. An effective system has been implemented to identify patients with physical health conditions or concerns and assure that they receive care either in our center or in another setting	75%
c. We have an effective system to help us identify whether a client has a relationship with a primary care clinician and, if so, to assure that our care is coordinated with that clinician	100%
d. We have defined the physical health conditions that we will particularly target for quality improvement, population management, and/or care management.	75%
e. We educate all patients and their family members on the benefits of integrated behavioral health and primary care	50%

Building Block 11 - Behavioral Health Integration Item Responses (CMHCs) Percentage “Completely Implemented” at Final Assessment	
Survey Question	**CMHCs**
f. We have an effective system for identifying and assisting patients with chronic physical health issues who are not improving with treatment	50%
g. We have an effective system for identifying and following up with patients with physical health issues who do not follow through with planned visits	50%
h. Protocols and work flows have been implemented for coordination between primary care and behavioral health clinicians	75%
i. Our primary care staff clinicians work closely as a team with the behavioral health staff and clinicians to provide integrated care	50%
j. Our practice utilizes warm handoffs and close collaboration between onsite primary care and behavioral health providers	75%
k. Training on integrated care is provided to all clinicians and staff joining our center	50%
l. We have developed collaborative agreements such as care compacts with specialty behavioral health clinicians and medical specialists, covering timely access, communication, handoffs, and coordination of services	75%
m. We systematically collect data to track the reach and outcomes of our integrated primary care services	75%

Figure 5 shows the change over time reported for the first Building Block 11 item by CMHCs (item A in Table 15 above). As with the primary care practices, CMHCs responded to this item on a 4-point scale from “none” to “completely.”

As shown below, half of the CMHCs at baseline reported they had completely implemented activities to improve their care of physical health conditions. By the final assessment, this had risen to 100%. And although the change over time was not statistically significant based on results of chi-square tests (χ^2=3.66, df=2, p=.453), the change over time was informative and indicated that CMHCs reported a positive influence on physical integration over time.

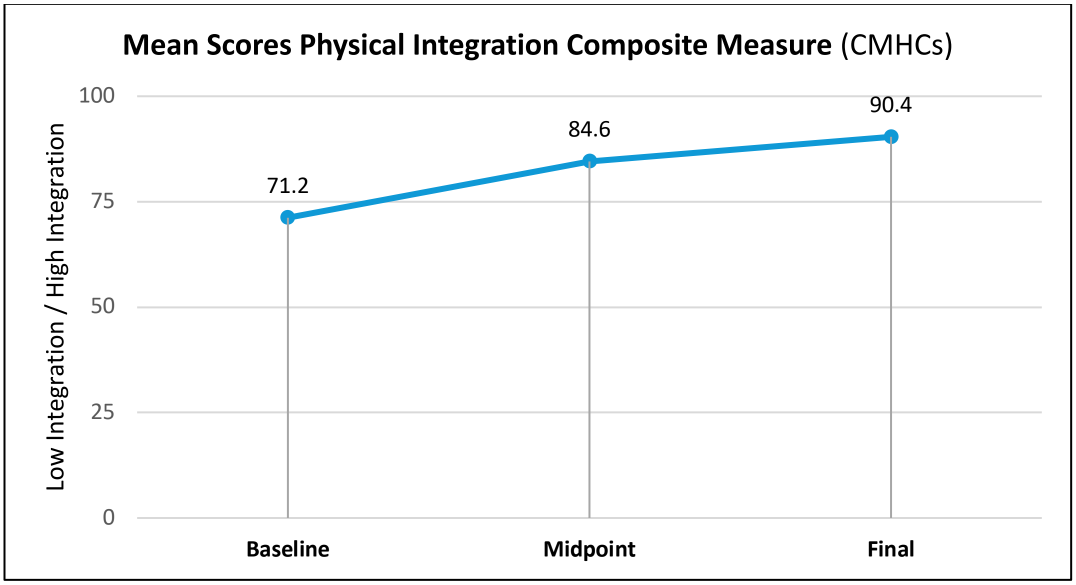
Figure 5. CMHCs Practice Monitor Building Block 11 Item A



We also examined effects of time period on the composite index (calculated by dividing the sum of recorded points by the total possible points) of physical health integration. Again, the small number of cases precluded reliability validation of the index, but we were still able to explore changes in mean scores over time and get a sense of any possible general trends.

As shown in Figure 6 below, mean scores for the composite measure increased markedly over time for CMHCs. We conducted t-tests and—while noting that the small number of sites meant broader generalized conclusions could not be determined— we found several noteworthy outcomes. First, while the difference between midpoint and final time periods was not statistically significant (largely because of the small sample size), we can report two other statistically significant results: (1) there was a statistically significant change in composite Building Block 11 scores from the baseline (M=71.20, SD=21.20) to midpoint (M=84.61, SD=10.99) assessments, indicating increased integration between the two time points (t = 2.77, df=6, p=.032); (2) results showed the largest increase occurred between baseline (M=71.20, SD=21.20) and the final time period (M=90.38, SD=6.28), again showing an overall positive trend in physical integration among CMHCs (t = 3.52, df=6, p=.012).

Figure 6. CMHC Composite Integration Scores



Integration by Practice Characteristics

The following section compares practice site reports of integration levels broken out by specific site characteristics. As noted in the introduction chapter of this report, SIM realized from the outset that a variety of characteristics would shape how practice transformation affected different practice sites. Among those factors were whether the site served adults, children, or both; whether the site was situated in an urban or rural area (based on RUCA code categorizations[23]); the size of the practice (based on the number of annual patient visits); and the degree to which the site served people who were typically underserved (as defined by the percentage of patients insured by Medicaid or uninsured). Please note that except for the RUCA designation, all characteristics were grouped based on data that practices provided on their initial SIM applications.

The tables that follow show the various characteristics used to describe SIM-participating practice sites, and they present the level of integration for each cohort and subgroup at different assessment points. For each practice characteristic, we present characteristic integration levels using two measures of integration. The first is the completeness of implementation for a single item from Building Block 11. As above, this was item A: “our practice is actively working to improve our care of behavioral health conditions.” Practice sites

---

[23] The rural-urban commuting area (RUCA) codes classify U.S. census tracts using measures of population density, urbanization, and daily commuting. The most recent RUCA codes are based on data from the 2010 decennial census and the 2006–10 American Community Survey. Whole numbers (1–10) delineate metropolitan, micropolitan, small town, and rural commuting areas based on the size and direction of the primary (largest) commuting flows. For more information, see https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes.aspx.

chose one of five ratings to describe their status for this item, and the percentage that reported "complete" are shown in the tables. Sites that rated themselves "complete" were coded "yes," and the other sites were coded "no." χ^2 tests were used to analyze change over time.

The second measure was the composite score as described above, with Building Block 11's recorded points divided by the total possible points. We analyzed change over time using ANOVA F-tests and paired-comparison t-tests.

Practice Site Findings by Practice Type

The change in the percentage of practice sites that reported complete implementation of item A was statistically significant for mixed primary care practices (those seeing both pediatric and adult patients) in all three cohorts. The percentage of mixed primary care practice sites reporting complete implementation tended to be lower than their overall cohort averages at baseline, and then increased at both the midpoint and final assessments for cohorts 1 and 2. Mixed primary care practices also reported a statistically significant improvement between their baseline and final assessments.

Adult primary care practices (those seeing only adults patients) also increased over all assessment periods, but those changes were not statistically significant for cohorts 1 and 2, likely because the numbers of adult primary care practice sites was small. The change was statistically significant for cohort 3 adult primary care sites.

Pediatric practice sites tended to report higher levels of integration of physical and behavioral healthcare sooner, and, as a result, their change over time was not statistically significant for cohorts 1 and 3.

For this measure, the percentage of practice sites "completely" implemented increased more from midpoint to final for cohort 1 whereas cohorts 2 and 3 increased more from baseline to midpoint/final.

For the composite percentage of maximum possible ratings in the second table, the increases in behavioral health integration were statistically significant for all cohorts overall and for all but one subgroup, adult primary care practice sites in cohort 1.

For the composite percentage of maximum possible ratings, the degree of increase from baseline to midpoint was greater than from midpoint to final for cohorts 1 and 2.

Table 16. Practice Monitor Building Block 11 Behavioral Health Integration Item A

Practice Monitor Building Block 11. Baseline, Midpoint, and Final					
Item A. % Completely Implemented		**Baseline**	**Midpoint**	**Final**	**P-value**
Cohort and Type	N	%	%	%	χ^2 test
Cohort 1	-	-	-	-	-
Adult	10	60.0%	60.0%	80.0%	.549
Mixed Primary Care	62	61.3%	67.7%	85.5%	.008*
Pediatric	20	80.0%	85.0%	100.0%	.122
Total Cohort 1	**92**	**65.2%**	**70.7%**	**88.0%**	**.001***
Cohort 2					
Adult	5	60.0%	80.0%	100.0%	.287
Mixed Primary Care	103	43.7%	75.7%	96.1%	<.001*
Pediatric	37	64.9%	94.6%	100.0%	<.001*
Total Cohort 2	**145**	**49.7%**	**80.7%**	**97.2%**	**<.001***
Cohort 3					
Adult	17	41.2%	-	100.0%	<.001*
Mixed Primary Care	45	44.4%	-	86.7%	<.001*
Pediatric	21	66.7%	-	85.7%	.139
Total Cohort 3	**83**	**49.4%**	**-**	**89.2%**	**<.001***

* Indicates P-value ≤ 0.05

Table 17. Practice Monitor Building Block 11 Behavioral Health Integration Composite Scores

Practice Monitor Building Block 11. Baseline, Midpoint, and Final					
Percent of Maximum Possible		**Baseline**	**Midpoint**	**Final**	**P-value**
Cohort and Type	N	%	%	%	F test
Cohort 1					
Adult	10	63.6%	73.4%	75.4%	.059
Mixed Primary Care	62	58.4%	69.0%	76.3%	<.001*
Pediatric	20	67.8%	78.5%	88.4%	<.001*
Total Cohort 1	**92**	**61.0%**	**71.6%**	**78.7%**	**<.001***
Cohort 2					
Adult	5	41.1%	67.5%	84.3%	.001*
Mixed Primary Care	103	51.6%	72.7%	84.8%	<.001*
Pediatric	37	57.8%	78.6%	88.4%	<.001*

Practice Monitor Building Block 11. Baseline, Midpoint, and Final					
Percent of Maximum Possible		**Baseline**	**Midpoint**	**Final**	**P-value**
Total Cohort 2	**145**	**52.8%**	**74.0%**	**85.7%**	**<.001***
Cohort 3					
Adult	17	54.1%	-	79.1%	<.001*
Mixed Primary Care	45	55.6%	-	76.0%	<.001*
Pediatric	21	55.4%	-	79.1%	<.001*
Total Cohort 3	**83**	**55.2%**	**-**	**77.4%**	**<.001***

* Indicates P-value ≤ 0.05

Practice Site Findings by Practice Size[24]

There were statistically significant differences along practice site sizes in all cohorts (Table 18 and Table 19 below). For this measure, the percentage of practice sites that reported their site had “completely” implemented increased more from midpoint to final for cohort 1 whereas cohorts 2 and 3 reported larger increases from the baseline to midpoint assessments.

For the composite percentage of maximum possible ratings in the second table, the increases in behavioral health integration were statistically significant for all cohorts overall and for all subgroups.

For the composite percentage of maximum possible ratings, the degree of increase from baseline to midpoint was greater than from midpoint to final for cohorts 1 and 2.

Table 18. Practice Monitor Building Block 11 Behavioral Health Integration Item A.

Practice Monitor Building Block 11. Baseline, Midpoint, and Final					
Item a. % Completely Implemented		**Baseline**	**Midpoint**	**Final**	**P-value**
Cohort and Size	N	%	%	%	
Cohort 1					
Small	18	77.8%	77.8%	100.0%	.096
Medium	27	48.1%	63.0%	74.1%	.145
Large	46	69.6%	71.7%	91.3%	.023
Total Cohort 1	**91**	**64.8%**	**70.3%**	**87.9%**	**.001***
Cohort 2					

---

[24] Small is defined as 0 to 5,999 annual patient visits; medium as 6,000 to 14,999 annual patient visits; and large as 15,000+ annual patient visits.

Practice Monitor Building Block 11. Baseline, Midpoint, and Final					
Item a. % Completely Implemented		**Baseline**	**Midpoint**	**Final**	**P-value**
Cohort and Size	N	%	%	%	
Small	67	55.2%	88.1%	97.0%	<.001*
Medium	49	46.9%	73.5%	98.0%	<.001*
Large	29	41.4%	75.9%	96.6%	<.001*
Total Cohort 2	**145**	**49.7%**	**80.7%**	**97.2%**	**<.001***
Cohort 3					
Small	39	51.3%	-	92.3%	<.001*
Medium	24	50.0%	-	79.2%	.034*
Large	20	45.0%	-	95.0%	.001*
Total Cohort 3	**83**	**49.4%**	-	**89.2%**	**<.001***

Table 19. Practice Monitor Building Block 11 Behavioral Health Integration Composite Scores

Practice Monitor Building Block 11. Baseline, Midpoint, and Final					
Item a. % Completely Implemented		**Baseline**	**Midpoint**	**Final**	**P-value**
Cohort and Size	N	%	%	%	
Cohort 1					
Small	18	59.2%	72.4%	78.9%	<.001*
Medium	27	54.6%	66.9%	72.8%	<.001*
Large	46	65.6%	73.6%	81.7%	<.001*
Total Cohort 1	**91**	**61.1%**	**71.4%**	**78.5%**	**<.001***
Cohort 2					
Small	67	55.6%	76.0%	86.1%	<.001*
Medium	49	46.7%	70.1%	82.9%	<.001*
Large	29	56.7%	76.2%	89.5%	<.001*
Total Cohort 2	**145**	**52.8%**	**74.0%**	**85.7%**	**<.001***
Cohort 3					
Small	39	56.8%	-	80.9%	<.001*
Medium	24	53.3%	-	71.4%	<.001*
Large	20	54.6%	-	77.8%	<.001*
Total Cohort 3	**83**	**55.2%**	-	**77.4%**	**<.001***

* Indicates P-value ≤ 0.05

Practice Site Findings by Practice Location[25]

There were statistically significant changes in the percentage of practice sites that reported complete implementation of efforts to improve their care of behavioral health conditions (item A) over time for all three cohorts overall as well as for all location subgroups except urban practice sites in cohort 1.

For this measure, the percentage of practice sites reporting this activity as "completely" implemented increased more between the midpoint and final assessment periods for cohort 1 whereas cohorts 2 and 3 increased more between their baseline to midpoint (cohort 2)/final (cohort 3) assessments.

For the composite score, the increases in behavioral health integration were statistically significant for all cohorts overall and for all subgroups.

For the composite percentage of maximum possible ratings, the degree of increase from baseline to midpoint was greater than from midpoint to final for cohorts 1 and 2.

Table 20. Practice Monitor Building Block 11 Behavioral Health Integration Item A.

Practice Monitor Building Block 11. Baseline, Midpoint, and Final					
Item a. % Completely Implemented		**Baseline**	**Midpoint**	**Final**	**P-value**
Cohort and Location	N	%	%	%	
Cohort 1					
Urban	69	68.1%	68.1%	89.9%	.105
Rural	23	56.5%	78.3%	82.6%	.003*
Total Cohort 1	**92**	**65.2%**	**70.7%**	**88.0%**	**.001***
Cohort 2					
Urban	89	41.6%	83.1%	98.9%	<.001*
Rural	56	62.5%	76.8%	94.6%	<.001*
Total Cohort 2	**145**	**49.7%**	**80.7%**	**97.2%**	**<.001***
Cohort 3					
Urban	65	49.2%	-	87.7%	.004*
Rural	18	50.0%	-	94.4%	<.001*
Total Cohort 3	**83**	**49.4%**	**-**	**89.2%**	**<.001***

* Indicates P-value ≤ 0.05

---

[25] Urban (RUCA codes 1–3) and Rural (RUCA codes 4–10)

Table 21. Practice Monitor Building Block 11 Behavioral Health Integration Composite Scores

Practice Monitor Building Block 11. Baseline, Midpoint, and Final					
Percent of Maximum Possible		**Baseline**	**Midpoint**	**Final**	**P-value**
Cohort and Location	N	%	%	%	
Cohort 1					
Urban	69	63.8%	71.6%	79.9%	<.001*
Rural	23	52.6%	71.4%	75.1%	<.001*
Total Cohort 1	**92**	**61.0%**	**71.6%**	**78.7%**	**<.001***
Cohort 2					
Urban	89	51.0%	76.5%	89.0%	<.001*
Rural	56	55.7%	70.1%	80.5%	<.001*
Total Cohort 2	**145**	**52.6%**	**74.2%**	**85.7%**	**<.001***
Cohort 3					
Urban	65	57.1%	-	77.6%	<.001*
Rural	18	48.4%	-	76.9%	<.001*
Total Cohort 3	**83**	**55.3%**	**-**	**77.4%**	**<.001***

* Indicates P-value ≤ 0.05

Practice Site Findings by Volume of Underserved Patients[26]

There were statistically significant changes in the percentage of practice sites that reported complete implementation of efforts to improve their care of behavioral health conditions (item A) for all three cohorts overall as well as for all underserved patient volume subgroups in cohorts 2 and 3. We found statistically significant changes in cohort 1 practice sites that saw a medium volume of underserved patients.

For this measure, the percentage of practice sites reporting “complete” implementation increased more from the midpoint to final assessment periods for cohort 1 whereas cohorts 2 and 3 increased more from baseline to their midpoint (cohort 2)/final (cohort 3).

For the composite score of overall achievement (Table 23), the increases in behavioral health integration were statistically significant for all cohorts overall and for all subgroups.

---

[26] Low is defined as 0% to 10% underserved patients; medium as 11% to 30%; high as 31% to 50%; and very high as over 50%.

For the composite percentage of maximum possible ratings, the degree of increase from baseline to midpoint was greater than from midpoint to final assessment periods for cohorts 1 and 2.

Table 22. Practice Monitor Building Block 11 Behavioral Health Integration Item A

Practice Monitor Building Block 11. Baseline, Midpoint, and Final					
Item A % Completely Implemented		**Baseline**	**Midpoint**	**Final**	**P-value**
Cohort and Volume Underserved	N	%	%	%	
Cohort 1					
Low	24	54.2%	54.2%	83.3%	.052
Medium	26	57.7%	65.4%	96.2%	.004*
High	13	76.9%	76.9%	84.6%	.854
Very High	27	77.8%	88.9%	88.9%	.415
Total Cohort 1	**90**	**65.6%**	**71.1%**	**88.9%**	**.001***
Cohort 2					
Low	34	32.4%	82.4%	97.1%	<.001*
Medium	28	50.0%	82.1%	96.4%	<.001*
High	30	56.7%	83.3%	96.7%	<.001*
Very High	53	56.6%	77.4%	98.1%	<.001*
Total Cohort 2	**145**	**49.7%**	**80.7%**	**97.2%**	**<.001***
Cohort 3					
Low	16	31.3%	-	100.0%	<.001*
Medium	28	57.1%	-	92.9%	.002*
High	15	26.7%	-	66.7%	.033*
Very High	24	66.7%	-	91.7%	.036*
Total Cohort 3	**83**	**49.4%**		**89.2%**	**<.001***

* Indicates P-value ≤ 0.05

Table 23. Practice Monitor Building Block 11 Behavioral Health Integration Composite Scores

Practice Monitor Building Block 11. Baseline, Midpoint, and Final					
Percent of Maximum Possible		**Baseline**	**Midpoint**	**Final**	**P-value**
Cohort and Volume Underserved	N	%	%	%	
Cohort 1					

Practice Monitor Building Block 11. Baseline, Midpoint, and Final

Percent of Maximum Possible		**Baseline**	**Midpoint**	**Final**	**P-value**
Cohort and Volume Underserved	N	%	%	%	
Low	24	54.2%	67.0%	73.4%	<.001*
Medium	26	58.9%	68.5%	78.0%	<.001*
High	13	61.5%	74.3%	82.7%	<.001*
Very High	27	70.4%	77.8%	83.3%	<.001*
Total Cohort 1	**90**	**61.5%**	**71.7%**	**79.0%**	**<.001***
Cohort 2					
Low	34	41.3%	71.0%	86.6%	<.001*
Medium	28	51.6%	78.8%	87.4%	<.001*
High	30	53.6%	70.4%	82.8%	<.001*
Very High	53	60.3%	75.6%	85.8%	<.001*
Total Cohort 2	**145**	**52.8%**	**74.0%**	**85.7%**	**<.001***
Cohort 3					
Low	16	47.6%	-	73.9%	<.001*
Medium	28	55.5%	-	78.0%	<.001*
High	15	53.3%	-	67.6%	.023*
Very High	24	61.1%	-	85.3%	<.001*
Total Cohort 3	**83**	**55.2%**	**-**	**77.4%**	**<.001***

* Indicates P-value ≤ 0.05

Provider Satisfaction and Burnout

PT4. To what extent are primary care and behavioral health providers satisfied with the experience of integrating primary and behavioral healthcare? Report burden? Does satisfaction increase and burden decrease over time?

This section summarizes results from Clinician and Staff Experience Survey (CSES) data as reported by cohorts 1, 2, and 3. Workplace satisfaction and burnout served as the key outcomes, or dependent variables, of interest. We explored the effects of cohort (1, 2, and 3) and practitioner role as either “behavioral,” “physical,” or “other,” (independent variables) on changes in workplace satisfaction and burnout. Additionally, we evaluated change over time by analyzing changes from across all available assessment periods: baseline (cohorts 1, 2, and 3), midpoint (cohorts 1, 2, and 3—though cohort 3's "midpoint" was actually its final because of cohort 3's by-design abbreviated SIM participation), and final (cohorts 1 and 2).

The measure of burnout was a single item in the CSES that reads as follows: "Using your own definition of burnout, please indicate which statement best describes your situation at work." Burnout rating choices ranged from 0 to 4, with higher scores indicating higher levels of burnout.

We used two measures to capture feelings of workplace satisfaction. The first is a single-item indicator that asks respondents to rate their level of agreement with the statement, "Overall, I am satisfied with my work in our practice." The second measure is a composite measure of satisfaction using all 15 CSES items by totaling all ratings and using a simple mathematical formula to create a 100-point scale. For both satisfaction measures, higher scores indicate higher levels of satisfaction.

Summary

Results of Clinician and Staff Experience Survey (CSES) data provide a multifaceted portrait of workplace satisfaction and burnout. In general, several key results stand out as particularly noteworthy based on statistical analysis. Two points are important in considering the findings. The first is that the number of completed surveys, across all time periods, was high (N=15,448), partially because the practice sites were required to participate in this survey (measured as at least 70% of staff completing the survey), and the University of Colorado Department of Family Medicine (UCDFM) employed a follow-up process to ensure participation. The high number of surveys completed is also seen as a positive indicator of staff enthusiasm and interest. The second is that the large number of respondents also provide high levels of statistical power to find statistical significance with relatively small differences that may or may not be meaningful. We can outline several key results as follows:

Overall, the average workplace satisfaction across all respondents was generally high, with large majorities (85%) of respondents agreeing or strongly agreeing that they are satisfied with the work they do at their practices. Most respondents reported either no burnout (20%) or occasional stress in the workplace (56%), but 7% reported high levels of burnout. The remaining 17% reported that they were gradually burning out.

We analyzed responses from those practices that had completed surveys across all assessment periods and assessed any possible change in burnout and practice satisfaction over the available time periods for each cohort (e.g., baseline, midpoint, final). In terms of burnout, results of chi-square tests indicated that time period effects are negligible and do not reach statistical significance for cohorts 2 and 3, with very little noticeable change taking place between assessment periods. However, statistically significant effects of time period did appear at cohort 1 (χ^2=28.46, df=8, p<.001). The overall pattern showed that percentages reporting “no burnout” declined from 21% at baseline, to 17% at midpoint, and then increased to 19% at the final assessment period. There was also a decline in respondents reporting “definitely burning out” from baseline (19%) to the final assessment time period (16%).

Unlike the burnout measure, we did find evidence of a time period effect on workplace satisfaction. For cohort 1, workplace satisfaction increased across time periods. The results of that chi-square test indicated that this increase was statistically significant (χ^2=30.41, df=8, p<.001). For cohort 2, results were somewhat mixed. While the highest level of agreement (“strongly agree”) in workplace satisfaction decreased across time periods, the percentage choosing “4” (“agree”) increased across time periods. Again, this result is statistically significant, based on results of a chi-square test for cohort 2 (χ^2=15.89, df=8, p<.05). Finally, cohort 3 is very similar from midpoint to final assessment periods but showed a general increase in satisfaction from the baseline score (χ^2=11.98, df=4, p<.05).

For the composite satisfaction measure, f-tests showed that the largest and only statistically significant increase in workplace satisfaction was for cohort 2 from baseline to the final assessment period (F=2.88, df=2,6,689, p<.05). Composite satisfaction scores remained constant across time periods for cohorts 1 and 3, with results of F-tests failing to reach statistical significance.

Practitioner Roles

Effects of practitioner roles were highly significant and influenced both burnout and workplace satisfaction for the entire sample. There was substantial evidence that physical health providers reported more negative workplace experiences when compared with behavioral health providers.[27] Specifically, our results indicated that physical providers experienced slightly lower levels of workplace satisfaction when compared to behavioral providers. Physical providers experienced higher levels of workplace burnout when compared to behavioral providers. Both differences (based on the entire aggregate sample across cohorts) were statistically significant and will be reported in detail with the accompanying graphs below.

When exploring changes in burnout and workplace satisfaction among practitioner types over time, results of t-tests indicated that scores for behavioral and physical providers did not change significantly over time in any cohort. This indicated that there was very little change in either burnout or workplace satisfaction over the time points for either provider type or across cohorts. However, there was slight (but not statistically significant at the .05 level) evidence that burnout scores increased for cohort 3 behavioral (p=.158) and physical health (p=.111)

---

[27] This role delineation item was added after cohort 1 and included only for cohorts 2 and 3.

providers over time; we discuss this phenomenon in detail below. Additionally, there was some evidence of lowering scores on the composite measure of satisfaction (p=.135), but only for physical health providers in cohort 2. Outside of these three marginally significant t-tests, the evidence overall pointed to minimal effects of time period on changing burnout and satisfaction scores among practitioner types.

Burnout Measure

For the measure of work-related burnout, the CSES included the following questionnaire items shown below. We assigned a specific numeral score for each possible response, as shown in Figure 7, that we then used for analysis. Higher scores corresponded to higher levels of burnout.

Figure 7. CSES Burnout Item

Clinician and Staff Experience Survey Burnout Item
Please indicate which statement best describes your situation for your work at your practice site: (Using your own definition of "burnout" select the most applicable response)
○ I enjoy my work. I have no symptoms of burnout. **(Score as 0)**
○ Occasionally I am under stress, and I don't always have as much energy as I once did, but I don't feel burned out. **(Score as 1)**
○ I am definitely burning out and have one or more symptoms of burnout, such as physical and emotional exhaustion. **(Score as 2)**
○ The symptoms of burnout that I'm experiencing won't go away. I think about frustrations at work a lot. **(Score as 3)**
○ I feel completely burned out and often wonder if I can go on. I am at the point where I may need some changes or may need to seek some sort of help. **(Score as 4)**

Satisfaction Measures

The CSES also included a series of questions designed to measure overall workplace experience of, or satisfaction with, work in the practices. The individual items (shown below) can either be analyzed as single-item indicators or as a 15-item composite index that, according to our analysis, scored high in internal consistency (reliability analysis, $\alpha = .85$). The satisfaction index allowed us to gauge an overall score based on the 15 items below and ranged from 0 to 100, with a higher score indicating a higher level of workplace satisfaction.

The results for workplace satisfaction shown in this report are based on both the composite satisfaction score calculated by UCDFM researchers, confirmed by TriWest, and the single-item indicator asking for level of agreement with the statement, "Overall, I am satisfied with my work in our practice" (item 1).

Figure 8. CSES Items Included in Composite

Clinician and Staff Experience Survey Items
For each of the following statements, please indicate a level of agreement or disagreement, as it applies to you and your coworkers for your work at your practice site, using a scale of 1 to 5 (1= Strongly Disagree, 2= Disagree, 3= Neutral, 4= Agree, and 5= Strongly Agree).
1) Overall, I am satisfied with my work in our practice:
2) My work in our practice is very stressful:
3) Work rarely encroaches on my personal life:
4) My practice does a great job in dealing with quality and safety issues:
5) Our staff and clinicians work together really well as a team:
6) Our clinicians have adequate time to spend with our patients during their office visits:
7) Our staff members have adequate time to spend with our patients during their office visits:
8) In providing care to our patients, our clinicians end up doing many things that the staff could take care of:
9) It is possible to provide high quality care to all patients in our practice:
10) I am frequently overwhelmed by the needs of our patients:
11) I am very satisfied with my career in healthcare:
12) Time pressures keep us from getting to know our patients as well as we need to:
13) I find my current work personally rewarding:
14) I am able to balance work and personal needs in my practice:
15) Our medical record system provides the information we need to provide high quality patient care:

Table 24 below shows the number of responses by cohort for each assessment period of data collection. The numbers of completed surveys shown were from those practice sites that had completed surveys for all assessment periods. This included all sites completing their SIM participation: 92 cohort 1 practice sites, 144 cohort 2 practice sites, and 83 cohort 3 practice sites. For the entire cross section of responses, the final data yielded a relatively large survey sample of 15,448. As expected, because of the smaller number of practice sites in cohort 3, a lower number of respondents were concentrated in cohort 3, with the largest number in cohort 2. Additionally, since cohort 3 only had two assessment points, there were only baseline and final assessments.

Table 24. CSES Respondent Breakout by Cohort and Assessment Period

Task Assessment Period	Cohort 1	Cohort 2	Cohort 3	Row Total
Apr. 2016–Nov. 2016	1,816 (Baseline)	0	0	1,816
Mar. 2017–Apr. 2017	2,175 (Midpoint)	0	0	2,175
Oct. 2017–Dec. 2017	0	2,656 (Baseline)	0	2,656
Feb. 2018–Apr. 2018	2,060 (Final)	0	0	2,060
July 2018–Aug. 2018	0	0	1,031 (Baseline)	1,031
Aug. 2018–Sep. 2018	0	2,236 (Midpoint)	0	2,236
Mar. 2019–Apr. 2019	0	2,222 (Final)	1,252 (Final)	3,474
Total	**6,051**	**7,114**	**2,283**	**15,448**

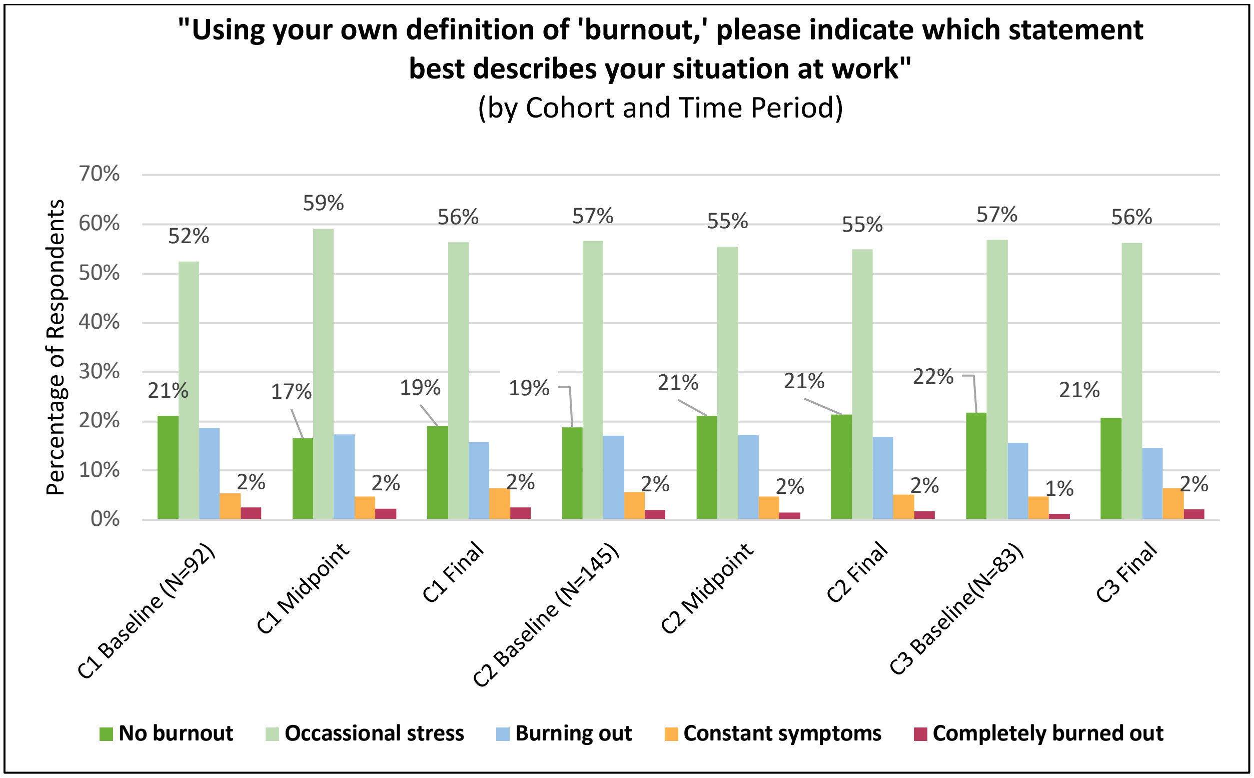
In terms of practitioner roles, the CSES provided several options for respondents to report their professional roles. First, for cohort 2 and cohort 3, the survey provided codes based on practitioner role that included medical, administrative, and behavioral categories. Additionally, respondents were able to choose "other" and write in a more specific description. We conducted a coding analysis of these write-in responses and, wherever possible, collapsed practitioner roles into the three categories. We placed individuals who indicated credentials in behavioral health (e.g., LPC, MSW) in the behavioral category and categorized people with medical degrees and certifications (e.g., M.D., PA) as physical health practitioners. This distinction in the data is limited for cohort 1. Since the behavioral health category was not available for cohort 1 respondents, we identified all cohort 1 behavioral health providers through self-reported, optional write-ins in the "other" category. Because of the indirectness of this method, some cohort 1 behavioral health providers may not have been captured.

Assessment Period Differences

Figure 9 below shows the distribution of responses to the CSES burnout question by cohort and time period within each cohort. As the chart clarifies, across all cohorts and time periods the highest percentage of responses indicated that practitioners and staff experienced occasional stress with no signs of burnout. This trend was followed by consistent percentages of practitioners and staff who reported feeling no burnout and high levels of workplace enjoyment.

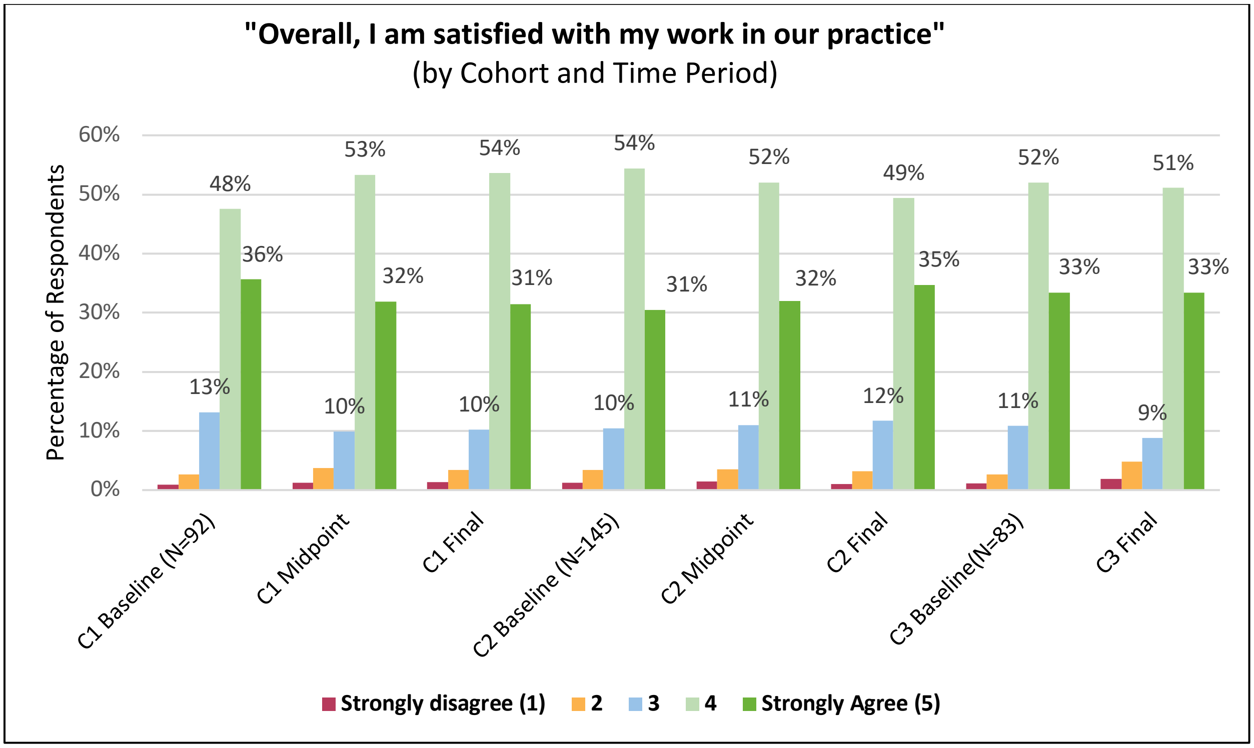
Figure 9 also allows us to consider possible variation based on time periods within each cohort. At first glance, the graph showed little in terms of fluctuations across time periods. Cohort 1 showed a decline in the percentage reporting no burnout at the midpoint but subsequently indicated an increase at the final assessment period. A series of chi-square tests to assess time period effects in each cohort showed that the only statistically significant effects of time period appeared at cohort 1 (χ^2=28.46, df=8, p<.001). All other chi square tests failed to reach statistical significance, indicating that time period effects were negligible for cohorts 2 and 3, with very little noticeable change taking place between assessment periods.

Figure 9. CSES Burnout Responses



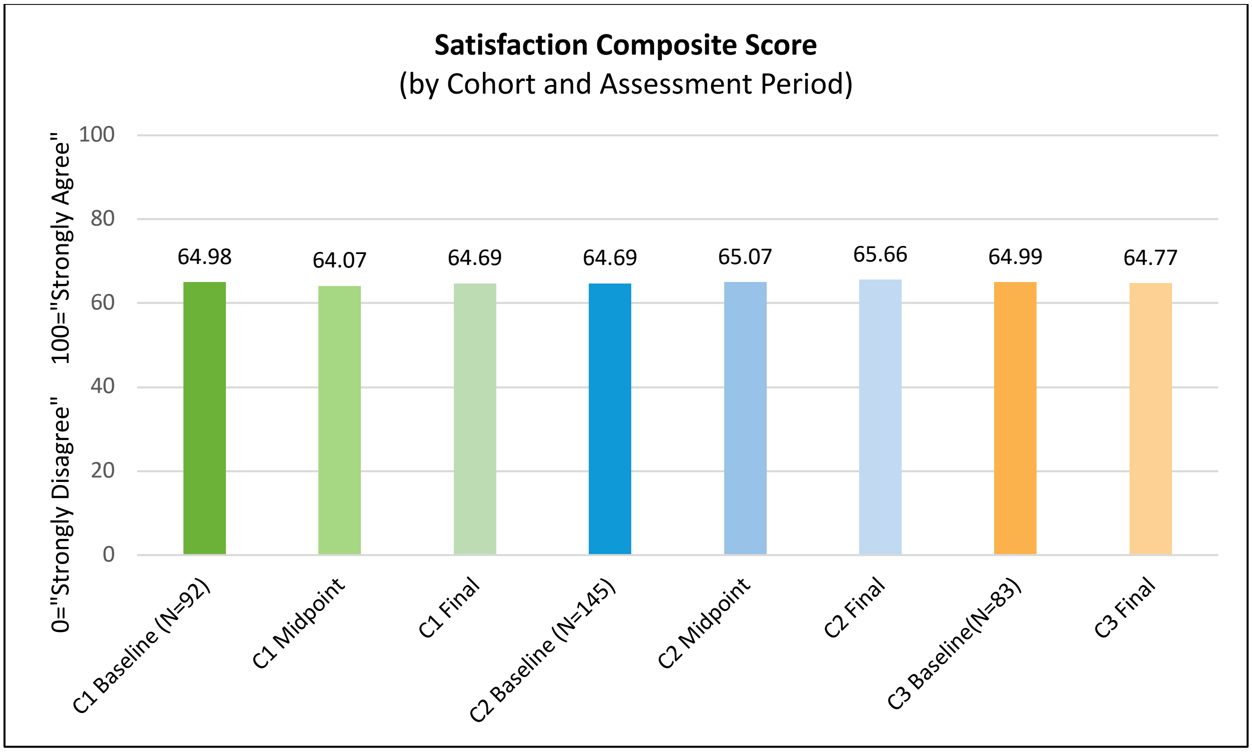
Unlike the burnout measure, we did find evidence of a time period effect on workplace satisfaction. For cohort 1, workplace satisfaction increased across time periods (higher percentages of "agree" and "strongly agree" responses). Again, we conducted chi-square tests to assess statistical significance in the variation in percentages. For cohort 1, the chi-square test indicated that this increase was statistically significant (χ^2=30.41, df=8, p<.001). For cohort 2, results were somewhat mixed. While the highest level of agreement ("strongly agree") in workplace satisfaction decreased across time periods, the percentage choosing "4" ("agree") increased across time periods. Again, results of chi-square tests showed this result to be statistically significant for cohort 2 (χ^2=15.89, df=8, p<.05). Finally, cohort 3 was very similar from baseline to final at the higher levels of satisfaction (4 and 5) but showed a significant decline in average agreement (3) from 11% to 9% between their two assessment periods (χ^2=11.98, df=4, 2, p<.05).

Figure 10. CSES Overall Satisfaction



For the composite satisfaction measure, F-tests showed that the largest and only statistically significant increase in workplace satisfaction was for cohort 2 from baseline to the final assessment period (F=2.88, df=2,6,689, p<.05). As was the case for burnout, satisfaction remained consistent for cohort 3 from baseline to the final assessment periods. Composite satisfaction scores remained constant across time periods for cohort 1 and cohort 3, with results of F-tests failing to reach statistical significance.

Figure 11. CSES Satisfaction Composite Scores by Cohort



Roles

This section examines burnout and workplace satisfaction as a function of respondent roles. Everyday workplace experiences generally differ in terms of stress levels, demands, and overall challenges. These differences might then reasonably translate to differences in burnout and satisfaction.

We close this section by considering the effects of time period *and* practitioner role on burnout and workplace satisfaction. To carry out this portion of the analysis, we conducted a series of t-tests evaluating changes in means scores across time periods for physical and behavioral providers. Table 25 shows mean scores, difference between the first and final time points, and statistical results (size of t-test, and p-value). For ease of interpretation, we have **bolded** changes in means that statistically were below a p-value of .20. And although this benchmark was above the standard .05–.10 cutoff, we noted that the small size of some of the collapsed cells in the analysis (a result of the small number of behavioral providers in the data) would impact statistical significance. With that caveat in mind, we determined these results may be illustrative and provide avenues for further discussion.

As is evident, few of the scores for behavioral and physical providers changed over time or even approached a .20 level of significance. This indicated that there was very little change in either burnout or workplace satisfaction over the time points for either provider type or across cohorts. Interestingly, there was slight evidence that burnout scores increased for cohort 3 behavioral and physical providers over time. Additionally, there was some evidence of lowering scores on the composite measure of satisfaction, but only for physical providers in cohort 2.

Outside of these three marginally significant t-tests, the evidence, overall, pointed to minimal effects of time period on changing burnout and satisfaction scores among practitioner types.

Table 25. Detailed Breakout of CSES Respondents

Means and T-Tests: Key Dependent Variables by Cohort, Time, and Practitioner Roles							
Role	**Cohort**	**Baseline Mean**	**Midpoint Mean**	**Final Mean**	**Difference**	**t-test**	**P-value**
Burnout							
Behavioral (N=98)	1	--[28]	1.16	1.04	-0.12	0.887	0.377
Physical (N=1,446)	1	--	1.16	1.15	-0.01	0.346	0.728
Behavioral (N=156)	2	1.20		1.06	-0.14	1.03	0.301
Physical (N=628)	2	1.03		1.14	0.11	-0.686	0.492
Behavioral (N=95)	**3**	**0.85**	**--**	**1.08**	**0.23**	**-1.42**	**0.158**
Physical (N=840)	**3**	**1.05**	**--**	**1.14**	**0.09**	**-1.59**	**0.111**
Workplace Satisfaction Single-Item indicator							
Behavioral (N=98)	1	--	4.27	4.25	-0.02	0.174	0.862
Physical (N=1,446)	1	--	4.06	4.08	0.02	-0.547	0.584
Behavioral (N=156)	2	4.22		4.21	-0.01	0.072	0.942
Physical (N=628)	2	4.33		4.21	-0.12	0.792	0.428
Behavioral (N=95)	3	4.08	--[29]	4.15	0.07	-0.422	0.673
Physical (N=840)	3	4.18	--	4.14	-0.04	0.555	0.578
Workplace Satisfaction Composite Measure							
Behavioral (N=98)	1	--	66.63	66.74	0.11	-0.052	0.958
Physical (N=1,446)	1	--	64.82	65.68	0.86	-1.17	0.239
Behavioral (N=156)	2	64.58		66.82	2.24	-0.979	0.328
Physical (N=628)	**2**	**69.44**		**65.42**	**-4.02**	**1.49**	**0.135**
Behavioral (N=95)	3	65.10	--	65.68	0.58	-0.204	0.838
Physical (N=840)	3	64.76	--	64.30	-0.46	0.438	0.661

---

[28] Cohort 1 Assessments did not include an item to distinguish behavioral and physical healthcare providers; however, we could identify several individuals who self-identified using the “other” option on the survey.

[29] Cohort 3 did not have a midpoint assessment.

Burnout and Satisfaction by Practice Characteristics

Practice Site Findings by Practice Type

For the analysis of practice site subgroup characteristics, we calculated percentages for each characteristic breakout at each assessment point. The first table below and the tables to follow show percentages for the main outcome variable of interest, either burnout or workplace satisfaction. We used statistical significance (p-values) as determined from chi-squared tests as a means of highlighting statistically significance differences in percentages over time. For all analysis we included only practice sites that had completed surveys at all assessment points. We included all surveys at each assessment period for those practices.

The first table shows the combined percentages who scored in the top 3 “burnout” categories: “definitely burning out,” “burnout symptoms will not go away,” and “completely burned out.” Chi-squared test performed used two levels of burnout (burned out or not burned out) maximum available reporting periods (three for cohorts 1 and 2; two for cohort 3).

Overall, the results suggested several patterns. First, percentages reporting burnout were highest for individuals working in mixed primary care settings. Second, the largest effects from changes over time were found in cohorts 1 and 2. In cohort 1, burnout scores declined at midpoint for adult primary care and pediatric practitioners, before returning to higher levels at the final assessment period. However, these percentages did not reach the level of statistical significance and should be taken cautiously. There was no change at cohort 1 for mixed primary care providers.

While there was little burnout effect over time for cohort 2, there were some notable trends for cohort 3. The findings shown at the lower portion of the table indicate that burnout scores declined for adult primary care providers (p=.066) but increased for mixed care providers between baseline and final assessment periods (p=.098).

Table 26. CSES Burnout by Practice Type

CSES Burnout. Baseline, Midpoint, and Final					
Percentage Burnout		**Baseline**	**Midpoint**	**Final**	**P-value**
Cohort and Practice Site Type	**N of Sites** **N of Surveys**	**%** **Column N**	**%** **Column N**	**%** **Column N**	
Cohort 1					
Adult	10	18.7%	12.8%	20.0%	0.167
	504	155	179	170	
Mixed Primary Care	61	29%	27.7%	26.2%	0.278
	4,110	1,179	1,523	1,408	

CSES Burnout. Baseline, Midpoint, and Final

Percentage Burnout		Baseline	Midpoint	Final	P-value
Cohort and Practice Site Type	N of Sites N of Surveys	% Column N	% Column N	% Column N	
Pediatric	20	22.4%	17.5%	21.78%	0.137
	1,383	450	451	482	
Total	91	26.4%	24.3%	24.66%	0.280
	5,997	1,784	2,153	2,060	
Cohort 2					
Adult	4	14.7%	16.2%	21.7%	0.435
	260	88	80	92	
Mixed Primary Care	99	26.2%	24.8%	24.9%	0.576
	5,207	1,878	1,663	1,666	
Pediatric	36	19.6%	18.7	19.3	0.950
	1,225	417	394	414	
Total	139	24.6%	23.4%	23.7%	0.598
	6,692	2,383	2,137	2,172	
Cohort 3					
Adult	17	25.6%	-	18.1%	0.066
	415	195	-	220	
Mixed Primary Care	42	21.4%	-	25.1%	0.098
	1,422	635	-	787	
Pediatric	21	17.6%	-	20.7%	0.424
	405	193	-	212	
Total	80	21.5%	-	23.1%	0.357
	2,242	1,023	-	1,219	

Results for workplace satisfaction are shown below. Overall, the results of chi-square tests indicated little change over time periods in terms of satisfaction across cohorts for all practice types. Several exceptions included a moderate “U-shape” change in cohort 2 for adult primary care practices, close to the .10 level of significance. This implied that for cohort 2 adult primary care respondents, satisfaction declined at midpoint but increased to its highest levels at the final time period. The only change that was statistically significant at the .05 level occurred among cohort 2 pediatric providers as percentages indicated a gradual decline in satisfaction for this group. Additionally, one overarching pattern seemed to be generally higher satisfaction percentages among pediatric providers compared to the other groups.

Table 27. CSES Highest Overall Satisfaction

CSES Satisfaction Item 1. Baseline, Midpoint, and Final					
Percentage Agree & Strongly Agree		**Baseline**	**Midpoint**	**Final**	**P-value**
Cohort and Practice Site Type	**N of Sites N of Surveys**	**% Column N**	**% Column N**	**% Column N**	
Cohort 1					
Adult	10	89.0%	91.6%	89.4%	0.687
	504	155	179	170	
Mixed Primary Care	61	81.5%	83.3%	83.7%	0.279
	4,110	1,179	1,523	1,408	
Pediatric	20	85.5%	88.9%	87.3%	0.318
	1,383	450	451	482	
Total	91	83.1%	85.2%	85.0%	0.157
	5,997	1,784	2,153	2,060	
Cohort 2					
Adult	4	86.3%	80.0%	90.2%	0.157
	260	88	80	92	
Mixed Primary Care	99	83.3%	83.1%	83.2%	0.975
	5,207	1,878	1,663	1,666	
Pediatric	36	91.6%	88.5%	86.2%	0.048*
	1,225	417	394	414	
Total	139	84.9%	84.0%	84.12%	0.635
	6,692	2,383	2,137	2,172	
Cohort 3					
Adult	17	82.0%	-	85.4%	0.347
	415	195	-	220	
Mixed Primary Care	42	85.0%	-	82.4%	0.192
	1,422	635	-	787	
Pediatric	21	90.1%	-	90.6%	0.424
	405	193	-	212	
Total	80	85.4%	-	84.5%	0.761
	2,242	1,023	-	1,219	

*Indicates P value ≤0.05

Practice Site Findings by Practice Size

Based on the table below, there is very little change in burnout because of assessment period. The only exception is a statistically significant level of change for cohort 2 among practitioners in large offices. Here, the data show a general decline in the percentages reporting high levels of burnout over time.

Table 28. CSES Burnout by Practice Size

CSES Burnout. Baseline, Midpoint, and Final					
Percentage Burnout		**Baseline**	**Midpoint**	**Final**	**P-value**
Cohort and Practice Site Size[30]	**N of Sites N of Surveys**	**% Column N**	**% Column N**	**% Column N**	
Cohort 1					
Small	18	23.6%	21.8%	20.0%	0.651
	733	233	261	239	
Medium	27	22.6%	22.2%	22.9%	0.970
	1,138	335	410	393	
Large	45	28.0%	25.4%	25.8%	0.137
	4,110	1,213	1,475	1,422	
Total	90	26.4%	24.4%	24.6%	0.289
	5,981	1,781	2,146	2,054	
Cohort 2					
Small	65	24.3%	24.2%	22.4%	0.710
	1,670	558	582	530	
Medium	46	21.1%	23.7%	23.1%	0.416
	2,526	832	829	865	
Large	28	27.7%	22.4%	25.3%	.043*
	2,496	993	726	777	
Total	139	24.6%	23.4%	23.7%	0.598
	6,692	2,383	2,137	2,172	
Cohort 3					
Small	38	20.0%	-	16.5%	0.294
	526	254	-	272	

---

[30] Small is defined as 0 to 5,999 annual patient visits; medium as 6,000 to 14,999 annual patient visits; and large as 15,000+ annual patient visits.

CSES Burnout. Baseline, Midpoint, and Final					
Percentage Burnout		**Baseline**	**Midpoint**	**Final**	**P-value**
Cohort and Practice Site Size[30]	**N of Sites N of Surveys**	**% Column N**	**% Column N**	**% Column N**	
Medium	24	21.9%	-	23.5%	0.602
	702	333	-	369	
Large	18	22.0%	-	25.9%	0.148
	1,014	436	-	578	
Total	80	21.5%	-	23.1%	0.357
	2,242	1,023	-	1,219	

*Indicates P value ≤0.05

The workplace satisfaction table below provides at least partial evidence that time periods have an overall greater impact on satisfaction levels. As presented in the table, more of the chi-square tests reach or hover near statistical significance. Workplace satisfaction for cohort 1 practitioners in large offices tended to increase over time, a statistically significant finding at the .10 level. For cohort 2, satisfaction for practitioners in medium size offices declined at midpoint but returned to its original level, a result that is at the .10 level. Finally, we can report several inverse—but statistically significant—trends for cohort 3. Although satisfaction tended to increase for practitioners in medium size offices, the result was the opposite for those working in large offices as the percentage reporting high satisfaction declined between the two time periods. Both effects were significant at the .10 and .05 levels, respectively.

Table 29. CSES Satisfaction – Item 1

CSES Satisfaction Item 1. Baseline, Midpoint, and Final					
Percentage Agree & Strongly Agree		**Baseline**	**Midpoint**	**Final**	**P-value**
Cohort and Practice Site Size	**N of Sites N of Surveys**	**% Column N**	**% Column N**	**% Column N**	
Cohort 1					
Small *	18	83.2%	88.8%	87.0%	0.181
	733	233	261	239	
Medium	27	88.6%	85.8%	85.2%	0.366
	1,138	335	410	393	
Large	45	81.7%	84.3%	84.6%	0.090
	4,110				
Total	90	83.2%	85.1%	85%	0.181

CSES Satisfaction Item 1. Baseline, Midpoint, and Final					
Percentage Agree & Strongly Agree		Baseline	Midpoint	Final	P-value
Cohort and Practice Site Size	N of Sites N of Surveys	% Column N	% Column N	% Column N	
	5,981	1,781	2,146	2,054	
Cohort 2					
Small *	65	86.0%	84.1%	81.7%	0.149
	1,670	558	582	530	
Medium	46	84.9%	82.2%	85.9%	0.103
	2,526	832	829	865	
Large	28	84.2%	85.8%	83.7%	0.525
	2,496	993	726	777	
Total	139	84.9%	84.0%	84.1%	0.635
	6,692	2,383	2,137	2,172	
Cohort 3					
Small *	38	88.1%	-	88.2%	0.987
	526	254	-	272	
Medium	24	82.2%	-	86.9%	0.083
	702	333	-	369	
Large	18	86.2%	-	81.1%	0.031*
	1,014	436	-	578	
Total	80	85.4%	-	84.5%	0.536
	2,242	1,023	-	1,219	

*Indicates P value ≤0.05

Practice Site Findings by Practice Location

In general, the two tables below show very little effect from time/assessment period on either burnout or satisfaction. None of the percentage differences (based on chi-square tests) were statistically significant for the burnout measure, which implied that burnout remained consistent across time periods for urban and rural sites in all three cohorts.

Table 30. CSES Satisfaction Item 1 by Location

CSES Burnout. Baseline, Midpoint, and Final					
Percentage Burnout		**Baseline**	**Midpoint**	**Final**	**P-value**
Cohort and Location	N of Sites N of Surveys	% Column N	% Column N	% Column N	
Cohort 1					
Urban[31]	68	25.7%	23.7%	24.7%	0.452
	4,863	1,331	1,674	1,678	
Rural	23	28.4%	26.5%	24.3%	0.404
	1,314	453	479	382	
Total	91	26.4%	24.3%	24.6%	0.280
	5,997	1,784	2,153	2,060	
Cohort 2					
Urban	85	25.3%	23.4%	24.2%	0.475
	4,567	1,664	1,405	1,498	
Rural	54	23.0%	23.3%	22.7%	0.958
	2,125	719	732	674	
Total	139	24.6%	23.4%	23.7%	0.598
	6,692	2,383	2,137	2,172	
Cohort 3					
Urban	62	22.0%	-	23.2%	0.527
	1,969	898	-	1,071	
Rural	18	17.6%	-	22.3%	0.335
	273	125	-	148	
Total	80	21.5%	-	23.1%	0.357
	2,242	1,023	-	1,219	

Satisfaction trends showed a similar outcome, with only cohort 1 and cohort 2 urban respondents showing modest effects that were not statistically significant. The percentage of highly satisfied cohort 1 respondents employed in urban settings increased gradually from the baseline time period, whereas the same urban group in cohort 2 saw their percentages modestly decline. Again, these findings were not statistically significant.

---

[31] Urban (RUCA codes 1–3) and Rural (RUCA codes 4–10).

Table 31. CSES Satisfaction Overall by Location

CSES Satisfaction Item 1. Baseline, Midpoint, and Final					
Percentage Agree & Strongly Agree		**Baseline**	**Midpoint**	**Final**	**P-value**
Cohort and Location	**N of Sites** **N of Surveys**	**%** **Column N**	**%** **Column N**	**%** **Column N**	
Cohort 1					
Urban	68	82.8%	85.1%	85.1%	0.145
	4,863	1,331	1,674	1,678	
Rural	23	84.1%	85.3%	84.5%	0.859
	1,314	453	479	382	
Total	91	83.1%	85.2%	85.0%	0.157
	5,997	1,784	2,153	2,060	
Cohort 2					
Urban	85	86.9%	85.4%	84.6%	0.164
	4,567	1,664	1,405	1,498	
Rural	54	80.2%	81.2%	82.9%	0.430
	2,125	719	732	674	
Total	139	84.9%	84.0%	84.1%	0.635
	6,692	2,383	2,137	2,172	
Cohort 3					
Urban	62	84.5%	-	83.6%	0.603
	1,969	898	-	1,071	
Rural	18	92.0%	-	90.5%	0.671
	273	125	-	148	
Total	80	85.4%	-	84.5%	0.536
	2,242	1,023	-	1,219	

Practice Site Findings by Volume of Underserved Patients[32]

Below we explore possible effects of assessment period on burnout scores for sites grouped into categories based on the percentage of typically underserved patients. The first general pattern in the data was the somewhat higher percentages of respondents in “high” and “very high” settings reporting burnout. One clear exception was at the cohort 3 baseline assessment period as the percentage of providers reporting burnout was lowest for “very high” underserved practice sites.

---

[32] Low is defined as 0% to 10% underserved patients; medium as 11% to 30%; high as 31% to 50%; and very high as over 50%.

In terms of assessment period effects, the effect was minimal with two exceptions below the .10 level of significance. For cohort 1 practice sites seeing a “low” volume of underserved patients, there was some evidence of a decline in burnout at the midpoint but a gradual return to original levels at the final assessment period. Additionally, there was a statistically significant effect (.10 level) for cohort 2 practice sites seeing a “very high” volume of underserved patients, whereas the general trend seemed to be a modest decline in burnout from the baseline assessment period.

Table 32. CSES Burnout by Volume of Underserved Patients

CSES Burnout. Baseline, Midpoint, and Final					
Percentage Burnout		**Baseline**	**Midpoint**	**Final**	**P-value**
Cohort and Level Underserved	**N of Sites N of Surveys**	**% Column N**	**% Column N**	**% Column N**	
Cohort 1					
Low	24	28.5%	22.5%	28.2%	0.092
	1,200	347	425	428	
Medium	27	25.1%	23.3%	21.9%	0.467
	1,599	461	545	593	
High	14	29.0%	23.8%	26.3%	0.222
	1,178	440	415	323	
Very High	26	24.7%	26.9%	24.3%	0.478
	1,946	518	742	686	
Total	91	26.4%	24.3%	24.6%	0.280
	5,997	1,784	2,153	2,060	
Cohort 2					
Low	32	22.4%	22.0%	23.1%	0.899
	1,793	606	557	630	
Medium	28	21.2%	23.2%	23.9%	0.643
	1,316	404	477	435	
High	30	25.0%	26.8%	22.0%	0.278
	1,259	416	440	403	
Very High	49	27.3%	22.4%	25.1%	0.082
	2,324	957	663	704	
Total	139	24.6%	23.4%	23.7%	0.598
	6,692	2,383	2,137	2,172	

CSES Burnout. Baseline, Midpoint, and Final					
Percentage Burnout		**Baseline**	**Midpoint**	**Final**	**P-value**
Cohort and Level Underserved	**N of Sites N of Surveys**	**% Column N**	**% Column N**	**% Column N**	
Cohort 3					
Low	16	23.4%	-	20.9%	0.423
	748	281	-	467	
Medium	27	23.6%	-	27.6%	0.234
	671	317	-	354	
High	14	21.3%	-	20.5%	0.848
	362	187	-	175	
Very High	23	16.3%	-	22.4%	0.101
	461	238	-	223	
Total	80	21.5%	-	23.1%	0.357
	2,242	1,023	-	1,219	

Finally, the last table below shows workplace satisfaction over the assessment periods for categories of practice sites based on percentage of typically underserved patients. Once again, the effect of assessment period was minimal but was highly significant at the .05 level for two categories. First, percentages of satisfied practitioners tended to increase for cohort 1 for respondents in practice sites serving “medium” levels of underserved populations. Second, we saw a statistically significant effect for cohort 2 practice sites seeing “low” volumes of underserved patients. In this case, however, the effect of assessment period seemed to be the opposite of that described above for cohort 1. In this case, percentages reporting high workplace satisfaction *decreased* moderately across assessment periods.

Table 33. CSES Satisfaction by Volume Underserved

CSES Satisfaction Item 1. Baseline, Midpoint, and Final					
Percentage Agree & Strongly Agree		**Baseline**	**Midpoint**	**Final**	**P-value**
Cohort and Level Underserved	**N of Sites N of Surveys**	**% Column N**	**% Column N**	**% Column N**	
Cohort 1					
Low	24	83.86%	85.1%	85.2%	0.835
	1,200	347	425	428	
Medium	26	82.6%	86.2%	88.0%	0.043*
	1,599	461	545	593	
High	13	83.86%	87.2%	85.4%	0.377

CSES Satisfaction Item 1. Baseline, Midpoint, and Final					
Percentage Agree & Strongly Agree		**Baseline**	**Midpoint**	**Final**	**P-value**
Cohort and Level Underserved	**N of Sites N of Surveys**	**% Column N**	**% Column N**	**% Column N**	
	1,178	440	415	323	
Very High	26	82.4%	83.1%	81.4%	0.710
	1,946	518	742	686	
Total	91	83.1%	85.2%	85.0%	0.157
	5,997	1,784	2,153	2,060	
Cohort 2					
Low	32	91.2%	88.3%	86.8%	0.044*
	1,793	606	557	630	
Medium	28	81.6%	80.7%	80.6%	0.916
	1,316	404	477	435	
High	30	80.7%	80.2%	84.1%	0.294
	1,259	416	440	403	
Very High	49	84.1%	85.2%	83.8%	0.750
	2,324	957	663	704	
Total	139	84.9%	84.0%	84.1%	0.635
	6,692	2,383	2,137	2,172	
Cohort 3					
Low	16	86.4%	-	85.4%	0.693
	748	281	-	467	
Medium	27	82.6%	-	79.6%	0.324
	671	317	-	354	
High	14	83.4%	-	86.8%	0.359
	362	187	-	175	
Very High	23	89.5%	-	22.4%	0.693
	461	238	-	223	
Total	80	85.4%	-	84.5%	0.536
	2,242	1,023	-	1,219	

*Indicates P value ≤0.05

CSES Qualitative Data: Workplace Satisfaction and Suggested Areas of Improvement

This section will provide an overview of systematic qualitative analysis of two open-ended questions available as part of the Clinician and Staff Experience Survey (CSES). We examined data from all three cohorts and employed advanced textual analysis techniques that allowed us to categorize key words and phrases that emerged from the written responses, and to then quantify the rate of specific response themes. The CSES included two open ended items.

- **Open-ended question 1**: After responding to satisfaction items “a-o” on the first page of the survey, partners are asked the following open-ended question: “Please provide comments for any of the above items that you think could be improved.”
- **Open-ended question 2**: The second question asks, “What is one specific suggestion you have for how your practice could increase your overall experience and satisfaction with your job?”

The table below shows the distribution of total survey respondents and the percentages that provided written replies to the two open-ended questions on the CSES. The analysis of this high volume of sentence and paragraph-length responses required a specific methodological approach (e.g., content or textual analysis). As the table shows, 37% of respondents provided responses to question item 1, whereas 44% responded to the second. Keeping in mind the large N size for the total sample, and the fact that many responses were multiple sentences (often written out as long paragraphs), we applied a newly emerging methodology known as “text mining,” using WordStat version 8. This approach allowed us to extract themes from longer responses by essentially “quantifying” all the available text. We were then able to present numerical information on recurring key words and phrases and to connect that information to specific responses provided by our partners.

Table 34. CSES Open-Ended Responses

CSES Total N and Percentage of Open-Ended Responses		
Category	**N**	**% of Total**
Total CSES Respondents	15,448	-
Valid Responses to Workplace Improvement Question	5,726	37%
Valid Responses to Additional Suggestions Question	6,874	44%

Below we provide some examples of open-ended responses.

Workplace Satisfaction Improvement

Workplace Improvement Text Mining: Key Words

Regarding responses about ways to improve workplace satisfaction, the highest occurring words, as depicted in the word cloud below, were “time,” “patients,” “work,” and “staff.” The presence of the singular “patient,” along with the highly recurring “patients,” indicated that responses to questions about workplace experiences that could be improved often revolve around patient-provider interactions. Below we will provide additional context to these responses.

Word counts provide an overall view of reappearing concerns based on single words. Our analysis allows us to dig deeper and find the underlying meaning of these repeating words. For instance, the appearance of the word “time” was associated with two main themes. The first was the desire for more “time off” or time to rest and manage personal life issues, whereas “time” also appeared repeatedly as a reference to time spent with patients. The word “patients” is most referenced in association with specific patient needs that could improve the workplace experience. The table below provides several illustrations from the data that add a much-needed “voice” from practitioners in the field.

Table 35. Workplace Improvement Keyword Context

Workplace Improvement Recurring Keyword Occurrences in Context		
Response Category: Time and Work Life Balance	**Role**	**Size and Location**
"It is hard to work 5 days a week. Having an extra day off would help with personal obligations. I realize that my position would be hard to accommodate a 4 day a week work week, but I do not think it is impossible. This is the only part of this job that I am unhappy with. Overall I love my job. I feel that the administration is accommodating when I need time off for school functions."	Medical Administrative	Large, rural
“Our practice is not good about flexibility and allowing [removed for anonymity] time to attend our children/family events. It's shameful…If they would learn to be flexible for families, job satisfaction would be much higher and people would want to stay, [and this practice] would not lose so much money hiring and training constantly!” [inserted to protect anonymity].	Medical Provider	Large, urban
Response Category: Time with Patients	**Role**	**Location**
"The amount of documentation and recording of data is what results in work encroaching on personal time. In order to get the quality time for patients, I end up documenting/charting and doing administrative work (such as emails, surveys, etc.) in my own time."	Not provided	Medium, urban

Workplace Improvement Recurring Keyword Occurrences in Context		
Response Category: Time and Work Life Balance	**Role**	**Size and Location**
"It would be nice to have more time for the complicated patients. It's very helpful" having a case manager speak with the complicated patients, in order to make sure all their psychological/ medical /social issues are addressed.	Mixed primary care provider	Medium, urban
"I personally feel that more time with patient's both from clinic staff and providers could outcome less use of ER and hospitalizations."	Medical provider	Large, urban
Response Category: Patients	**Role**	**Location**
"I think the behavioral health clinician (BHC) could be involved in more patient care. I think we are missing opportunity to involve her in supporting patients with behavioral and emotional concerns. Improving the teamwork between the providers and BHC is another area for growth."	Medical administrative	Large, rural
"Managed medical care has made time with patients a major issue. While I tend to have plenty of time with my patients (I'm a psychologist), I frequently hear our patients complain about the very fast paced care they receive from the medical providers."	Behavioral Health Provider	Large, urban

Workplace Improvement Text Mining: Key Phrases

The next word cloud and accompanying table below showed recurring phrases for the workplace improvement question. We saw that the phrases involving the themes of "patient care," "medical records," and "quality care" appeared at a high frequency.

Below, we can see the specific context of the top two key phrases. Issues such as integrated care, obstacles of administrative tasks, and cultural competency surface as critical factors in improving patient care. The phrase "medical records" is related to the additional obstacles created by medical record keeping processes, but also alludes to improvements made in this area.

Table 36. Workplace Improvement Phrases in Context[33]

CSES Top Two Key Phrases in Context		
Response Category: Patient Care	**Role**	**Size and Location**
"As one could anticipate, the transition to EMRs has reduced face to face contact with patients and increased the burden of record keeping at the detriment to patient care. It may be necessary at this point in time, but I would hope that the pendulum swings toward more patient centered care, rather than reporting and metrics."	Medical Administrative	Large, rural
“Being on a provider team we all have that goal to help our patients as well as reaching our bonus. In reaching our bonus we have to see a certain number of patients per day to do so. I don't feel that is it is good patient care limiting the time we spend with them just to make out goal for the bonus. We should be rewarded for the care of our patients and not by the number of patients we can see in a day.”	Medical Provider	Large, rural
“I believe that focusing more on effective integrated care practices, like having well attended integrated care meetings and integrated care treatment plans will help immensely with patient care and decrease confusion between clinicians/staff. Additionally, a training focus on cultural competency, motivational interviewing, trauma informed care would be helpful in order to improve client well-being and break down barriers experienced by clients in the medical settings. Finally, it appears that many people are overburdened and stressed at work- sometimes to the point of complete burn out. Really focusing on implementing practical policies and procedures could help to ensure staff are able to complete their tasks and hopefully have time for professional development in order to find satisfaction in their jobs.”	Not provided	Medium, urban

---

[33] Responses have been selected and presented to protect the identities of respondents and practices

Response Category: Medical Records	Role	Location
"Things with medical records change often and we find out we are doing it wrong. Would like a clear guideline for what we send to and request from who."	Front Office Staff	Small, rural
"Increased amount of time for comprehensive care is a necessity! We should not be providing superficial, convenient or urgent care unless it is in addition to addressing other medical (Chronic diseases management, preventative care) issues. Urgent care and productivity models have reduced the focus on the whole patient and reduced the ability to tackle complex issues that require time and rapport building. Doctors should not be reduced to only being decision makers but should have the opportunity to engage in true healing time with patients, that's where your outcome will be changed! There needs to be a better way to "give credit" to docs for everything they do in the exam room and through their administrative tasks like consults, telephone calls, medical records reviews and inbox tasks. I spend 3+ hours a night handling all this daily and this does not often include the time I am also covering for colleagues."	Medical Provider	Large, rural
"Our medical records have certainly improved over the years and it would be helpful that when pts make an apt for a hospital or ER f/u to ensure we have their records. This greatly improves quality and continuity of care."	Administrative, Mixed Primary Care	Large, urban

Workplace Experiences and Satisfaction

Suggestion to Improve Workplace Experiences and Satisfaction Text Mining: Key Words

Regarding additional suggestions related to improving workplace experiences and satisfaction, results of text mining indicated several distinct patterns. First, the words "staff," "work," "time," and "patients" had the highest occurrences. These responses suggest that issues related to patients and the importance of time spent with patients was a salient concern for respondents. As an example, the following informative excerpt provides some context for this interpretation:

> *"Focus on customer service at the front and improve training for our front desk staff, allow for more **time** in the room with **patients** or set expectations with **patients** from the beginning about what we can reasonably accomplish in the **time** allowed, and involve clinical staff in all decisions involving our work prior to rolling new things out."*

Below, we placed the top two recurring words in a broader context. Suggestions are clearly interconnected and show the extent to which satisfaction and positive workplace experiences are associated with very specific perceptions of relationships, policies, and lines of communication. Additionally, the responses from partners below show the extent to which partners enjoy their workplace experiences at the same time that they can shed light on workplace obstacles and constraints.

Table 37. Additional Suggestions Recurring Key Words in Context

CSES Additional Suggestions Top Two Key Words in Context		
Response Category: Staff	**Role**	**Size and Location**
“Change the style of the management team. They are not an entity that supports the staff equally in a positive way and are often unavailable or available with grudge. It does not provide a confident environment for the staff that ensures them that the management is approachable to questions or concerns.”	Medical provider	Large, rural
“Better communication between management and providers and staff, and transparency in decision making processes.”	Administrative staff	Large, rural
“We need help with coverage. Life doesn't always go as planned, and it would be extremely comforting for staff and for patients knowing that everything is getting taken care of and things aren't falling through the cracks. However, I love the current staff like a second family!”	Medical provider	Large, rural
Response Category: Work	**Role**	**Location**
“This is more of a systemic challenge and I realize it isn't necessarily a feasible option at this time, but the one thing that would absolutely improve satisfaction and mitigate risks of burnout in my practice would be to increase the number of BH staff. I love my work and love my job, I love working in BH, I feel satisfied, and I even accept that the BH team/program may continue to look the same as it does now for the foreseeable future due to budgetary and other constraints. AND, at the same time, I think that eventually increasing BH staff --even if doing so is more of a long-range vision or goal for our practice would make a monumental difference in the team's ability to manage the work load, maintain a sustainable work/life balance, and provide optimal care for patients (especially during busy times).”	Behavioral health provider	Large, urban
“I truly enjoy my job and really enjoy the clients that I help every day and I feel that the department I work in makes a difference. I do feel that [practice] needs to find a way to give raises and give the employee incentive. Most of the employees have not seen a raise come through for five years. We do make a difference because most of us love our jobs.”	Administrative specialist	Medium, rural
“Leadership needs to have a paradigm shift in their thinking of how to make the clinics and system work.”	Medical provider	Medium, rural

Suggestion to Improve Workplace Experiences and Satisfaction Text Mining: Key Phrases.

Regarding additional suggestions to improve workplace experiences and satisfaction, recurring phrases, again, revolve around patient and office environment issues. Key phrases included "patient care," "support staff," "front desk," and "time with patients." These phrases captured a central theme of improvement suggestions that revolved around (a) the needs and quality of encounters with patients and (b) the everyday workplace processes under which practitioners, administrators, and staff do their work. Additionally, phrases such as "work flow," "team building," and "team work" also appeared repeatedly, reinforcing the idea that higher satisfaction and enhanced workplace experiences were associated with both patient experiences and service delivery as well as the dynamics of the everyday office environment.

The specific responses below added an important context to the numerical results of text mining and indicate both frustrations and successes among respondents. One overriding theme was the perceived challenge of balancing patient care with operational processes such as record keeping and paperwork.

Also evident was the important role of support staff in providing overall integrated physical and behavioral healthcare. The excerpts show that staff provide an essential component of overall care. For several respondents, the ability of the practice to create a "team-centered" environment that supported practitioners was critical to accomplishing their main goal of providing high-quality patient care.

Table 38. Additional Suggestions Recurring Phrases in Context

CSES Additional Suggestions Top Two Key Phrases in Context		
Response Category: Patient Care	**Role**	**Size and Location**
"Minimize nonclinical pressures… It doesn't help patient care. It decreases patient access Because we are taking time 'meeting measures…' It does not improve the practice or lives of the people that actually provide patient care. This is one of the issues that DIRECTLY DRIVES PHYSICIAN BURNOUT."	Administrative Medical Support	Small, urban
"Work provides me with good challenges for patient care, but not too overwhelming. I am often able to complete all my work at work, which also contributes to my quality of life."	Not provided	Large, urban
"We need to be able to focus on patient care, not focus on "checking the boxes". Healthcare has become soooooooooooo overregulated."	Medical provider	Large, rural
Response Category: Support staff	**Role**	**Location**

CSES Additional Suggestions Top Two Key Phrases in Context		
“Listen to your physicians. Copic has on-line seminars about EMR clearly citing studies that show physicians cannot possibly come close to seeing the number of patients in a day that were seen prior to EMR and provide good quality of care. I think a lot of our physicians are proving this wrong but are sacrificing their own health to do so. Many middle management levels in the practice who have no idea what we do in a day. Eliminate call center - instead give every physician 2 support staff - bring care back to the team caring for the physicians. Would estimate it may cut down on volume of calls by 50%. Be prepared to let staff go when they [don’t meet standards of work ethic and pt care required to do this job.”	Not provided	Large, urban
“Adding more support staff, so that we can concentrate on patient care instead of trying to figure out where her Medical Records Request is or the status. Resending request for medical records.... Add another Nurse Practitioner that would see any type of patient on any giving day. Regardless if its a Pregnant patient or a gynecological patient. Supporting staff that sticks around when you need them. I know there responsibilities have been added on, so it's hard to get help when you need the most.”	Clinical staff	Large, urban
“More support staff around behavioral health”	Administrative support	Large, rural

Patient Engagement

SIM cohort 1 practice sites and CMHC programs were encouraged to prioritize building blocks and milestones using the Milestone Activity Inventory (MAI) to meet their self-identified needs and plans for becoming more integrated. The practice site and CMHC program experiences and lessons learned from that approach led to the evolution of the MAI to become the Milestone Attestation Checklist (MAC) for the CMHC programs and cohorts 2 and 3. With the change to the MAC, building blocks were revised and practice sites and CMHCs were given requirements for milestone activities. As part of that process, Building Block 5, Patient-Team Partnership, remained in the MAC as an optional focus, with revisions that retained questions about Patient and Family Advisory Councils (PFAC) but not about patient and family surveys. The Practice Monitor has been used to assess Building Block 5 Patient-Team Partnership activities.

The Practice Monitor employs seven items used to assess implementation of activities related to partnering with patients to include patient and family input in ongoing improvement activities such as through patient advisory groups and using patient surveys. Practice Monitor Building Block 5 items also ask about involving patients and families in the management of their care and health and linking families with community resources.

This section reports on the completeness of implementation of two of the items, items a. and b. in the table below, and reports on the composite percent of maximum possible implementation of all seven items combined. The remainder of this section presents percentage completeness of implementation for the two items and the composite scores at different assessment points.

Figure 12. Practice Monitor Building Block 5 Patient-Team Partnership Items

Practice Monitor Building Block 5 Items
a. A system has been implemented for including patient and family input in ongoing improvement activities (such as patient advisory groups or patients and family members on QI teams)
b. A patient experience survey is administered regularly (monthly or quarterly) and the data used to monitor and improve practice performance
c. Patients and families are actively linked with community resources to assist with their self-management goals.
d. Patients and families are provided with tools and resources to help them engage in the management of their health between office visits
e. Personalized shared care plans are developed collaboratively with patients and families
f. Personalized shared care plans are regularly reviewed to monitor patient progress in accomplishing their goals and adjusted when appropriate
g. Our practice has implemented and regularly uses shared decision-making tools or aids for at least two health conditions, decisions, or tests

Practice sites chose one of five ratings to describe themselves for each item, and the percentage choosing the highest rating “completely implemented” are shown in the tables. We coded sites that rated themselves as “completely” “yes,” and we coded the other sites “no.” We used χ^2 tests to analyze change over time. We analyzed change over time for the composite measure by using ANOVA F-tests and paired-comparison t-tests.

All cohorts showed statistically significant increases in the percentage of maximum possible completion of patient-team partnership activities. CMHCs also showed notable increases. However, because of the small numbers of CMHCs, we did not analyze those changes for statistical significance.

There were statistically significant changes in the percentage of practice sites that reported complete implementation of efforts to include patient and family input in ongoing improvement activities (item A) for cohorts 2 and 3. There were no statistically significant changes for cohort 1.

There were statistically significant changes in the percentage of practice sites that reported complete implementation of efforts to administer a patient experience survey regularly (monthly or quarterly) and use it to monitor and improve practice performance (item B) over time for all cohorts.

The increase in percentage of practice sites that “completely” implemented a PFAC varied between cohorts. The increase in practice sites for cohort 1 came between midpoint and final assessments. Cohort 2 increased slightly more from midpoint to the final assessment. Cohort 3 only had a baseline and a second or final assessment and increased almost as much during that time as cohort 2.

For item B, the percentage of practice sites implementing patient surveys increased more from baseline to the midpoint/final assessment. This was also the case for the percentage of maximum possible implementation for patient-team partnership overall composite measure. All cohorts improved on this measure from approximately 58% to 71% for cohort 1 to 89.9% for cohort 2.

The CMHCs also showed noticeable improvement in their implementation of activities to improve in patient-team partnership. Because of the small number of CMHCs, we did not conduct statistical significance tests.

Table 39. PCP Practice Monitor Building Block 5 – Items A, B, and Composite Scores

Practice Site. Practice Monitor Building Block 5. Baseline, Midpoint, and Final					
% Complete and Composite Score		**Baseline**	**Midpoint**	**Final**	**P-value**
Cohort – Item a. PFAC	N of Sites ↓	%	%	%	χ^2 test
Cohort 1	92	23.9%	23.9%	34.8%	.163
Cohort 2	145	17.9%	32.4%	53.1%	<.001*
Cohort 3	83	14.5%		43.4%	<.001*
Cohort Item b. Survey	N of Sites ↓	%	%	%	χ^2 test
Cohort 1	92	38.0%	48.9%	56.5%	.042*
Cohort 2	145	35.9%	49.0%	53.1%	.009*
Cohort 3		27.7%		56.6%	<.001*
Cohort % of Maximum	N of Sites ↓	%	%	%	F test
Cohort 1	92	58.4%	66.4%	71.0%	<.001*
Cohort 2	145	57.8%	74.2%	89.9%	<.001*
Cohort 3	83	57.2%		79.2%	<.001*

*Indicates P value ≤0.05

Table 40. CMHC Practice Monitor Building Block 5 – Items A, B, and Composite Scores

CMHC. Practice Monitor Building Block 5. Baseline, Midpoint, and Final					
% Complete and Composite Score		**Baseline**	**Midpoint**	**Final**	**P-value**
Item a. PFAC	N of CMHCs ↓	%	%	%	χ^2, F test
Item a. PFAC	4	0.0%	50.0%	50.0%	NA
Item b. Survey	4	0.0%	25.0%	50.0%	NA
Composite Score	4	45.5%	83.9%	89.3%	NA

PT9. To what extent are patients in SIM practice sites and bi-directional programs satisfied with the experience of primary and behavioral healthcare?

Integration of physical and behavioral health treatment ensures that fewer people are lost in the process of referral to external services, health issues are identified earlier, interventions are initiated sooner, and overall care is better coordinated. The SIM Framework diagram in the "Introduction and Approach" section shows the levels of integration and how they are associated with SIM efforts. As illustrated in the diagram, integral to the SIM framework, as well as the Quadruple Aim, has been a focus on components related to improving patient experience of care.

The systematic coordination of physical and behavioral healthcare often enhances the patient experience by focusing on whole-person healthcare and by equipping practices with the capacity to connect patients to the appropriate services based on their broad spectrum of needs. As highlighted in the operational plan, from the outset of the SIM initiative, there was a recognition of the essential role that patients play in healthcare reform. By engaging patients as partners in their healthcare, patients will feel empowered to participate in decision making, which links to improved patient experience and better health outcomes. As patients are more informed about their holistic healthcare needs and related services, they are better positioned to seek access to appropriate, timely, and needed care. For example, one practice related a patient experience story involving a patient receiving care for diabetes. During this treatment, a newly established screening process in the clinics helped to diagnosis the patient with depression. The clinical team was able to then help this patient with diabetic medications, counseling for depression, a food source, and financial assistance for his medication.

Building Block 5, Patient-Team Partnership, has milestones and activities that are relationship-based to partner with and support patients and their families. To implement Building Block 5 milestones, practice sites and CMHCs must gain an understanding of both patient and family perspectives on the care experience. Advanced efforts in this building block build on the patient-team partnership to help shape practice site and CMHC operations by soliciting and acting on feedback from surveys and by engaging patients and families in advisory roles to set and address clinic improvement priorities. Gathering this feedback and establishing ways to communicate information back to the patients and families are goals of Building Block 5.

TriWest asked cohort 1 SIM practice sites that did conduct surveys to share their aggregated findings. In total, 15 practice sites responded to requests and sent data, and 15 practice sites did not respond to our requests. One site responded to inform us it had not created a patient survey. Some sites informed us that their patient surveys had been conducted either by a larger organization (such as a health system) or by their Practice Transformation Organizations. Of the practice sites that sent us data, reporting formats varied, which made aggregating results difficult.

We received a copy of the survey tool from only one site; however, in most cases the summative data indicated which survey instrument was used by the practice site. Those included the Woodward Patient Satisfaction Survey (five sites), the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) (three sites), and CONNECT by the National Research Corporation (two sites). Two sites purchased an automated intake platform to engage their patients. This online application, Phreesia, has a patient survey component. Three other sites appear to have created an individualized survey tool.

We requested survey data for 2016, 2017, and 2018 with the intent to analyze results for significant changes over time. However, most sites submitted data from only one year. These years spanned from 2013 to current quarter of 2018. Four practice sites provided data from two years (2017 and 2018). The total number of patients who responded to surveys varied widely among practice sites and years, from two respondents to 1,184 respondents.

The most common content areas for the data summaries were satisfaction with care, satisfaction with provider, and whether the respondent would recommend the practice site. Although there was commonality in content areas, the questions, scales, and response rates were different among the surveys, making aggregating data across the practice sites that submitted survey data difficult.

Summary of Cohort 1 Survey Results

As described above, differences across the practice site survey summaries precluded a summary of survey results across all practice sites. As a result, we summarize data below for two content areas and one year (2017). The content areas are the overall experience with the practice site (eight sites) or provider (two sites) and whether respondents would recommend the site. Across practice sites providing 2017 data, 10 of the 15 sites provided data specifically related to patients' overall experience, and 10 of the 15 provided data specifically on whether the patient would recommend the practice site. The other sites provided data either on earlier years, on 2018 (see below), or in a manner that was not comparable.

There were four or fewer sites providing 2018 data on those questions. Since the 2017 data are more prevalent, and because so few sites reported 2018 data, we report the 2017 data here. Reporting change over time did not seem valid using these data.

The overall experience rating represents either the percentage of "satisfied" responses or a combination of the two highest rating choices, "excellent or very good" and "great or good." The average across the 10 sites reporting this level of satisfaction was 82.8%. The question about recommending the practice to others was "yes" or "no." Across the 10 sites reporting, 89.6% responded, "Yes."

Considering the small number of data points, this summary does not include reporting by subgroups of practices.

CAHPS Survey Findings

Each year, HCPF administers the Child and Adult Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Patient-Centered Medical Home (PCMH) surveys in practice sites with a high proportion of Medicaid members. For 2018, HCPF worked with the SIM office to assist in gathering patient data from some SIM practice sites. HCPF sampled 20 cohort 1 SIM practice sites and reported a series of findings regarding patient satisfaction. This *2018 Colorado Patient-Centered Medical Home Survey Report*, issued by the Colorado Department of Health Care Policy and Financing,[34] provided summaries of two health status questions across 20 cohort 1 SIM practice sites (2,063 respondents): overall general health status and general overall mental or emotional health status. The pattern of health status ratings, particularly with more adults rating their physical health and mental health as fair or poor (41.8% and 35.5% respectively), provides direction for action for the second general conclusion above: that practice sites serving adults should especially be aware that individuals with poor health who may need the most care and services are less satisfied with their providers and the health system.

---

[34] https://www.colorado.gov/pacific/sites/default/files/2018COPCMHSurveyReport.pdf

Table 41. HCPF Satisfaction Survey

2018 HCPF Satisfaction Survey Summary (20 cohort 1 SIM practice sites; n=2,063)					
Health Status Questions	**Excellent**	**Very Good**	**Good**	**Fair**	**Poor**
Child Practice Sites					
Overall General Health Status	42.3%	32.2%	19.8%	5.3%	0.4%
Overall Mental Health Status	42.6%	29.2%	20.1%	6.9%	1.2%
Adult Practice Sites					
Overall General Health Status	6.3%	18.4%	33.4%	30.2%	11.6%
Overall Mental Health Status	11.6%	20.5%	32.3%	26.8%	8.7%

This report provides the ratings for the “overall” survey items only—referred to in the HCPF report as “2018 Colorado SIM aggregate ratings.” The child Colorado SIM aggregate results were weighted using the child-eligible population for each practice site. The report compared the case-mix-adjusted SIM practice site results with the Colorado SIM aggregate to determine whether the SIM practice site results differed significantly statistically from the Colorado SIM aggregate[35].

We provide ratings for parents/caretakers in Table 42 below. Table 43 includes ratings for adult respondents. Each table shows SIM aggregate average scores for 2018. Provider ratings in the table below illustrate the “top-box” ratings summarized. Parents or caretakers of child members were asked to rate their child’s provider on a scale of 0 to 10, with 0 being the “worst possible” and 10 being the “best possible.” Top-box rates were reported and were defined as the percentage of responses with a rating of “9” or “10” on that rating scale. On scales with a “Never” to “Always” choice, top-box rates were reported as the percentage of responses of “Always.” The HCPF report provides ratings for each individual practice site and groups, with indications of practice sites that differed significantly from the overall.[36]

---

[35] The full survey methodology is explained in the full report.

[36] 2018 Colorado Patient-Centered Medical Home Survey Report. State of Colorado Department of Health Care Policy & Financing, August 2018. Denver, Colorado.

Table 42. Child Practice Sites Survey Summary (2018)[37]

Child Practice Sites Survey Summary	Percentage Rating 9 or 10
Global Ratings	
All Health Care *	75.9%
Provider *	74.7%
Specialist Seen Most Often	71.6%
Composite Ratings (of multiple questions)	
Getting Timely Appointments/Care/Information *	64.4%
How Well Providers Communicate with Child *	81.5%
How Well Providers Communicate with Parents/Caretakers *	81.5%
Office Staff Were Helpful, Courteous, and Respectful *	70.3%
Providers Discussed Child Development with Parents/Caretakers *	68.3%
Providers Discussed Child's Health Practices with Parents/Caretakers *	62.6%
Providers' Use of Information to Coordinate Patient Care	71.2%
Individual Items	
Received Care During Evenings, Weekends, or Holidays *	43.3%
Received Information on Evening, Weekend, or Holiday Care*	82.5%
Received Reminders about Child's Care from Provider Office *	69.6%
Saw Provider within 15 Minutes of Appointment	35.7%

The set of ratings for adult respondents includes two ratings of interest to SIM sites integrating physical and behavioral health: whether the provider office staff spoke with the patient about issues of stress or behavioral health issues and whether patients received health and mental healthcare at the same place.

Table 43. Adult Practice Sites Survey Summary (2018)[38]

Adult Practice Sites Survey Summary	Percentage Rating 9 or 10
Global Ratings	
All Health Care	63.3%
Health Plan (Medicaid/Health First Colorado)	60.8%
Provider	66.5%
Specialist Seen Most Often	66.3%

---

[37] Colorado SIM aggregate rating values are shown in the table. * indicates statistically significant differences exist between 2018 individual practice sites and the Colorado SIM Aggregate.

[38] * indicates statistically significant differences between 2018 individual practice sites.

Adult Practice Sites Survey Summary	**Percentage Rating 9 or 10**
Composite Ratings (of multiple questions)	
Customer Service	62.6%
Getting Timely Appointments/Care/Information *	47.4%
How Well Providers Communicate wWith Patients	74.9%
Office Spoke Withwith Patient About Stress or Behavioral Health Issues *	52.8%
Office Staff Were Helpful, Courteous, and Respectful	67.4%
Providers' Use of Information to Coordinate Patient Care *	62.8%
Talking wWith You About Taking Care of Your Own Health *	54.1%
Individual Items	
Received Care During Evenings, Weekends, or Holidays *	23.4%
Received Health and Mental Health Care at Same Place *	72.9%
Received Reminders About Care from Provider Office	77.4%
Saw Provider Within 15 Minutes of Appointment *	34.4%

The 2018 Colorado Patient-Centered Medical Home Survey Report, issued by the Colorado Department of Health Care Policy & Financing,[39] contained six recommendations for the SIM practice sites participating in the CAHPS (these are quoted directly from the report):

- "Respondents reported that when they contacted their provider's office during regular office hours, they did not receive an answer to their medical question within the same day. This indicates an opportunity for improvement in communication skills and timely access to care for the Colorado SIM practices."
- "Respondents reported that when their provider ordered a blood test, x-ray, or other test, no one from their provider's office followed up to give them the results. This indicates an opportunity for improvement in communication skills for the child Colorado SIM practices."
- "Respondents reported that clerks and receptionists at their child's provider's office were not as helpful as they thought they should be. This indicates an opportunity for improvement in communication skills for the child Colorado SIM practices."
- "Respondents reported that their child's provider did not always seem informed and up-to-date about the care their child received from specialists. Also, respondents reported that their child's provider did not seem to know important information

---

[39] https://www.colorado.gov/pacific/sites/default/files/2018COPCMHSurveyReport.pdf

about their child's medical history. This indicates an opportunity for improvement in coordination of care for the child Colorado SIM practices."

- "Respondents reported that they were not able to get the care they needed from their provider's office during evenings, weekends, or holidays. Also, respondents reported that when they needed care right away, they did not obtain an appointment with their provider as soon as they thought they needed. These indicate an opportunity for improvement in access and timely access to care for the adult Colorado SIM practices."
- "Respondents reported that they had not spoken with anyone from their provider's office about prescription medicines they were taking. Also, Respondents reported that information in written materials or on the Internet about how the Medicaid/Health First Colorado program works did not provide them with the information they needed. This indicates an opportunity for improvement in communication skills and access to information for the adult Colorado SIM practices."

Small Grants Program Outcomes

SIM created a competitive small grants process as part of the practice transformation efforts. During the SIM initiative a total of $3 million in funding was make available from Colorado Health Foundation and the Center for Medicare and Medicaid Innovation (CMMI) to advance behavioral health integration. In total the SIM office made awards to 108 practices in cohorts 1, 2, and 3. The SIM program implementation manager and SIM small grants administrator met regularly with funders to ensure that the evolution of the small grants program aligned with the strategic goals. This small grant program for practices was to help integration efforts across the state. Each practice was eligible for up to $40,000. The funds were meant to be foundational and help behavioral health integration efforts in ways that the practice would not have had the means to implement.

Impact on Patients

In the annual small grant reports for cohort 2 and cohort 3, grantees were asked how the small grant funded activities had "an impact on patients." Most grantees were able to provide an example, if not several patient stories. The few grantees who were not able to report a patient impact story indicated that there had been a delay in implementing their activities and it was too early to collect patient outcomes and stories. The following are some of the patient impact stories provided by grantees for more patient experience stories and to hear the patient voices from interview podcasts see the data hub.

“… a patient came into the dental office in [name of town] in pain and distraught. Her tooth pain was an acute issue needing attention, but our dental assistant soon discovered this patient’s husband had committed suicide just days before and she was still trying to cope with the loss. Our behavioral health specialist, [name of BHP], was able to talk to the patient in the dental chair and let her share her story, which then allowed us to finish her appointment and relieve her physical pain.” (Cohort 2 grantee)

“We had a 13yr old who was struggling with obesity and anxiety. We had tried many times in the past to help with both these issues with no success. Once we had [name of the BHP] on our staff to assist in this patient's appointments, the patient successfully lost over 30lbs, began treatment for anxiety and was able to return to school and social life.” (Cohort 2 grantee)

“…a 17-year-old high school student began receiving integrated behavioral service after her Wellness exam with her PCP when she screened positive on her PHQ-9/GAD at our office. [patient name] experienced untreated symptoms of depression and anxiety for most of her junior year in high school. But she avoided hospitalization for her symptoms by engaging in services with our therapist. The therapist met her weekly to discuss her symptoms and to help her build essential skills. She learned coping and communication skills to help improve her relationship with her mother and to effectively manage her symptoms of depression and anxiety. [patient name] had tried to gain access to our local Mental Health Center but was unable to afford the weekly sessions. After notable progress and a few set-backs, [patient name] describes herself as a stable and accomplished person who graduated from high school in May 2019 and plans to start college in the fall. She attributes her successes to her therapist who worked with her through her symptoms. ‘He helped me with my battle of depression and has shaped me into a better person with the help of his counseling - I am so grateful and don't know what I would have done without him.’” (Cohort 2 grantee)

“We had a diabetic patient who was diagnosed with depression and multiple concerns using the PAID survey. We were able to start medication, ordered counseling, found a food source for him, financial assistance for his medication and sources of transportation for him to use. He recently screened as having his depression controlled and his prior concerns had resolved with the resources he had been provided.” (Cohort 3 grantee)

“As an example of how integrated behavioral health can support patients beyond the clinic walls: One patient at [clinic name] reported a recent fall by her husband, resulting in a head injury. The BHP discussed home-based health with the patient and the potential benefit of home modifications to prevent future accidents. However, such modifications were unlikely to be reimbursed by the patient’s insurance. In response, the BHP contacted one of [clinic name’s] care coordinators to request follow-up on attaining the necessary resources. This is not a typical therapy session but illustrates how BHPs can provide patients with tools to examine their own situations and seek out resources that they did not know of before.” (Cohort 3 grantee)

“X and Y are a couple in my practice who have had a tumultuous relationship for many years. Recently, I received a copy of an ER report, indicating that there had been a domestic altercation, and Y had punched X forcefully enough to result in some (non-critical) physical injuries. (Legal and safety issues for all parties were resolved appropriately.) X and Y then each came in for medical appointment, separate from one another, but both within about a week of the incident. X initially insisted they were fine, and just wanted a recheck on the physical injuries. Y however was more immediately shaken up by the incident and was very interested in meeting with a counselor; after a warm hand-off, Y has continued in weekly therapy and is starting to practice new, healthier coping skills. Another couple of weeks later, X then reached back out, now also interested in counseling. At this point, I was very grateful for the grant, and specifically the flexibility our SIM grant had supported, in allowing [practice name] to partner with a small independent Behavioral Health group practice rather than hiring one individual. For a single individual therapist, treating both members of this couple separately could have been an ethical dilemma. However, with a partner organization, I was simply able to schedule X on a different day and facilitate another warm handoff to a different therapist who was also a good fit for X, and now both parties are engaging in appropriate, conflict-free individual therapy as they navigate these challenges in their lives. Neither of these individual Patient Members would have been comfortable or very willing to see a therapist without my individual endorsement and a warm hand-off, and the fact that I was able to arrange for both of their needs to be met separately and without an ethical conundrum, all within the walls of what is still a small solo family physician's practice has been absolutely amazing!” (Provider from a Cohort 3 Grantee)

Summary and Conclusions

- **We presented key findings at the start of this chapter that detail significant improvements in integration across all three primary care practice cohorts and the CMHCs.** While relatively few other results of our analyses proved to be statistically significant, these findings or practices’ progress are not meaningless. Practice transformation efforts are immensely complex and dynamic initiatives that rely on wide swaths of stakeholder engagement and participation to enact measurable changes, many of which are not likely to be observable in the short term. Even large-scale projects require time that often exceeds the length of the program’s activities to create changes that may be captured and quantified statistically. SIM is one of these projects.

- We identified statistically significant differences across and within the primary care practice cohorts, and these findings may guide future evaluation efforts and contribute to the growing body of knowledge and literature around integrated care and best practices. For example, **practices showed significant growth in their movement toward greater integration and achievement of the building blocks for advanced primary care.**

- Some findings were not statistically significant, but they provide valuable insight into SIM progress. **In aggregate, all cohorts and practice sites improved over the course of SIM participation as evidenced by continued growth and achievement on the various practice assessments: the Practice Monitor, milestone achievements, and provider and staff feedback surveys.** This is a valuable foundation of good habits as practices continue to develop operations that will support further and additional efforts to offer high-quality, integrated care across Colorado.

Lessons Learned from Small Grants Practice Sites

Better able to provide integrated care when seed monies are available. Small grant recipients provided numerous examples of how the funding had supported integration efforts. Examples of the kinds of support included recruiting and hiring onsite BHPs; contracting with an external BHP using care contracts; creating new referral sources; increased screening with new technologies; providing integrated care trainings and education for staff; and helping with stigma reduction due to the convenience and comfort of warm handoffs – as one grantee said, “In their mind they are walking into primary care that offers counseling.”

Change to process, workflows and culture take time and resources. When considering the challenges experienced while implementing small grant activities, grantees reported delays. This was a prominent theme among many grantees. Although progress and successes were occurring, it took longer than anticipated to work out new processes, complete renovations, or to engage all staff in the culture change. Several grantees suggested that knowing this is a good lesson to share with others about to embark on practice transformation efforts.

Capital expenditures laid the foundation for ongoing work. Small grants that funded office renovations and technology solutions in particular were seen as foundational. These one-time capital expenditures allowed practices to initiate new activities that they might not otherwise have been able to begin. Whether it was a new online screening software, iPads for patient use, or a remodeled space for the onsite BHP, practices reported small grants paved the way and sustainability could be managed by practices.

Data-driven change requires data. Although this may seem obvious, until a practice had gone through the process of determining what data they needed to have – if it was available, if it is valid, if they could analyze it to craft a story, then sharing those stories with practice staff; was not intuitive. Practice sites who implemented a PDSA process noted how good data can power change and increase quality care:

"…[data] helped us understand the volume of depression with comorbid physical conditions in the clinic which has allowed us to make numbers-based decisions in regard to staffing and new program launches." (Cohort 1 Practice, Small Grant Recipient)

Demonstrating both a financial and non-financial ROI. Many grantees shared success stories, and these are both monetary and non-monetary in nature. Some sites reported financial gains as they were better able to utilize the per-member, per-month (PMPM) reimbursements, other found new behavioral health revenue streams. Most of the return on investment were non-financial and included improved patient satisfaction, successful implementation of team-based care, increased access to behavioral health services, and greater awareness of and commitment to integrated care services.

"This project has been extraordinary. We have integrated an LPC [licensed professional counselor] into out practice and are able to provide our patients with quick access to behavioral health treatment as well as to track their progress with an electronic screening that is also integrated in our EHR. We use a shared EHR and have direct access to the notes of our LPC and as he is on-site, we are able to discuss our patients and their needs in real time. We have been able to prevent multiple ER visits due to his onsite presence and have been able to watch many patients respond very positively to our co management." (Cohort 2 Practice, Small Grant Recipient)

3 Practice Transformation: Access to Care

Introduction

As discussed in the Introduction to this report, the overall SIM goal contained two primary components:

- Access to comprehensive care that integrates physical and behavioral health
- Using increasingly value-based payment models

SIM's goals aligned with the "Quadruple Aim"[40] of:

- Better experience of care (including increasing access to integrated physical and behavioral healthcare services)
- Lower costs
- Improved population health
- Reduced provider burnout

The Agency for Healthcare Research and Quality (AHRQ) defines "access to care" as "the timely use of personal health services to achieve the best health outcomes."[41] In explaining this definition, AHRQ lists four components:

- **Coverage**: Facilitates entry into the healthcare system. Uninsured people are less likely to receive medical care and more likely to have poor health status.
- **Services**: Having a usual source of care is associated with adults receiving recommended screening and prevention services.
- **Timeliness**: Ability to provide healthcare when the need is recognized.
- **Workforce**: Capable, qualified, culturally competent providers.

Access to care is part of both a better patient experience of care and links to overall improvements in population health. The SIM driver diagram mapped a path whereby SIM activities supported "access to integrated physical and behavioral healthcare services…" and "remove[d] barriers to accessing care." These barriers to care can include availability and quality of outpatient care, cost, stigma, and workforce issues, including provider burnout, which impact availability and quality.

---

[40] The SIM began with a focus on the "Triple Aim" of lower costs, better care, and better patient experience, then elected to add a focus on workforce during its initial planning year.

[41] https://www.ahrq.gov/research/findings/nhqrdr/chartbooks/access/elements.html

To measure changes in access to care, we have taken both direct and indirect approaches. For direct measurement, we report on the use of screening and prevention services in the Clinical Outcomes chapter. In the Cost and Utilization chapter, we report on changing use of outpatient follow up after psychiatric hospitalization. As described in AHRQ's definition of access to care, increasing use of screening and follow up corresponds to increased access to care. The chapter on workforce describes provider satisfaction and burnout. We summarize these direct measurements in the conclusion of this chapter. Given the nature of the data available to us, we are unable to directly measure for patients attributed to practice sites participating in SIM any changes in insurance coverage, timeliness, or any location specific factors such as accessibility of clinics to public transportation or weekend and evening hours, all of which affect access to primary care.

Although the SIM office and evaluation work committee initially explored several approaches to measuring changes in access, the SIM office selected four indirect measures for use with APCD data. These are AHRQ prevention quality indicators that count inpatient hospital admissions that are avoidable with good access to primary and specialty outpatient care. Reductions in these types of admissions generally correspond to improvements in access to care. Change in these measures is the focus of this chapter. We describe each measure and their advantages and limitations below.

Improving Access to Care Through Practice Transformation and Public Health Drivers

This SIM initiative influenced access to care primarily through practices' transformation efforts at primary care practice sites and CMHCs. SIM practice transformation assistance included supporting 319 primary care practice sites and four community mental health centers (CMHCs) to advance physical and behavioral healthcare integration within their sites. SIM facilitated this work by funding practice coaching, disseminating achievement-based payments and small grants to some primary care practice sites, addressing workforce pipeline issues, investing in health information technology, payment reform and identifying and working to address barriers to integration, including information-sharing policies and regulations.

We encourage readers to see the Practice Transformation chapter in the Final SIM Process Evaluation Report for details regarding SIM efforts to support physical and behavioral health integration in participating practice sites and CMHCs.

As mentioned, access to care also impacts population health. In addition to practice site and CMHC efforts, SIM also addressed access to care through work in its population health primary driver. This included funding for Local Public Health Agencies (LPHAs) and Behavioral Health Transformation Collaboratives (BHTCs), as well as Regional Health Connectors (RHCs), to target specific access barriers, including stigma reduction, workforce development, and awareness about the importance of comprehensive and integrated physical and behavioral healthcare. We encourage readers to review the Population Health chapter of the Final SIM Process Evaluation Report for details regarding these efforts.

Evaluation Questions

SIM office staff and workgroup members collaborated with us (TriWest) to develop and refine the following evaluation question. The question examines the degree to which SIM activities contributed to increased access to care with the 319 primary care practice sites and four CMHCs over the course of the initiative:

PT4. Do patients attributed to SIM-participating practices have better access to primary care relative to patients attributed to comparison practices? Better access to behavioral healthcare relative to patients attributed to comparison practices?

SIM staff and workgroup members worked with us to formulate an additional evaluation question specifically around barriers and access to care:

PT3.1. Did communities with better coordination and alignment to SIM goals experience improved access to care and/or improved health outcomes (compared to those with less coordination and alignment)?

This second question proved difficult to answer because the data available to link specific SIM activities and efforts to these outcomes were too limited to make definitive connections. We worked closely with the SIM office, and the Population Health and Evaluation workgroups, to attempt to create a coordinated community systems index (CCSI) that would help to assign quantitative assessments of the levels of coordination and alignment within each Colorado Health Statistics Region (HSR). These quantitative metrics could then be used as independent variables in examining pre/post differences in access to care measures in HSRs throughout the state. However, all coordination/alignment measures were either very qualitative or subjective in nature (e.g., survey or interview results with RHCs or LPHAs) or so closely related to the access to care outcomes that the workgroups ultimately could not endorse the CCSI.

However, we can discuss changes in access to care across Colorado that occurred before and during SIM implementation. We outline these changes more fully in the Methods section below.

Access to Care Measures Data and Methods

We measured the following two questions: “Do patients attributed to SIM-participating practices have better access to primary care relative to patients attributed to comparison practices?” and “Do these patients have better access to behavioral healthcare relative to patients attributed to comparison practices?” To respond to these questions, we need measures of access to care for SIM patients. The four components of access to care (coverage, services, timeliness, workforce) described previously were not easily measured for a program that involved hundreds of primary care practice sites and multiple CMHCs. Although the chapters on clinical outcomes, cost and utilization, and workforce do contain some direct measurement of these components, to comprehensively measuring changing access, the SIM office selected four indirect measures in its initial operations plan, all of which could be calculated with claims data.

These measures were based on inpatient admissions for conditions that could be prevented through appropriate access to outpatient primary and specialty care. A decrease in these potentially avoidable admissions would correspond to improvements in access to care. The selected measures include three adult prevention quality indicators (PQI) and one pediatric quality indicator (PDI). AHRQ provides this description of the PQIs and PDI:[42]

- **Prevention Quality Indicators (PQI)**. Indicators representing hospital admission rates for common ambulatory care-sensitive conditions. Hospitalization for these types of conditions can often be avoided with appropriate use of high quality, community-based primary care services.
- **Pediatric Quality Indicators (PDI)** can provide a check on children’s primary care access or outpatient services in a community by using patient data found in a typical hospital discharge abstract or dataset.

As shown in the list below, these composite indicators do not directly measure access to care or quality of care. They, however, comprise the types of inpatient hospital admissions that often occur when a patient may not be receiving quality primary care. As such, lower incidence of these indicators likely provides a good way to determine whether an individual had access to high-quality primary care.

There are three adult indicators and one pediatric indicator. The adult indicators include one each for acute and chronic conditions and an overall composite combining the two. The acute and chronic measures are also composite measures, combining admissions based on different medical conditions. To see the link between an inpatient admission and access to primary care, we have listed the conditions that make up the composite measures for consideration of how

---

[42] https://www.qualityindicators.ahrq.gov/

they represent acute or chronic conditions that would lead to hospitalization less often with better primary care treatment.

Prevention Quality Acute Composite (PQI 91)

- Dehydration Admission (PQI 10)
- Bacterial Pneumonia Admission (PQI 11)
- Urinary Tract Infection Admission (PQI 12)

Prevention Quality Chronic Composite (PQI 92)

- Diabetes Short-Term Complications Admission (PQI 1)
- Diabetes Long-Term Complications Admission (PQI 3)
- Chronic Obstructive Pulmonary Disease (COPD) or Asthma in Older Adults (PQI 5)
- Hypertension Admission Rate (PQI 7)
- Heart failure Admission Rate (PQI 8)
- Uncontrolled Diabetes Admission Rate (PQI 14)
- Asthma in Younger Adults Admission Rate (PQI 15)
- Lower Extremity Amputation among Patient with Diabetes Rate (PQI 16)

Prevention Quality Overall Composite (PQI 90)

PQI 90 is a combination of all measures that make up PQIs 91 and 92. Because analysis of PQI 91 and 92 reveal the same relationships as the overall composite PQI 90, we only report PQI 90 results in this chapter.

- Prevention Quality Acute Composite (PQI 91)
- Prevention Quality Chronic Composite (PQI 92)

Pediatric Quality Overall Composite indicator (PDI 90)

- Asthma Admission (PDI 14)
- Diabetes Short-Term Complications Admission (PDI 15)
- Gastroenteritis Admission (PDI 16)
- Urinary Tract Infection Admission (PDI 18)

Each indicator has a numerator and denominator. The numerator is a count of inpatient admissions unique to each indicator. The denominator is the age-specific population that is potentially subject to each category of admission. The indicators are therefore interpreted as rates of inpatient admissions per relevant population. To make changes in these rates more apparent in time series graphs, we converted them to a rate per 100,000 people.

We calculate these indicators for two groups: SIM primary care practice-attributed patients (broken down by cohort 1 or cohort 2)[43] and matched patients attributed to non-SIM primary care National Provider Identifiers (NPIs). With this data structure, we were able to identify changes in these measures for patients as their practice sites started to participate in SIM, and we could compare any changes to matched patients at non-SIM practice sites.[44] This approach helped identify whether any observed change in these measures was associated with the SIM intervention—an association that would impact patients attributed to SIM practice sites only—or other factors unrelated to the intervention, which would impact both SIM- and non-SIM-attributed patients.

To perform these calculations, we used an AHRQ-supplied SQL program and APCD data. The program counts the number of qualifying inpatient hospital admissions for the SIM-practice-attributed or comparison patients. We used the number of patients in each category as a denominator to calculate a rate per patient and converted this to a per 100,000 for ease of interpretation.

An important limitation of these measures is that they do not adequately measure access to behavioral healthcare specifically. We address access to behavioral healthcare by analyzing use of behavioral health services in the chapter on cost and utilization.

These were not indicators intended to apply to the individual SIM primary care practice sites or CMHCs. Instead, they were intended as on overall SIM accountability metric used for CMMI reporting. For measures more closely related to the work of the practice sites, please see the clinical quality measures (CQMs) reported in the Clinical Outcomes chapter of this report

Access to Care Comparison: Patients Served in SIM Primary Care Practice Sites Versus. Comparison Patients

PQI 90 - Prevention Quality Overall Composite

PQI 90 combines the admission categories of PQIs 91 and 92; it is an overall composite. Because the chronic (PQI 92) admissions occur at two to three times the rate as the acute (PQI 91)

---

[43] Cohort 3 is not presented in this analysis because the data available for analysis includes claims from 2015 through June of 2018. Cohort 3 began SIM participation in the fall of 2018.

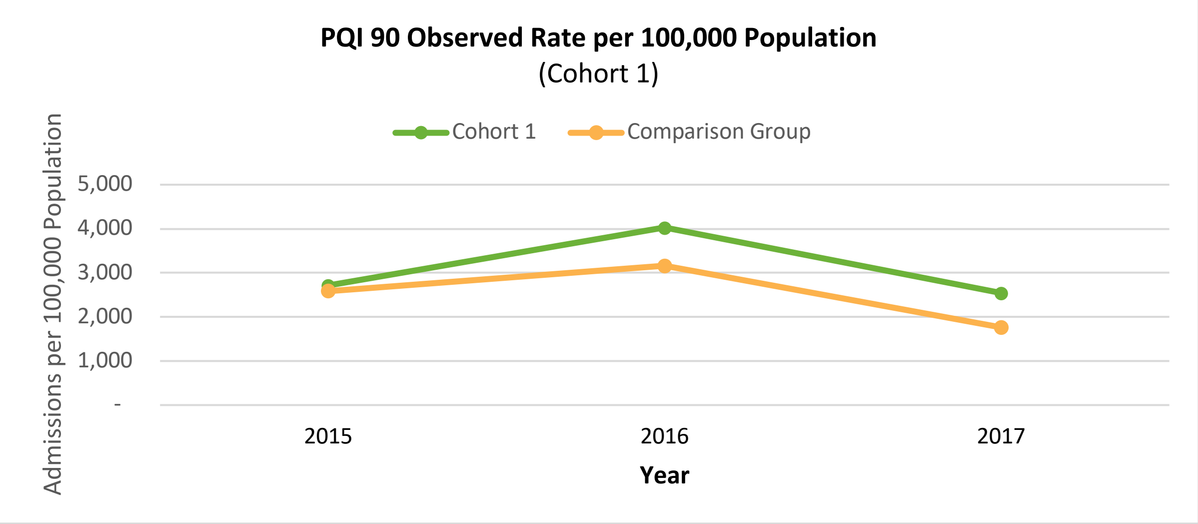
[44] The process used to develop a matched comparison group of patients is described in chapter 2 (Methodology).

admission, the chronic admission dominates the overall composite. Our analysis of PQI 91 and 92 show the same relationship to the SIM intervention as the overall composite, and we therefore only present analysis of PQI 90.

Table 44. PQI 90 Observed Rates (Cohort 1)

PQI 90 - Prevention Quality Overall Composite Observed Rate per 100,000 Population		
Year	**Cohort 1 Patient Observed Rate**	**Matched Comparison Patient Observed Rate**
2015	2,719	2,586
2016	4,029	3,162
2017	2,543	1,763

Figure 13. PQI 90 Observed Rate (Cohort 1)



Using the PQI 90 as a summative access to care indicator for adults, the time series graphs (Figure 13 and following) indicate that the matching process based on 2015 data, including Milliman Advanced Risk Adjusters (MARA) scores, resulted in a comparison group with very similar composite indicator values at baseline. Over the next two years of SIM participation, the patients attributed to cohort 1 practice sites developed slightly higher indicator values than did the comparison group, which corresponds to more inpatient admissions for the included preventable conditions.

Several reviewers of this analysis expressed concern that problems in Medicaid data caused the uptick in 2016. We divided these data by line of business, and all four lines (Medicaid, commercial, Medicare Advantage, and Medicare fee-for-service) experienced increases from 2015 to 2016. Medicaid experienced a large decline in 2017 unmatched by the other lines of business, and Medicaid is the component driving the aggregate 2017 decline. It would therefore be more accurate to describe Medicaid's data as causing a 2017 decline than describing, instead, the Medicaid data causing a 2016 increase.

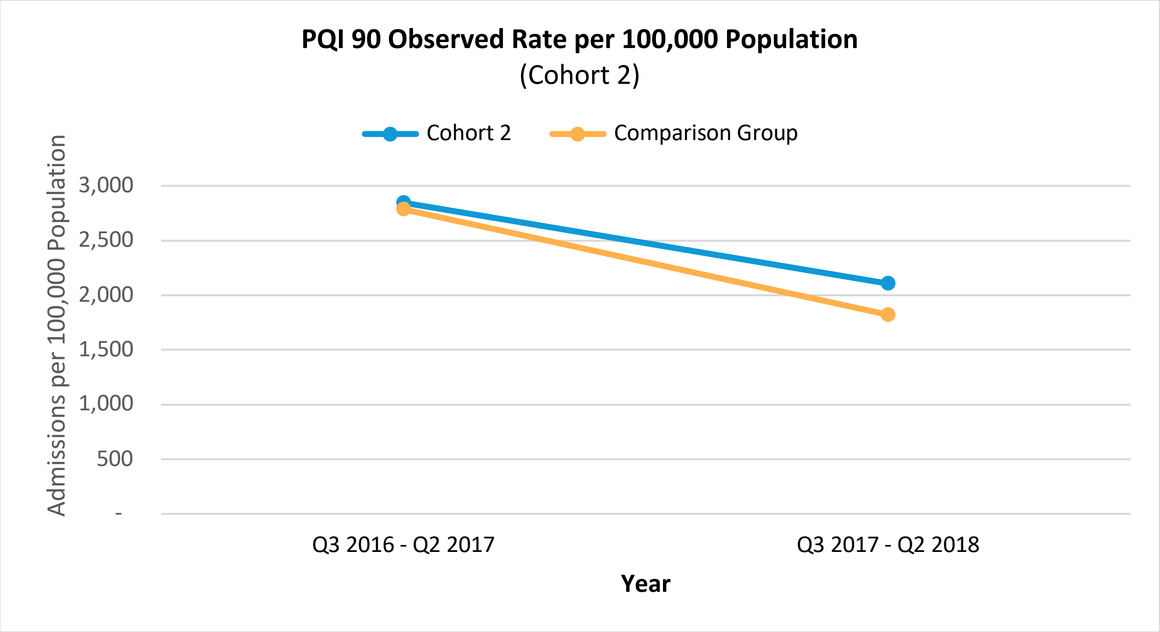
Of greater importance in understanding the impact of SIM on access to care, the 2016 increase is apparent in both the cohort 1 patients attributed to SIM practice sites and the comparison group. Because we constructed the comparison group to have an equal number of patients by payer type, any abnormality in the 2016 or 2017 Medicaid data affects both groups equally.

Based on the analysis of changes over time for each group, the access to care indicators for cohort 1 attributed primary care patients did not improve relative to the comparison group.

Table 45. PQI 90 Observed Rates (Cohort 2)

PQI - 90 Prevention Quality Overall Composite Observed Rate per 100,000 Population		
Year	**Cohort 2 Patient Observed Rate**	**Matched Comparison Patient Observed Rate**
Q3 2016–Q2 2017	2,846	2,785
Q3 2017–Q2 2018	2,106	1,821

Figure 14. PQI 90 Observed Rate (Cohort 2)



Cohort 2's PQI 90 values were well-matched to the comparison group at baseline, and the comparison group showed greater decline in inpatient admissions (Figure 14). Therefore, while the cohort 2 patients saw improvements in access to care as measured by the PQI 90 indicator, this analysis does not support the hypothesis that the improvement was a result of SIM participation.

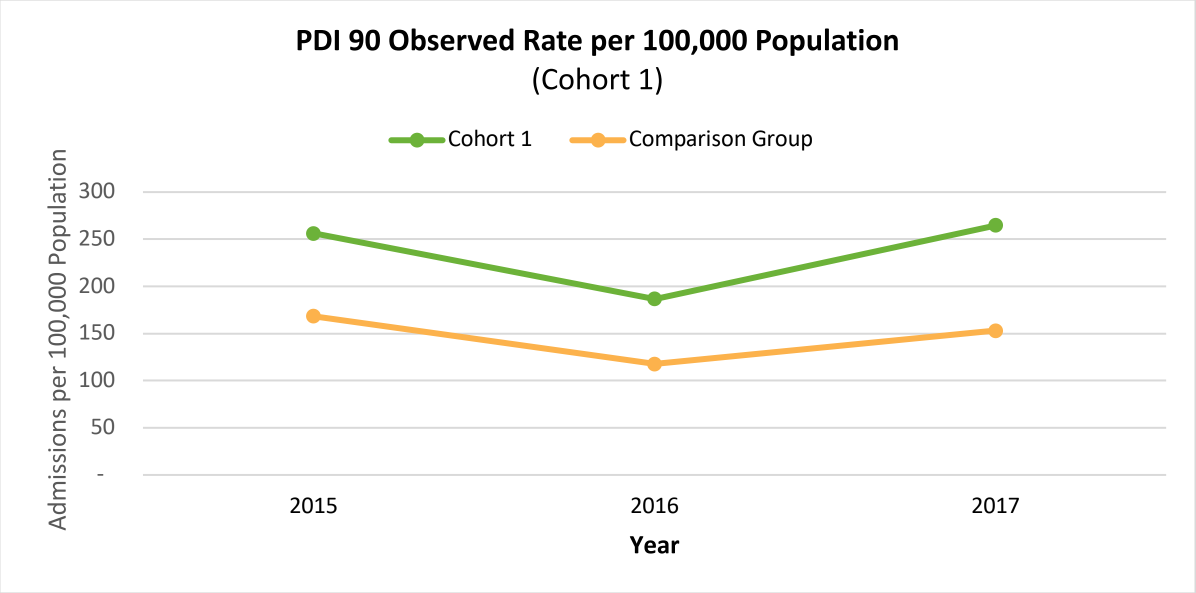
PDI 90 – Pediatric Quality Overall Composite

For children and youth, the SIM evaluation used a single AHRQ access to care indicator. The pediatric quality overall composite indicator (PDI 90) is an aggregate count of admissions for the population of 6 to 17 years of age for one or more of the conditions listed previously.

Table 46. PDI 90 Observed Rates (Cohort 1)

PDI 90 - Pediatric Quality Overall Composite Observed Rate per 100,000 Population		
Year	**Cohort 1 SIM Patient Observed Rate**	**Matched Comparison Patient Observed Rate**
2015	256	168
2016	187	118
2017	265	153

Figure 15. PDI 90 Observed Rates (Cohort 1)

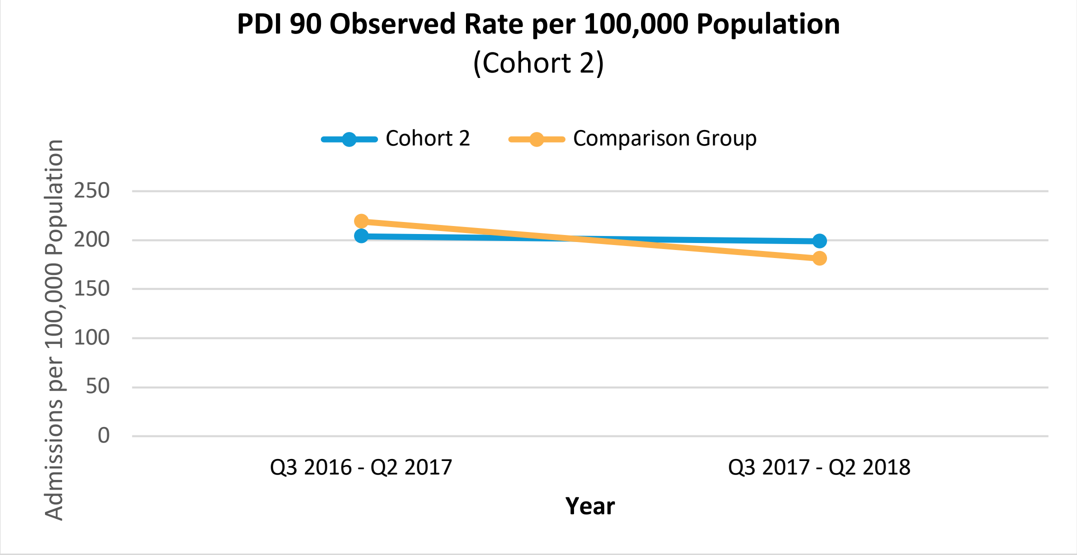


For cohort 1, matching in 2015 resulted in a comparison group with much lower indicator values (Figure 15). This gap did not narrow over the two years of cohort 1's participation. Because there was not any relative improvement, the SIM intervention did not appear to improve this indicator measure relative to the comparison group for cohort 1 pediatric patients.

Table 47. PDI 90 Observed Rates (Cohort 2)

PDI 90 - Pediatric Quality Overall Composite Observed Rate per 100,000 Population		
Year	Cohort 2 SIM Patient Observed Rate	Matched Comparison Patient Observed Rate
2017 Q3	204	219
2018 Q2	199	181

Figure 16. PDI 90 Observed Rates (Cohort 2)



For PDI 90, the cohort 2 patients and comparison group had similar baseline and Year 1 values, with almost no change from baseline (Figure 16). The conclusion for cohort 2 is the same as the other access to care results: relative to the comparison group, the SIM intervention did not appear to reduce the number of inpatient admissions that are used to calculate these indicators.

Summary for Adult and Pediatric Indicators from the SIM Cohort Analysis

The SIM Evaluation Workgroup chose these indirect indicators early on in the initiative, as explained at the start of this chapter, as a method of using available claims data to measure the possible effects of SIM practice transformation activities on access to care for patients attributed to primary care practices participating in SIM.

By using pre-SIM baseline year claims available in the APCD, we were able to report change over time in the AHRQ prevention quality indicators for the patients attributed to the SIM primary care practice sites. Adding the same change over time for the matched comparison group allowed us to identify whether any changes were likely to be associated with the SIM intervention, or if they were instead the result of forces affecting both SIM and non-SIM primary care practices with patients represented in the APCD.

The trend lines of the SIM and comparison observed rates in time series graphs, with data presented in the accompanying tables, led us to conclude that the SIM intervention did not result in improvements to the adult and pediatric indicators. The adult composite indicators showed some improvements after the start of the implementation, but improvements for the comparison groups were equally large or larger. The pediatric indicators showed increased admissions for cohort 1 after the start of SIM, but the same increase was observed in the comparison group. For cohort 2, the pediatric indicators showed almost no change during the first year of the intervention[45].

Colorado Health Access Survey (CHAS)[46]

In the introduction to this chapter, we discussed SIM population health efforts and a related evaluation question:

PT3.1. Did communities with better coordination and alignment to SIM goals experience improved access to care and/or improved health outcomes (compared to those with less coordination and alignment)?

Although we could not identify adequate data to link SIM efforts (specifically, "better coordination and alignment") to the "large P" Colorado population, the initiative's goal was to take steps toward improving population access to care. The SIM office and the Evaluation Workgroup identified the Colorado Health Access Survey (CHAS) as a mechanism for assessing access to care for Coloradans and worked with the Colorado Health Institute (CHI), who administers the survey, to add behavioral health related items. TriWest has pulled from those items, and others on the survey, to describe the current state of access to care in Colorado.

We did not expect that the SIM model test would be able to demonstrate significant impacts on these measures in such a short time period, particularly given that the most recent data available were from early in SIM (2017). However, the data below serve as a starting point for any efforts to assess whether Coloradans' access to care improved post-SIM, when 2019 data were available.[47]

---

[45] No statistical significance tests were performed because there was no actual evidence of any improvement for the SIM practice patients, much less whether improvement might be significant.

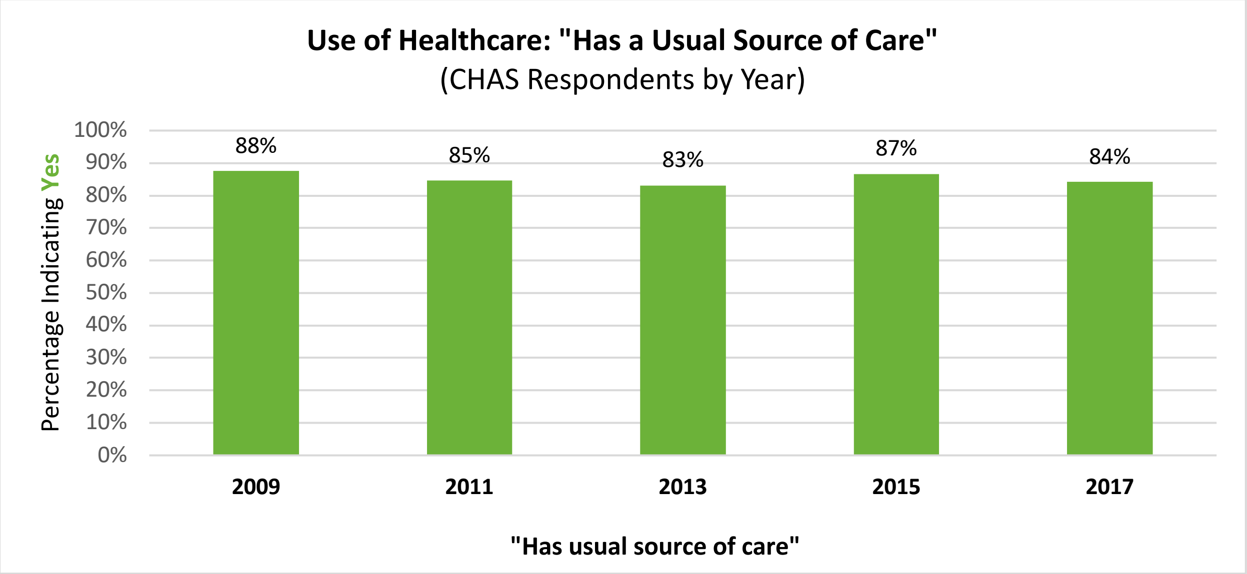
[46] Courtesy of the Colorado Health Institute, www.ColoradoHealthInstitute.org, 303 E. 17th Avenue, Suite 930, Denver, CO 80203. 7/1/2019.

[47] The CHAS is administered statewide every two years.

Access to Physical Healthcare in Colorado

As reported in the introduction section of this chapter, AHRQ listed reporting a "usual source of care" that would be associated with receiving the recommended screening and prevention services as one component of good access to care. The graph (Figure 17) below illustrates trends in Coloradans' reporting of this component between 2009 and 2017. As shown, the percentage of individuals reporting a usual source of care fluctuated slightly but remained consistent over those years.

Figure 17. CHAS Respondents: "Usual Source of Care" (2009–2017)



Timeliness of appointments (which relates to provider availability) and cost were two additional components. Again, the trends shown in Figure 18 and Figure 19 were generally stable over the latter three survey periods. A decline in these indicators (or an increase in the indicator above) in the 2019 data might suggest a possible correlation between SIM activities and these components of access to care. However, as stated previously, making any causal links was not possible, given that SIM was a statewide initiative and no comparison data were available.

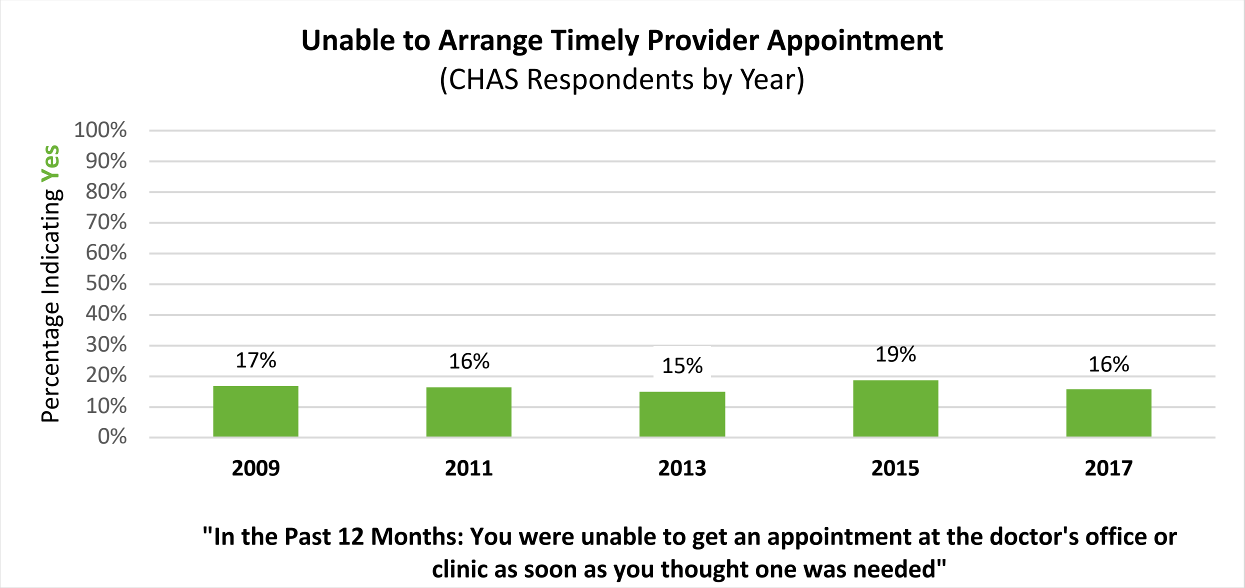
We carried out a two-sample test of proportions to estimate levels of statistical significance between two-assessment time periods for the trend shown by the graph above.[48] We

---

[48] Without access to raw CHAS survey data, we calculate statistical significance using total, weighted N size based on Colorado population parameters that TW analysts had access to. This creates large N sizes that often translate to unvarying levels of statistical significance. While this may seem problematic, it is important to note, as many public health researchers have noted, that small-percentage levels of change within state populations can often translate to large numbers of individuals being impacted by that percentage change. For instance, a 1.5% change in a population of 5 million is a small percentage but a large N. Regardless of this general observation, we address the issue of consistent statistical significance with large samples in several ways. First, for the analysis below, all levels

computed Z-scores to assess statistical significance in percentage change between the SIM baseline (2015) and 2017 time periods. Regarding responses to the access to care question shown by the graph, the 2017 score of 84% (z-score -108.98, p<.001) indicates that there is a statistically significant decline in the percentage of Coloradans who reported having a usual source of care.

Figure 18. CHAS Respondents: Ability to Arrange Timely Provider Appointment (2009–2017)

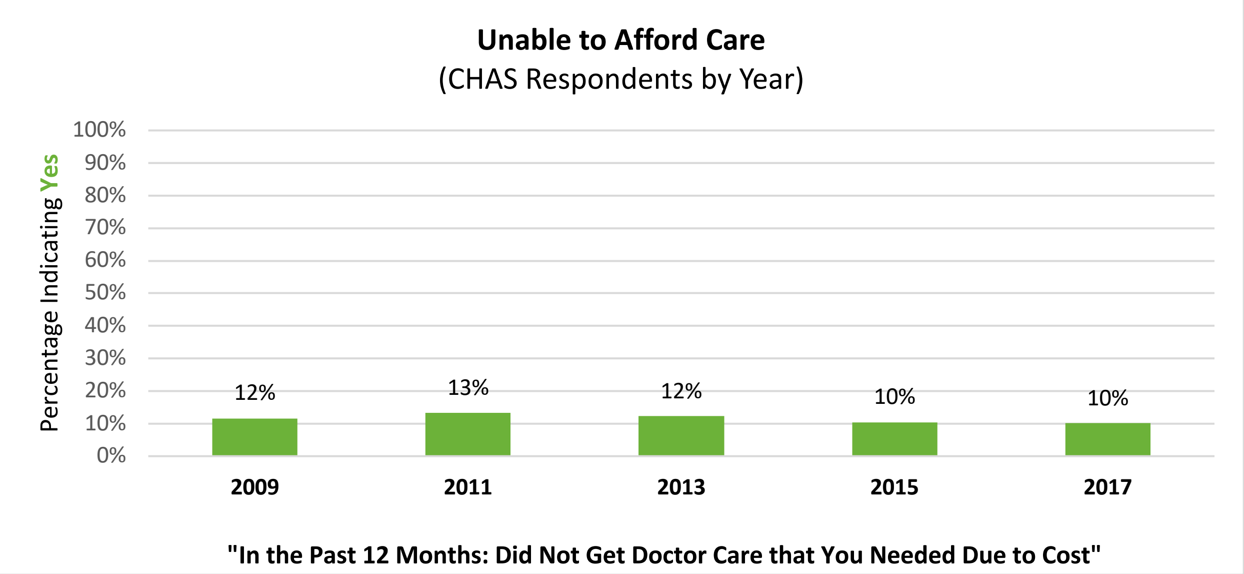


Between the SIM baseline year (2015) and 2017, we see a statistically significant decline in the percentage of Coloradans who reported that they were unable to arrange an appointment when one was needed (z-score -130.36, p<.001).

---

of significance are raised above the standard .95 level to .99. Next, we show z-scores to assess the magnitude and direction of change.

Figure 19. CHAS Respondents: Ability to Afford Care (2009–2017)

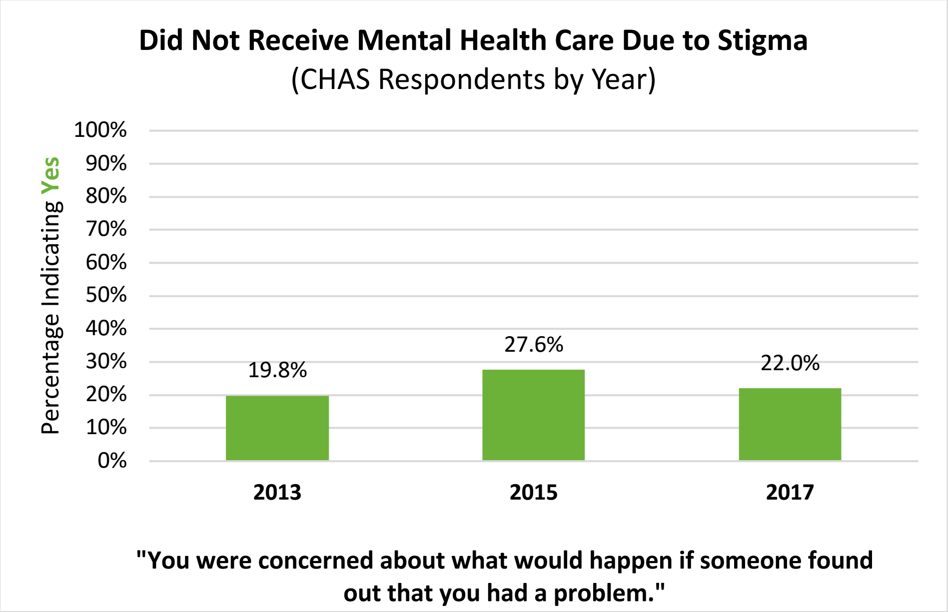


For the affordability measure, we again caution against over-generalizing based on statistical significance. Since the weighted data used for this analysis are based on the entire Colorado population, it is not surprising that statistical significance is once again reached. Although there is a statistically significant decline from 10.35% in 2015 to 10.13% in 2017 (z-score -11.83, p<.001) in the percentage of Coloradans who reported an inability receive medical care because of cost, for this large sample size, the size of the z-score is considerably smaller than those shown above. This indicates a modest percentage decline, yet one impacting relatively large numbers of Coloradans.

Stigma and Access to Behavioral Health Services

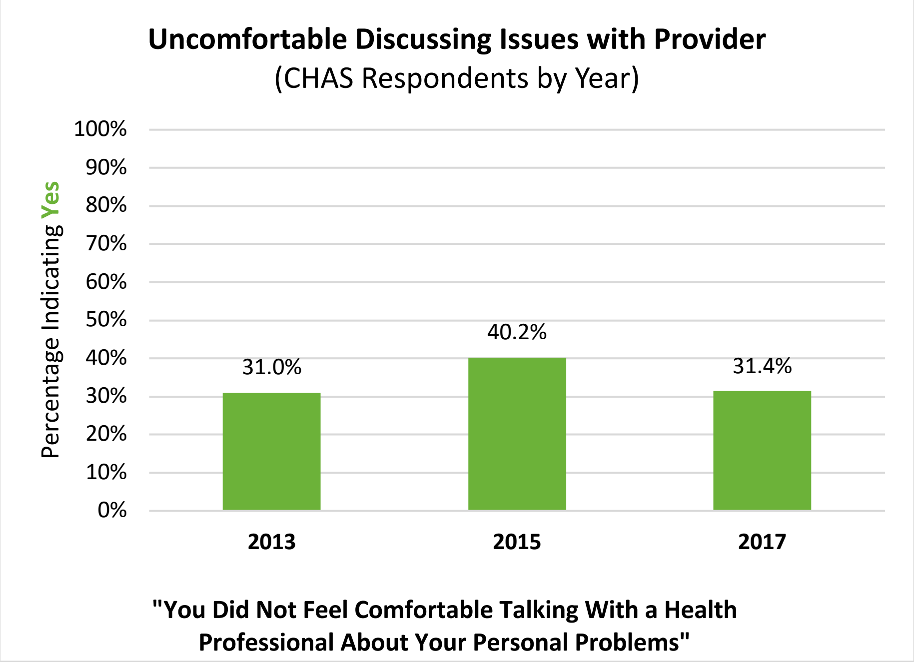
A distinct barrier to accessing care is stigma, or a fear of seeking out resources caused by anticipated shame associated with certain conditions, treatments, or services. Combatting this stigma has been a challenge, and we discuss stigma's deterrent effect on seeking services in the Population Health chapter of this report and in the SIM Final Evaluation Report. Practice sites, key informants, and workgroups have all noted its prevalence—particularly regarding behavioral healthcare and substance use disorders. The CHAS data are further evidence of this difficulty: approximately 20% of respondents over 2013–2017 indicated that they did not receive mental healthcare out of concern someone would learn of their problems (Figure 20). An even higher percentage—at least 30% in each survey period—reported they were not comfortable discussing personal problems with a healthcare provider (Figure 21). Because so many practice sites, LPHAs, and BHTCs opted to target reducing stigma and increasing behavioral health and substance use disorder services during their SIM participation, results from 2019 and beyond may be reviewed to better measure and understand longer-term SIM impacts on these challenges.

Figure 20. CHAS Respondents: Did Not Receive Care Because of Stigma



Unlike the generally modest level of change discussed above, there is a substantial and statistically significant change in levels of perceived stigma in healthcare settings between baseline and 2017. The graph below shows that the 2017 score of 22% (z-score -58.11, p<.001) corresponds to a considerable decline in the percentage of Coloradans who reported concerns over being stigmatized as a result of a mental health condition.

Figure 21. CHAS Respondents: Comfort Discussing Issues with Provider

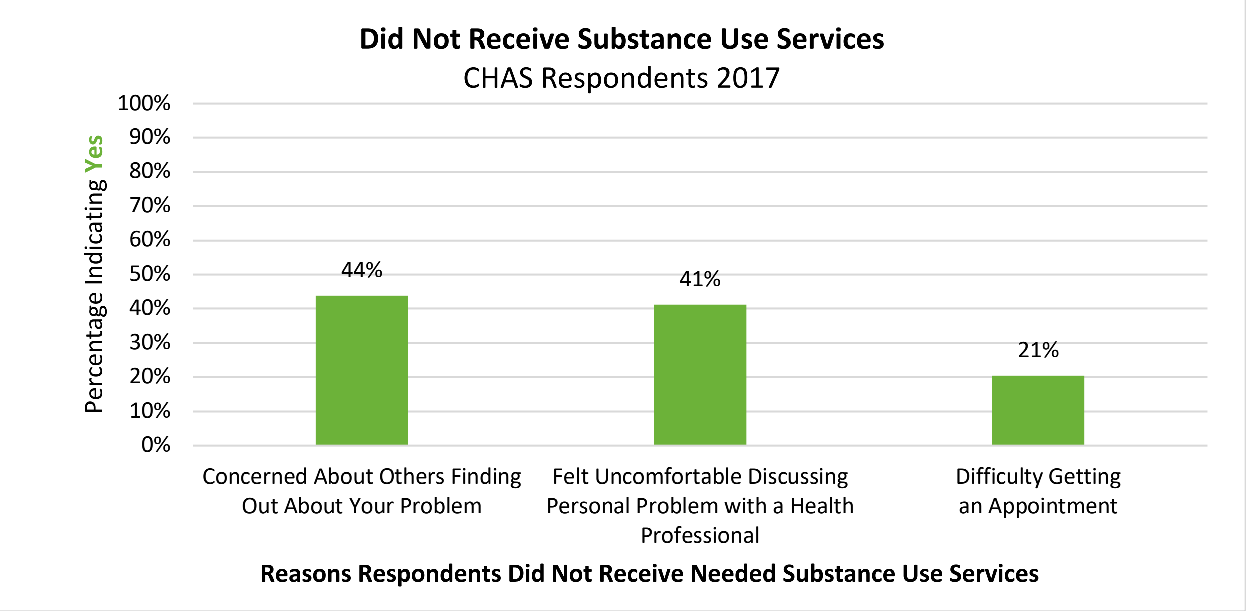


With funding from the SIM office, CHI added questions regarding substance use services in the 2017 CHAS questionnaire (Figure 22). These items reflected possible factors in Coloradans not seeking substance use treatment services. Approximately twice as many individuals (44%) reported that they did not receive substance use services because of stigma (i.e., concern over others finding out, being uncomfortable discussing substance use with their providers) than those who did not receive services because of difficulty getting an appointment (21%).

Similarly, concerning the level of comfort discussing personal problems with health professionals, the change in percentages over time can, once again, be interpreted as a sign of positive short-term trends in Colorado. The graph below shows that the 2017 score of 31.4% (z-score -82.54, p<.001) is a substantial drop in the percentage of Coloradans who were uncomfortable talking with a health professional about their personal problems.

Although we caution against making causal claims with aggregate data that lack a more extensive time series to assess longer-term trends, the two preceding graphs and results of two-sample tests of proportions, provide evidence of a reduction of feelings of stigma among Coloradans seeking behavioral health services. It is plausible that SIM activities, and the emphasis on behavioral health integration, made a positive impact on patient experiences while they are treated in behavioral health facilities. However, this data was collected early in the SIM effort, so it is unlikely that the changes observed are entirely due to Sim efforts.

Figure 22. CHAS Respondents: Reception of Substance Use Services (2017)



Effectiveness of the Colorado Healthcare System

In addition to the above responses, CHAS presented data on views of the Colorado health system. Below, we present the percentages of Coloradans responding “strongly agree” to the following statements:

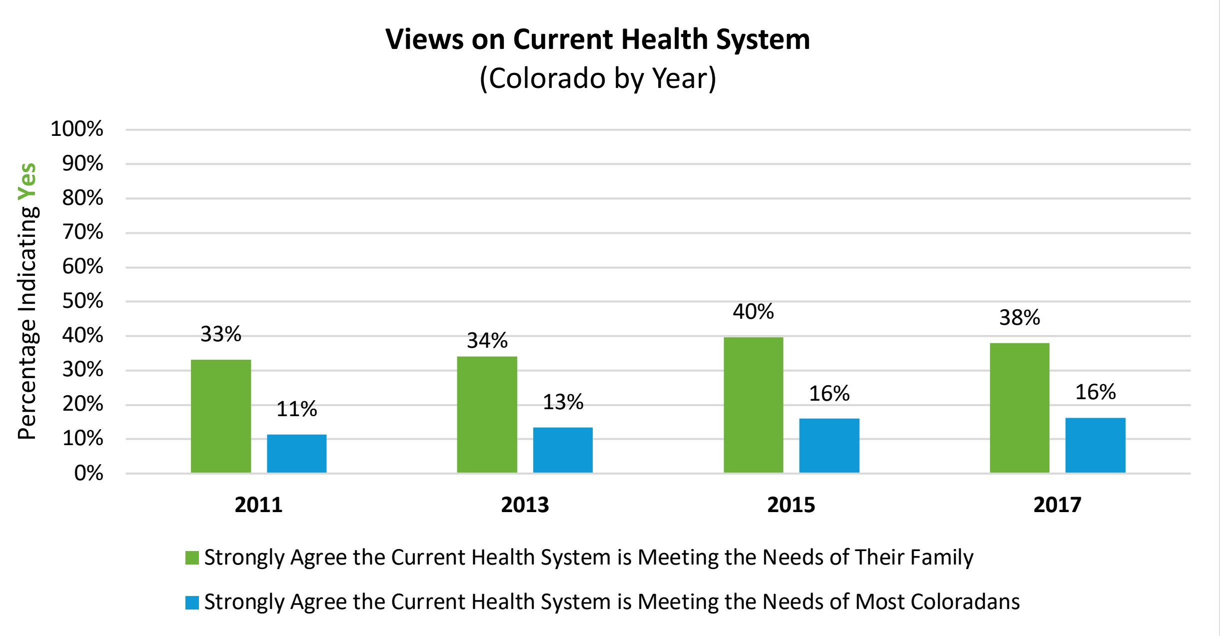
- “The current health system is meeting the needs of most Coloradans.”
- “The current health system is meeting the needs of their family.”

Regarding these topics, the considerably higher percentages of respondents who agree that the Colorado healthcare system meets the needs of their families—as opposed to the lower percentages that agree that this is the case for all Coloradans—stands out. This contrast indicates that Coloradans are generally satisfied with their own personal healthcare access, but they simultaneously consider that the current system does not provide comparable access to all Colorado residents.

Second, the graph below (Figure 23) indicates slight overall improvement on both measures between 2011 and 2017. However, the increases are uneven and these results are an insufficient source for forming conclusions regarding significant improvement or long-term change.

Once again, we observed no real changes between the SIM baseline year and 2017. Once 2019 data are available, it may be possible to see more change after full SIM implementation, even if that change cannot be directly attributed to SIM efforts.

Figure 23. CHAS Respondents: Views on Current Health System



Once again, the large sample size leads us to focus on the size and direction of z-scores as opposed to solely relying on statistical significance when estimating from large samples. First, when considering differences in level of agreement that the current health system meets the needs of their family, the 2017 score of 38% (z-score -56.67, p<.001) indicates that there is a measurable decline from 40% in the proportion of Coloradans who agree that the health system meets their family's needs. Less pronounced is the difference from those responding that the healthcare system meets the needs of most Coloradans. The lower size of the z-score on this question for the 2017 percentage (z-score 6.57, p<.001) indicates a very slight increase (based on the positive z-score) but a change less that 1% between baseline and 2017.

Summary and Conclusions

Both SIM evaluation questions presented in this chapter asked about changes or improvements in access to care, either across Colorado ("large P" population) or for patients attributed to practices participating in SIM ("small p" population).

In analyzing changing access to care for patients attributed to primary care practices participating in SIM, we used both direct and indirect approaches. For direct measurement, we examined changes in screening associated with the "services" component of AHRQ definition of access to care. We drew these from the analysis in other chapters (see Clinical Outcomes), and they include statistically significant improvements in measures related to depression screening, adult obesity screening, and fall risk screening. Cohort 1 also had statistically significant improvements in *Hemoglobin A1c Control* and *Asthma Medication Management* measures. Cohort 2, likewise, showed an increase in the *Adolescent Obesity Screening* measure. However, the other screening measures did not show statistically significant change.

In addition to these screening services, we also are able to report on changing use of preventative services. As discussed more completely in the Cost and Utilization chapter, Milliman calculated one practice-site-level measure related to prevention services: the percentage of psychiatric admissions with a follow-up outpatient visit within 30 days. This is an outpatient service linked to access to behavioral healthcare. Unfortunately, it did not increase on average.

This same measure was calculated for patients attributed to primary care practices sites participating and not participating in SIM. The patients attributed to SIM sites showed decreasing follow up, whereas the comparison patients showed increasing follow up. The difference was not statistically significant.

We had limited capacity to measure other components of access, including coverage, timeliness, or workforce. Coverage references health insurance. Our data source, the APCD, includes claims by commercial insurance companies, Medicaid, and Medicare. Patients without any type of coverage are not included in our data, and we are unable to analyze changing insurance coverage. Timeliness is often analyzed through review of clinic hours, days until an appointment, and practices accepting new patients. With hundreds of primary care practice sites participating in SIM, we were unable to access information related to these measures.

Our one quantitative measure for access to care from a workforce (provider availability) perspective is based on a survey of SIM practice staff and clinicians regarding their work satisfaction. Increased burnout, had it occurred, would have hindered access to care. Results of these surveys, presented in the Practice Transformation: Integration Efforts chapter show, however, that SIM work in integrating physical and behavioral healthcare did not increase provider burnout, thus not impeding access to care. The same chapter documents substantial programmatic activities undertaken through SIM to improve the availability of qualified behavioral health providers. Despite these efforts, an insufficient number of qualified providers are available to meet the behavioral health needs of patients in primary care practices.

In addition to measuring access to care directly, we used four AHRQ indirect indicators. These are based on inpatient hospital admissions for conditions that could be treated successfully in primary care settings, such as dehydration or diabetes complications. These measures were selected by the Evaluation Workgroup during the SIM planning phase and were never intended for use at an individual practice level. Instead, we use them in this evaluation as a measure of the overall SIM impact on access to care. Comparing changes in this measure between patients served by SIM practice sites and a comparison group of patients in the state allows for an isolation of effects that can be directly attributed to SIM. However, SIM practices were never asked to monitor these metrics or to report them.

A decrease in these admissions corresponds to improvements in access to outpatient care. Analysis of the SIM and comparison group time series shows no evidence that the SIM intervention resulted in improvements to the adult and pediatric indicators for patients attributed to SIM-participating practice sites. For both cohorts, the adult indicator (PQI 90) showed improvement (i.e., declining admissions) after the start of the implementation, but improvements for the comparison group were equally large or larger. The pediatric indicator (PDI 90) showed increased admissions after the start of SIM for cohort 1, decreases for cohort 2, and decreases for both cohorts' comparison groups.

We saw few changes in the statewide measures of access to care (large P) as measured by the CHAS. These findings were expected based on (1) the results above showing that these measures did not seem to improve significantly for the SIM population, specifically, and (2) the fact that statewide data for all measures were only available through 2017, which was early in SIM implementation. Data points for 2019 and later will be better potential indicators of longer-term effects of full SIM implementation.

Additionally, CHAS data showed general patterns of stability over time. Large majorities of Coloradans reported positive mental and general health, whereas much lower percentages indicated a lack of access to mental health counseling services or experiences of discrimination in medical settings. Finally, based on evidence from CHAS data, views towards the Colorado health system are improving slightly over time. It is notable that the number of individuals reporting concerns about mental health stigma (as a barrier to treatment) declined significantly between 2015 and 2017. Although data are not available to directly attribute this change to SIM efforts, the finding is encouraging.

In summary, patients attributed to primary care practices participating in SIM experienced increasing screening associated with improvements in access to care. Statewide survey results do not demonstrate improvements in access attributable to SIM, and comparison group analysis does not show improvement for patients attributed to SIM practices in either follow up after psychiatric hospitalization or the AHRQ indicators that indirectly measure access to quality outpatient care.

4 Clinical Outcomes

Introduction

SIM identified primary drivers to further the Quadruple Aim of better care, lower costs, improved population health, and reduced provider burnout.[49] Clinical quality measures (CQMs) offered one way to measure the degree to which practice transformation efforts in SIM-participating sites lead to **better care** for patients. SIM-participating primary care practice sites and community mental health centers (CMHCs) reported these measures quarterly in order to track processes and outcomes for their populations. These measures focused primarily on increased screening of patients for important physical and behavioral health conditions. They also included two measures related to positive health outcomes.

As discussed more fully in the SIM Final Process Evaluation Report, these measures played an important foundational role in supporting value-based payment (VBP) structures. To participate fully in VBPs, practice sites needed to demonstrate the value they provided to their patients in the form of quality care that leads to better health outcomes. Being able to reliably report these measures and use them to improve care provided practice sites with an important tool in being compensated for providing value and high-quality care, rather than being paid on a fee-for-service basis. As such, the CQMs aimed to contribute to the SIM aim of lower costs. The SIM implementation made significant strides in assisting practice sites in this area. The Practice Level Health Information Technology chapter in the SIM Final Process Evaluation Report details the work done during the SIM model test period.

This Clinical Outcomes chapter focuses on the use of CQMs as a way to describe the SIM progress on the outcome of “better care” within the SIM Quadruple Aim. This relates directly to practice transformation outcomes and the following evaluation question:

PT6. What specific transformation factors (level of integration, milestone targets, data quality, clinician and staff experience, etc.) most influence outcomes?

The chapter begins with a discussion of the methods used to define, compile, and analyze these measures. It also discusses the role of “proxy” clinical outcome measures, compiled from claims data, and the relationship between these and practice-reported measures. Following the Proxy Measures section, we discuss changes over time in each of the 14 identified SIM CQMs. We then analyze the relationship between the outcomes and several transformation factors. The

---

[49] SIM began with a focus on the “Triple Aim” of lower costs, better care, and better patient experience, then elected to add a focus on workforce during its initial planning year..

chapter ends with a separate analysis of the four SIM-participating CMHCs and a summary table of CQM descriptive statistics.

Methods

The SIM office selected these CQMs to measure changes in the quality of care provided by SIM-supported primary care practice sites and community mental health centers (CMHCs). These CQMs are described in Table 48 and are taken from the *Colorado SIM Clinical Quality Measure Specifications Guidebook*.[50] Although most of these measures were related to the process of care, some provided indicators of better patient outcomes. These are noted in the table below and within the discussion of changes over time. The SIM office also commissioned related proxy CQMs based on claims data, described in depth below. For ease of comparison, the table also includes a brief description of proxy measures with the respective SIM CQMs. A table of numerator and denominator descriptions used to calculate each measure is included at the end of this chapter (Table 71).

We present CQMs related to behavioral health first, followed by CQMs with a smaller behavioral health component, such as *Hemoglobin A1c Control*. Measures specialized to geriatric patients or children and youth are presented and analyzed last. Some practice sites provided data on depression screening and remission using a measure from CPC+. None of those data allow for analysis over time, and we therefore exclude them from this chapter.

The table below provides definitions for the SIM CQMs and related Proxy CQMs. However, these measures are not necessarily similar or comparable, as there are sometimes significant differences between them. The analyses below for each measure discussed the degree to which these measures can or cannot be compared.

Table 48. CQM and Proxy Measures with Descriptions

Measure Title	Description	
Depression Screening (SIM/QPP) **NQF 0418** **CMS 2v5**	**SIM CQM**	Percentage of patients aged 12 years and older screened for clinical depression using an age-appropriate standardized tool AND follow-up plan documented.
	Proxy CQM[51]	Percentage of patients aged 12 years and older screened for depression AND with new depression diagnosis.

---

[50] Retrieved from http://www.practiceinnovationco.org/wp-content/uploads/vfb/2016/06/FINAL_SIM-CQM-GUIDEBOOK_20160609.pdf

[51] For detailed information on SIM and Proxy CQMs, please see Table 70 at the end of this chapter.

Measure Title	Description	
Maternal Depression Screening **NQF 1401** **CMS 82v3**	SIM CQM	Percentage of children who turned six months of age during the measurement year, who had a face-to-face visit between the clinician and the child during the child's first six months, and who had a maternal depression screening for the mother at least once between 0 and 6 months of life.
	Proxy CQM	Percentage of mothers receiving screening for maternal depression by a clinician in the first six months following delivery.
Alcohol and Other Drugs Screening **NQF 0004** **CMS 137v4**	SIM CQM	Percentage of patients aged 13 years and older with a new episode of alcohol and other drug (AOD) dependence who received the following (two rates are reported): - Percentage of patients who initiated treatment within 14 days of the diagnosis - Percentage of patients who initiated treatment and who had two or more additional services with an AOD diagnosis within 30 days of the initiation visit
	Proxy CQM	Percentage of patients aged 18 years and older screened for illicit drug use.
Tobacco Use Screening	SIM CQM	Percentage of patients aged 18 years and older who were screened for tobacco use one or more times within 24 months AND who received cessation counseling intervention if identified as a tobacco user.
	Proxy CQM	Percentage of patients aged 18 years and older screened for tobacco use.
Alcohol Use Screening	SIM CQM	Percentage of patients aged 18 years and older who were screened for unhealthy alcohol use using a systematic screening method at least once within the last 24 months AND who received brief counseling if identified as an unhealthy alcohol user.
	Proxy CQM	Not available for this measure.
Diabetes: Hemoglobin A1c Control **NQF 0059** **CMS 122v4**	SIM CQM Outcome	Percentage of patients 18–75 years of age with diabetes who had hemoglobin A1c > 9.0% during the measurement period.
	Proxy CQM	Diabetes A1c Control: Percentage of patients with Type 1 or Type 2 diabetes who had hemoglobin A1c testing at clinician encounter.
Hypertension Management **NQF 0018**	SIM CQM Outcome	Percentage of patients 18–85 years of age who had a diagnosis of hypertension and whose blood pressure was adequately controlled (< 140/90mmHg) during the measurement period.

Measure Title	Description	
CMS 165v4	Proxy CQM	Percentage of patients 18–85 years of age with essential hypertension dispensed at least 90-day supply of therapeutic medications.
Adult Obesity Screening **NQF 0421** **CMS 69v4**	SIM CQM	Percentage of patients aged 18 years and older with BMI documented during the current encounter or during the previous six months AND with a BMI outside of normal parameters, a follow-up plan is documented during the encounter or during the previous six months of the current encounter.
	Proxy CQM	Percentage of patients aged 18 years and older diagnosed with obesity receiving two or more clinician encounters and one or more tests for diabetes (blood glucose or HA1c).
Asthma Medication Management **NQF 0036** **CMS 126v4**	SIM CQM	Percentage of patients 5–64 years of age who were identified as having persistent asthma AND were dispensed appropriate medication that they remained on for at least 75% of their treatment period.
	Proxy CQM	Percentage of patients 5–64 years of age who were identified as having persistent asthma AND were dispensed appropriate medications that they remained on for at least 75% of their treatment period.
Fall Risk Screening **NQF 0101** **CMS 139v4**	SIM CQM	Percentage of patients aged 65 years and older who were screened for future fall risk.
	Proxy CQM	Percentage of patients aged 65 years and older receiving annual health evaluation with assumed assessment of future fall risk during the encounter.
Developmental Screening **NQF 1448**	SIM CQM	Percentage of children screened for risk of developmental, behavioral, and social delays using a standardized screening tool in the 12 months preceding their first, second, or third birthday.
	Proxy CQM	Percentage of children turning one, two, or three years of age during the measurement period receiving developmental screening in the 12-month period preceding the first, second, or third birthday.
Adolescent Obesity Screening **NQF 0024** **CMS 155v4**	SIM CQM	Percentage of patients 3–17 years of age who had an outpatient visit with a PCP or OB/GYN AND who had evidence of the following during the measurement period (three rates are reported): - Percentage of patients with height, weight, and BMI percentile documentation - Percentage of patients with counseling on nutrition - Percentage of patients with counseling for physical activity

Measure Title	Description	
	Proxy CQM	Not available for this measure.
Depression Remission (CPC+)	SIM CQM Outcome	Adult patients aged 18 years and older with major depression or dysthymia AND an initial PHQ-9 score > 9 who demonstrate remission at 12 months defined as PHQ-9 score less than 5. This measure applies to both patients with newly diagnosed and existing depression whose current PHQ-9 score indicates a need for treatment.
	Proxy CQM	Not available for this measure.

CQM reporting requirements varied depending on practice type. Primary care practice sites seeing only adult patients were required to report on six primary measures, whereas pediatric practice sites were required to report on four primary pediatric measures. Mixed family practice sites, or those sites seeing both pediatric and adult patients, may have chosen to report from either set of CQMs. These are minimum reporting standards, and practice sites could submit more CQMs than required. Furthermore, practice sites were not required to submit the same CQMs across quarters or throughout SIM. Each quarter, sites were asked to report on a 12-month "trailing" year, so, for example, the fourth quarter of 2016 should reflect data for the whole 2016 calendar year (CY). Not all practice sites were able to report using these reporting standards, and some had to modify their CQM specifications. We discuss this more thoroughly in the HIT Chapter of the Process Evaluation Report.

As discussed in previous reports, practice sites reported technical difficulties early in their CQM collection, and many sites continued to report challenges. Practice sites and the SIM office advised that, based on feedback from sites and common national practice,[52] that Q4 data are the most accurate and complete. As such, we have limited our analyses to Q4 data in 2016, 2017, and 2018. Additionally, one electronic health record (EHR) vendor announced widespread issues with its Q4 2018 data. A total of seven practice sites (five from cohort 2 and two from cohort 3) indicated they used this vendor, and we have removed those sites from this analysis.

Because of variances in cohort onboarding, requirements, and length of participation, close alignment across CQMs reported and other data assessment periods (particularly with in the Practice Monitor and Milestone assessments collected in SPLIT) is not possible. Further, using the Q4 CQM submissions representing calendar year performance means that the measurement periods do not align with a true "baseline" or "final" measure for any of the cohorts. This is because SIM cohorts begin and end in different time periods throughout any

---

[52] See, for example, https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/ClinicalQualityMeasures.html

given calendar year. For example, cohort 2 began participation in September of 2017, so although we use 2017 as the "baseline" calendar year, it is not a true baseline (i.e., the whole of 2017 did not completely precede implementation) because four months of program participation occurred during the year. Thus, these yearly measures represent a snapshot along the course of SIM participation for practice sites and only capture changes during segments of their overall SIM work. For this analysis of outcomes, we show change over time by presenting data as reported in three different "assessment periods." These periods are described as follows:

- **Assessment Period 1** is listed as "1" on the following graphs. This is the best baseline measure available. For cohort 1, this is CY 2016. For cohort 2 and the CMHCs, this is CY 2017. For cohort 3, Assessment Period 1 is CY 2018. No additional data are presented other than for the first assessment period. Because of a very small sample (n=4) and lack of comparability to primary care practice sites, we present CMHC data independently in a table after the primary care results (Table 60). Results for the CMHCs are not plotted on graphs.[53]
- **Assessment Period 2** is listed as "2" on the graphs. This is the best final measure available. For cohort 1, Assessment Period 2 is CY 2017. For cohort 2 and the CMHCs, this is CY 2018.
- **Assessment Period 3** is listed as "3" and is an additional assessment period added because nearly half of cohort 1 practice sites elected to report CQMs after their participation in SIM ended. This period is for 2018 and includes cohort 1 only. These data are most useful in ascertaining whether gains made during active participation were retained post-SIM.

Proxy Clinical Quality Measures

In addition to reporting on the EHR-based CQMs, we are also able to report proxy CQMs calculated by CIVHC and based on claims and encounters contained in the APCD. We briefly described the proxy CQMs above. Proxy CQMs differ from EHR-based CQMs for several reasons.

First, claims and encounters contained somewhat different data than did practice site EHRs. In some cases, CIVHC modified the proxy measure to use a claim field that closely related to, but was not a direct replacement of, an EHR-based data field[54]. The APCD contained information

---

[53] There are significant differences in the populations served and the interventions delivered by CMHCs versus primary care. Comparing CQMs between the two groups is, therefore, inappropriate. Readers should focus on change over time rather than group comparisons.

[54] Center for Improving Value in Health Care. (2017). *State Innovation Model (SIM) proxy measures.* Retrieved from https://www.civhc.org/wp-content/uploads/2017/07/Spot-Analysis-Series-SIM-Measures.pdf

unavailable to EHRs, such as prescriptions that are filled as opposed to written. But in other cases, diagnostic information stored in the EHR (e.g., weight) were not submitted as part of the claim. Please see CIVHC's documentation of the proxy CQMs for more details on the exact fields used.

Next, practice sites did not necessarily submit claims for services that have a very low billing rate. For example, payments for depression screening were low enough that the cost of completing billing may have exceeded revenue paid[55].

A third difference between EHR-based CQMs and proxy CQMs is the patient population included. Practice sites (presumably) reported on all patients served. We calculated the proxy measures for patients attributed to all SIM practice sites by CHVHC, and the two patient populations did not coincide perfectly. The proxy measures are not based on the average of each site's reported values, but instead the average over all SIM-attributed patients (by cohort). The proxy measures are consistent with other products generated from the APCD—such as Milliman cost and utilization reports and AHRQ "Access to Care" measures (with all of the caveats and limitations already discussed)—whereas the CQMs were drawn from the electronic health records of patients seen by the practice sites.

Finally, we calculated the proxy CQMs using the same methodology for each practice site, and every site with attributed patients has a value for each period covered. Practice sites implemented reporting CQMs at somewhat different times and with varying fidelity to the reporting standard. See the HIT section of this report for additional details.

We include the proxy CQMs values on the table following each CQM graph below. Unlike with the CQMs, we can report for the baseline (year before SIM participation) for each cohort. This is useful for identifying change starting with a true pre-SIM baseline value. We also include statewide values for periods corresponding to those of cohort 1 (e.g., 2015 is baseline for both cohort 1 and the first statewide value). The statewide values may be calculated from populations of patients who are not comparable to SIM primar- care-attributed patients, as will be apparent by comparing baseline (pre-SIM) values for each group. The greatest value in the statewide values is the ability to identify changes in trends over time caused by factors independent of the SIM program.

---

[55] Center for Improving Value in Health Care. (2017). *State Innovation Model (SIM) proxy measures.* Retrieved from https://www.civhc.org/wp-content/uploads/2017/07/Spot-Analysis-Series-SIM-Measures.pdf

Changes in Clinical Quality Measure Reporting

Limitations

Multiple challenges with CQM data reporting may have created results that require caution when being interpreted. As already discussed, the measurement timeframes did not align well with intervention periods, making it difficult to assume change over time correlated to SIM efforts. Further, changes in reporting requirements for cohorts were necessary and vital in setting a solid foundation for practice sites to make progress in reporting these measures and better participating in VBPs.[56] But a secondary effect of these changes, along with flexibility in reporting requirements,[57] was inconsistency in reporting from period to period. SIM practice sites may, for example, have reported on the maternal depression measure in period 2 but not in period 1, meaning that we cannot make a comparison over time for that site. This significantly reduced the overall number of sites included in the analysis of change over time. Finally, these were measures reported directly by the practice sites, and data quality might have been reduced in some cases because of EHR or practice workflow issues that result in low accuracy. We report quality of CQMs and the improvement in this quality over time in the Practice Health Information Technology chapter of the SIM Final Process Evaluation Report.

CQM Change-Over-Time Analysis

The following graphs show change over time in each of the CQMs.[58] We present a separate plot for each measure, and each plot presents three pieces of information:

- **Dots on each vertical axis**. To better visualize the distribution of practice site results, each dot along the vertical assessment period axis represents an individual practice site value. Darker clusters indicate a higher concentration of sites at that value. For Assessment Period 1, we include data from all cohorts. For Assessment Period 2, we exclude cohort 3 data, and for Assessment Period 3, we exclude cohort 2 and 3 data.
- **Reporting Periods 1 and 2**. This is the average CQM value for SIM practice sites in the specified cohort across only the sites **reporting on that measure in both the first and second assessment periods**. We calculate this average by averaging the value reported by each practice site.

---

[56] We recognize this information is redundant to issues already discussed, but they bear repeating to avoid inappropriate interpretations of the data.

[57] This was good for practice sites as they increased their competencies around CQM reporting but affected the evaluation negatively.

[58] Because of insufficient reporting data, we include a summary analysis of the “Alcohol Use Screening” CQM but do not include a plot or table.

- **Reporting Periods 2 and 3**. This is the average CQM value for SIM practice sites in **second and third assessment periods**. We calculate this average by averaging the value reported by each practice site.

The tables that follow the plots contain information about each respective CQM and analysis of change over time. The “Cohort 1”, “Cohort 2”, and “Cohort 3” rows present the average of each practice site’s value for the CQM in the specified cohort, broken out by assessment period. This average includes all practice sites reporting in that single period, regardless of whether they reported in any other period. The cohort averages are disproportionately influenced by patients from smaller practice sites because smaller sites are weighted equally to larger sites in calculating the average across practice sites. The “N” associated with each assessment period is the number of practice sites reporting the given value. We include more detailed information on all reporting practice sites, including the standard deviation by cohort and period, in a summary table at the end of the chapter (see Table 70).

Because the difference between the practice site averages over time is influenced by different sites reporting in each period, we also report the average limited to those practice sites reporting in both the first and second period, or both the second and third period (“Practice Sites Reporting Periods 1 and 2” and “Practice Sites Reporting Periods 2 and 3”). By limiting to those sites reporting in both periods, we ensure that any change is the result of differing practice site performance and not a result of a changing mix of practice sites. Because half of cohort 1 and all of cohort 2 have data for only assessment periods 1 and 2 and cohort 3 only has data for Assessment Period 1, we made this calculation for reporting periods 1 and 2 only for cohorts 1 and 2, and we excluded it for cohort 3. The “N” associated with each assessment period is the number of practice sites reporting the given value.

“Difference/P-Value” shows the change in the practice site average for sites reporting both assessment periods, followed by the P-Value associated with that change. An asterisk (*) indicates statistically significant differences at a P-Value of < 0.05. Because the change over time is measured with the same practice sites in each period, we based this P-Value on a paired T-test.

At the bottom of each table, we provide the values of the proxy CQMs previously described. The “N” reported represents the number of SIM-attributed patients included in the calculation. Individual practice sites play no role in these numbers.

We also provide a “summary” of results following the chart, table, and narrative for each CQM when there are any significant or seemingly meaningful changes to report.

Depression Screening

The following plot (Figure 24) shows individual (reporting) practice sites' depression screening rate at each reporting period. The dots representing practice site values show a very wide distribution of depression screening rates, with some sites reporting rates at or close to 0% and some having rates at nearly 100% in each reporting period. The standard deviation does decrease in each cohort's last reporting period (see Table 70).

The first row of the table (Table 49) reports the average practice site value for each of the three potential reporting periods. The value for cohort 1 increases for each period, from 45.4% to 54.7% to 66.1%. This change is somewhat misleading. Since different practice sites report each period, the increase reflects a changing mix of practice sites, not necessarily a gradual increase across periods. We control for this variability by reporting the change in average CQM scores between periods for practice sites reporting in each period of the change. The graph's dark green line shows the average rate of screening for all SIM practice sites reporting in both the first and second reporting period (December 2016 and December 2017).

We report the two averages for cohort 1 on the second row of the table, along with the difference in the average value (an increase of 13.2 percentage points). A T-test for a difference in means has a P-Value less than 0.01, indicating this difference is statistically significant. The increase for cohort 1 was large and was not caused by random factors related to sampling. A smaller number of practice sites reported in period 3, and the third row of the table lists the average value for sites reporting in both period 2 and period 3. As indicated by the light green line on the graph, the change between these periods for the 37 practice sites that reported in both periods was small—and with a P-Value of 0.97, the change was not statistically significant.

Based on the relatively large change between period 1 and 2—and the small change between period 2 and 3—the improvement in CQMs for cohort 1 occurred almost exclusively between 2016 and 2017. On average, practice sites were able to implement depression screening during the SIM participation phase and maintain this level of screening post SIM.

Cohort 2 practice sites had similar but smaller average increases between assessment periods 1 (2017) and 2 (2018). The difference was 6.1%, and a difference-in-means test yielded a P-Value less than 0.01, which is highly significant). Cohort 2 practice sites started with higher average values than did cohort 1 sites and, despite having a smaller increase, cohort 2 ended the second assessment period with a higher value. Table 49 reports the practice-site average for assessment periods 1 and 2 for both all practice sites that reported in any period and those practice sites that reported in both periods.

Figure 24. Depression Screening

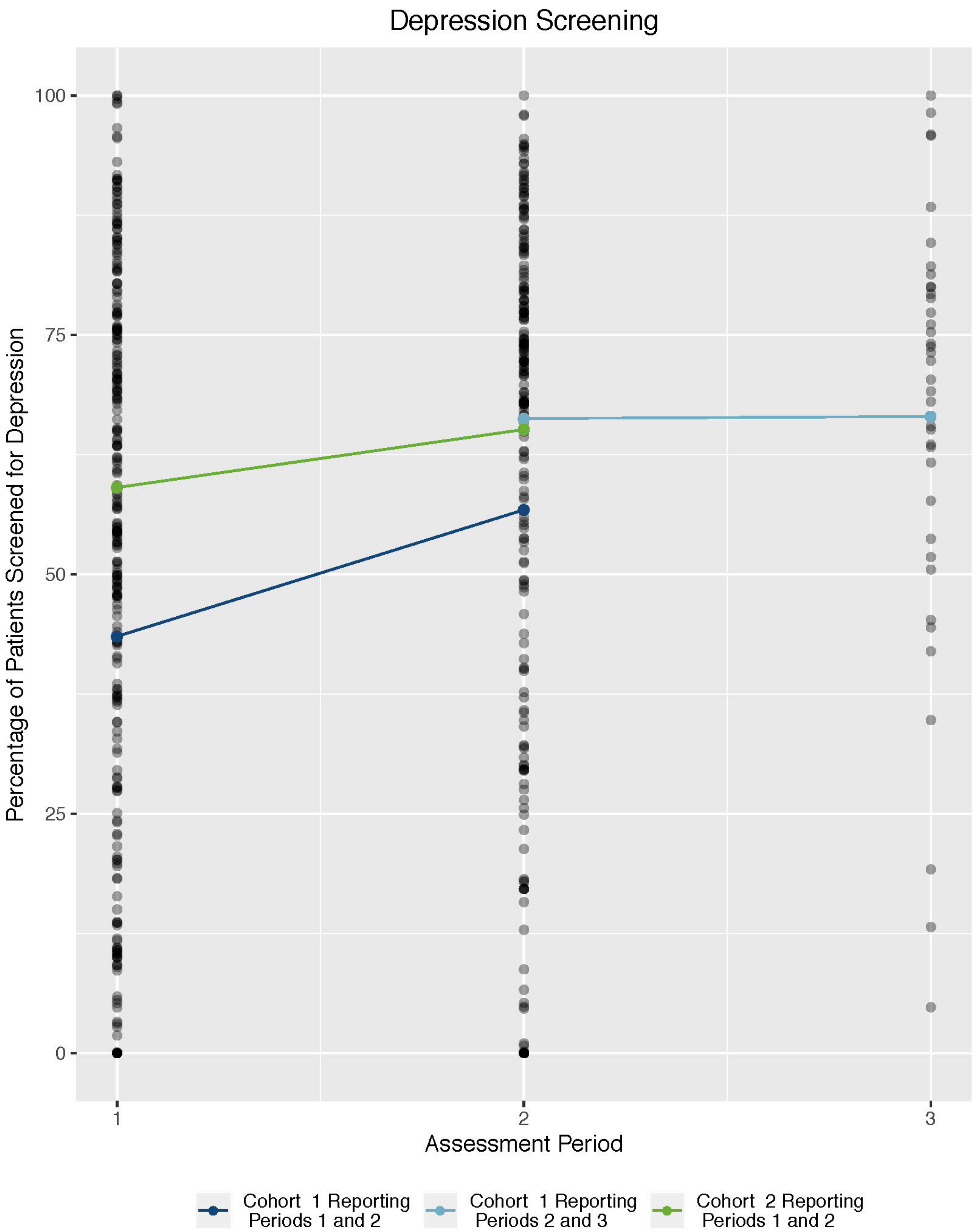


Table 49. Depression Screening

Depression Screening[59]	Baseline		First		Second		Third	
	%	N	%	N	%	N	%	N
Cohort 1 (all reporting)	-	-	45.4%	82	54.7%	74	66.1%	38
Cohort 1 Practice Sites Reporting Periods 1 and 2	-	-	43.5%	69	56.7%	69	-	-
Difference/P-Value	-	-	-	-	13.2% 0.00*	-	-	-
Cohort 1 Practice Sites Reporting Periods 2 and 3	-	-	-	-	66.3%	37	66.5%	37
Difference/P-Value	-	-	-	-	-	-	0.2% 0.91	-
Cohort 2 (all reporting)	-	-	53.3%	112	64.6%	120	-	-
Cohort 2 Practice Sites Reporting Periods 1 and 2	-	-	59.1%	101	65.1%	101	-	-
Difference/P-Value	-	-	-	-	6.1% 0.00*	-	-	-
Cohort 3 (all reporting)	-	-	51.9%	69	-	-	-	-
Proxy CQM Cohort 1	10.9%	6,284	10.8%	8,815	11.7%	9,556	-	-
Proxy CQM Cohort 2	12.0%	4,704	12.9%	8,007	-	-	-	-
Proxy CQM Cohort 3	13.0%	3,913	-	-	-	-	-	-
Proxy CQM Statewide**	9.6%	50,257	11.3%	42,527	12.4%	59,366	-	-

Cohort 3 has only one reporting period (December 2018). Its average value falls between cohorts 1 and 2.

The last five rows of Table 49 report results for the Proxy CQMs. These proxy measures report the percentage of screened patients who had a new diagnosis of depression. They measure prevalence of patients diagnosed with depression. The CQM instead measures the percentage of patients screened for depression. The proxy measure should not therefore be compared to the CQM. Its value lies in presenting information on changes over time in rates of depression, both in the SIM cohorts and the broader population of Coloradans included in the APCD.

As compared to the statewide baseline value, patients attributed to all three cohorts had higher rates of diagnosed depression in the pre-SIM baseline year. Over time, the statewide

---

[59] A single * indicates statistical significance (P-Value ≤ 0.05). A double ** indicates that statewide proxy CQM assessment periods correspond with cohort 1.

prevalence increased substantially, whereas the rates of the SIM cohorts experienced smaller change.

Summary of Change in Depression Screening

- **Cohort 1**. Large and statistically significant improvement during SIM participation, gains were maintained post-SIM participation.
- **Cohort 2**. Small but statistically significant increase.

Maternal Depression Screening

Both cohorts 1 and 2 showed increased maternal screening between the assessment periods (Table 50). However, since only a small number of practice sites reported this measure, the results are not statistically significant. More of the gain for cohort 1 practice sites occurred between the first and second assessment periods, during the time they were participating in SIM, yet the post-SIM period showed additional gains. In this sense, gains for cohort 1 were more than sustained post SIM; there were additional gains.

The plot following (Figure 25) shows individual (reporting) practice sites' maternal depression screening rate at each reporting period. Each dot along the vertical axis represents an individual site's reported rate. As was the case with depression screening, the distribution of each site's rate, represented by the dots on each time period line on the plot, is wide. The distribution narrowed over the time periods, with lower standard deviation in periods 2 and 3. Fewer sites reported very low values in later periods.

Figure 25. Maternal Depression Screening

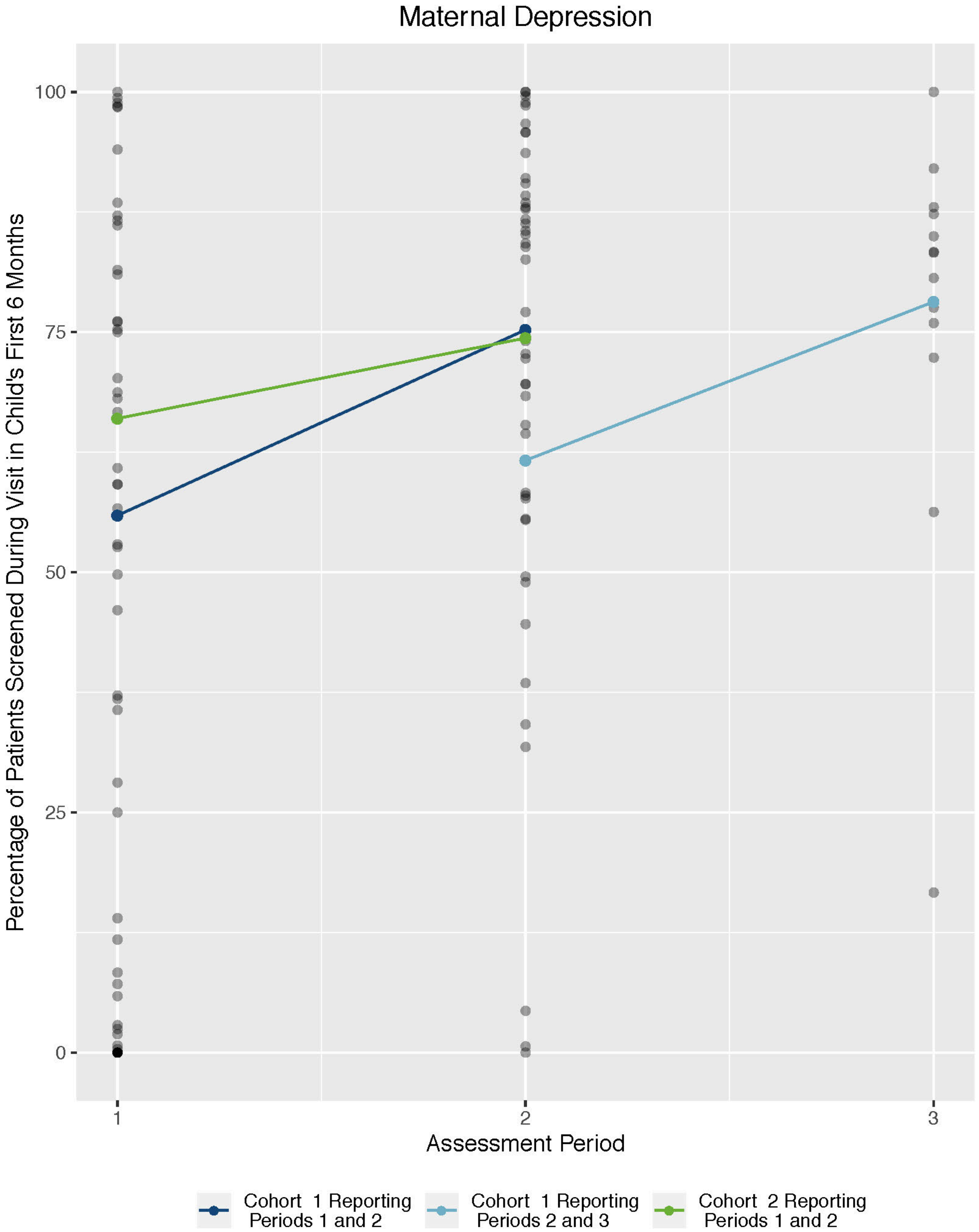


Table 50. Maternal Depression Screening

Maternal Depression[60]	Baseline		First		Second		Third	
	%	N	%	N	%	N	%	N
Cohort 1	-	-	51.3%	12	72.4%	16	76.8%	13
Cohort 1 Practice Sites Reporting Periods 1 and 2	-	-	55.9%	11	75.2%	11	-	-
Difference/P-Value	-	-	-	-	19.3% 0.06	-	-	-
Cohort 1 Practice Sites Reporting Periods 2 and 3	-	-	-	-	61.6%	7	78.2%	7
Difference/P-Value	-	-	-	-	-	-	16.5% 0.16	-
Cohort 2	-	-	66.0%	17	69.6%	31	-	-
Cohort 2 Practice Sites Reporting Periods 1 and 2	-	-	66.0%	17	74.4%	17	-	-
Difference/P-Value	-	-	-	-	8.4% 0.14	-	-	-
Cohort 3	-	-	28.3%	21	-	-	-	-
Proxy CQM Cohort 1	18.7%	8,999	30.7%	10,613	31.7%	7,525	-	-
Proxy CQM Cohort 2	13.6%	4,823	17.6%	4,781	-	-	-	-
Proxy CQM Cohort 3	8.7%	1,487	-	-	-	-	-	-
Proxy CQM Statewide**	8.2%	42,558	13.1%	44,176	12.0%	49,192	-	-

The final five rows of Table 50 report the proxy CQMs. SIM practice sites have consistently and considerably exceeded statewide proxy rates for this measure. Furthermore, SIM rates have consistently improved from baseline, whereas the statewide rate improved between periods 1 and 2 before dropping again between periods 2 and 3. This difference may reflect that SIM practice sites, having once identified this measure as a priority for improvement, were able to increase screening relatively easily and to considerable effect. **In summary, the proxy results have different magnitudes than the CQMs have but tell the same story of increased maternal depression screening during the SIM intervention period, whereas the statewide rate did not show increases. These results suggest the possibility that the SIM intervention is responsible for increasing maternal depression screening rates.**

---

[60] A single * indicates statistical significance (P-Value ≤ 0.05). A double ** indicates that statewide proxy CQM assessment periods correspond with cohort 1.

Summary of Maternal Depression Screening

- Small numbers of sites reported, making statistical significance difficult to achieve.
- Trends in both the practice-site-reported CQMs and the proxy CQMs do seem to indicate a possible increase in maternal depression screening for cohorts 1 and 2 that was not seen in the average state rate. This is a potential indicator that the SIM intervention had a positive influence on this measure.

Alcohol and Other Drug Screening

Despite its title, this CQM measures the percentage of patients with a new SUD diagnosis who receive treatment. Most practice sites did not report this measure. The most pressing barrier to reporting this measure is the practice site EHRs: practice sites in all cohorts said they targeted AOD through screenings and interventions, but their EHRs did not support the workflows needed for CQMs. The EHRs may not be built to capture AOD, or the practice may not be able to afford buying the upgrades that sometimes have the fields.

Although Table 51 and Figure 26 summarize results for those practice sites reporting, the small number of reporting sites—some of which reported zero in the denominator—suggests the results are not credible.

Data gathered from CHITAs through field notes, and in narratives that accompany CQM reporting, show that practice sites have considerable difficulty reporting on this measure. By far, the most pressing barrier is the EHRs. Practice sites in all cohorts said they targeted AOD through screenings and interventions, but their EHR did not support the CMS workflows needed for CQMs. The EHR may not be built to capture AOD, the practice site may not be able to afford buying the upgrades that sometimes have the fields, and one site in cohort 2 noted “provider resistance” but did not provide further context. We discuss barriers and limitations in greater detail in the Practice Health Information Technology chapter of the SIM Final Process Evaluation Report.

Figure 26. Alcohol and Other Drug Screening

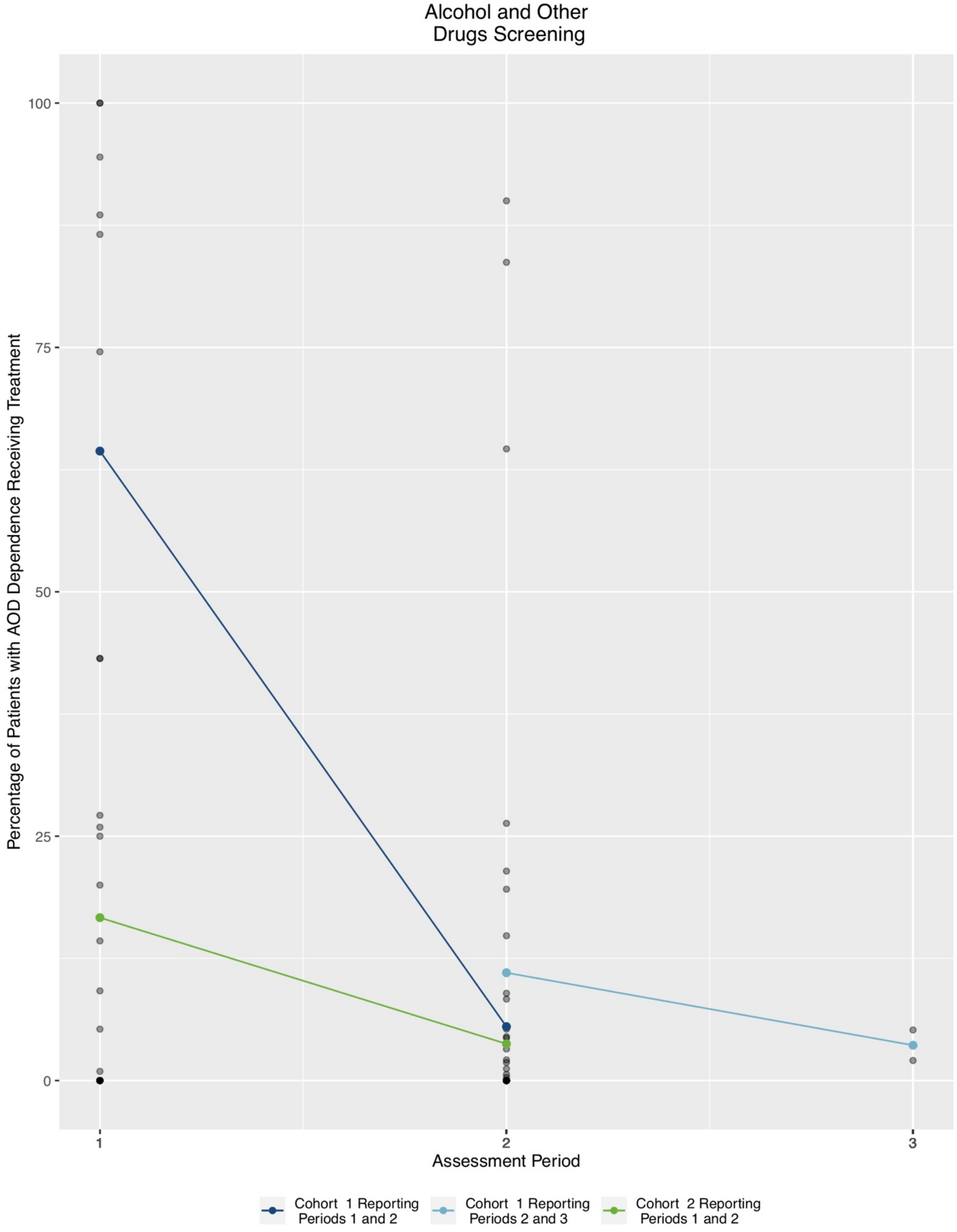


Table 51. Alcohol and Other Drug Screening

Alcohol and Other Drug Screening [61]	Baseline		First		Second		Third	
	%	N	%	N	%	N	%	N
Cohort 1	-	-	57.4%	6	9.7%	20	3.6%	2
Cohort 1 Practice Sites Reporting Periods 1 and 2	-	-	64.4%	4	5.5%	4	-	-
Difference/P-Value	-	-	-	-	-58.9% 0.12	-	-	-
Cohort 1 Practice Sites Reporting Periods 2 and 3	-	-	-	-	11.0%	2	3.6%	2
Difference/P-Value	-	-	-	-	-	-	-7.4% 0.56	-
Cohort 2	-	-	16.4%	15	8.0%	21	-	-
Cohort 2 Practice Sites Reporting Periods 1 and 2	-	-	16.7%	6	3.8%	6	-	-
Difference/P-Value	-	-	-	-	-12.9% 0.45	-	-	-
Cohort 3	-	-	16.8%	10	-	-	-	-
Proxy CQM Cohort 1	18.8%	94,761	12.8%	141,035	5.7%	150,378	-	-
Proxy CQM Cohort 2	10.8%	101,039	5.0%	104,865	-	-	-	-
Proxy CQM Cohort 3	4.9%	51,286	-	-	-	-	-	-
Proxy CQM Statewide**	13.8%	1,089,412	10.4%	1,349,711	4.6%	1,581,051	-	-

The final five rows of Table 51 include results for the proxy CQMs. Unlike the actual CQM, the proxy CQM does not measure the linkage to treatment and instead measures the frequency of screening by primary care practice sites. The declining rates for both SIM practice sites and statewide are difficult to understand. Although it is conceivable that practice sites were not successful in performing this type of screening, it seems very unlikely that as practice sites implemented integrated care, they decreased screening rates by over 50% and that the statewide rate also decreased by 50% during the same period. The greater likelihood is that these decreases resulted from incompleteness in CIVHC's ability to report on SUD-related diagnoses caused by limitations associated with using claims data to calculate these measures.

---

[61] A single * indicates statistical significance (P-Value ≤ 0.05). A double ** indicates that statewide proxy CQM assessment periods correspond with cohort 1.

Summary of Alcohol and Other Drug Screening

- The number of sites reporting on this metric was too small to allow conclusions about change over time.

Tobacco Use Screening

This CQM measured screening and receipt of tobacco cessation counseling. Cohort 1 did not report on the measure during the Assessment Period. For Assessment Period 2, almost all practice sites reported a value of 100%. For the 28 practice sites that provided data post SIM, the average value was lower than 100%, and the decline was statistically significant.

Cohort 2 reported for both first two assessment periods, but all practice sites that reported in the first assessment period (n = 93) reported a value of 100%—except for three sites that reported zero in the numerator and denominator. Because the first-period cohort 2 value was 100, the change from first to second period was also a decline, and the change was statistically significant.

Despite declines, the overall percentage of patients screened for tobacco use in the second time period remained at the SIM target for the measure when combining all three cohorts (93%).

Because all practice sites from cohort 1 and 2 reported 100% in their initial data and then showed similar declines, we are suspicious that the initial values represent a misunderstanding of the CQM metric definitions. We have greater confidence in the data reported in the last period for each of the two cohorts and suggest caution in interpreting trends over time.

Figure 27. Tobacco Use Screening

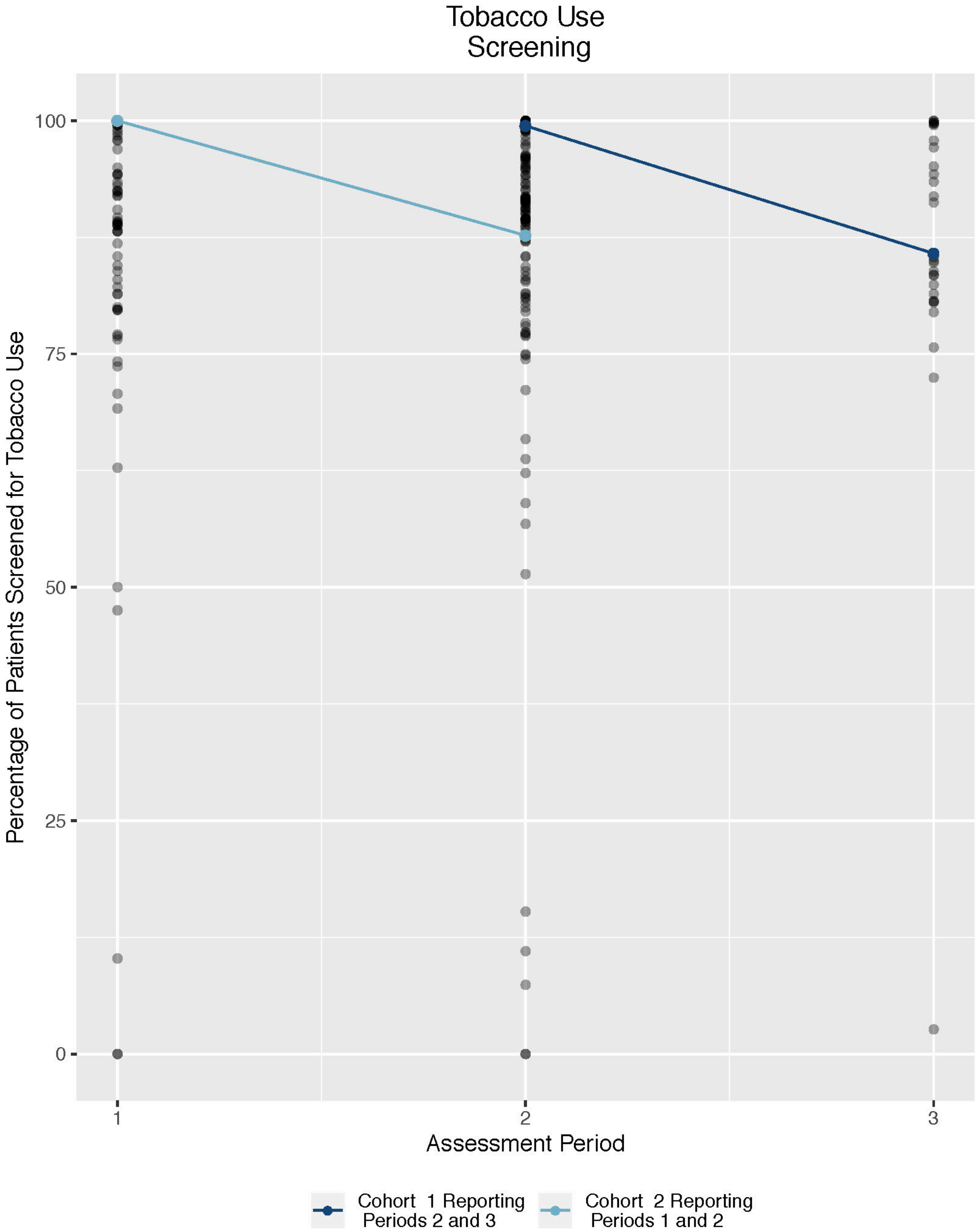


Table 52. Tobacco Use Screening

Tobacco Use Screening[62]	Baseline		First		Second		Third	
	%	N	%	N	%	N	%	N
Cohort 1	-	-	-	-	99.8%	67	85.8%	28
Cohort 1 Practice Sites Reporting Periods 2 and 3	-	-	-	-	99.5%	28	85.8%	28
Difference/P-Value	-	-	-	-	-	-	-13.7% <0.01*	-
Cohort 2	-	-	100.0%	93	84.9%	107	-	-
Cohort 2 Practice Sites Reporting Periods 1 and 2	-	-	100.0%	84	87.7%	84	-	-
Difference/P-Value	-	-	-	-	-12.3% <0.01 *	-	-	-
Cohort 3	-	-	82.5%	59	-	-	-	-
Proxy CQM Cohort 1	23.8%	94,761	24.4%	141,035	26.0%	150,378	-	-
Proxy CQM Cohort 2	22.0%	101,039	23.9%	104,865	-	-	-	-
Proxy CQM Cohort 3	24.2%	51,286	-	-	-	-	-	-
Proxy CQM Statewide**	23.8%	1,089,412	23.3%	1,349,711	24.2%	1,581,051	-	-

In support of our questioning of the CQM data from cohorts 1 and 2, the proxy CQM results do not show any decline over time for these cohorts. The values are substantially below the CQMs, which is likely caused by the low reimbursement rates for tobacco use screening, with few practice sites finding it worthwhile to make claims despite almost universal screening. It is important to note, however, that the proxy CQM only measures whether patients were screened, not whether a positive screening resulted in treatment. As with the *Depression Screening* CQM, the CIVHC proxy measures are much higher and may reflect an encounter-based payment system. The increase over time for the CIVHC proxy measures may reflect the effect of the SIM intervention.

Summary of Change in Tobacco Use Screening

- For both cohorts, initial reporting of 100% screening rates by almost all practice sites raises questions about the reliability of the initial data and our ability to detect change over time.

---

[62] A single * indicates statistical significance (P-Value ≤ 0.05). A double ** indicates that statewide proxy CQM assessment periods correspond with cohort 1.

Summary of Tobacco Use Screening

- Cohorts 1 and 2 had statistically significant declines over time, but all practice sites reported 100% screening at baseline. The final screening rate still met the SIM target for this measure for this cohort.
- Practice-site-reported CQMs for *Tobacco Screening* greatly exceeded proxy CQM rates. This was likely caused by low reimbursement rates, which means practice sites probably did not bill for this service.

Alcohol Use Screening

The overall percentage of SIM patients being screening for alcohol use in practice sites that reported on this measure increased, though not significantly, from the first and second assessment periods. Only two cohort practice sites chose to report this measure in the third reporting period (which was after SIM participation when measures were not required). Those two sites reported significant increases in the number of patients screened for alcohol use. CMHCs only reported on this measure during the first reporting period.

We discuss difficulties with this measure in the Practice HIT chapter of the SIM Final Process Evaluation report. Because so few sites reported on this measure, we did not include a plot and table for this CQM.

Hemoglobin A1c Control

This CQM reports the percentage of patients with a diabetes diagnosis who have hemoglobin A1c levels above 9% during the reporting period. As such, *Hemoglobin A1c Control* is the one CQM in which a **decrease** in the percentage of patients was the better outcome. It is also one of two CQMs that represented a health outcome.

The following plot (Figure 28) shows individual (reporting) practice sites' rate of patients with poor hemoglobin A1c control at each reporting period. Although there were some practice site outliers that reported high rates, most sites reported rates below 50%, and most were clustered near the average percentage across all reporting time periods. The dark green line corresponds to a statistically significant decline for cohort 1 between assessment periods 1 and 2. The cohort 1 change from period 2 to period 3 was essentially zero, indicating that cohort 1 practice sites maintained gains made during participation in SIM. Most of cohort 2 sites that reported on this measure reported both periods. The average value from period 1 to period 2 increased slightly, but the increase was not statistically significant.

Figure 28. Hemoglobin A1c Control

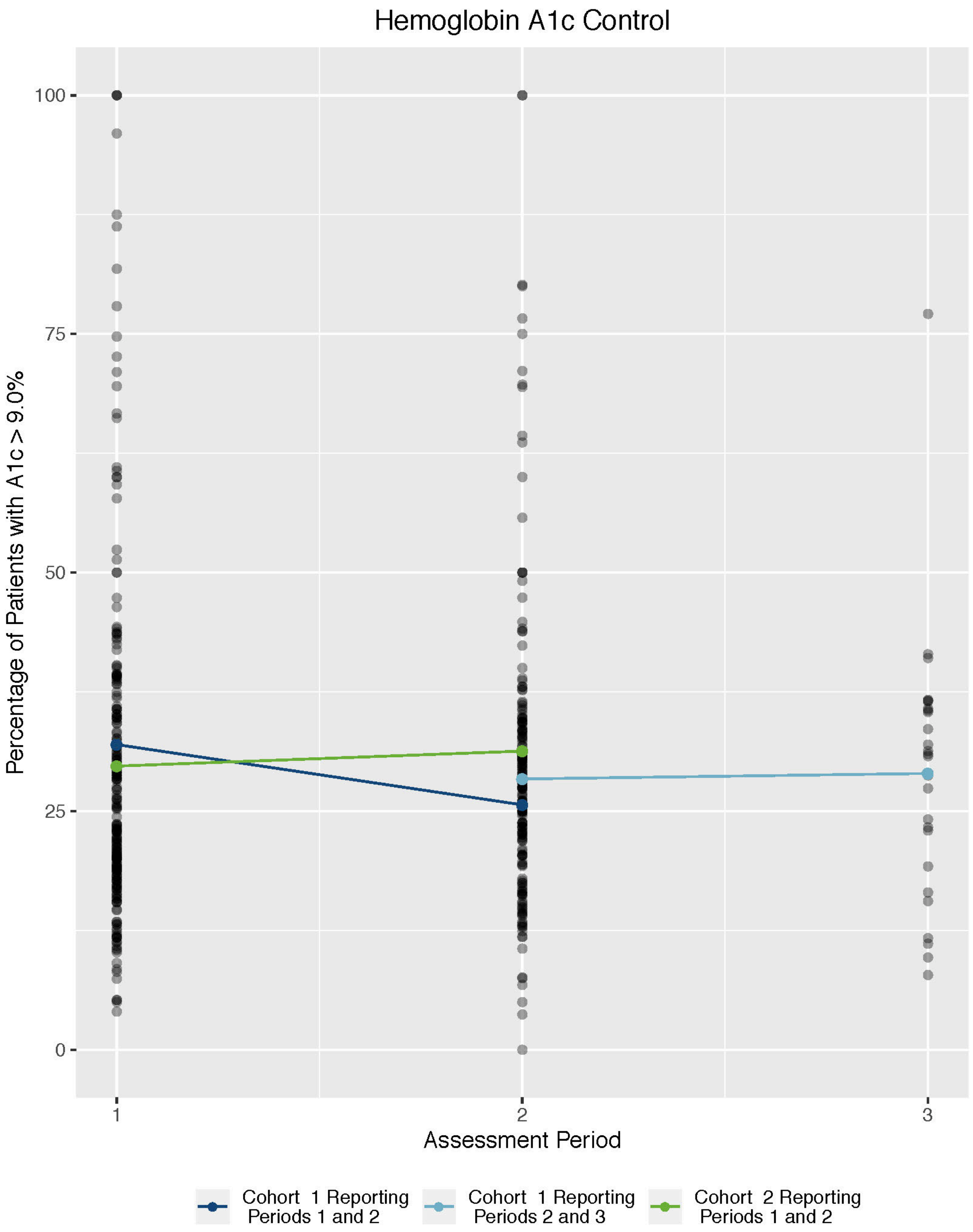


Table 53. Hemoglobin A1c Control

Hemoglobin A1c Control[63]	Baseline		First		Second		Third	
	%	N	%	N	%	N	%	N
Cohort 1	-	-	31.6%	61	26.1%	67	29.0%	27
Cohort 1 Practice Sites Reporting Periods 1 and 2	-	-	32.0%	59	25.7%	59	-	-
Difference/P-Value	-	-	-	-	-6.3% <0.01*	-	-	-
Cohort 1 Practice Sites Reporting Periods 2 and 3	-	-	-	-	28.4%	27	29.0%	27
Difference/P-Value	-	-	-	-	-	-	0.6% 0.84	-
Cohort 2	-	-	29.5%	88	32.5%	102	-	-
Cohort 2 Practice Sites Reporting Periods 1 and 2	-	-	29.7%	87	31.3%	87	-	-
Difference/P-Value	-	-	-	-	1.6% 0.40	-	-	-
Cohort 3	-	-	30.8%	53	-	-	-	-
Proxy CQM Cohort 1	65.2%	14,155	75.4%	15,941	76.2%	17,636	-	-
Proxy CQM Cohort 2	78.1%	10,330	79.1%	11,761	-	-	-	-
Proxy CQM Cohort 3	77.2%	6,601	-	-	-	-	-	-
Proxy CQM Statewide**	69.3%	167,937	76.8%	163,437	77.6%	185,206	-	-

For the proxy CQMs, careful attention must be paid to this measure: the SIM CQM measure looks at patients with **uncontrolled** diabetes defined as a hemoglobin A1c > 9.0, whereas the proxy measure looks at patients with diabetes who had hemoglobin A1c testing at an encounter. For the SIM measure, a **lower rate is desired** because it indicates fewer patients with uncontrolled diabetes. Conversely, for the proxy measure, a higher rate is desired because it indicates more of the population was screened.

The cohorts and statewide rates improved over all included periods from baseline to the second reporting period, indicating an increase in overall screening. Because the SIM rate increases were similar to the statewide increases, there does not appear to be a large change attributable to the SIM intervention. This is consistent with the SIM CQM rate change, in which only cohort 1 experienced improved performance.

---

[63] A single * indicates statistical significance (P-Value ≤ 0.05). A double ** indicates that statewide proxy CQM assessment periods correspond with cohort 1.

Summary of Hemoglobin A1c Control

- Cohort 1 had a statistically significant improvement (decline in patients with poor A1c control), but cohort 2 did not.
- Comparisons between practice-site-reported versus proxy measures are problematic because the measures are not aligned.

Hypertension Management

Hypertension Management is the second (along with poor A1c control) patient-outcome CQM. The percentage of SIM patients diagnosed with hypertension who successfully controlled their blood pressure remained virtually unchanged across any reporting period. However, in isolating the small number of cohort 1 practice sites that reported in both the second and third periods (n=27), we found a small but statistically significant improvement in the measure.

As plotted in Figure 29, there were practice site outliers (i.e., some very low, some very high), but most values remain clustered close to the average. This measure had much less fluctuation across practice sites than other measures had. See the summary table (Table 70) at the end of the chapter for a comparison of standard deviations.

Figure 29. Hypertension Management

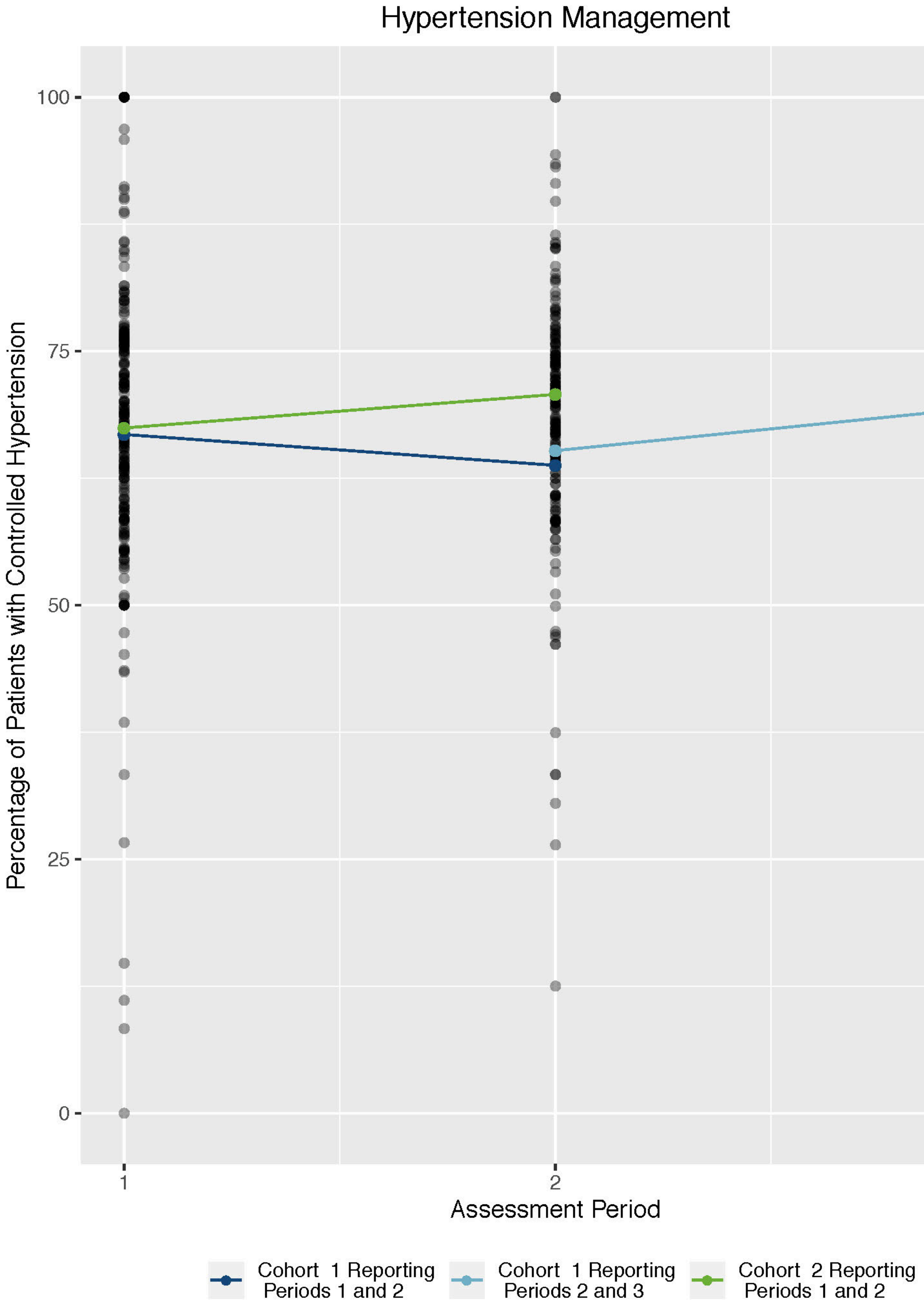


Table 54. Hypertension Management

Hypertension Management[64]	Baseline		First		Second		Third	
	%	N	%	N	%	N	%	N
Cohort 1	-	-	67.2%	55	65.3%	66	69.5%	27
Cohort 1 Practice Sites Reporting Periods 1 and 2	-	-	66.8%	51	63.8%	51	-	-
Difference/P-Value	-	-	-	-	-3.0% 0.12	-	-	-
Cohort 1 Practice Sites Reporting Periods 2 and 3	-	-	-	-	65.2%	27	69.5%	27
Difference/P-Value	-	-	-	-	-	-	4.3% <0.01	-
Cohort 2	-	-	67.6%	93	70.0%	100	-	-
Cohort 2 Practice Sites Reporting Periods 1 and 2	-	-	67.4%	88	70.8%	88	-	-
Difference/P-Value	-	-	-	-	3.3% 0.07	-	-	-
Cohort 3	-	-	66.7%	55	-	-	-	-
Proxy CQM Cohort 1	55.9%	40,527	55.4%	50,576	53.6%	53,042	-	-
Proxy CQM Cohort 2	55.1%	34,351	52.5%	36,627	-	-	-	-
Proxy CQM Cohort 3	51.2%	18,366	-	-	-	-	-	-
Proxy CQM Statewide**	47.0%	534,227	49.0%	548,963	50.4%	566,290	-	-

The proxy CQM uses the number of patients with a hypertension diagnosis as its denominator and the number of patients with a 90-day supply of medication as its numerator. As such, the proxy CQM is more of a process measure than is the SIM CQM. But like the SIM CQM, the proxy measure also showed little change. The SIM cohorts saw small decreases over time whereas the statewide proxy measure increased slightly. Although we do not report on statistical significance for proxy measure changes, the small magnitude of all changes could indicate generally no impact of SIM on this measure.

Summary of Change in Hypertension Management

- Both cohorts experienced very little change in this CQM over time.

---

[64] A single * indicates statistical significance (P-Value ≤ 0.05). A double ** indicates that statewide proxy CQM assessment periods correspond with cohort 1.

Adult Obesity Screening

Both cohorts 1 and 2 showed statistically significant improvements in levels of adult obesity screening. Cohort 1’s change was between the second and third period and was based on a relatively small number of practice sites. Cohort 2’s improvement, shown as a dark blue line in Figure 30, occurred between the first and second assessment period. Although there were changes in the average values, the assessment periods 1 and 2 plots show considerable variation across practice sites, with some sites having reported percentages below 25% and a few having reported that close to 100% of adult patients were screened for obesity.

Figure 30. Adult Obesity Screening

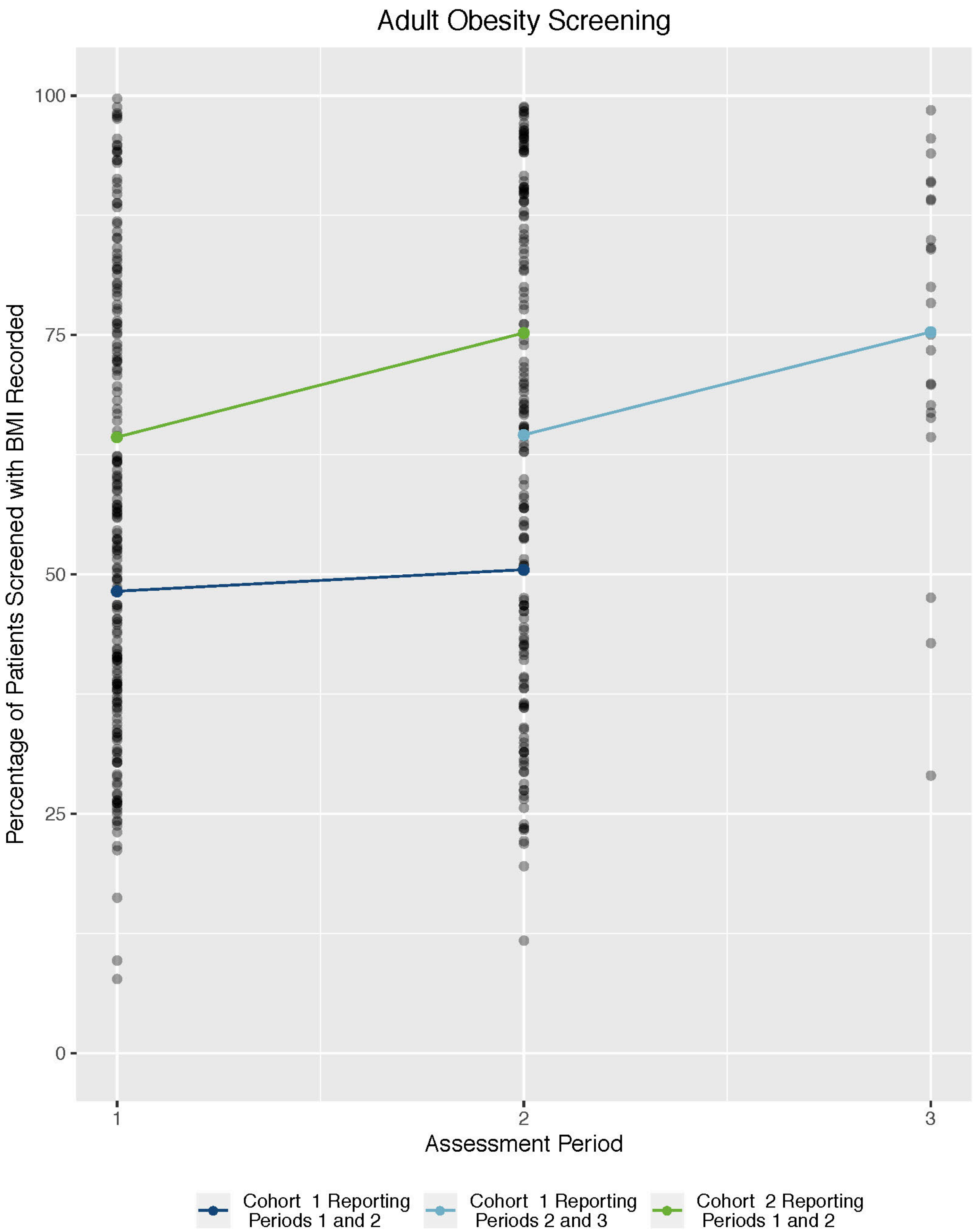


Table 55. Adult Obesity Screening

Adult Obesity Screening[65]	Baseline		First		Second		Third	
	%	N	%	N	%	N	%	N
Cohort 1	-	-	48.9%	64	48.6%	65	75.3%	23
Cohort 1 Practice Sites Reporting Periods 1 and 2	-	-	48.2%	59	50.5%	59	-	-
Difference/P-Value	-	-	-	-	2.3% 0.36	-	-	-
Cohort 1 Practice Sites Reporting Periods 2 and 3	-	-	-	-	64.6%	23	75.3%	23
Difference/P-Value	-	-	-	-	-	-	10.7% <0.01*	-
Cohort 2	-	-	64.1%	75	72.3%	94	-	-
Cohort 2 Practice Sites Reporting Periods 1 and 2	-	-	64.3%	71	75.2%	71	-	-
Difference/P-Value	-	-	-	-	10.9% <.01*	-	-	-
Cohort 3	-	-	53.9%	48	-	-	-	-
Proxy CQM Cohort 1	65.2%	14,155	48.2%	15,405	48.9%	18,678	-	-
Proxy CQM Cohort 2	49.8%	10,002	46.7%	12,908.	-	-	-	-
Proxy CQM Cohort 3	53.9%	5,114	-	-	-	-	-	-
Proxy CQM Statewide**	69.3%	167,937	50.7%	148,983	51.2%	178,493	-	-

The SIM CQM uses the count of patients with a primary care visit as the denominator and the number with both screening and a follow-up plan if needed as the numerator. It is therefore the prevalence of correct screening and follow up. The proxy measure uses the number of patients with an obesity diagnosis in the denominator and the number with a specific type of treatment in the numerator. It is therefore a process measure; it reports the percentage of patient who need treatment and receive it. In contrast to the SIM CQMs, the proxy CQMs show declining rates for SIM practice sites. Proxy measures decreased between the baseline and first reporting periods for all groups that reported in both periods. More data are necessary for a better understanding of these trends, although, as is the case with other proxy measures, low reimbursement rates may mean that practice sites are not billing for obesity screening.

---

[65] A single * indicates statistical significance (P-Value ≤ 0.05). A double ** indicates that statewide proxy CQM assessment periods correspond with cohort 1.

Adult Obesity Screening Summary

- Cohorts 1 and 2 both saw statistically significant improvements in this measure during their participation in SIM.

Asthma Medication Management

There was less variation across practice sites in the percentage of patients with asthma where the condition was being well-managed with medication. Most sites tended to report fairly high percentages, with a few lower outliers in all reporting periods.

Both cohorts 1 and 2 saw improvement over time. Cohort 1 had statistically significant increases between assessment periods 1 and 2 and maintained the gains post SIM. Cohort 2, which started Assessment Period 1 with higher average values, had improvement over time but the change was not statistically significant.

Figure 31. Asthma Medication Management

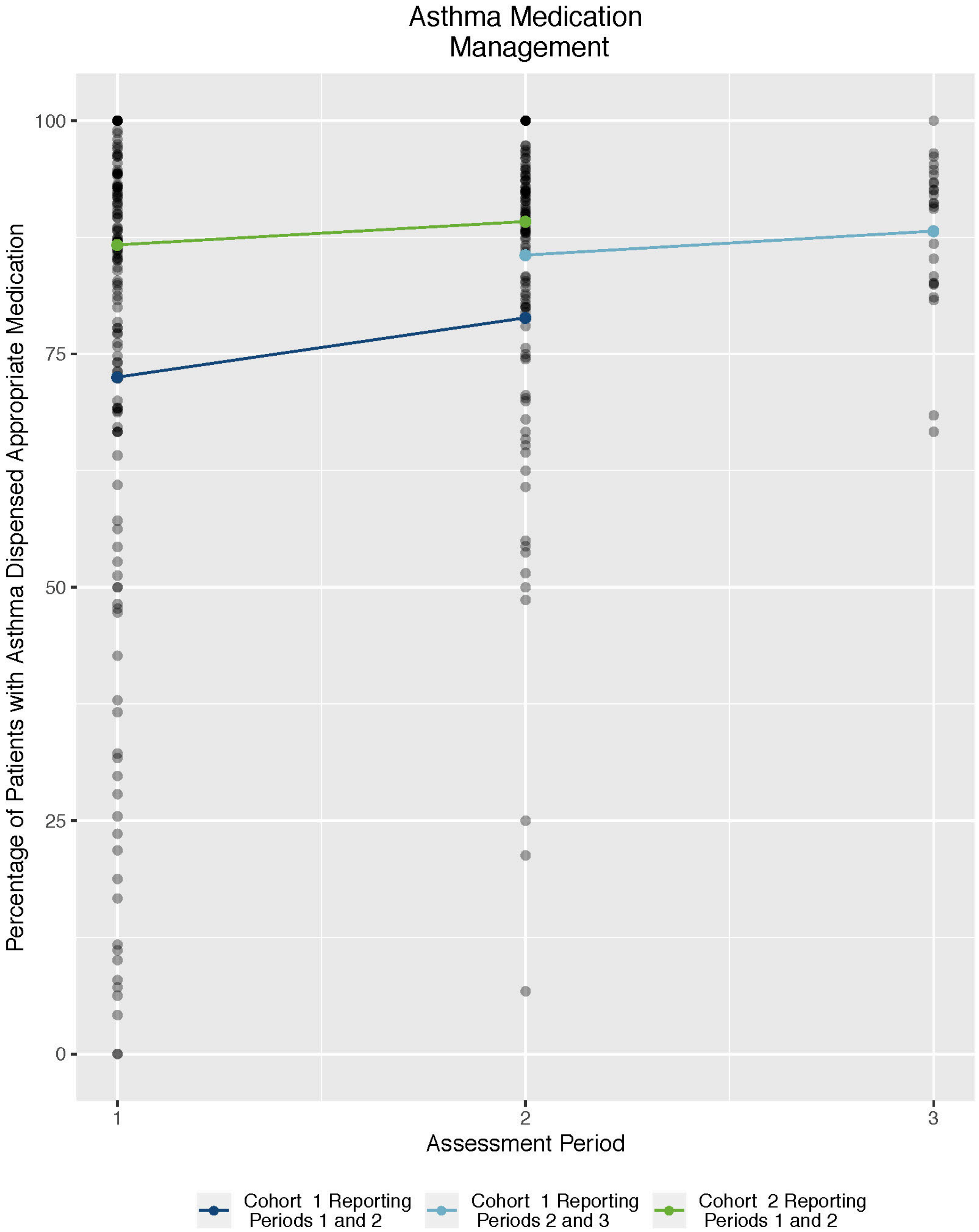


Table 56. Asthma Medication Management

Asthma Medication Management[66]	Baseline		First		Second		Third	
	%	N	%	N	%	N	%	N
Cohort 1	-	-	68.1%	67	78.3%	38	88.2%	25
Cohort 1 Practice Sites Reporting Periods 1 and 2	-	-	72.5%	34	78.9%	34	-	-
Difference/P-Value	-	-	-	-	6.4% 0.02*	-	-	-
Cohort 1 Practice Sites Reporting Periods 2 and 3	-	-	-	-	85.6%	25	88.2%	25
Difference/P-Value	-	-	-	-	-	-	2.6% 0.408	-
Cohort 2	-	-	83.2%	57	86.6%	65	-	-
Cohort 2 Practice Sites Reporting Periods 1 and 2	-	-	86.7%	46	89.2%	46	-	-
Difference/P-Value	-	-	-	-	2.5% 0.34	-	-	-
Cohort 3	-	-	75.7%	22	-	-	-	-
Proxy CQM Cohort 1	9.9%	5,240	8.5%	5,403	8.2%	5,164	-	-
Proxy CQM Cohort 2	9.1%	3,621	7.2%	3,644	-	-	-	-
Proxy CQM Cohort 3	9.4%	1,602	-	-	-	-	-	-
Proxy CQM Statewide**	10.2%	45,210	9.6%	39,948	7.9%	39,689	-	-

The proxy measures had values approximately one eighth of the SIM CQM, highlighting the problem in comparing levels between the CQMs and their proxies. The time trends also moved in the opposite direction: the SIM CQMs increased over time while the proxies decreased. Given the extremely low values, the matching statewide decline, and the increases in the SIM CQMs, these proxy measures do not provide helpful information in identifying if the gains made by the SIM practice sites were part of statewide gains or were caused by SIM participation.

Asthma Medication Management Summary

- Both cohorts 1 and 2 in the practice-site-reported CQM measure had increases over time, but cohort 2's gains were not statistically significant. This may be caused by

---

[66] A single * indicates statistical significance (P-Value ≤ 0.05). A double ** indicates that statewide proxy CQM assessment periods correspond with cohort 1.

very high starting scores. The proxy measure showed the opposite pattern for the two cohorts, and the declining rates were matched by declines in the statewide rate.

Fall Risk Screening

Cohorts 1 and 2 demonstrated statistically significant increases in fall risk screening between periods 1 and 2. For cohort 1, the change between period 2 and 3 was also positive but not statistically significant because of the small number of practice sites that reported in period 3. A wide variance in reported practice site values occurred across all reporting periods.

Figure 32. Fall Risk Screening

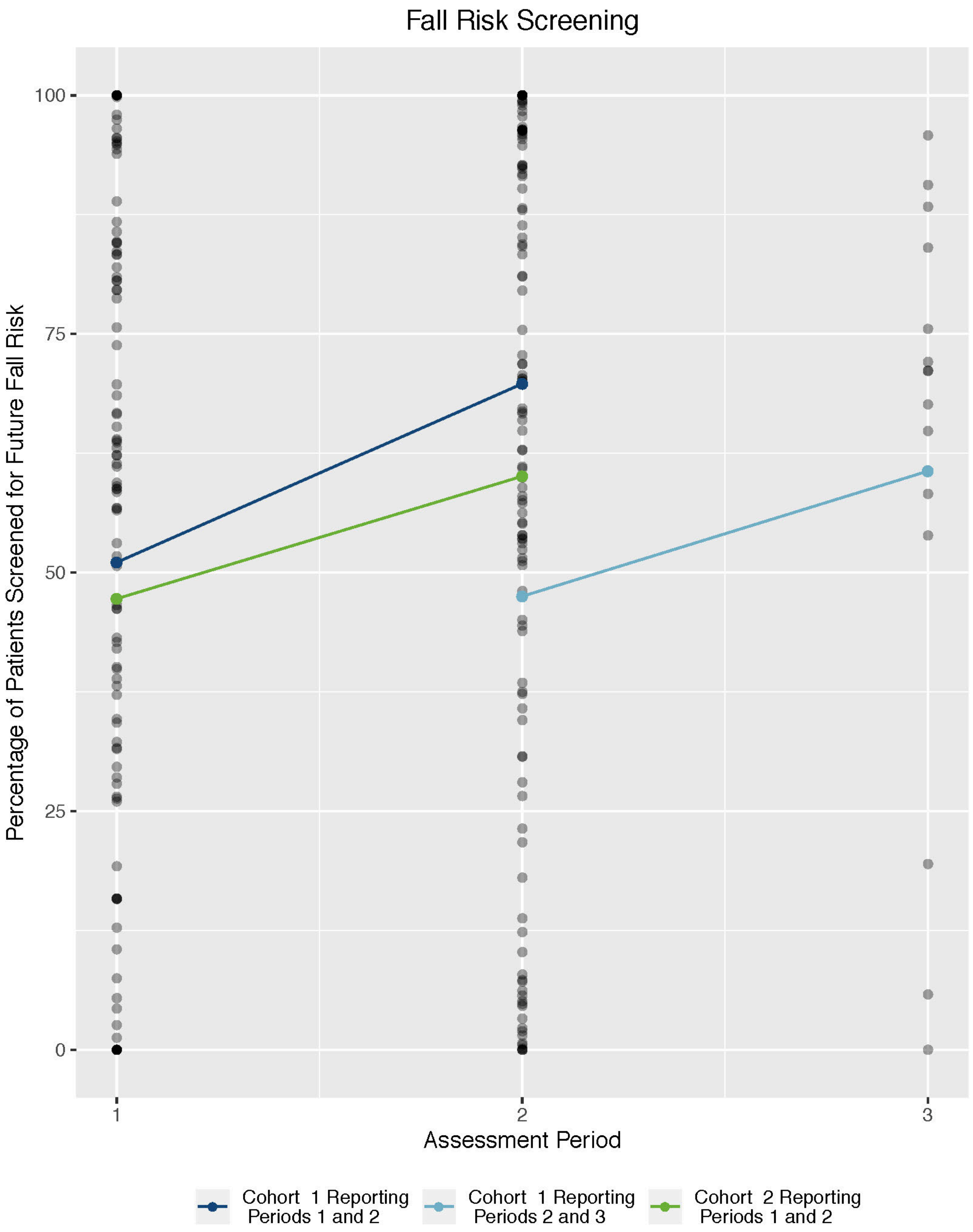


Table 57. Fall Risk Screening

Fall Risk Screening[67]	Baseline		First		Second		Third	
	%	N	%	N	%	N	%	N
Cohort 1	-	-	46.1%	24	63.4%	49	61.2%	15
Cohort 1 Practice Sites Reporting Periods 1 and 2	-	-	51.1%	21	69.8%	21	-	-
Difference/P-Value	-	-	-	-	18.7% <0.01*	-	-	-
Cohort 1 Practice Sites Reporting Periods 2 and 3	-	-	-	-	47.5%	13	60.6%	13
Difference/P-Value	-	-	-	-	-	-	13.1% 0.05*	-
Cohort 2	-	-	48.1%	51	55.6%	64	-	-
Cohort 2 Practice Sites Reporting Periods 1 and 2	-	-	47.3%	50	60.1%	50	-	-
Difference/P-Value	-	-	-	-	12.8% <0.01*	-	-	-
Cohort 3	-	-	56.6%	38	-	-	-	-
Proxy CQM Cohort 1	33.8%	34,207	34.9%	39,494	37.7%	46,546	-	-
Proxy CQM Cohort 2	35.2%	27,993	40.6%	29,922	-	-	-	-
Proxy CQM Cohort 3	51.9%	17,133	-	-	-	-	-	-
Proxy CQM Statewide	33.80%	529,200	36.20%	528,450	40.40%	549,752		

The fall risk proxy CQM measure is a true proxy, essentially measuring wellness visits for older patients with the assumption that fall risk screening occurred. Both the time path and level of screening rates were similar for the cohorts and the statewide measure, with steady increases over time. These increased rates match the pattern of increasing SIM CQM rate but do not provide evidence that the increases were driven by SIM since the statewide proxy rates also increased.

Fall Risk Screening Summary

- Both cohorts 1 and 2 reported statistically significant improvements in practice-site-reported fall screenings during their participation in SIM.
- Cohort 1 had additional increase post SIM.

---

[67] A single * indicates statistical significance (P-Value ≤ 0.05). A double ** indicates that statewide proxy CQM assessment periods correspond with cohort 1.

Developmental Screening

For both cohorts 1 and 2, the average practice site result of patients under three years of age who were screened for developmental or behavioral delays increased over the first reporting period. Cohort 1 had a very small increase between the assessment periods, and only four practice sites reported in both periods. These factors contributed to the lack of statistical significance. Cohort 2 had more practice sites that reported both periods (n = 14) and a larger increase, but the change was also not statistically significant. The number of practice sites that reported in both periods was very small. Despite the low number of reports, those sites that submitted screening data clustered far above the SIM target of 16%.

Figure 33. Developmental Delay Screening

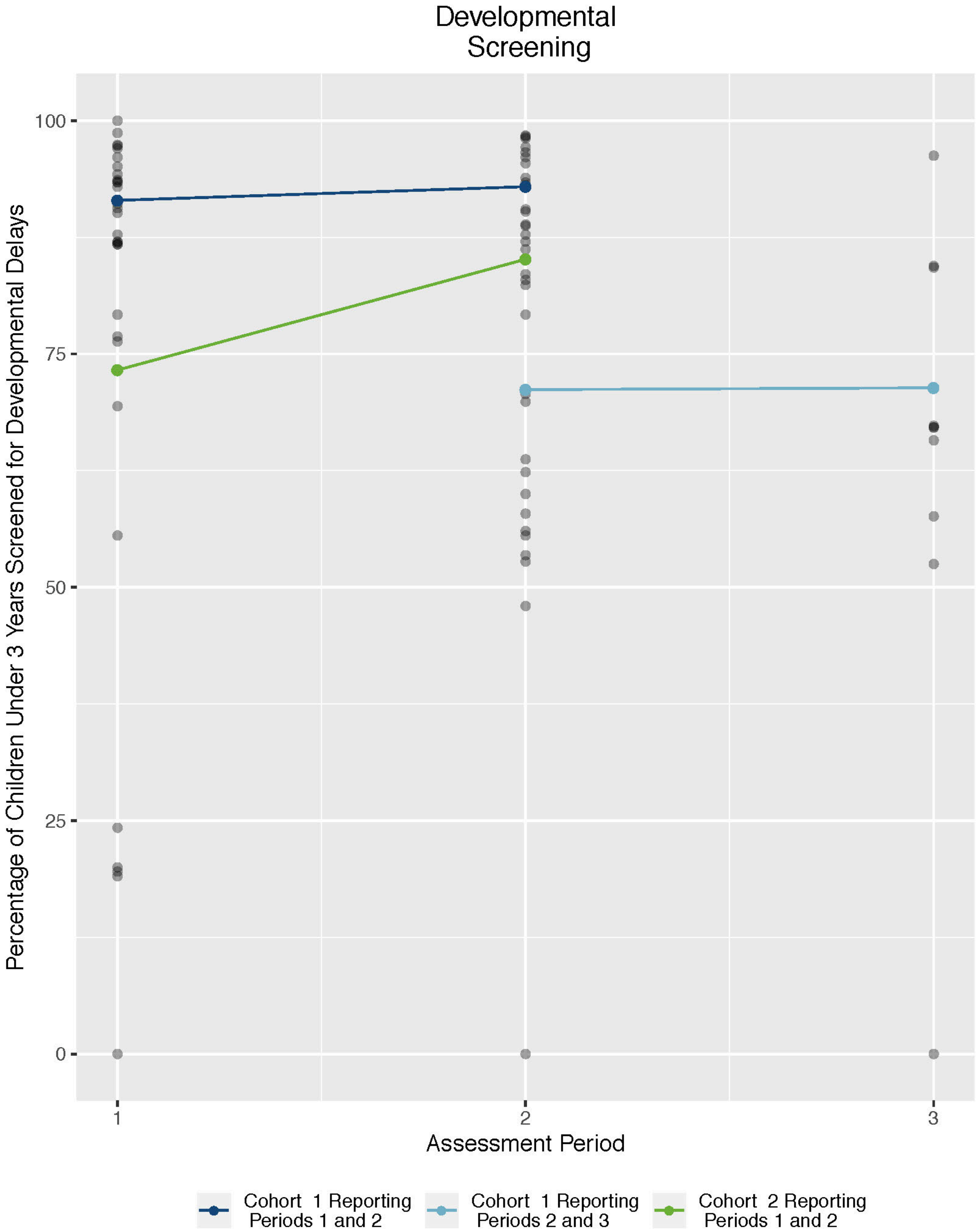


Table 58. Developmental Delay Screening

Developmental Screening[68]	Baseline		First		Second		Third	
	%	N	%	N	%	N	%	N
Cohort 1	-	-	90.5%	5	78.0%	14	64.2%	10
Cohort 1 Practice Sites Reporting Periods 1 and 2	-	-	91.5%	4	92.9%	4	-	-
Difference/P-Value	-	-	-	-	1.5% 0.78	-	-	-
Cohort 1 Practice Sites Reporting Periods 2 and 3	-	-	-	-	71.2%	9	71.4%	9
Difference/P-Value	-	-	-	-	-	-	0.2% 0.96	-
Cohort 2	-	-	73.3%	14	77.6%	20	-	-
Cohort 2 Practice Sites Reporting Periods 1 and 2	-	-	73.3%	14	85.2%	14	-	-
Difference/P-Value	-	-	-	-	11.9% 0.10	-	-	-
Cohort 3	-	-	74.4%	11	-	-	-	-
Proxy CQM Cohort 1	74.8%	8,901	71.0%	10,021	77.0%	10,009	-	-
Proxy CQM Cohort 2	75.5%	4,920	76.4%	5,018	-	-	-	-
Proxy CQM Cohort 3	65.2%	1,424	-	-	-	-	-	-
Proxy CQM Statewide**	66.4%	45,518	67.8%	45,604	72.8%	46,403	-	-

Statewide proxy CQMs showed gradual increases over time. SIM cohorts generally also showed increases and were at higher levels than were the statewide measure. The proxy measures were of a similar magnitude to the SIM CQMs, supporting the conclusion that screening was widespread and increased over time.

Developmental Delay Summary

- Practice-site-reported CQM values for developmental delay screening showed small increases over time. However, too few practice sites reported to draw meaningful conclusions.

---

[68] A single * indicates statistical significance (P-Value ≤ 0.05). A double ** indicates that statewide proxy CQM assessment periods correspond with cohort 1.

- SIM cohorts generally showed increases over time in the proxy CQM for development screening, but these increases were matched by increases in the statewide measure.

Adolescent Obesity

The overall cohort average percentage of children aged 3–17 screened for obesity increased substantially for cohort 2 between the first and second assessment period. This improvement was statistically significant.

For cohort 1 there was first a decline between the first and second period, followed by an increase between the second and third. The small number of reporting practice sites contributed to the lack of a statistically significant change. It is visually apparent that practice sites clustered tightly around 100%, with a long tail of each distribution declining as low as zero.

There were no proxy measures for this CQM.

Figure 34. Adolescent Obesity Screening

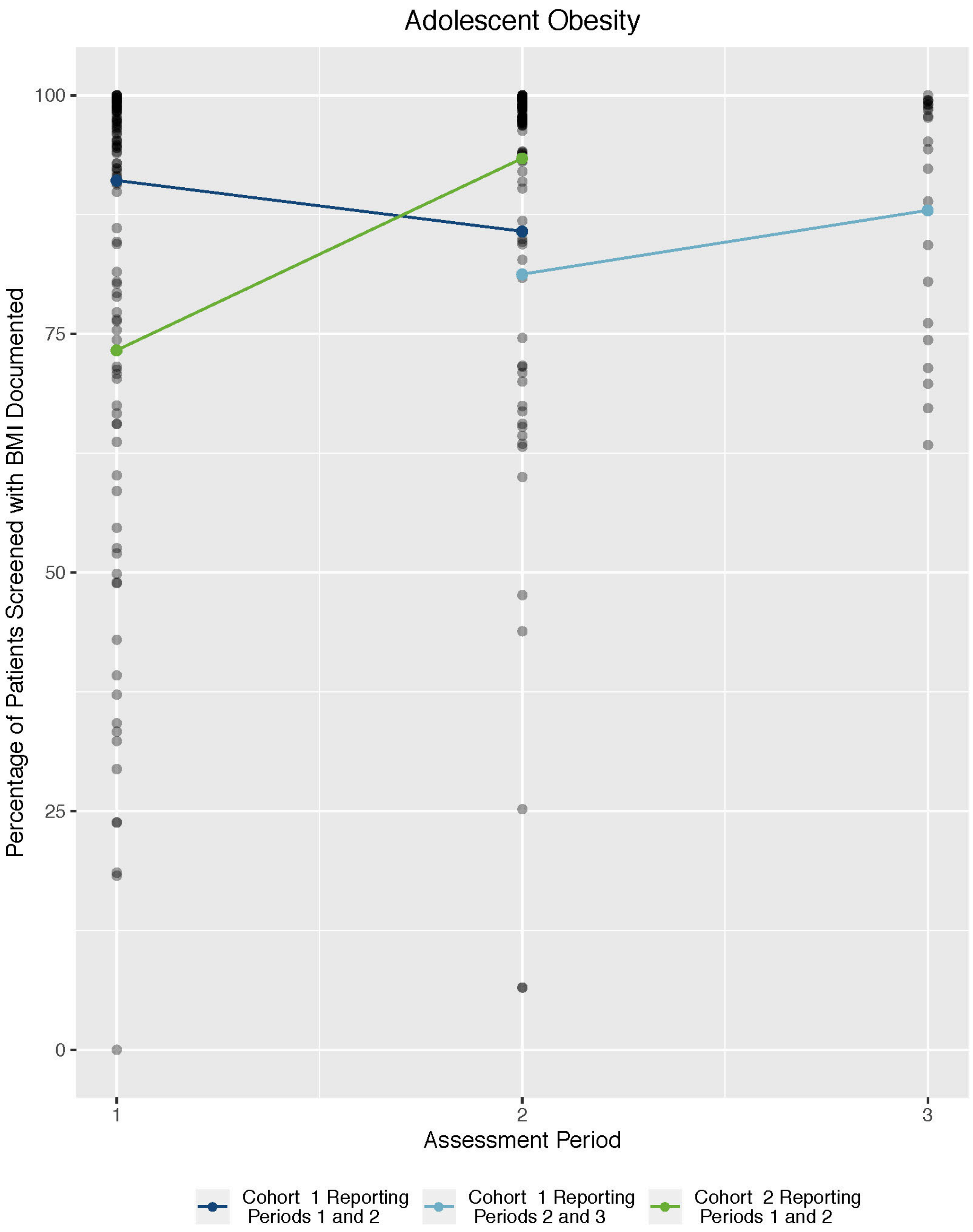


Table 59. Adolescent Obesity Screening

Adolescent Obesity[69]	Baseline		First		Second		Third	
	%	N	%	N	%	N	%	N
Cohort 1	-	-	89.9%	22	83.0%	27	88.5%	22
Cohort 1 Practice Sites Reporting Periods 1 and 2	-	-	91.1%	19	85.7%	19	-	-
Difference/P-Value	-	-	-	-	-5.3% 0.10	-	-	-
Cohort 1 Practice Sites Reporting Periods 2 and 3	-	-	-	-	81.3%	17	88.0%	17
Difference/P-Value	-	-	-	-	-	-	6.7% 0.15	-
Cohort 2	-	-	74.8%	49	90.4%	60	-	-
Cohort 2 Practice Sites Reporting Periods 1 and 2	-	-	73.3%	46	93.4%	46	-	-
Difference/P-Value	-	-	-	-	20.1% <0.01*	-	-	-
Cohort 3	-	-	82.5%	30	-	-	-	-

Summary of Adolescent Obesity Screening Changes

- There was a statistically significant improvement in this measure for cohort 2 practice sites.

CMHC CQM Values

The four CMHCs reported CQMs at two points during their four years of SIM participation: at the end of the second (2017) and third (2018) implementation years. Not all CMHCs reported all CQMs, and in order to preserve anonymity, we are only including averages for CQMs in which at least two sites reported.

Because our largest sample size is four, we do not show time series graphs or calculate any statistical tests. Between the two assessment periods, the average value for the CMHC CQMs showed improvement for *Depression Screening* and *Alcohol and Other Drugs Screening*. The average declined for *Hypertension Management*, *Adult Obesity Screening*, and *Asthma Medication*. *Hemoglobin A1c Control* increased, which represents a larger proportion of clients with hemoglobin A1c above recommended levels. *Tobacco Use Screening* was only reported in one period.

---

[69] A single * indicates statistical significance (P-Value ≤ 0.05). A double ** indicates that statewide proxy CQM assessment periods correspond with cohort 1.

The declining CMHC CQM measures must be interpreted in the context of a much later initial value than existed for primary care practice sites. The first value we received for the CMHCs was at the end of Year 2. Change, therefore, does not reflect the potentially large changes that might have occurred from an earlier baseline period.

Table 60. CQM Measures: Second and Third Assessment Periods (CMHCs)

CQM	Second Assessment Period		Third Assessment Period	
	%	N	%	N
Depression Screening	61.20%	4	65.50%	4
Alcohol and Other Drugs Screening	31.20%	3	40.10%	3
Tobacco Use Screening	71.50%	4	-	-
Hemoglobin A1c Control	21.70%	4	28.10%	4
Hypertension Management	75.40%	4	71.90%	4
Adult Obesity Screening	68.30%	4	62.80%	4
Asthma Medication Management	94.10%	2	88.60%	2

CQM Target Values

The SIM office established program-wide targets for the CQMs. No individual practice site was expected to achieve these targets; targets were instead for all patients served by SIM primary care practice sites. Because of this, we aggregated all CQM values across SIM primary care practice sites for each assessment period. This process involved combining cohorts from different calendar years corresponding to each cohort’s assessment period. As an example, we based Assessment Period 1 on 2016 for cohort 1 and based it on 2017 for cohort 2.

We did not average all reporting practice sites. Instead, we added up across practice sites the numerators and denominators used to by sites to calculate their own value of each metric. The resulting CQM value is for **patients served in SIM sites, not an average for SIM practice sites themselves**.

CQMs with targets achieved or surpassed by period 2 or period 3 include *Depression Screening*, *Maternal Depression Screening*, *Tobacco Use Screening*, *Hemoglobin A1c Control*, *Hypertension Management, Adult Obesity Screening*, *Fall Risk Screening*, and *Developmental Screening*. *Tobacco Use Screening* declined below the target in period 3 but also had a small portion of practice sites that report the measure (all cohort 1). *Maternal Depression Screening*, *Hypertension Management*, and *Fall Risk Screening* did not meet the target by Assessment

Period 2 but were satisfied by the smaller number of practice sites reporting in Assessment Period 3.

CQMs with assessment-wide values that did not meet or surpass the target values include *Alcohol and Other Drugs Screening* and *Alcohol Use Screening*. We note that these are two of the three CQMs related to substance use disorder and that the third, *Tobacco Use Screening*, met but did not exceed its target.

Table 61. CQM Means and Targets (All Cohorts)

CQM	Target	Period 1		Period 2		Period 3	
		Mean	N	Mean	N	Mean	N
Depression Screening	52%	49.70%	272	61.20%	196	70.50%	38
Maternal Depression	76%	60.40%	131	66.00%	51	80.50%	13
Alcohol and Other Drugs Screening	94%	63.00%	189	40.90%	163		
Tobacco Use Screening	93%	92.30%	160	93.00%	177	87.20%	28
Alcohol Use Screening	94%	34.70%	17	34.10%	10		
Hemoglobin A1c Control	30%	30.80%	238	30.00%	173	28.70%	27
Hypertension Management	70%	68.40%	245	69.10%	171	70.60%	27
Adult Obesity Screening	55%	59.50%	219	60.20%	159	77.90%	23
Asthma Medication Management	65%	74.80%	174	80.70%	110	89.00%	26
Fall Risk Screening	64%	63.70%	180	15.50%	115	64.20%	15
Developmental Screening	16%	90.50%	116	85.50%	36	77.40%	10
Adolescent Obesity	95%	87.20%	170	87.20%	89	90.40%	22

Relationship Between Practice Transformation and Changes in Clinical Quality

PT6. What specific transformation factors (level of integration, milestone targets, data quality, clinician and staff experience, etc.) most influence outcomes?

In the previous section, we reported change over time in the CQMs. In this section, we measure the effect of increased integration, improvements in the staff survey, and improvements in HIT on the CQMs. We address these items with four interrelated analysis:

1. **For Assessment Period 1, and each cohort separately, did practice sites with higher practice transformation scores also have higher CQM scores?** We used a cross-sectional (between practice site) regression based on the first assessment period CQMs and first assessment transformation scores to measure this relationship. This approach does not address change over time.
2. **For Assessment Period 2 (and 3 separately when available), and each cohort separately, did practice sites with higher transformation scores also have higher CQM scores?** This analysis duplicates the previous analysis but focuses on the end of the SIM intervention. We anticipated that practice sites with high practice transformation scores would, by the end, also achieve high CQM scores. As with the previous analysis, this approach is based on a cross-sectional regression and does not address change over time.
3. **For practice sites that reported CQMs in assessment periods 1 and 2, or 2 and 3, did sites with greater improvement in practice transformation factors also have greater improvement in CQMs?** This analysis is based on regressions run over changes in CQMs and the transformation measures and, therefore, explicitly analyzes change over time.
4. **As a variation on the third analysis approach, we considered whether controlling for practice site characteristics changed the relationship between changes in CQMs and transformation factors.** Practice site characteristics include practice site type (pediatric, adult, mixed), urbanicity (urban vs. rural), size, and proportion of underserved population. Each of these is a categorical variable.

In the tables that follow, we report results from each of these analyses separately for each CQM and cohort. Cohort 1 had three periods of data, resulting in two changes. Cohort 2 had two periods of data, and one change. Cohort 3 had only Assessment Period 1 data and no change over time. We also performed these same three analyses for the primary care practice sites in aggregate. Results did not differ, so we are excluding reporting any aggregate results.

Regarding community mental health centers, the small sample size prevents us from reporting statistical tests on CMHCs.

Effect of Integration on CQMs

In the first section of the table below (Table 62), we present, for each assessment period and cohort, the results from a regression, with the *Depression Screening* CQM serving as the dependent variable and integration measure serving as the independent variable. We summarize each result with three numbers. The first, labeled “Effect,” is the estimated change in the *Depression Screening* CQM from a one-percentage-point increase in the integration measure. In the case of cohort 1, Assessment Period 1 (2016), each percentage-point increase in integration associated with a 0.41 percentage-point increase in the *Depression Screening* CQM.

The P-Value is a measure of statistical significance calculated with paired T-tests, and we designate any results with a P-Value less than or equal to 0.05 as statistically significant.

“N” represents the number of practice sites with complete data. In this case, practice sites with higher integrations scores also had higher CQM scores, and the result was statistically significant.

The other cohorts had very similar results in their first assessment period. Those practice sites that were more integrated also had higher rates of depression screening. By the second assessment period, at the end of SIM participation, these results broke down. Practice sites with higher levels of integration did not report statistically significant higher rates of depression screening. A subset of practice sites reported in the third assessment periods, which corresponds to post SIM. The effect size was larger than it was in Assessment Period 2, and results were close to significant.

The row labeled “1–2” represents the change-over-time analysis. In this case, practice site increases of one percentage point between assessment periods 1 and 2 associated with an increase of 0.050 in the practice site Depression Screening CQM. This is a small change, and it is not statistically significant. Cohort 2 had similar results.

For cohort 1, the period 2 to 3 changes in the degree of integration and CQM reporting were negatively related (-0.144), which means that practice sites with increases in integration had decreases in the *Depression Screening* CQM. This result is also not statistically significant.

Based on this analysis, changes over time in the level of integration were uncorrelated with changes in the *Depression Screening* CQM (*Depression Screening* was the CQM with the largest number of reporting practice sites). The other CQMs generally did not have statistically significant relationships between the CQM and integration, either in first differences, or cross sectionally. The exception to this is the *Fall Screening* CQM. For cohort 1, increased integration caused increases in fall screening. But for cohort 2 there was a non-significant decreasing relationship.

Table 62. Relationship Between Integration Score and CQM results

Assessment Period	Cohort 1			Cohort 2			Cohort 3		
	Effect	P-Value[70]	N	Effect	P-Value	N	Effect	P-Value	N
Depression Screening									
1	0.41	0.01*	82	0.43	<0.01*	112	0.55	0.02*	69
2	0.23	0.37	74	0.03	0.84	120	-	-	-
3	0.49	0.07	38	-	-	-	-	-	-
1–2	0.05	0.76	69	0.10	0.34	101	-	-	-
2–3	-0.14	0.39	37	-	-	-	-	-	-
Maternal Depression Screening									
1	0.15	0.76	12	0.18	0.63	17	0.40	0.48	21
2	0.35	0.40	16	-0.27	0.55	31	-	-	-
3	0.73	0.10	13	-	-	-	-	-	-
1-2	-1.39	0.27	11	-0.23	0.46	17	-	-	-
2-3	-0.83	0.39	7	-	-	-	-	-	-
Alcohol and Other Drug Screening									
1	1.13	0.32	6	0.14	0.75	15	-0.42	0.54	10
2	0.48	0.11	21	0.06	0.85	21	-	-	-
3	0.20	-	2	-	-	-	-	-	-
1–2	0.72	0.65	4	0.46	0.67	6	-	-	-
2–3	0.63	-	2	-	-	-	-	-	-
Hemoglobin A1c Control									
1	-0.09	0.46	61	0.08	0.40	88	-0.15	0.45	53
2	-0.03	0.74	68	-0.05	0.68	102	-	-	-
3	-0.48	0.02*	27	-	-	-	-	-	-
1–2	-0.13	0.31	59	-0.05	0.56	87	-	-	-
2–3	-0.81	<0.01*	27	-	-	-	-	-	-
Hypertension Management									
1	-0.03	0.75	55	0.09	0.22	93	0.20	0.28	55
2	0.12	0.18	67	-0.11	0.20	100	-	-	-
3	0.16	0.14	27	-	-	-	-	-	-
1–2	-0.05	0.71	51	0.02	0.85	88	-	-	-
2–3	-0.24	0.03*	27	-	-	-	-	-	-
Adult Obesity Screening									
1	0.13	0.29	64	0.36	0.00*	75	0.21	0.43	48
2	0.28	0.09	65	-0.12	0.43	94	-	-	-
3	0.33	0.27	23	-	-	-	-	-	-
1–2	-0.08	0.60	59	0.21	0.03*	71	-	-	-

---

[70] A single * indicates statistical significance (P-Value ≤ 0.05).

Assessment Period	Cohort 1			Cohort 2			Cohort 3		
	Effect	P-Value[70]	N	Effect	P-Value	N	Effect	P-Value	N
2–3	-0.01	0.97	23	-	-	-	-	-	-
Asthma Medication Management									
1	0.01	0.95	67	0.10	0.48	57	0.57	0.15	22
2	-0.12	0.63	38	-0.24	0.03*	65	-	-	-
3	0.41	<0.01*	25	-	-	-	-	-	-
1–2	-0.11	0.51	34	0.27	0.11	46	-	-	-
2–3	0.04	0.89	25	-	-	-	-	-	-
Fall Risk Screening									
1	0.22	0.57	24	0.25	0.51	51	1.30	<0.01*	38
2	0.16	0.61	50	0.34	0.18	64	-	-	-
3	0.69	0.21	15	-	-	-	-	-	-
1–2	0.87	0.01	21	-0.22	0.12	50	-	-	-
2–3	0.62	0.18	13	-	-	-	-	-	-
Developmental Screening									
1	0.03	0.87	5	-0.42	0.50	14	-0.52	0.41	11
2	-0.19	0.67	14	-0.11	0.82	20	-	-	-
3	-0.44	0.82	10	-	-	-	-	-	-
1–2	-1.33	0.11	4	-0.31	0.49	14	-	-	-
2–3	-0.29	0.42	9	-	-	-	-	-	-
Adolescent Obesity									
1	-0.11	0.55	22	-0.18	0.39	49	-0.15	0.61	30
2	-0.15	0.63	28	-0.13	0.62	60	-	-	-
3	-0.47	0.06	22	-	-	-	-	-	-
1–2	0.33	0.12	19	0.29	0.28	46	-	-	-
2–3	-0.02	0.97	17	-	-	-	-	-	-

Because practice transformation is based on the expectation that increased integration should result in improvements in clinical quality, we attempted to use other variations of this analysis to identify evidence of a statistically significant relationship. We focused on the first CQM, *Depression Screening*, since that measure is directly related to primary and behavioral health integration and had some of the largest sample sizes.

Our first analysis variation used a different modeled relationship between changes in integration and changes in clinical quality. Previously, we assumed a simple linear model. Because we were not able to find a statistically significant relationship, we tried to use a quadratic model, which allows changes in integration to have an increasing or decreasing effect on depression screening. Neither cohort 1 or 2 had a statistically significant quadratic relationship between changes in depression screening and changes in integration.

We next tried to control for practice site characteristics (i.e., type, urbanicity, size, proportion of typically underserved patients). Adding these characteristics to the cross-sectional (i.e., at a single time period) regressions yielded some statistically significant results. For example, in cohort 1's first period, the practice sites characterized as medium or very high with the underserved variable had statistically significant lower average depression screening rates as compared to practice sites characterized as high. Because these results do not involve changes in integration or the CQM, we do not report the cross-sectional results. Most of the cross-sectional results were not significant.

Of greater relevance for assessing the effect of changes of integration on changes in the CQM values, we added these same practice site variables to the change-over-time regressions. This form of analysis allows the impact of changes in integration on depression screening or other measure to vary by practice site type. By pooling all practice site types together, we potentially miss any statistically significant relationships—since a positive relationship for one group of practice sites could be offset by a negative relationship in a different group.

The structure of this analysis is to interact (multiply) the practice site characteristics variables with the integration variable. However, only the urban and practice site size characteristic variables were statistically significant. For cohort 2 only, adding the categorical variable related to location (urban, rural) to the regression made the integration measure significant at the 0.09 significance level. We generally follow the convention of identifying variables as statistically significant if their P-Value is less than or equal to 0.05. In this case, adding the degree of urbanization of the practice site made the integration measure close to statistically significant (the urban variable was not statistically significant itself, but controlling for it allowed the integration measure alone to be close to significant).

The other variable closest to statistical significance, for cohort 1 only, was practice size. Interacting this variable with the integration variable yielded the result that for small practice sites, the effect of changes of integration on the *Depression Screening* measure were larger than for large practice sites. As small-size sites increased their degree of integration over time, their improvement in depression screening was larger than was improvement for large practice sites. This result had a P-Value of 0.07.

No other use of the practice site characteristics variables yielded statistically significant results. Because the change over time results did not yield a statistically significant relationship, and because adding practice site characteristics did not reveal a statistically significant relationship, we conclude that no robust relationship exists between changes in the integration measure and changes in the CQM reporting rates. Based on the lack of findings for the practice site characteristics variables, we did not attempt to interact them with the other practice transformation variables.

The Effect of Staff Survey Results (Increased Clinician/Staff Satisfaction) on CQMs

The second transformation factor we analyzed was staff satisfaction as measured in the Clinician and Staff Experience Survey (CSES). Completed within SPLIT, practice staff and providers responded to 20 items on the CSES that measure the respondent's overall satisfaction with their work, how stressful they found their work, their assessment of quality of patient visits, and work-life balance. University of Colorado Department of Family Medicine (UCDFM) then calculated a composite score of 1–100 for all measures. Because multiple staff and providers completed the CSES, we aggregated these scores for a practice-level average. We calculated this average for a baseline and for final scores as well as the average change over SIM participation. We hypothesized that by providing more integrated physical and behavioral healthcare, practitioners would be more fully able to meet the needs of their patients, which would lead to improved staff satisfaction. One concern of this hypothesis is a possible feedback loop in which increased satisfaction resulted in more provider engagement and even better provision of care.

We report results as we have previously. We first look at cross-sectional relationships between CQMs and staff satisfaction scores to consider whether practice sites with higher satisfaction scores also had higher CQMs. We then examine change over time in each and determine whether increasing satisfaction scores correlated with increased CQM scores. These second type of results are of greater importance in assessing the relationship between staff satisfaction and CQMs.

For cohort 1, we did not observe any statistically significant cross-sectional relationships between staff satisfaction and CQMs. The result with the lowest P-Value is the *Hemoglobin A1c Control* CQM: it negatively correlated with staff satisfaction at a P-Value of 0.07. As noted above, lower values for this specific CQM indicate improvement, so the negative relationship indicates that practice sites with higher levels of staff satisfaction also had patients with better control over hemoglobin A1c.

For cohort 2, we found several statistically significant cross-sectional relationships. During the second assessment period, cohort 2 practice sites had a negative relationship with depression screening; firms with higher levels of staff satisfaction performed depression screening at a slightly lower rate. The magnitude of this difference is only one percentage point and is therefore not particularly meaningful. This same result occurred for the *Asthma Medication Management* CQM. Cohort 2 practice sites with higher staff survey results had higher levels of developmental screening and adolescent obesity screening. The magnitude of these results was also small.

Neither cohort had statistically significant relationships between changes in staff satisfaction over time and changes in CQMs. With the exception of *Maternal Depression Screening* and *Hemoglobin A1c Control* measures, the direction of change was what we expected to observe: increases in staff satisfaction had positive correlations with increases in the CQMs, but magnitudes were small and not statistically significant.

Table 63. Relationship Between Clinician and Staff Experience and CQM Results

Assessment Period	Cohort 1			Cohort 2		
	Effect	P-Value	N	Effect	P-Value	N
Depression Screening						
Assessment Period 1	-0.05	0.923	82	-0.51	0.237	112
Assessment Period 2	0.18	0.710	73	-1.01	0.000*	120
Assessment Periods 1 to 2	0.21	0.527	68	0.31	0.269	101
Maternal Depression Screening						
Assessment Period 1	-1.45	0.368	12	0.63	0.495	17
Assessment Period 2	0.23	0.851	16	0.17	0.764	31
Assessment Periods 1 to 2	-1.87	0.426	11	0.07	0.936	17
Alcohol and Other Drug Screening						
Assessment Period 1	-0.61	0.840	6	0.14	0.904	15
Assessment Period 2	0.42	0.65	21	1.63	0.009*	21
Assessment Periods 1 to 2	-	-	-	11.02	0.095	6
Hemoglobin A1c Control						
Assessment Period 1	-0.60	0.073	61	-0.14	0.680	88
Assessment Period 2	-0.20	0.288	67	0.15	0.574	102
Assessment Periods 1 to 2	0.18	0.487	58	0.19	0.520	87
Hypertension Management						
Assessment Period 1	0.05	0.797	55	0.07	0.796	93
Assessment Period 2	0.01	0.945	66	0.03	0.866	100
Assessment Periods 1 to 2	0.36	0.123	50	0.20	0.499	88
Adult Obesity						
Assessment Period 1	0.01	0.985	64	-0.36	0.428	75
Assessment Period 2	-0.28	0.378	64	-0.60	0.083	94
Assessment Periods 1 to 2	0.17	0.587	58	-0.47	0.135	71
Asthma Medication Management						
Assessment Period 1	-0.34	0.567	67	-0.95	0.011*	57

Assessment Period	Cohort 1			Cohort 2		
	Effect	P-Value	N	Effect	P-Value	N
Assessment Period 2	-0.84	0.104	37	0.04	0.847	65
Assessment Periods 1 to 2	0.04	0.934	33	-0.49	0.236	46
Fall Risk Screening						
Assessment Period 1	1.14	0.337	24	1.47	0.115	51
Assessment Period 2	-0.63	0.276	50	0.37	0.604	64
Assessment Periods 1 to 2	0.42	0.641	21	0.25	0.546	50
Developmental Screening						
Assessment Period 1	-0.52	0.491	5	3.20	0.014*	14
Assessment Period 2	-0.22	0.801	14	-0.02	0.977	20
Assessment Periods 1 to 2	-	-	-	-0.10	0.938	14
Adolescent Obesity Screening						
Assessment Period 1	-0.52	0.400	22	1.38	0.007*	49
Assessment Period 2	0.82	0.195	27	-0.25	0.378	60
Assessment Periods 1 to 2	1.00	0.259	19	-0.48	0.467	46

The Effect of HIT Improvements on CQMs

Changes in HIT had the potential for driving changes in the CQMs through two processes. First, HIT improvements may have facilitated the providing of the screening and case planning measured with the CQMs. Second, for practice sites that had already performed this type of screening, HIT data quality improvements may have facilitated reporting CQM results more accurately. We do not attempt to untangle these alternative pathways relating HIT and CQMs. Instead we attempt to identify any aggregate effects.

Our analysis used practice sites' self-reported data quality of data elements and CQMs as recorded in the SPLIT HIT assessments. Data elements are pieces of basic information (i.e., patient date of birth, gender, medications, screenings performed)[71] that are necessary and foundational to accurately capture and report on CQMs. Practice sites rated their data elements on a color-coded, 3-point scale: green (3 = data consistently captured), yellow (2 = data inconsistently captured), and red (1 = data not captured). The CQMs in the HIT assessment reflect the outcome CQMs. However, rather than reflecting numerator/denominator rates, the HIT assessment of CQMs only asks practice sites to rate their data quality. These CQMs are rated on a slightly different scale than are the data elements: green (3 = data captured and trusted), yellow (2 = data captured but not trusted), and red (1 = data not captured).

---

[71] See Practice HIT chapter for additional information

Although the assessment also allowed practice sites to select “blue” or “black” data (indicating the sites did not see those populations), we limited our analysis to sites seeing those populations—therefore reporting a particular CQM as green, yellow, or red—and we adjusted our denominators. We operationalized data quality by creating quality scores for data elements and CQMs: we aggregated the percentage of items practice sites reported as “green” for both a baseline and final score, and we calculated an overall change score to compute the difference between baseline and final scores.

For cohort 1, there are several statistically significant cross-sectional relationships. Some of these have the anticipated relationship that practice sites with higher HIT scores also had higher CQMs. But other statistically significant cross-sectional relationships had the opposite relationship in which practice sites with higher HIT scores had lower CQMs.

CQMs with statistically significant positive correlations to HIT scores include *Depression Screening* and *Falls Screening*. CQMs with negative correlations include *Asthma Screening* and *Fall Screening* (Assessment Period 1). *Hemoglobin A1c Control* had a statistically significant positive relationship, which is contrary to our expectations, since lower hemoglobin scores indicate improvement.

For changes over time, cohort 1 increases in HIT scores positively correlated with increases in depression screening. But they negatively correlated with adolescent obesity screening.

For cohort 2, we see the same mixed pattern in cross-sectional statistically significant regressions. Positively correlated CQMs and HIT scores include *Depression Screening*, *Adult Obesity Screening*, and *Asthma Screening* in period 1 and *Adolescent Obesity Screening* in period 2. Negatively correlated cross-sectional results include *Depression Screening* and *Asthma Screening* (both in Assessment Period 2).

For changes over time, cohort 2 increases in HIT scores positively correlated with increases in *Adult Obesity Screening* measures (P-Value = 0.058) and negatively correlated with *Adolescent Obesity Screening* measures.

Table 64. Relationship Between HIT Improvements and CQM Results

Assessment Period	Cohort 1			Cohort 2		
	Effect	P-Value	N	Effect	P-Value	N
Depression Screening						
Baseline	-3.13	0.773	81	37.24	0.002*	111
Final	33.83	0.019*	74	-27.06	0.018*	120

Assessment Period	Cohort 1			Cohort 2		
	Effect	P-Value	N	Effect	P-Value	N
Baseline to Final	17.29	0.006*	68	-2.06	0.753	100
Maternal Depression Screening						
Baseline	0.39	0.992	12	-5.01	0.815	17
Final	-16.92	0.541	16	-9.60	0.701	31
Baseline to Final	-20.37	0.480	11	-0.63	0.967	17
Alcohol and Other Drug Screening						
Baseline	206.99	0.143	6	26.27	0.446	15
Final	-5.27	0.77	20	4.72	0.857	21
Baseline to Final	19.21	0.88	4	77.39	0.409	6
Hemoglobin A1c Control						
Baseline	23.83	0.004*	60	-9.41	0.333	88
Final	-2.11	0.723	67	-3.91	0.680	102
Baseline to Final	1.90	0.706	57	-5.24	0.526	87
Hypertension Management						
Baseline	3.66	0.483	55	0.31	0.968	93
Final	-5.85	0.321	66	-7.43	0.277	100
Baseline to Final	2.00	0.693	50	-0.84	0.920	88
Adult Obesity Screening						
Baseline	13.58	0.125	63	22.09	0.066	75
Final	3.03	0.780	64	18.28	0.130	94
Baseline to Final	-4.00	0.523	57	18.05	0.058	71
Asthma Medication Management						
Baseline	14.96	0.284	65	25.94	0.023*	57
Final	-50.35	0.008*	38	-24.03	0.016*	65
Baseline to Final	5.08	0.421	33	13.13	0.190	46
Fall Risk Screening						
Baseline	-53.57	0.026*	24	-20.90	0.419	51
Final	44.66	0.008*	49	-1.92	0.935	64
Baseline to Final	-9.64	0.476	20	14.77	0.157	50
Developmental Screening						
Baseline	-11.22	0.461	5	36.26	0.141	13
Final	-50.72	0.149	14	22.76	0.329	20

Assessment Period	Cohort 1			Cohort 2		
	Effect	P-Value	N	Effect	P-Value	N
Baseline to Final	-14.39	0.48	4	17.62	0.299	13
Adolescent Obesity Screening						
Baseline	9.82	0.428	21	-15.58	0.291	49
Final	-16.70	0.396	28	31.26	0.020*	60
Baseline to Final	-20.68	0.012*	18	-26.00	0.045*	46

Overall, the change over time analysis shows that increases in HIT potentially impact both *Depression Screening* and *Adult Obesity Screening* measures but appear to reduce the *Adolescent Obesity Screening* CQM. This last result is contrary to our expectations and warrants further investigation.

Clinical Quality Care Measures by Practice Site Characteristic

The following tables contain the average reported values of clinical quality care measures for all practice sites in cohort 1 and 2 that reported values for assessment periods 1 and 2, broken out by site characteristics. In many instances, there were not enough practice sites with a given characteristic reporting for a given measure, and we did not provide data in any case in which the number of reporting practice sites was five or fewer. In some cases, low reporting numbers may be a result of practice sites not being required to report. For example, pediatric clinics would not need to report a rate for adult obesity screening because they do not see adult patients.

Table 65. CQM by Practice Site Location

CQM	Urban			Rural		
	Baseline	Final	Difference (N)	Baseline	Final	Difference (N)
Depression Screening	51.5%	59.3%	7.8% (105)	54.8%	65.6%	10.8% (65)
Hemoglobin A1c Control	29.5%	28.3%	-1.2% (96)	32.7%	30.4%	-2.3% (50)
Hypertension Management	68.6%	68.3%	-0.3% (90)	64.7%	68.1%	3.4% (49)
Adult Obesity Screening	54.5%	61.3%	6.8% (81)	61.2%	68.5%	7.3% (49)
Alcohol and Other Drugs Screening	12.4%	6.0%	-6.4% (6)	-	-	-

CQM	Urban			Rural		
	Baseline	**Final**	**Difference (N)**	**Baseline**	**Final**	**Difference (N)**
Tobacco Use Screening	100.0%	86.5%	-13.5% (48)	100.0%	89.3%	-10.7% (36)
Asthma Medication Management	80.5%	85.7%	5.2% (40)	80.8%	84.0%	3.2% (40)
Fall Risk Screening	64.8%	74.2%	9.4% (42)	24.6%	46.7%	22.1% (29)
Maternal Depression	60.6%	73.5%	12.9% (23)	-	-	-
Developmental Screening	86.9%	92.6%	5.7% (13)	-	-	-
Adolescent Obesity	82.2%	91.6%	9.4% (44)	70.7%	90.3%	19.6% (21)

Overall, rural practice sites appeared to see greater gains and less declines than urban practice sites. Greater positive change was seen in *Depression Screening*, *Hypertension Management*, *Adult Obesity Screening*, *Fall Risk Screening*, and *Adolescent Obesity Screening* for rural practice sites. *Hemoglobin A1c Control* also saw greater improvements for rural sites, bearing in mind that this metric measures the percentage of people whose hemoglobin is above recommended levels, making a more negative number an improvement. *Tobacco Use Screening* saw less decline for rural than for urban practice sites.

Asthma Medication Management also saw improvements in both rural and urban practice sites, though more so in urban areas. Urban areas also had higher final rates of *Hypertension Management, Fall Risk Screening,* and *Adolescent Obesity Screening*, despite not seeing large improvements from baseline in those measures. More urban practice sites reported on CQMs than did rural ones, meaning that greater improvements in CQMs for rural sites may be a result of practice sites with less improvement not reporting. Nonetheless, evidence suggests that rural practice sites had greater improvement than urban sites had.

Table 66. CQM by Proportion of Patients Underserved

	Depression Screening	Hemoglobin A1c Control	Hypertension Management	Adult Obesity Screening	Tobacco Use Screening
Low					
Baseline	46.3%	25.2%	69.0%	47.7%	100.0%
Final	59.9%	24.4%	69.5%	52.6%	92.9%
Difference(N)	13.6% (29)	-0.8% (40)	0.5% (35)	4.9% (31)	-7.1% (23)

	Depression Screening	Hemoglobin A1c Control	Hypertension Management	Adult Obesity Screening	Tobacco Use Screening
Medium					
Baseline	47.7%	29.5%	67.6%	46.9%	100.0%
Final	60.0%	25.8%	67.0%	53.9%	88.8%
Difference(N)	12.3% (36)	-3.7% (37)	-0.6% (33)	7% (29)	-11.2% (20)
High					
Baseline	59.7%	33.7%	63.0%	67.8%	100.0%
Final	68.5%	31.0%	68.5%	76.5%	88.4%
Difference(N)	8.8% (35)	-2.7% (27)	5.5% (26)	8.7% (27)	-11.6% (20)
Very High					
Baseline	54.6%	35.3%	68.1%	64.4%	100.0%
Final	60.0%	35.4%	67.9%	71.9%	80.3%
Difference(N)	5.4% (70)	0.1% (41)	-0.2% (44)	7.5% (42)	-19.7% (21)

Table 67. CQM by Proportion of Patients Underserved (Continued)

	Asthma Medication Management	Fall Risk Screening	Maternal Depression	Developmental Screening	Adolescent Obesity
Low					
Baseline	63.6%	63.9%	65.3%	-	-
Final	67.3%	73.7%	85.0%	-	-
Difference(N)	3.7% (6)	9.8% (28)	19.7% (7)	-	-
Medium					
Baseline	76.3%	56.0%	-	-	65.6%
Final	80.0%	75.7%	-	-	92.9%
Difference(N)	3.7% (12)	19.7% (20)	-	-	27.3% (13)
High					
Baseline	83.3%	18.3%	-	-	78.5%
Final	85.4%	32.6%	-	-	84.5%
Difference(N)	2.1% (24)	14.3% (13)	-	-	6% (10)
Very High					
Baseline	83.1%	24.5%	58.8%	63.2%	82.1%
Final	88.7%	44.6%	67.4%	79.6%	92.1%
Difference(N)	5.6% (38)	20.1% (9)	8.6% (14)	16.4% (9)	10% (41)

No clear trend emerged regarding the change over time on practice sites by proportion of underserved patients. Practice sites with a low underserved population (under 10% of patients insured by Medicaid or uninsured) had more improvement than did sites with a higher underserved population in *Depression Screening* and *Maternal Depression*. They also did comparatively well in *Hemoglobin A1c Control*. They saw less of a decline in *Tobacco Use Screening* than did other practice sites.

For practice sites with a medium proportion of underserved patients (11–30% of all patients), *Hemoglobin A1c Control* and *Adolescent Obesity Screenings* saw great improvements; however, *Hypertension Management* and *Tobacco Use Screenings* both saw declines from baseline to final. Practice sites with high and very high underserved populations (31–50% and greater than 51% respectively) also each had areas where they performed comparatively well, such as in *Adult Obesity Screening*, and some areas where they did not, such as *Tobacco Use Screening*. However, these practice sites did not have particularly low percentages for each CQM in general, as might be expected given the populations they serve. Given this, it would appear that SIM acts equitably on practice sites regardless of the payer type for the service (Medicaid and self-pay).

Table 68. CQM by Practice Size

CQM	Small			Medium			Large		
	Baseline	Final	Difference (N)	Baseline	Final	Difference (N)	Baseline	Final	Difference (N)
Depression Screening	51.2%	60.1%	8.9% (65)	48.5%	61.0%	12.5% (50)	58.5%	64.3%	5.8% (55)
Hemoglobin A1c Control	32.5%	30.3%	-2.2% (51)	27.4%	29.2%	1.8% (46)	31.8%	27.5%	-4.3% (49)
Hypertension Management	65.0%	69.1%	4.1% (54)	67.1%	68.0%	0.9% (39)	69.9%	67.2%	-2.7% (46)
Adult Obesity Screening	58.0%	68.8%	10.8% (53)	51.4%	57.7%	6.3% (38)	61.2%	63.6%	2.4% (39)
Alcohol and Other Drugs Screening	-	-	-	16.7%	6.4%	-10.3% (6)	-	-	-
Tobacco Use Screening	100.0%	84.4%	-15.6% (40)	100.0%	91.5%	-8.5% (25)	100.0%	89.7%	-10.3% (19)
Asthma Medication Management	76.7%	83.1%	6.4% (28)	83.5%	83.0%	-0.5% (24)	82.2%	88.1%	5.9% (28)
Fall Risk Screening	36.5%	55.8%	19.3% (28)	57.3%	70.3%	13% (26)	54.4%	63.5%	9.1% (17)

CQM	Small			Medium			Large		
	Baseline	Final	Difference (N)	Baseline	Final	Difference (N)	Baseline	Final	Difference (N)
Maternal Depression	-	-	-	57.8%	73.6%	15.8% (16)	77.6%	82.6%	5% (8)
Developmental Screening	-	-	-	83.5%	91.4%	7.9% (11)	-	-	-
Adolescent Obesity	80.3%	90.2%	9.9% (28)	72.8%	94.6%	21.8% (16)	80.5%	89.8%	9.3% (21)

Larger practice sites tended to show slightly less improvement in CQMs than small and medium size practice sites. Large practice sites showed the least improvement or most decline in the areas of *Depression Screening, Hypertension Management, Adult Obesity Screening*, *Fall Risk Screening*, *Maternal Depression Screening*, and *Adolescent Obesity Screening*. Small practice sites showed the most improvement as a result of SIM in the areas of *Hypertension Management*, *Adult Obesity Screening, Asthma Medication Management*, and *Fall Risk Screening*. Medium size practice sites showed the most improvement in the areas of *Depression Screening*, *Maternal Depression Screening*, and *Adolescent Obesity Screening*.

Table 69. CQM by Practice Type

CQM	Pediatrics			Mixed Primary Care			Adult Primary Care		
	Baseline	Final	Difference (N)	Baseline	Final	Difference (N)	Baseline	Final	Difference (N)
Depression Screening	51.5%	60.8%	9.3% (50)	53.7%	62.5%	8.8% (114)	45.6%	55.1%	9.5% (6)
Hemoglobin A1c Control	-	-	-	30.9%	29.5%	-1.4% (136)	20.0%	15.4%	-4.6% (9)
Hypertension Management	-	-	-	67.1%	68.3%	1.2% (127)	67.0%	69.6%	2.6% (10)
Adult Obesity Screening	-	-	-	57.4%	64.5%	7.1% (116)	48.0%	47.0%	-1.0% (9)
Alcohol and Other Drugs Screening	-	-	-	28.6%	4.0%	-24.6% (9)	-	-	-
Tobacco Use Screening	-	-	-	100.0%	87.4%	-12.6% (77)	-	-	-
Asthma Medication Management	76.2%	82.3%	6.1% (22)	82.4%	85.8%	3.4% (58)	-	-	-

CQM	Pediatrics			Mixed Primary Care			Adult Primary Care		
	Baseline	**Final**	**Difference (N)**	**Baseline**	**Final**	**Difference (N)**	**Baseline**	**Final**	**Difference (N)**
Fall Risk Screening	-	-	-	44.7%	59.3%	14.6% (63)	77.3%	91.6%	14.3% (8)
Maternal Depression	66.7%	74.8%	8.1% (24)	-	-	-	-	-	-
Developmental Screening	76.7%	86.2%	9.5% (17)	-	-	-	-	-	-
Adolescent Obesity	81.7%	89.5%	7.8% (45)	71.2%	94.8%	23.6% (20)	-	-	-

We could not deduce any major trends based on practice type, in part because both pediatrics and adult primary care practice sites had low levels of reporting. Compared to mixed family practice sites, pediatric units had slightly more improvement in *Asthma Medication Management*. However, this improvement may be caused by only practice sites with great improvements tending to report. Mixed primary care sites showed greater improvements than pediatric clinics in *Adolescent Obesity Screening* as opposed to adult-serving units' decline. Adult primary care saw greater improvements in *Depression Screening*, *Hemoglobin A1c Control* and *Hypertension Management* than other practice site types. One reason practice sites did not report on many measures is that particular CQMs were outside their purview (e.g., the *Adult Obesity* measure is irrelevant to a pediatrics clinic).

Summary and Conclusions

The practice transformation component focuses on reporting transformation efforts at SIM primary care and CMHC sites. This chapter relates to outcomes from these efforts, reporting on statistically significant increases in the use of screening measures, referred to as clinical quality measures (CQMs), and analyzing the effect of practice transformation factors on changes in the CQMs.

SIM practice sites showed statistically significant improvement in six of the 14 practice-site-reported CQMs during participation in SIM. Both cohorts had improvements in *Depression Screening* and *Fall Risk Screening* measures. Cohort 1 also had improvements in *Hemoglobin A1c Control* and *Asthma Medication Management* measures, and cohort 2 showed an increase in *Adolescent Obesity Screening* and *Adult Obesity Screening.*

Two other measures showed encouraging results for both cohorts. There were positive trends for *Maternal Depression Screening*, although differences were not significant, likely a result of smaller sample sizes. Likewise, both proxy and practice-site-reported measures had consistent increases over time. The same was true for *the Developmental Screening* measure.

Because of reporting issues with both the *Alcohol and Other Drug* and the *Alcohol Screening* measures, we cannot make any conclusions about changes over time.

We observed no statistically significant improvements in any of the other measures. Additionally, only one measure showed declines over the reporting periods (*Tobacco Use Screening*), but this decline is likely related to anomalies in data reporting and not a meaningful trend.

As measured by the reported CQMs, these results show some improvements in clinical quality over the course of SIM implementation. Although the lack of a comparison group means that we cannot definitively state that SIM alone contributed to these improvements, the results remain encouraging.

The proxy CQMs have the potential to provide insight into whether improvements in the SIM CQMs were caused by SIM involvement or resulted from statewide trends unrelated to SIM. A pattern of improvements in both the SIM CQMs and the SIM cohort proxy CQMs with no corresponding improvement in statewide proxy CQMs would support the hypothesis that SIM participation drove changes in CQM. However, we did not observe any examples with this pattern. Generally, the proxy CQMs for the SIM cohorts matched the statewide trends.

The relationship between practice transformation factors and CQMs is complex. Rather than pursuing cross-sectional correlations, we focused our analysis on correlations between improvements in the transformation factors and improvements in the CQMs. The staff survey results show little relationship to changes in CQMs, perhaps because the staff survey results had very little change over time. Changes over time in integration have a statistically significant relationship to changes in *the Hemoglobin A1c Control* and *Hypertension Management* CQMs. Both had negative relationships. And although we anticipated that improvements in integration would decrease the number of patients with hemoglobin above the recommended level, improved integration should not have reduced the number of patients with correctly managed hypertension. Our data do not allow us to explore reasons for this outcome, such as practice sites with the most improvement in integration drawing in those patients with disproportionately difficult to control cases of hypertension.

Improvements in HIT over time correlated with improvements in the *Depression Screening* CQM. But they also correlated with worsening values for the *Adolescent Obesity Screening* measure. Based on these conflicting results, the transformation factor least associated with improvements in CQMs is the staff survey. HIT and integration improvements were associated with changes in the CQMs, albeit in inconsistent ways.

Most clinical quality measures did demonstrate average improvement, although only some of these were statistically significant changes. We attempted to determine which of the practice transformation factors (integration, staff satisfaction, or HIT) drove most of the change but were unable to identify a clear pattern.

Supporting Data Summary Table

Although the previous scatter plots provide visual information about the distribution of CQMs in each assessment period, we do not break out the plotted points by cohort. The following table provides the standard deviation as a measure of variability for each CQM by cohort and period. We present these values for all practice sites that reported, and we indicate with an asterisk any statistically significant change (p<0.05) from the previous period. We base this calculation on only those practice sites that reported in both periods.

The change in in the average *Fall Risk Screening* score for cohort 1 from assessment periods 2 to 3 was positive for the 13 practice sites that reported in both periods but negative for all sites reporting both periods. With this exception, the direction of change was the same for all statistically significant measures calculated with scores from all reporting practice sites versus those sites reporting both periods. See tables in the body of this chapter for summary information on change over time in these measures.

Table 70. Clinical Quality Measure Results by Cohort and Assessment Period[72]

CQM	Cohort	Assessment Period 1			Assessment Period 2			Assessment Period 3		
		Mean	N	Standard Deviation	Mean	N	Standard Deviation	Mean	N	Standard Deviation
Depression Screening	1	45%	82	27.77	55%*	74	31.09	66%	38	22.35
	2	53%	112	31.29	65%*	120	23.71	-	-	-
	3	52%	69	28.39	-	-	-	-	-	-
Maternal Depression	1	51%	12	31.33	72%	16	22.49	77%	13	20.88
	2	66%	17	29.65	70%	31	27.91	-	-	-
	3	28%	21	35.87	-	-	-	-	-	-
Alcohol and Other Drug	1	57%	6	44.9	10%	20	22.4	4%	2	2.21
	2	16%	15	27.97	8%	21	20.04	-	-	-

---

[72] A single * indicates statistical significance (P-Value ≤ 0.05).

CQM	Cohort	Assessment Period 1			Assessment Period 2			Assessment Period 3		
		Mean	N	Standard Deviation	Mean	N	Standard Deviation	Mean	N	Standard Deviation
Screening	3	17%	10	31.03	-	-	-	-	-	-
Tobacco Use Screening	1	-	-	-	100%	67	1.16	86%*	28	18.36
	2	100%	93	0	85%*	107	20.16	-	-	-
	3	83%	59	21.62	-	-	-	-	-	-
Alcohol Use Screening	1	-	-	-	9%	5	9.31	59%	2	0.44
	2	21%	2	28.63	36%	5	39.76	-	-	-
	3	51%	8	30.57	-	-	-	-	-	-
Hemoglobin A1c Control	1	32%	61	18.32	26%*	67	11.56	29%	27	13.76
	2	30%	88	18.56	33%	102	18.33	-	-	-
	3	31%	53	19.05	-	-	-	-	-	-
Hypertension Management	1	67%	55	10.51	65%	66	11.18	70%*	27	7.12
	2	68%	93	15.76	70%	100	13.24	-	-	-
	3	67%	55	17.55	-	-	-	-	-	-
Adult Obesity Screening	1	49%	64	18.15	49%	65	20.27	75%*	23	17.59
	2	64%	75	23.01	72%*	94	21.93	-	-	-
	3	54%	48	23.97	-	-	-	-	-	-
Asthma Medication Management	1	68%	67	30.83	78%*	38	22.91	88%	25	8.24
	2	83%	57	20.53	87%	65	11.67	-	-	-
	3	76%	22	22.73	-	-	-	-	-	-
Fall Risk Screening	1	46%	24	33.78	63%*	49	32.89	61%*	15	29.9
	2	48%	51	35.85	56%*	64	34.12	-	-	-
	3	57%	38	33.74	-	-	-	-	-	-
Developmental Screening	1	91%	5	5.39	78%	14	18.07	64%	10	26.19
	2	73%	14	33.89	78%	20	23.47	-	-	-
	3	74%	11	28.74	-	-	-	-	-	-
Adolescent Obesity	1	90%	22	16.51	83%	27	20.24	88%	22	12.61
	2	75%	49	25.94	90%*	60	18.81	-	-	-

CQM	Cohort	Assessment Period 1			Assessment Period 2			Assessment Period 3		
		Mean	N	Standard Deviation	Mean	N	Standard Deviation	Mean	N	Standard Deviation
	3	83%	30	24.28	-	-	-	-	-	-

Table 71. Proxy and SIM CQM Descriptions, Numerators, and Denominators

Definition	Proxy	SIM
Depression Screening		
Description	Percentage of patients aged 12 years and older with new depression diagnosis.	Percentage of patients aged 12 years and older screened for clinical depression using an age-appropriate standardized tool AND follow-up plan documented.
Numerator	Number of patients screened for depression with new diagnosis of depression during clinician encounter in the measurement period. New diagnosis means a patient has not been previously diagnosed with depression during the preceding 12 months from the date of the diagnosis in the measurement period.	Patients screened for clinical depression on the date of the encounter using an age appropriate standardized tool AND if positive, a follow-up plan is documented on the date of the positive screen.
Denominator	Number of patients aged 12 years and older with one or more clinician encounters during the measurement period and screened for depression. Exclude those with an existing diagnosis of depression or bipolar depression during the 12-month lookback period preceding the date of depression diagnosis.	All patients aged 12 years and older before the beginning of the measurement period with at least one eligible encounter during the measurement period. Denominator Exclusions: Patients with an active diagnosis for Depression or a diagnosis of Bipolar Disorder. Denominator Exceptions: Patient Reason(s) Patient refuses to participate OR Medical Reason(s) Patient is in an urgent or emergent situation where time is of the essence and to delay treatment would jeopardize the patient's health status OR situations where the patient's functional capacity or motivation to improve may impact the accuracy of results of standardized depression assessment tools (e.g., certain court appointed cases or cases of delirium).
Maternal Depression Screening		

Definition	Proxy	SIM
Description	Percentage of mothers receiving screening for maternal depression by a clinician in the first six months following delivery.	Percentage of children who turned six months of age during the measurement year, who had a face-to-face visit between the clinician and the child during the child's first six months, and who had a maternal depression screening for the mother at least once between 0 and 6 months of life.
Numerator	Number of mothers receiving depression screening within 6 months of childbirth, at child or maternal clinician encounter.	Children with documentation of maternal screening or treatment for postpartum depression for the mother.
Denominator	Number of live births within 6 months of the beginning of the measurement period and 6 months after the start of the measurement period.	Children with a visit who turned 6 months of age in the measurement period.
Alcohol and Other Drugs Screening		
Description	Percentage of patients 18 years and older screened for illicit drug use.	Percentage of patients aged 13 years and older with a new episode of alcohol and other drug (AOD) dependence who received the following (two rates are reported): - Percentage of patients who initiated treatment within 14 days of the diagnosis - Percentage of patients who initiated treatment and who had two or more additional services with an AOD diagnosis within 30 days of the initiation visit
Numerator	Number of patient 18 years and older screened for illicit drug use or misuse of prescription drugs during the measurement period.	Numerator 1: Patients who initiated treatment within 14 days of the diagnosis. Numerator 2: Patients who initiated treatment and who had two or more additional services with an AOD diagnosis within 30 days of the initiation visit.
Denominator	Number of patients aged 18 years and older seen for at least two visits or at least one preventive visit during the measurement period.	Patients age 13 years of age and older who were diagnosed with a new episode of alcohol or drug dependency during a visit in the first 11 months of the measurement period. Denominator Exclusions: Patients with a previous active diagnosis of alcohol or drug dependence in the 60 days prior to the first episode of alcohol or drug dependence.
Tobacco Use Screening		

Definition	Proxy	SIM
Description	Patients 18 years and older screened for tobacco use.	Percentage of patients aged 18 years and older who were screened for tobacco use one or more times within 24 months AND who received cessation counseling intervention if identified as a tobacco user.
Numerator	Number of patients 18 years and older screened for tobacco use during the measurement period.	Patients who were screened for tobacco use at least once within 24 months AND who received tobacco cessation intervention if identified as a tobacco user.
Denominator	Number of patients aged 18 years and older seen for at least two visits or at least one preventive visit during the measurement period.	All patients aged 18 years and older seen for at least two visits or at least one preventive visit during the measurement period. Denominator Exceptions: Documentation of medical reason(s) for not screening for tobacco use (e.g., limited life expectancy, other medical reason).
Diabetes: Hemoglobin A1c Control		
Description	Percentage of patients with Type 1 or Type II diabetes who had hemoglobin A1c testing at clinician encounter.	Percentage of patients 18–75 years of age with diabetes who had hemoglobin A1c > 9.0% during the measurement period.
Numerator	Number of patients aged 18 to 75, diagnosed with type 1 & Type 2 Diabetes receiving Hemoglobin A1c Testing during the measurement period.	Patients whose most recent HbA1c level (performed during the measurement period) is >9.0%.
Denominator	Number of patients aged 18 to 75 years, diagnosed with Type 1 or Type 2 diabetes prior to the beginning of the measurement period with one or more clinician encounters during the measurement period.	Patients 18-75 years of age with diabetes with a visit during the measurement period.
Hypertension Screening		
Description	Percentage of patients 18 to 85 years of age with essential hypertension dispensed at least 90-day supply of therapeutic medications.	Percentage of patients 18–85 years of age who had a diagnosis of hypertension and whose blood pressure was adequately controlled (<140/90mmHg) during the measurement period.

Definition	Proxy	SIM
Numerator	Number of patients 18 to 85 years of age with diagnosis of essential hypertension dispensed at least 90-day supply of therapeutic medication during the measurement period.	Patients whose blood pressure at the most recent visit is adequately controlled (systolic blood pressure <140 mmHg and diastolic blood pressure <90 mmHg) during the measurement period.
Denominator	Number of patients 18 to 85 years with diagnosis of essential hypertension within six (6) months of the beginning of the measurement period or 12 months preceding the start of the measurement period, and with one or more clinician encounters during the measurement period. Exclude patients diagnosed with ESRD, chronic kidney disease, dialysis, kidney transplant, pregnancy.	Patients 18-85 years of age who had a diagnosis of essential hypertension within the first six months of the measurement period or any time prior to the measurement period. Denominator Exclusions: Patients with evidence of end stage renal disease (ESRD), dialysis or renal transplant before or during the measurement period. Also exclude patients with a diagnosis of pregnancy during the measurement period.
Adult Obesity Screening		
Description	Percentage of patients aged 18 years and older diagnosed with obesity receiving 2 or more clinician encounters and 1 or more tests for diabetes (blood glucose or HA1c).	Percentage of patients aged 18 years and older with BMI documented during the current encounter or during the previous six months AND with a BMI outside of normal parameters, a follow-up plan is documented during the encounter or during the previous six months of the current encounter.
Numerator	Number of patients aged 18 years and older diagnosed as obese receiving 2 or more clinician encounters and 1 or more tests for diabetes (blood glucose or HA1c) during the measurement period.	Patients with a documented BMI during the encounter or during the previous six months, AND when the BMI is outside of normal parameters, a follow-up plan is documented during the encounter or during the previous six months of the current encounter.

Definition	Proxy	SIM
Denominator	Number of patients aged 18 years and older diagnosed as obese during the six months preceding the start of the measurement period or during the first six months of the period. Excluding those patients who were pregnant during the measurement period.	All patients 18 and older on the date of the encounter with at least one eligible encounter during the measurement period Denominator Exclusions: Patients who are pregnant; patients receiving palliative care; patients who refuse measurement of height and/or weight or refuse follow up. Denominator Exceptions: Patients with a documented Medical Reason: Elderly Patients (65 or older) for whom weight reduction/weight gain would complicate other underlying health conditions such as the following examples: illness or physical disability; mental illness, dementia, confusion; nutritional deficiency, such as vitamin/mineral deficiency; patients in an urgent or emergent medical situation where time is of the essence and to delay treatment would jeopardize the patient's health status.
Asthma Medication Management		
Description	Percentage of patients 5-64 years of age who were identified as having persistent asthma and were dispensed appropriate medications that they remained on for at least 75% of their treatment period.	Percentage of patients 5–64 years of age who were identified as having persistent asthma AND were dispensed appropriate medication that they remained on for at least 75% of their treatment period.
Numerator	Number of patients 5-64 years of age who were identified as having persistent asthma and were dispensed appropriate medications that they remained on for at least 75% of their treatment period, during the measurement period.	The number of patients who achieved a proportion of days (PDC) of at least 75% for their asthma controller medications during the measurement year.
Denominator	Number of patients 5 to 64 years of age with persistent asthma, with one or more clinician encounters during the measurement period. Excluding those patients with emphysema, COPD, cystic fibrosis, acute respiratory failure during or prior to the measurement period.	Patients 5-64 years of age with persistent asthma and a visit during the measurement period. Exclusions: Patients with emphysema, COPD, chronic bronchitis, cystic fibrosis, or acute respiratory failure during or prior to the measurement period. Exclude any patients who have no asthma controller medications dispensed during the measurement period or who use hospice services any time during the measurement period.
Fall Risk Screening		

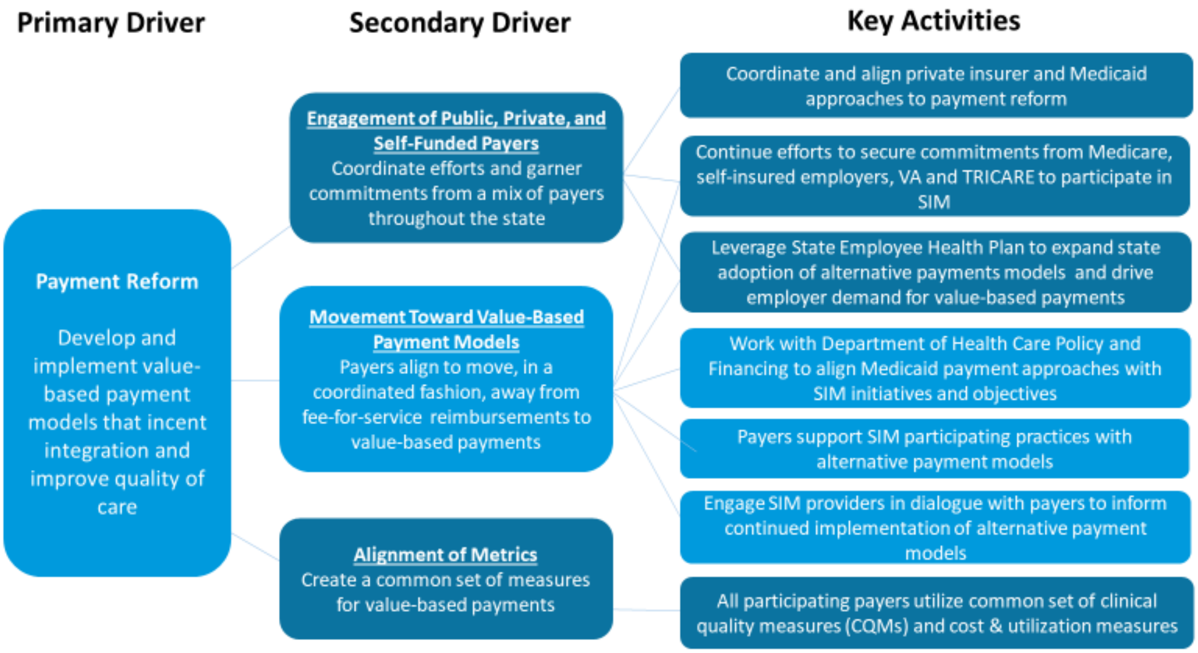
Definition	Proxy	SIM
Description	Percentage of patients 65 years and older receiving annual health evaluation with assumed assessment of future fall risk during the encounter.	Percentage of patients aged 65 years and older who were screened for future fall risk.
Numerator	Number of patients 65 years and older who received preventive/annual health encounter visit with assumed fall risk screening during the visit.	Patients who were screened for future fall risk at least once within the measurement period.
Denominator	Number of patients 65 years or older who received one or more face-to-face clinician encounters.	Patients aged 65 years and older with a visit during the measurement period. Denominator Exceptions: Documentation of medical reason(s) for not screening for fall risk (e.g., patient is not ambulatory).
Developmental Screening		
Description	Percentage of Children turning one, two or three years of age during the measurement period receiving developmental screening in the 12-month period preceding the first, second or third birthday.	Percentage of children screened for risk of developmental, behavioral, and social delays using a standardized screening tool in the 12 months preceding their first, second, or third birthday.
Numerator	Number of children turning one, two or three years of age during the measurement period receiving developmental screening in the 12-month period preceding the first, second or third birthday.	Children in Denominator 4 who had a screening for risk of developmental, behavioral, and social delays using an age appropriate standardized screening tool that was documented in the 12 months preceding their first, second, or third birthday.
Denominator	Number of children turning one, two or three years of age during the measurement period.	All patients who turn one, two, or three years old during the measurement period with at least one eligible outpatient encounter during the measurement period.

5 Payment Reform

Introduction

The Payment Reform primary driver emphasizes the role of value-based payment models (VBPs) in incentivizing integrating physical and behavioral healthcare, which should then lead to improvements in the overall quality of care. This driver focuses on three areas: (1) the engagement of payers in discussions around value based payment (VBPs) and encouraging continued efforts and commitments to support VBPs, (2) helping to align payers and support movement away from fee-for-service payment structures, and (3) the alignment of measures that can be used to demonstrate value in care delivery.

Figure 35. Payment Reform Driver Diagram



This chapter summarizes overall outcomes in each of these areas. Please note that although some of this information is in the Process chapter as well, we have attempted to reduce redundancy while emphasizing the importance of the data that appear in both chapters. However, more in-depth discussions of each of these can be found in the Final SIM Process Evaluation Report. Specifically, the Payment Reform chapter of that report discusses efforts to engage payers and to encourage movement toward VBPs. Within the overall payment reform driver, SIM staff and workgroup members helped to develop and refine six evaluation questions. One of these questions is more process-oriented, and we discuss it in the Final SIM Process Evaluation Report. The remaining three questions were designed to assess SIM progress in payment reform:

PR1. To what extent were value-based payment models implemented?
PR3. What is the cost of integration transformation efforts to SIM primary care practice sites? (Is the cost sustainable for practice sites? Are practices willing to absorb unreimbursed costs?)
PR4. To what extent did the utilization of services and total cost of care (including out-of-pocket costs to consumers) differ over time for consumers attributed to SIM participating practices? Was this different compared to consumers in comparison practices?

We answer the last of these questions in part within the Cost and Utilization chapter of this SIM Outcome Evaluation Report

We discuss the first two, around the extent of VBP implementation and the cost of integration efforts, in this chapter. However, answering these questions is hindered because data to support these evaluation questions were extremely limited. Below, we briefly discuss challenges with gathering data from providers[73] about the way in which SIM practice sites were supported by VBPs. Additionally, as part of the evaluation, we attempted to develop a methodology for analyzing and comparing specific APMs with the goal of identifying best models based on practice characteristics. We were ultimately unsuccessful because of these data challenges.[74] After repeated discussions with the SIM Evaluation Workgroup and the University of Colorado Department of Family Medicine (UCDFM), we decided to rely on closeout surveys conducted with practices to explore the degree they prepared for additional alternative payment models. Practices improved their ability to succeed with APMs throughout SIM. Some of these efforts included learning ways to better negotiate payer contracts, modify business practices, and plan for the sustainability of those changes.

An additional evaluation question was proposed early in the evaluation that was designed to, as is illustrated in the above SIM Driver Diagram (Figure 35), correlate movement toward value-based payment structures with improved quality of care and outcomes. Because the data have lacked standardization and were challenging to collect and validate, the extent of and nature of VBP implementation in SIM practice sites cannot be directly answered. However, key informants did offer their perspectives regarding the degree to which they either believe or have observed how the movement towards VBPs linked to better outcomes. We discuss these findings in this chapter.

---

[73] Please refer to the Payment Reform chapter of the Process Evaluation Report for a full description of data limitations.

[74] Please see the Methodology chapter of this document for a more detailed discussion of data collection challenges.

Data for This Chapter

The original SIM evaluation planning called for assessing the degree to which the implementation of VBPs influenced improvements in practice-level outcomes (e.g., clinical quality and access to care) and for assessing the costs of implementing specific VBP models. However, full data to support these analyses proved unavailable. The SIM office worked closely with payers to support their reporting of high-level descriptions of VBPs implemented within each of the SIM primary care practices. This chapter presents the high-level descriptions that were collected from payers and shows the change in value-based payments over time, as we are able. However, these data are not sufficient to create a quantitative measure that can be used as a discrete variable to assess the degree to which specific APMs might be associated with improvements in outcomes. We include more details about these data limitations in the following section.

This chapter also presents qualitative data gathered from key informant interviews (KIIs) regarding perceptions of the practice transformation coaches working with practices to prepare and participate in VBPs and the perceptions of the stakeholders and payers participating in the Multi-Payer Collaborative (MPC). We discuss the specific individuals interviewed and the resulting limitations of these data in that section. The Methodology chapter of this report contains further details on the KIIs. One important caveat is that the KIIs were not practice clinicians or staff who were directly impacted by VBPs but rather the PFs and CHITAs who worked closely with practice sites. KIIs represent a broader overall view of state efforts. Practice perspectives are provided via closeout surveys.

Finally, we used data from a closeout survey from practices to provide a practice perspective on the sustainability of VPBs within their individual practice sites. We provide more details on closeout surveys in the Methodology chapter of this report.

Moving Towards Value-Based Payments

PR1. To what extent were value-based payment models implemented?

As mentioned previously, the needed data were not available to fully answer this evaluation question, although we did include data below with some discussion of general trends and patterns. The Final SIM Process Evaluation Report details the work done over the life of the SIM initiative to move towards value-based payments and outlines some of the challenges to implementing these models, both at the individual practice site level and as a larger system reform effort. The primary trend we could observe and report is an increase in patients supported by Medicaid’s APM. This increase coincides with the expansion of that APM and may not indicate growth in the already-existing APMs. Although we cannot speak to the degree to which this is the result of many other changes in Medicaid happening concurrent to SIM versus the SIM implementation itself, between 2015 and 2018 the number of patients supported went from approximately 114,000 to nearly 434,000, almost quadrupling. However, it is difficult to assess how much, if at all, SIM-participating primary care practice sites shifted towards more value-based payments.

At the start of SIM implementation, VBPs (also referred to as "alternative payment models" [APMs]) were organized into categories based on the Health Care Payment Learning and Action Network (HCPLAN) framework,[75] as described in the following table. The SIM Final Process Evaluation Report discusses the initial frustration or surprise from some sites that there was not a single or automatic "SIM-specific APM" available upon enrollment. But this approach would not have been feasible given competition and antitrust laws governing the multiple health plans supporting the SIM initiative. Each practice did ultimately receive a different mix of payment support from participating MPC payers, an approach which promoted a degree of flexibility that was crucial to ensuring payer buy-in and retention and that supported sustainability.

Table 72. Alternate Payment Model (APM) Categories

Alternate Payment Model	Description
APM 1: Fee-for-Service – No Link to Quality	▪ Traditional FFS with no link to quality
APM 2: Fee-for-Service – Link to Quality	▪ Based on traditional FFS payments but subsequently adjusted based on infrastructure investments, quality data, or based on performance on cost and quality metrics
APM 3: Fee-for-Service Architecture	▪ Based on FFS architecture ▪ Payments are based on quality considerations, including cost performance against a target, shared savings payments, shared risk payments, and bundled payments
APM 4: Population-Based Payment	▪ Payments based on care within a defined or overall budget or for meeting care goals for a population of patients/members

---

[75] The HCPLAN is a public-private partnership established to promote healthcare payment reform. More information at https://hcp-lan.org/

Although most SIM practice sites received payer support through APMs, interviews and assessments consistently identified challenges in working with payers broadly and commercial payers specifically. As part of the evaluation effort, SIM collaborated with health plans to obtain payment support data, with the objective of being able to obtain a more comprehensive understanding of the landscape of value-based payment in Colorado[76] as well as tracking the progress of APM adoption statewide and for SIM practice sites. The original evaluation plan also intended to use these data to answer an evaluation question relating to whether the implementation/increase of payments within certain APM/VBP categories influenced outcomes.

Consistently obtaining these data from all payers participating in SIM has been challenging. Among these challenges were communication gaps among payer representatives, a lack of standardization of reportable payer data, system limitations, competing priorities, and the inability of some payers to share data containing amounts they were paying to each practice. Through discussing this work with other SIM states and other entities, we recognize that the issues identified by the SIM office are not unique to Colorado.

Although the aforementioned challenges have led to gaps in data, difficulty in standardizing across payer organization, and problems verifying the information provided (especially for commercial health plans), we provide a comprehensive scope of the data obtained in Table 73 and Table 74, below. We do, however, include the payer data as documentation of the support payers have provided to SIM practices and their patients and to increase the payment reform body of knowledge through efforts such as the MPC.

Medicaid provided the most complete payer data (summarized in Table 73 below). The table contains the number of SIM patients whose care Medicaid supported by each APM level and year, with total dollar amounts paid to Medicaid-supported practices. Data from private health plans reflecting APM support for SIM practice sites throughout the SIM initiative are included in Table 74 below. Although these data have limitations and gaps, they give some insight into the scope and variation of payment models that SIM practices were being supported by. In the tables below, note that an asterisk denotes data that payers did not provide.

---

[76] It was never assumed that it would be possible to get a fully comprehensive picture of the extent of VBPs in Colorado because not all payers participate in the Multi-Payer Collaborative.

Table 73. APM Support for SIM Practice Sites - HCPF[77]

HCPF/Medicaid APM Support for SIM Practice Sites – All Cohorts, All Years					
	APM Category	**2015**	**2016**	**2017**	**2018**
Number of Attributed Beneficiaries	APM 1	-	-	-	-
	APM 2	114,257	145,079	307,379	433,995
	APM 3	-	-	-	-
	APM 4	-	2,208	2,208	-
	Total	**114,257**	**147,287**	**309,587**	**433,995**
Total Payments Per APM	APM 1	**-**	**-**	**-**	**-**
	APM 2	-	$2,575,507.20	$5,767,311.47	$5,175,447.30
	APM 3	-	-	-	-
	APM 4	-	$37,094.40	$35,504.64	-
	Total	**-**	**$2,612,601.60**	**$5,802,816.11**	**$5,175,447.30**

Table 74. APM Support for SIM Practice Sites – Private Payers

Private Payers APM Support – All Cohorts, All Years					
	APM Category	**2015**	**2016**	**2017**	**2018**
Number of Attributed Beneficiaries	APM 1	6,156	-	24,492	-
	APM 2	-	-		-
	APM 3	108,265	11,875	219,500	683
	APM 4	16,822	7,500	26,089	19,647
	Total	**131,243**	**19,375**	**270,081**	**20,330**
Total Payments Per APM	APM 1	*	*	*	*
	APM 2	*	*	*	*
	APM 3	*	*	*	*
	APM 4	*	*	*	*
	Total	*	*	*	*

Payer data provide some insight into the actual status of APMs from SIM's preimplantation period (2015). Five (out of six) payers provided baseline (2015) information of the payment model(s) used with each of the SIM cohort 1 primary care practice sites, using the APM

---

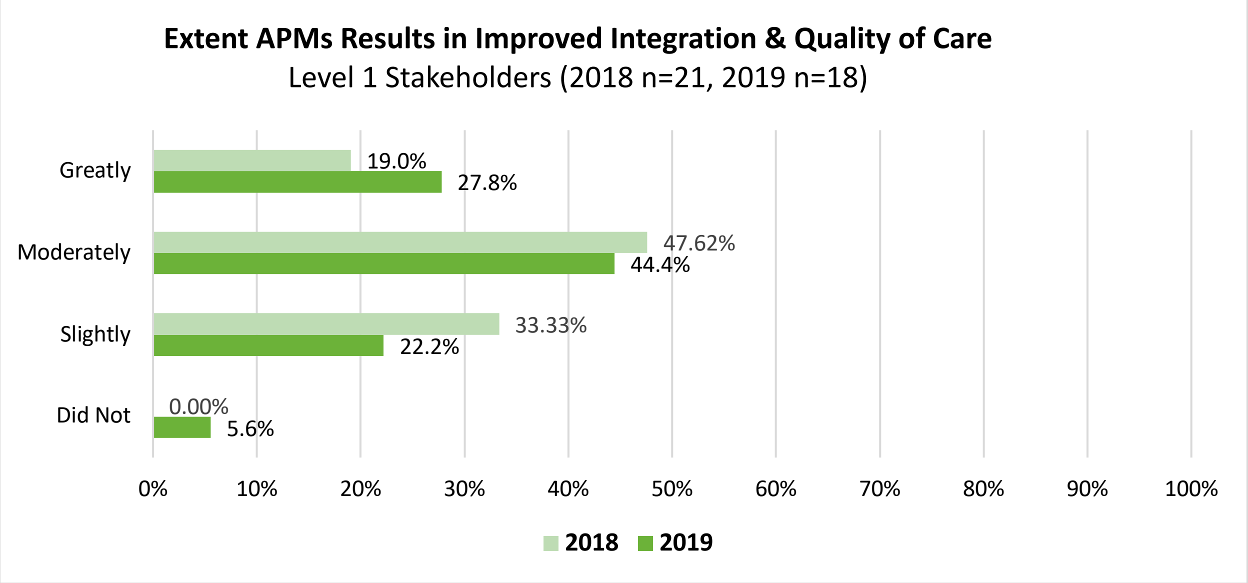
[77] HCPF summited these as fiscal year figures.

categorization developed by the Health Care Payment Learning and Action Network (HCPLAN).[78] Two commercial payers provided information on 2016, and three commercial payers provided information for 2017 and 2018.

Value-Based Payments Effects on Integration and Quality of Care

The graph below (Figure 36) shows key informants' (i.e., SIM office, other Governor's office staff, workgroup members, vendor partners) assessments regarding the degree to which implementation of APMs led to improved integration and quality of care. These key informants provide a very high-level view. Not directly impacted by VBPs, these stakeholders describe their perceptions, based on the work they are doing with SIM, with providers and practices, and in some cases, with payers.

Figure 36. Extent APMs Resulted in Improved Integration and Improved Quality of Care



Key informants acknowledged that it may be too early to discern the full impact APMs have on integration and quality of care. As one stakeholder reported, it is hard to tell which came first: the offer of an APM that led to integration *or* a practice site hiring a behavioral health professional, integrating, and then becoming eligible for an APM.

Of the 18 stakeholders who commented on the extent APMs led to integration and greater quality of care, nine (41%) discussed the challenge of lacking more concrete data (CQM reporting, for example) to confirm their response rating (based on data they had seen at the time of the interview). They desired to see outcomes that supported their expectation of a correlation.

---

[78] Alternative Payment Model Framework and Progress Tracking Work Group. (2016).

Cost of Integration and Sustainability for SIM Primary Care Practice Sites

PR2. What is the cost of integration transformation efforts to SIM primary care practice sites? (Is the cost sustainable for practice sites? Are practices willing to absorb unreimbursed costs?)

The introduction to this chapter and the Methodology chapter of this report discuss challenges in determining the costs of integration transformation efforts to SIM primary care practice sites. More detail is also available in the Final SIM Process Evaluation Report. However, as discussed, closeout surveys do provide practice site insight into the degree to which VBPs were implemented during their SIM-supported transformation activities and whether movement toward more value-based payments and greater integration were sustainable at the time the site ended SIM participation.

Responses from these surveys on the sustainability of integrated care suggest primary care sites are cautiously optimistic about the levels of VBPs they were receiving at the time they concluded their participation in SIM.

Figure 37 below shows responses from those practice sites that had hired an onsite behavioral health provider and were offering integrated care at the time they completed the closeout survey. Of these, 27% of cohort 1, 27% of cohort 2, and 37% of cohort 3 practice sites report that integrated care is sustainable with current revenue streams and structures. A further 28% of cohort 1, 43% of cohort 2, and 24% of cohort 3 sites report that integrated care is sustainable, but not strictly with their current revenue streams. The survey did not require additional clarification of these sources, but when an open-ended question asked for additional feedback immediately after answering revenue questions, 32 cohort 2 sites reported outside grant funding supported their integrated services.

Figure 37. Financial Sustainability of BHP and Integrated Care

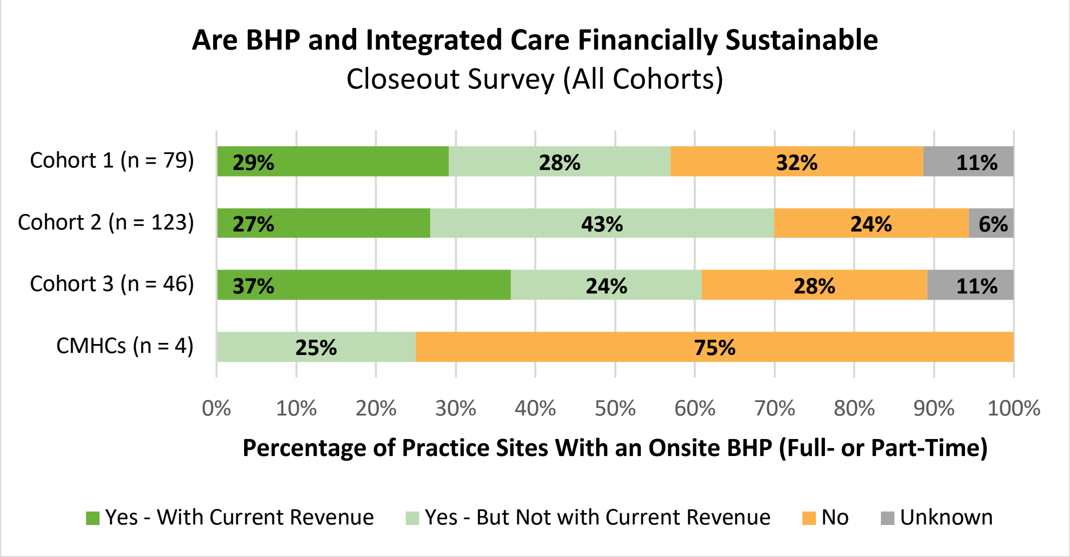
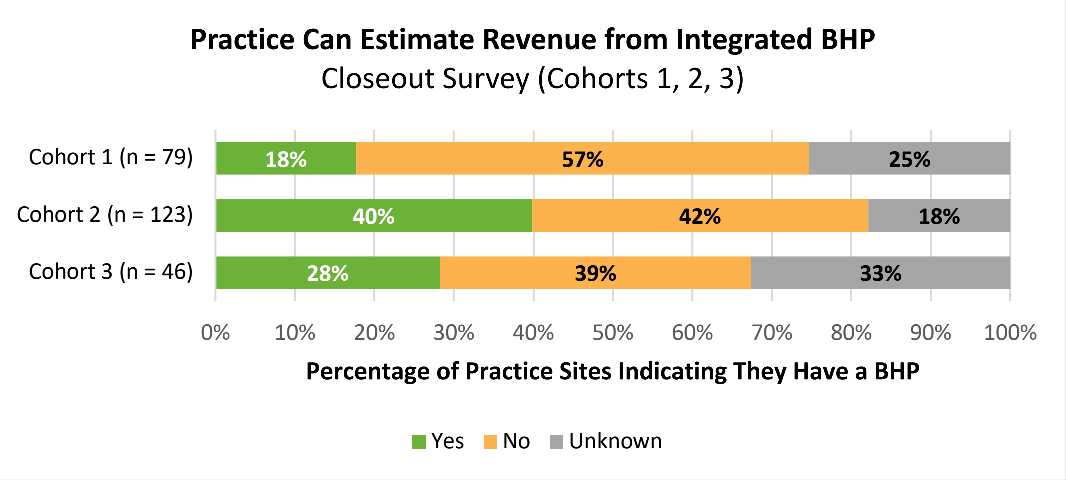


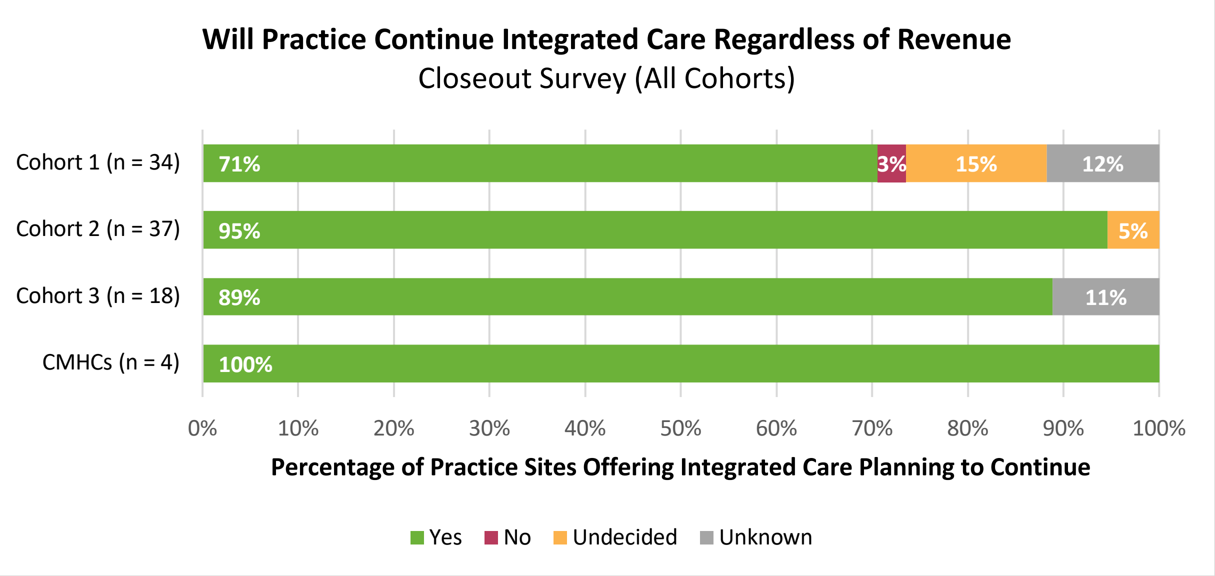
Figure 38. Practice Site Ability to Estimate Revenue from Integrated BHP



Responses to whether these practice sites could estimate revenue from their integrated BHP were mixed, with approximately 18% of cohort 1, 40% of cohort 2, and 28% of cohort 3 sites reporting they were able to estimate revenue (see Figure 38). Being able to determine or estimate the revenue generated from integrated care, presumably from some form of a value-based payment, is a vital component to maintaining and increasing integrated care within the practice. Without knowing whether revenue can cover costs and generate sufficient payments, it is difficult for providers to make investments towards more integrated care delivery.

Even though revenue uncertainty might deter some practices from expanding integrated care efforts, some practice sites did indicate a belief in the value of integrated care, even if it could not currently be fully supported by existing payment models. A subset of SIM practice sites already offering integrated care and that reported at closeout that they were planning to continue largely indicated that they were willing to offer it regardless of revenue (Figure 39 below). There were too few practice sites to draw generalizable conclusions, but these responses suggest that—once practice sites have worked through the challenges to practice transformation towards integration—they are reluctant to discontinue this care.

Figure 39. Practice Site Plans to Continue Integrated Care



An important part of evaluation question PR2 asked whether costs of integrated care were sustainable and whether practice sites were (at least at the time they ended their SIM participation) willing to absorb some reimbursed costs. The data presented above suggests that many sites are willing to absorb some costs—at least in the short term—and that many are willing, and even eager, to sustain their SIM-supported integration efforts.

Summary and Conclusions

Although payment reform has been, and will continue to be, a primary determining factor in the feasibility of integrated physical and behavioral healthcare, challenges loom around data, payment models, education, and communication. Many practice sites reported they will seek to continue offering integrated care, and SPLIT assessments indicated that practice sites from all cohorts have improved their data literacy and data usage to support these transformations.

Still, though, many practices and providers were frustrated over the perceived lack of results from their efforts and from a perceived disinterest among payers to support those efforts. SIM did create spaces for stakeholders—including payers—to collaborate, but actionable results of those efforts were moderately fruitful at best, and there were strong perceptions that payers were not held accountable to the same degree practices were. Additional work around expanding VBPs and increasing clarity and transparency would help correct misperceptions and raise practices' confidence in good-faith payer relationships.

Payer data did not allow for many conclusions regarding the degree to which private payers may have increased value-based payments as a result of the SIM implementation. Based on HCPF-reported data, we observe an increase in patients supported by Medicaid's APM. Between 2015 and 2018 the number of patients supported went from approximately 114,000 to nearly 434,000, almost quadrupling. However, it is difficult to assess how much, if at all, there was a shift towards more value-based payments for SIM-participating primary care practice sites.

6 Cost and Utilization

Introduction

One of the Colorado Quadruple Aims[79] that the SIM initiative targeted is lower costs. Integrating physical and behavioral health and, from that effort, improving the quality of care is designed to reduce unnecessary emergency department visits, hospitalizations, or otherwise worsening of chronic health conditions. All these have a substantial impact on healthcare costs, so reducing incidences can greatly reduce overall costs of care. Based on this logic, we expect that outpatient primary and behavioral healthcare utilization and costs will increase over time as more individuals receive more prevention and disease management care, including behavioral health screenings and referrals with their primary care doctors. This increase in outpatient primary care spending should then lead to decreases in spending for either crisis care (e.g., emergency department visits) or in hospitalizations and other expenses that result from a lack of preventative care or poor disease management.[80]

This chapter examines changes in healthcare costs and the types of healthcare utilization associated with higher spending. The SIM model put forth in the Colorado SIM proposal sought to reduce these costs in both the short and long term and to demonstrate the association between integration, better care, and lower costs.

For the three primary care cohorts and the four CMHCs, the SIM office contracted with Milliman, an actuarial consulting firm, to calculate nine practice site measures related to cost and utilization, described in the following section. Milliman also calculated 16 cost-related measures and seven utilization measures for each attributed patient (member) in the APCD. We used this patient-level data to calculate outcomes for the matched groups of SIM and non-SIM-attributed patients described in the Methodology chapter. Please see that chapter for a complete discussion of the APCD and the limitations of that data set.

There does remain an outstanding issue with APCD data that is particularly important for this cost and utilization analysis. In January 2019, monthly Medicaid cost PMPM calculations produced by HCPF and CIVHC were compared. The results were different and raised questions

---

[79] SIM began with a focus on the "Triple Aim" of lower costs, better care, and better patient experience, then elected to add a focus on workforce during its initial planning year.

[80] For example, see the "Discussion" section of the following report. Breslau, J., Sorbero, M. J., Kusuke, D., Yu, H., Scharf, D. M., Hackbarth, N. S., & Pincus, H. A. (2019, March). Primary and behavioral health care integration program: Impacts on health care utilization, cost, and quality. Retrieved from https://aspe.hhs.gov/system/files/pdf/260996/PBHCIP.pdf

about the quality of the Medicaid data in the Colorado APCD and concerns about the use of the data for the SIM evaluation.[81]

There are two evaluation questions addressed by the analysis of cost and utilization data.

PR4. To what extent did the utilization of services and total cost of care differ over time for consumers attributed to SIM participating practices? Was this different compared to consumers in comparison practices?

This evaluation question requires that we examine change over time for patients attributed to SIM-participating practice sites and for patients attributed to non-SIM providers. We will examine change over time for both the SIM primary care practices and separately for all patients attributed to these practices in aggregate. We do not have data on comparison practices and will instead rely on a matched comparison group of patients attributed to non-SIM primary care NPIs.

PT6. What specific transformation factors (level of integration, milestone targets, data quality, clinician and staff experience, etc.) most influence outcomes?

This evaluation question requires that that we compare changing practice site characteristics and accomplishments to outcomes. Our analysis must therefore be conducted at the practice-site level. The clinical quality measures (CQMs) "outcomes" referenced in this question are addressed in the Clinical Outcomes chapter,[82] and we use the same transformation factors used in that chapter to determine which factors most influenced outcomes. We end the chapter with supporting tables, including summary statistics for most variables as well as breakout tables by practice site characteristic.

Milliman Cost and Utilization Metrics

For each of the reported metrics, we will summarize the practice site average value by assessment period and separately for practice sites from cohort 1 and cohort 2. Because of data limitations, cohort 3 had only baseline values, and we did not include these practice sites in our analysis. Cost metrics include cohort 2 data with and without the inclusion of Medicare data. The Medicare data for cohort 2 related to the utilization of services are currently complete through the end of Q2 2018; however, Medicare Part D prescription data, which are applicable

---

[81] For a full discussion of this issue, please see pages 30–33 of the following report: Center for Improving Value in Health Care. (2019, May). *CO APCD data quality: Payer submission and processing discoveries*. Retrieved from https://www.civhc.org/wp-content/uploads/2019/07/Data-Discovery-May-2019_19.65_public-update.pdf

[82] The CQMs are also discussed in the Methodology chapter of this report.

only to cost analysis, are only complete through the end of 2017. For all of the metrics except total cost of care and out-of-pocket expenses, we have calculated the cohort 2 first-year results using data from July 1, 2017, to June 30, 2018, and baseline results from the prior 12-month period. For the cohort 2 total cost of care and out-of-pocket expenses, we have calculated results with and without Medicare to account for the final two quarters of missing Medicare prescription costs. Because cohort 3 does not have a full year of data after SIM implementation, we did not calculate cohort 3 practice site averages for this chapter.

Following the presentation of each metric's practice site averages, we present associated patient-level outcome results. Practice site averages for these results are only for SIM practice sites, and without any type of comparison practice data, it is not possible to determine if reported changes resulted from participation in SIM or from factors unrelated to SIM. The comparison group analysis allows us to make this determinization. We also have more patient-level than practice-level outcomes and will group patient-level outcomes with associated practice-level outcomes.

We first present the measures related to cost but without corresponding utilization measures. We then present change over time utilization of services for behavioral health conditions and treatments and their costs. We end with utilization of services for physical health related conditions and treatments. Because the SIM primary care practice sites focused on providing integrated care (e.g., progressing along the building blocks and activities identified for their sites), we anticipated the largest changes in utilization of services for behavioral health-related conditions.

A parallel effort to integrate physical healthcare into four community mental health centers (CMHCs) also occurred. Since the population of patients served by these CMHCs was markedly different from the population served by the three primary care cohorts, we have done analysis of costs and utilization for the CMHC population separately. We present these findings in the "CMHC Cost and Utilization" section near the end of this chapter. Based on the logic of adding primary care in a specialty behavioral health setting, we anticipate that outpatient primary care treatment costs of CMHC patients will rise but that emergency department and inpatient hospitalization costs related to comorbid medical conditions will decline.

Change Over Time (Summary of Results)

Because of SIM's focus in primary care sites to integrate behavioral healthcare, we hypothesized that we would observe more improvement in behavioral health metrics than in physical health cost and utilization metrics. And *practice site* measures were consistent with this prediction. For the two cohorts in aggregate, the rate of *Psychiatric Emergency Department Admissions, Psychiatric 30-Day Readmissions*, the (all inpatient) *Emergency Department Admissions,* and *30-Day Readmissions* all showed statistically significant improvements over time for at least one cohort. The rate of *Psychiatric Admissions,* (all inpatient) *Admissions,* and *Percentage of Psychiatric Inpatient Admissions with Follow Up,* showed some statistically significant improvements. Two of the four measures that showed improvement are related to behavioral health conditions. Results were not uniformly positive or statistically significant, but all changes we identified fall into these four measures. *Out of Pocket Expenditures for Consumers* and *Total Cost of Care* also showed statistically significant increases. However, given the rising cost of medical care, this increase was not unanticipated.

Because practice sites would have experienced some change in cost and utilization regardless of participation in SIM, we cannot attribute observed changes specifically to participation in SIM. To estimate the effect of SIM participation, we need to examine change for patients attributed to SIM practice sites in comparison to similar patients attributed to non-SIM providers. We did this using the comparison groups described in the introductory chapter.

Our comparison of changes over time for patients attributed to SIM practice sites to changes for patients attributed to non-SIM providers yielded some differences between the two groups, and the pattern of relative improvement or decline was inconsistent. For cohort 1 using two-year changes (2015 to 2017), statistically significant improvements for SIM-practice-site-attributed patients include lower costs of other medical services (not included elsewhere), greater declines in psychiatric ED visits, greater declines in the cost of ED visits, lower increases in the cost of non-psychiatric inpatient admissions, and greater declines in the cost of non-psychiatric readmissions

Statistically significant changes that represent better outcomes for patients attributed to comparison providers include, for patients attributed to cohort 1 SIM practice sites, higher primary care professional service cost changes, higher cost of psychiatric inpatient readmissions, greater increases in psychiatric hospital readmissions, greater increases in non-psychiatric hospital discharges, and lower decreases in non-psychiatric ED visits.

For cohort 2 SIM-practice-attributed patients, statistically significant improvements from the Quarter 3 (Q3) 2016 through Quarter 2 (Q2) 2017 baseline to Q3 2017 through Q2 2018 Year 1, as compared to the matched comparison group, include greater declines in primary care professional service costs, greater declines in non-psychiatric prescription drug costs, greater declines in the cost of other medical services, greater declines in psychiatric ED visits and costs, and greater declines in the cost of non-psychiatric ED visits.

For the total cost of care, because we were missing Medicare prescription drug costs for the last six months (Q1 and Q2 of 2018) of cohort 2’s first year, we calculated change over time with two methods. First, we compared equivalent periods of complete data: the baseline period of Q3 and Q4 2016 compared to Q3 and Q4 of 2017. This approach resulted in smaller increases in the total cost of care for patients attributed to cohort 2 SIM practice sites, and this result was statistically significant. This result also represents a cost savings from SIM.

The second approach was to use the full year of baseline and Year 1 data, recognizing that the Year 1 data were missing prescription drug costs for Medicare for both SIM-practice-site- and non-SIM-practice-site-attributed patients. The SIM group again had a lower increase in the total cost of care, but the result was not statistically significant.

Statistically significant changes for cohort 2 SIM-practice-site-attributed patients that represent poorer outcomes include increases in psychiatric inpatient hospital admissions and costs and increases in non-psychiatric inpatient admissions.

The most aggregated measure is the total cost of care. In comparison to patients attributed to non-SIM providers, we do not observe robust, statistically significant declines in the total cost of care. For cohort 1, using baseline to Year 1 or baseline to Year 2, we do not have statistically significant results. For cohort 2 we can detect a statistically significant improvement (relative decline) in the total cost of care, but this is based on only six months of data. In that sense, evidence for cost savings resulting from SIM participation is not robust across cohorts and periods of data included. We also do not see a consistent pattern of higher primary care and other outpatient costs and reduced ED and inpatient costs.

Transformation Factors Influencing Cost and Utilization Outcomes—Summary

The evaluation question related to transformation factors references their influence on outcomes. We are able to report on the relationship between the level of practice integration, clinician/staff job satisfaction, HIT improvements, and the Milliman-calculated practice site cost and utilization outcomes. Cross-sectional analysis demonstrated that integration was related to several outcomes such as total cost of care in cohort 2 when Medicare data were excluded, with practice sites with higher levels of integration having lower costs. But there did not appear to be any relationship between changes in integration over time or changes in any cost and utilization measure. Although cross-sectional correlations are interesting, causal relationships of influence are more likely to be identified by correlating changes over time in integration with changes over time in outcomes.

We describe the measurement of clinician and staff satisfaction in the Practice Transformation chapter. No statistically significant results occurred for cohort 1. For cohort 2, statistically significant results include the correlation of improvements over time in staff satisfaction with increased rates of follow ups with psychiatric patients after they were discharged—and a correlation with increasing inpatient readmission rates.

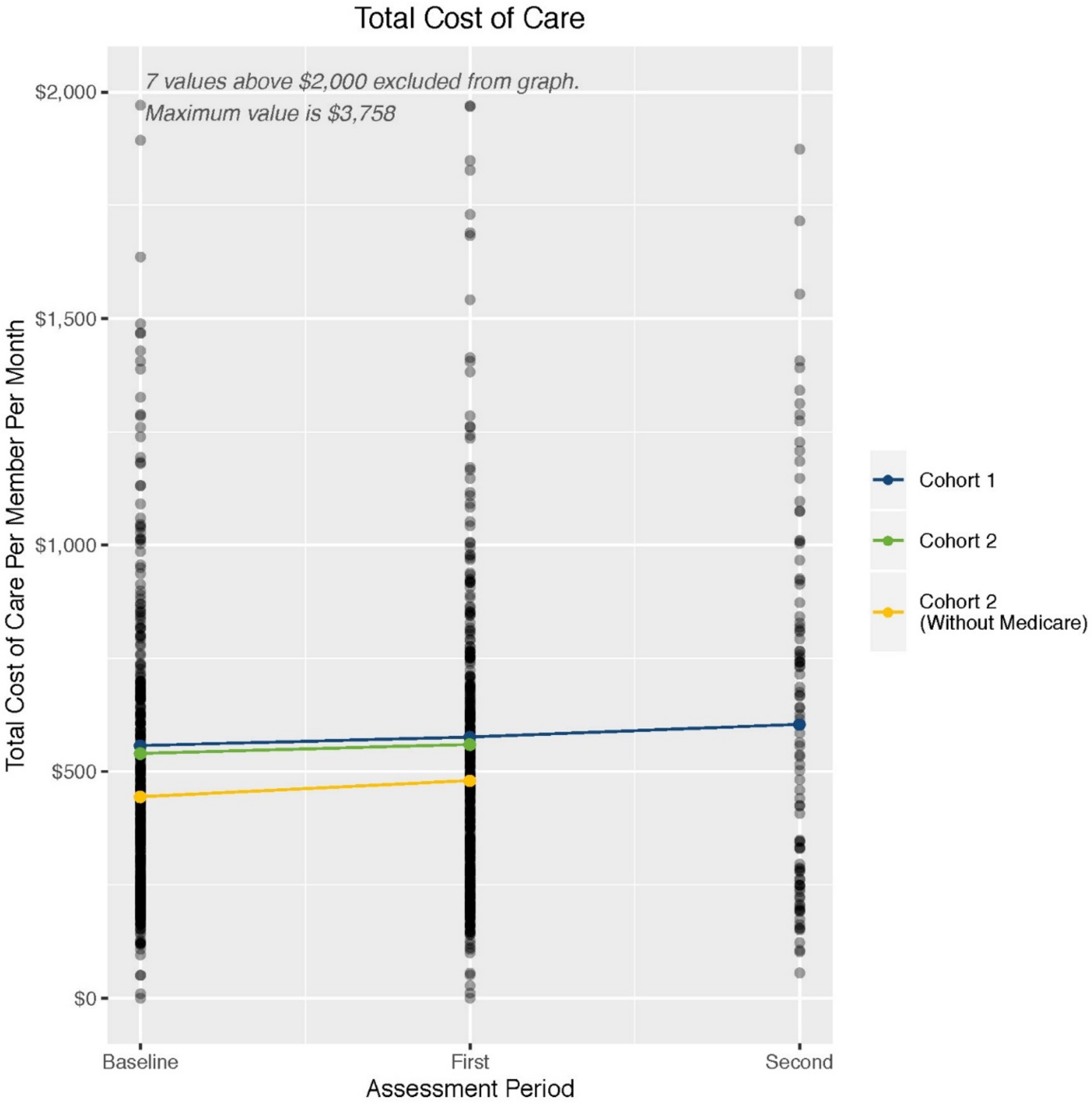
HIT improvements only had a statistically significant correlation with changes in one measure: the inpatient readmission rate for cohort 1. As with staff survey results, the correlation was positive and greater gains in HIT were associated with increasing readmission rates. Based on this analysis, we cannot state conclusively that changes in any of the practice transformation factors had a strong relationship to changes the cost and utilization metrics. Although many metrics had statistically significant cross-sectional correlations to the transformation factors, these did not remain when we examined change over time.

Detailed Analysis of Individual Measures

Total Cost of Care Per Member Per Month

This is the total amount allowed by payers, calculated on the per member per month (PMPM) rate, but using a year of data. As compared to out-of-pocket expenditures, this measure more accurately reflects the cost of medical services provided to patients attributed to SIM sites since it reflects all care, not only the portion patients pay for. As with other outcomes reported in this chapter, we can report on both averages across SIM practice sites and a comparison of total cost of care for patients attributed to providers both participating and not participating in SIM (the SIM and comparison groups). Because we have annual data for this measure, we represent practice site averages for both cohorts on a single time series graph.

Figure 40. Total Cost of Care



Cohort 1 and both versions of cohort 2 showed increases between each assessment period. However, increasing cost of care must considered in the context of generally rising medical costs.

As the distribution of practice site values reveal, each plotted with its own circle, many practice sites had values far from the average, and a small number of sites had values several times the practice-site average. Using paired T-tests, we found that most of the changes over time were positive and statistically significant.

Table 75. Total Cost of Care

Total Cost of Care						
	Baseline		First		Second	
	Cost	N	Cost	N	Cost	N
Cohort 1	$557.50	95	$576.31	95	$604.36	95
Practice Sites Reporting Baseline and First Period	$557.50	95	$576.31	95	-	-
Difference/ P-Value	-	-	$18.80 0.077	-	-	-
Practice Sites Reporting First and Second Periods	-	-	$576.31	95	$604.36	95
Difference/ P-Value	-	-	-	-	$28.05 0.007*	-
Cohort 2	**$540.10**	**145**	**$560.00**	**145**	**-**	**-**
Practice Sites Reporting Both Periods	$540.10	145	$560.00	145	-	-
Difference/ P-Value	-	-	$19.90 0.003*	-	-	-
Cohort 2 without Medicare	$444.57	143	$480	143	-	-
Practice Sites Reporting Both Periods	$444.57	143	$480.15	143	-	-
Difference/ P-Value	-	-	$$35.58 <0.001*	-	-	-

*P-Value ≤ 0.05

The total cost of care had statistically significant increases for cohorts 1 and 2. However, given the generally rising cost of care, these increases may have happened regardless of participation in SIM or may have been larger absent SIM. To address whether SIM participation reduced cost increases, we use the comparison groups described in the Methodology chapter to estimate what cost increase would have occurred for patients attributed to SIM practice sites had they been served by non-SIM providers.

In the three tables that follow, we report results for patients attributed to providers participating in SIM (the SIM group) and matched comparison patients attributed to providers not participating in SIM. We report separately for cohorts 1 and 2. For cohort 1, we report changes from 2015 pre-SIM baseline to Year 1 2016 and for changes from baseline to Year 2 2017.

For each group (SIM and comparison) we report baseline and follow-up average costs, the difference over time, and the difference in difference. For example, the SIM group experienced an increase in average total cost of care of $33.13, which is the difference between the 2015 and 2016 average totals of cost of care. The comparison group experienced a larger increase: $35.86. The difference in these, labeled “Difference in Difference,” is -$2.71. The value is negative because the comparison group had a larger increase in the total cost of care.

We base these figures on simple averaging of costs for each group in each period. Because our patient-level cost and utilization data set does not include data for patients without eligibility during a quarter, costs have the potential for being higher or lower because of differences in eligibility and may therefore not correctly reflect true cost differences. To control for this, we count the number of quarters of data for each patient in both the baseline and follow-up year. When measuring change between the baseline (e.g., 2015) and follow-up year (e.g., 2017), we control for the changing number of quarters of data in each period by using a regression-based difference-in-difference test. The column “Difference in Difference” is the simple change in average, whereas the column “After Controlling for Quarters of Data” is the same difference in difference but based on a regression, allowing us to control for changing eligibility. The P-value corresponds to the regression-based difference in difference. In the case of total cost of care, controlling for quarters of data increases the size of the savings for the SIM practice attributed group to $3.52 per patient per month beyond that of the comparison group. This difference was not statistically significant (see Table 76).

Although examining changes in the total cost of care helps describe difference, additional context is provided by a consideration of component categories of cost. The provision of integrated care is not anticipated to reduce all types of care costs; with integration, patients should receive more outpatient behavioral healthcare, and that category of costs should increase. These increased expenditures should be offset by reductions in the types of costs associated with untreated mental health conditions, such as emergency department and inpatient psychiatric costs. Because much of the SIM program also related to improved screening and treatment of non-behavioral health medical conditions, we anticipate seeing this same increase for outpatient treatment costs and reductions in ED and inpatient costs.

The following three tables present subcategories of cost. Other subcategories, such as those related to inpatient hospitalization and ED use, follow in later tables.

Table 76. Patient Level Analysis: Cohort 1 Total Cost of Care (2015–2016)

		2015	2016	Difference	Difference in Difference	After Controlling for Quarters of Data	P-Value
Total Cost of Care	SIM	$425.25	$458.39	$33.14			
	Comparison	$437.95	$473.80	$35.86	-$2.71	-$3.52	0.238
Non-Psychiatric Outpatient Hospital Costs	SIM	$126.44	$141.13	$14.70			
	Comparison	$130.39	$145.58	$15.19	-$6	-$9	0.701
Psychiatric Outpatient Hospital Costs	SIM	$0.66	$0.65	-$0.01			
	Comparison	$0.91	$0.72	-$0.19	$0.18	$0.17	0.022
Primary Care Professional Services Costs	SIM	$29.59	$27.59	-$2.00			
	Comparison	$30.82	$28.20	-$2.63	$0.63	$0.55	<0.001*
Non-Psychiatric Professional Specialty Care Costs	SIM	$15.91	$15.86	-$0.05			
	Comparison	$16.46	$16.88	$0.42	-$0.47	-$0.48	0.087
Non-Psychiatric Professional Specialty Care Costs	SIM	$6.29	$7.81	$1.52			
	Comparison	$6.58	$7.75	$1.17	$0.35	$0.33	0.044*
Non-Psychiatric Prescription Drug Costs	SIM	$64.12	$73.86	$9.74			
	Comparison	$70.96	$78.89	$7.93	$1.81	$1.61	0.120
Psychiatric Prescription Drug Costs	SIM	$18.75	$19.23	$0.48			
	Comparison	$18.75	$18.59	-$0.16	$0.64	$0.59	0.004*

		2015	2016	Difference	Difference in Difference	After Controlling for Quarters of Data	P-Value
Cost of Other Medical Services	SIM	$127.42	$131.71	$4.29			
	Comparison	$117.47	$122.37	$4.90	-$0.61	-$0.83	0.254

*P-Value ≤ 0.05

In the previous table (Table 76), cost savings that result from practice sites participating in SIM would be identified as *negative* values in the *Difference in Difference* and *After Controlling for Quarters of Data* columns. Only those negative values that have P-values equal to or below 0.05 would generally be considered statistically significant. There are several statistically significant differences, including *Psychiatric Outpatient Hospital Costs, Primary Care Professional Services Costs, Non-Psychiatric Professional Specialty Care Costs, and Psychiatric Prescription Drug Costs.* These statistically significant difference in costs are all positive; the SIM group had larger increases or smaller decreases for each of these.

Because these cost changes cover a period in which practice sites were just beginning to implement integrated care, these results are less compelling than longer-term cost changes. In the table below, we report the same difference-in-difference estimates for cohort 1 and the change from 2015 to 2017.

Table 77. Patient Level Analysis: Cohort 1 Total Cost of Care (2015–2017)

		2015	2017	Difference	Difference in Difference	After Controlling for Quarters of Data	P-Value
Total Cost of Care	SIM	$425.25	$478.15	$52.89			
	Comparison	$437.95	$489.54	$51.59	$1.30	-$0.65	0.853
Non-Psychiatric Outpatient Hospital Costs	SIM	$126.44	$155.55	$29.11			
	Comparison	$130.39	$158.50	$28.12	$0.99	$0.42	0.853
Psychiatric Outpatient Hospital Costs	SIM	$0.66	$0.61	-$0.05			
	Comparison	$0.91	$0.68	-$0.23	$0.18	$0.17	0.097
Primary Care	SIM	$29.59	$26.78	-$2.81			

		2015	2017	Difference	Difference in Difference	After Controlling for Quarters of Data	P-Value
Professional Services Costs	Comparison	$30.82	$27.62	-$3.20	$0.39	$0.28	0.003*
Non-Psychiatric Professional Specialty Care Costs	SIM	$15.91	$16.15	$0.24			
	Comparison	$16.46	$16.73	$0.27	-$0.03	-$0.07	0.748
Non-Psychiatric Professional Specialty Care Costs	SIM	$6.29	$7.31	$1.02			
	Comparison	$6.58	$7.32	$0.74	$0.28	$0.25	0.311
Non-Psychiatric Prescription Drug Costs	SIM	$64.12	$80.39	$16.27			
	Comparison	$70.96	$84.45	$13.49	$2.78	$2.39	0.074
Psychiatric Prescription Drug Costs	SIM	$18.75	$18.15	-$0.60			
	Comparison	$18.75	$17.70	-$1.04	$0.45	$0.37	0.213
Cost of Other Medical Services	SIM	$127.42	$133.26	$5.84			
	Comparison	$117.47	$125.40	$7.93	-$2.09	-$2.57	0.004*

*P-Value ≤ 0.05

For this period, controlling for the number of quarters of data for each patient, the SIM group had lower increases in cost than the comparison group ($0.65 dollars per month). This difference in difference is not statistically or practically significant. The only statistically significant results are lower costs for the SIM group in *Cost of Other Medical Services* and higher costs in *Primary Care Professional Services Costs*. This increase in primary costs is consistent with the results we anticipated.

A general conclusion for cohort 1 SIM-practice-site-attributed patients is that relative to a matched comparison group, the total cost of care did not change. Primary care costs increased slightly, but the costs of other services decreased.

For cohort 2 we have cost and utilization data for only the first year of SIM participation, approximately July 2017 –June 2018. The next table provides the same difference-in-difference cost changes for cohort 2 as previously presented for cohort 1. Because Medicare prescription drug data are not available for 2018, we calculate the four measures that use that data based on six-month comparisons of Q3 and Q4 2016 to Q3 and Q4 2017. The four measures are the total cost of care, out-of-pocket expenses, non-psychiatric prescription drug costs, and psychiatric prescription drug costs. For ease of comparison to cohort 1, we report these on an annual basis by doubling the six-month values.

We also calculated these measures using the full 12-month periods and observed declines from 2016/17 baseline to 2017/18 treatment year consistent with the missing prescription drug costs. Using the full year of data, which is missing six months of Medicare prescription drug costs, we found a difference in difference in the total cost of care of $2.95 per patient per month, with the SIM group having a smaller increase. This result was not statistically significant (P-value 0.516). By limiting the data to the six-month periods with complete Medicare prescription drug costs, the difference in difference increased to $12.61 per member per month, with the SIM group again having the smaller increase. This result was statistically significant (P-value 0.032). We include these results in the subsequent tables.

Table 78. Patient-Level Analysis: Cohort 2 Total Cost of Care (2016/2017–2017/2018)

		2016/2017	2017/2018	Difference	Difference in Difference	After Controlling for Quarters of Data	P-Value
Total Cost of Care	SIM	$522.21	$537.79	$15.58			
	Comparison	$520.98	$549.13	$28.16	-$12.58	-$12.61	0.032*
Non-Psychiatric Outpatient Hospital Costs	SIM	$167.17	$177.98	$10.81			
	Comparison	$164.55	$174.90	$10.35	$0.46	$1.06	0.726
Psychiatric Outpatient Hospital Costs	SIM	$0.41	$0.54	$0.13			
	Comparison	$0.43	$0.50	$0.07	$0.06	$0.06	0.574
Primary Care Professional Services Costs	SIM	$31.78	$30.57	-$1.21			
	Comparison	$32.00	$31.22	-$0.78	-$0.43	-$0.33	0.003*
Non-	SIM	$19.03	$19.19	$0.16			

		2016/2017	2017/2018	Difference	Difference in Difference	After Controlling for Quarters of Data	P-Value
Psychiatric Professional Specialty Care Costs	Comparison	$19.63	$19.44	-$0.19	$0.36	$0.41	0.330
Non-Psychiatric Professional Specialty Care Costs	SIM	$10.18	$11.13	$0.95			
	Comparison	$9.62	$10.35	$0.74	$0.22	$0.26	0.414
Non-Psychiatric Prescription Drug Costs	SIM	$89.54	$92.96	$3.42			
	Comparison	$93.62	$101.27	$7.65	-$4.22	-$4.23	0.055
Psychiatric Prescription Drug Costs	SIM	$21.52	$19.34	-$2.19			
	Comparison	$21.70	$19.07	-$2.63	$0.44	$0.44	0.247
Cost of Other Medical Services	SIM	$142.26	$139.86	-$2.40			
	Comparison	$136.36	$136.38	$0.02	-$2.42	-$2.10	0.037*

*P-Value ≤ 0.05

The cohort 2 SIM group had a smaller increase in costs as compared to its matched comparison group, but the difference was both practically small and not statistically significant. The SIM group did have a statically significant decline relative to the comparison group in Primary Care Professional Services Costs. This decline was unexpected; we had anticipated an increase in primary care costs. The cohort 2 SIM group also had larger declines in *Non-Psychiatric Prescription Drug Costs* and *Cost of Other Medical Services*.

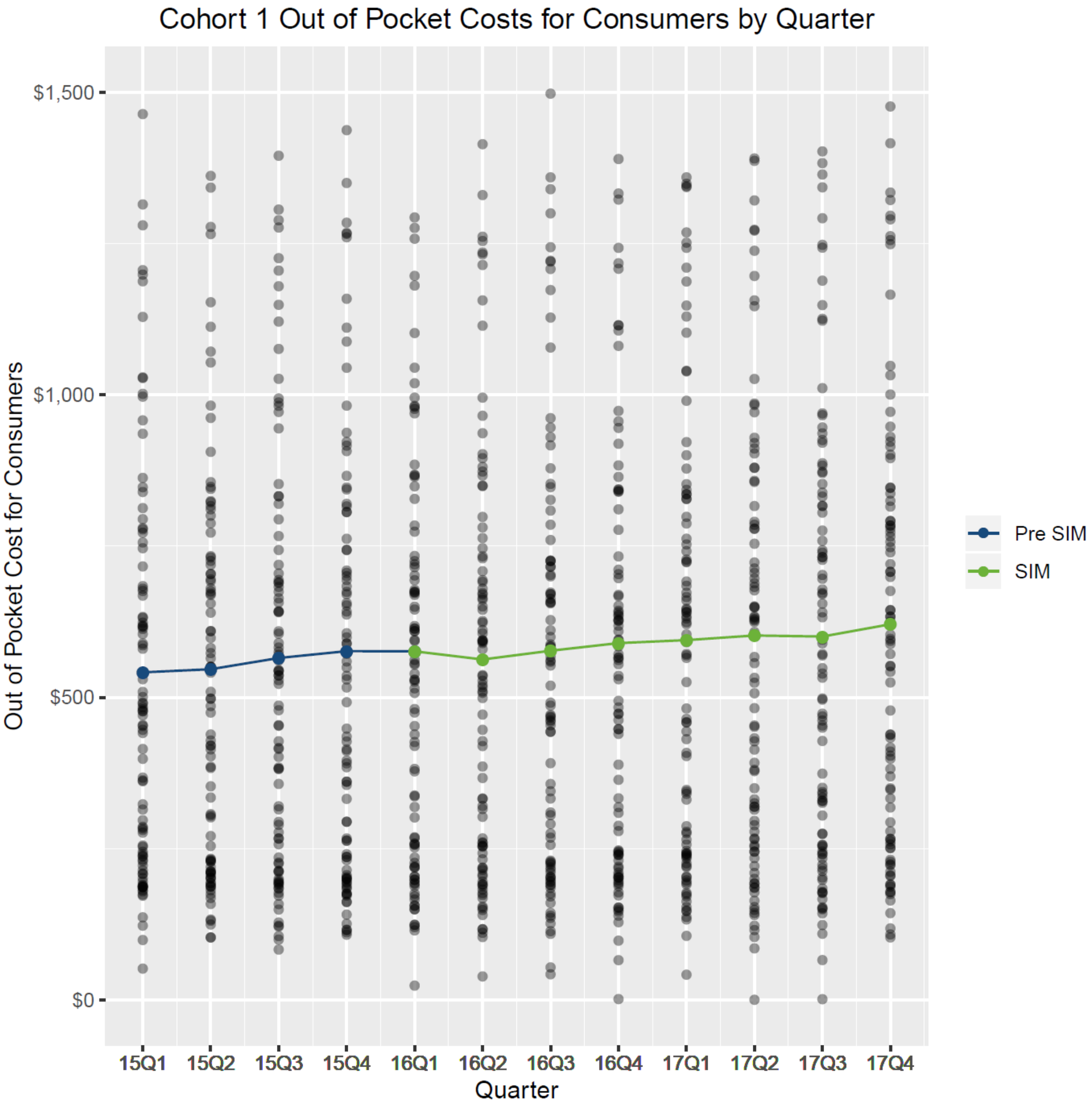
In summary, both cohorts did not have statistically significant total cost savings relative to the comparison groups. Some subcategories of cost did have significant cost differences, but these were not large enough to result in savings in the total cost of care.

Out-of-Pocket Expenditures for Consumers

Unlike the other cost and utilization outcomes, we report out-of-pocket costs on a quarterly basis. For cohort 2, we are using a July 1–June 30 assessment year. And although the other outcomes can be calculated on an annual basis by adding up the corresponding quarters, out-of-pocket costs are reported only for patients who have positive costs, and this number of patients varies by quarter. We cannot therefore report for the July 1–June 30 period on an annual basis.

To aid identifying trends in the data with the larger number of periods, we are graphing change over time separately for cohort 1 and cohort 2. In the next time series graph, each cohort 1 practice site out-of-pocket costs for consumers is represented as a gray dot. the colored dots and connecting lines are average for all practice sites each quarter. The blue line represents baseline values, while the green line is for the first two years of SIM participation.

Table 79. Out-of-Pocket Costs by Quarter (Cohort 1)



To compare the average practice site value over time, we measure the change from the baseline first quarter to the corresponding first quarter in the first or second year of cohort 1's SIM participation. As is visually apparent in the time series graph, the period from Q1 baseline (2015) to Q1 2016 for cohort 1 shows a small $35 quarterly increase, but that increase is not statistically significant. The change from Q1 baseline to Q1 2017 is somewhat larger, at $53 per quarter, and is statistically significant. The table below reports average values for Q1 of each year, and the P-value from paired T-tests of the change between years.

Table 80. Cohort 1 Out-of-Pocket Costs

Cohort 1 Quarterly Out-of-Pocket Costs						
2015 Q1 Mean	2016 Q1 Mean	2017 Q1 Mean	Mean Change 2015-2016	P-Value	Mean Change 2015-2017	P-Value
$541.32	$576.07	$594.61	$34.75	0.100	$53.29	<0.001*

*P-Value ≤ 0.05

To determine whether these rising out-of-pocket costs are greater or lower than what would have occurred absent the SIM program, we use our comparison group. In the table below (Table 81), we report the annual patient liability (out-of-pocket costs) for patients attributed to cohort 1 practice sites and for patients in the matched comparison group in the baseline year of 2015. The baseline mean is an average over all patients in each group, including those with no costs. Based on the patient-level data, patients attributed to SIM practice sites experienced an $8 increase in out-of-pocket costs over the 2015 to 2017 period. Matched comparison patients experienced a $19 increase during the same period. The SIM patients therefore had an $11 lower increase in costs as compared to the comparison patients.

These figures are based on simple averaging of costs for each group in each period. As described previously, the two groups do not necessarily have the same number of quarters of data in each year. To control for this, we count the number of quarters of data for each patient in both the baseline and follow-up year. The column "Difference in Difference" is the simple change in average, while the column "After Controlling for Quarters of Data" is the same difference in difference but based on a regression, allowing us to control for changing eligibility. The P-value corresponds to the regression-based difference in difference. In the case of out-of-pocket costs, controlling for quarters of data increases the size of the savings from for the SIM-practice-site-attributed group to $13 per patient per year. This difference was not statistically significant.

Table 81. Patient-Level Analysis: Annual Out-of-Pocket Costs (Cohort 1)

		2015	2017	Difference	Difference in Difference	After Controlling for Quarters of Data	P-Value
Out-of-Pocket Expenses	SIM	$467.71	$475.73	$8.02			
	Comparison	$466.25	$485.32	$19.07	-$11.05	-$13.34	0.178

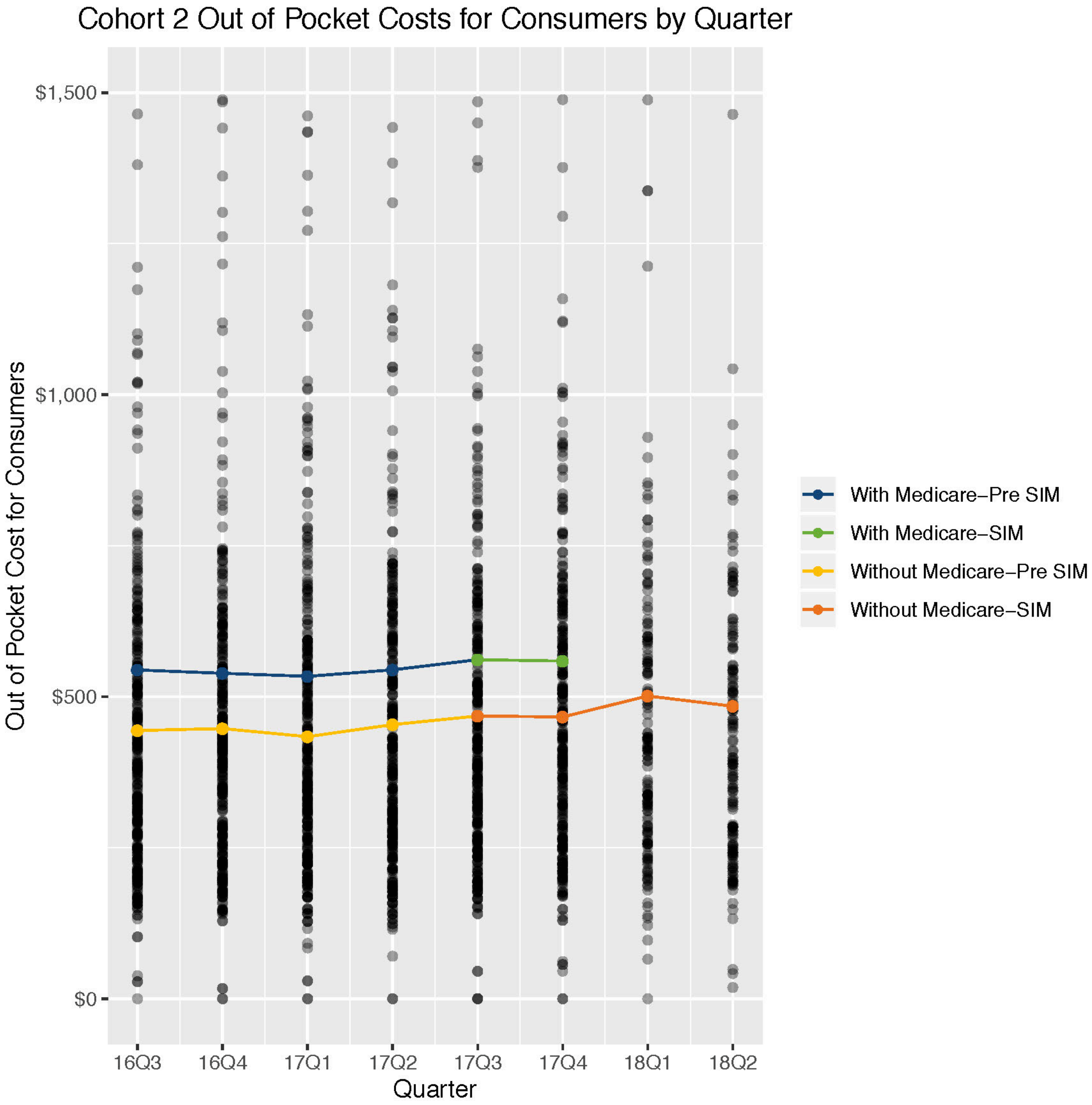
The practice site baseline average of $541 is similar to the SIM group 2015 average of $468, but these two out-of-pocket averages are calculated in a very different manner. For the practice site data from Milliman, we have a quarterly average, whereas the patient-level data reported in the previous table is an annual average. This difference should make the patient-level average much higher. We balance this disparity by modifying the way we compute “average.” For the patient level average, we averaged each patient’s annual out-of-pocket expenses. When Milliman calculated the quarterly average out-of-pocket expenses for practice sites, they used only patients with positive (non-zero) expenses. Because many patients have no out-of-pocket expenses in a given quarter, Milliman used a much smaller denominator than they would have if all patients had been included. This makes the quarterly averages much higher. These two offsetting methodological differences, one causing the average to be lower and the other causing the average to be higher, are one reason the practice-site average is not the same as the patient-level average.

One other difference is that the patient-level averages we calculated are simple averages: total out-of-pocket expenditures for all SIM-attributed patients divided by the number of patients. The practice site averages calculated by Milliman are an average for each practice. We then average across practice sites to create an average of averages. Because we count each practice site equally in making this across-site average, a site with few patients is counted equally as a site with many patients. As such, all patients are not counted equally in this average of averages. Because of these differences in computational approach, trends should only be compared over time, not compared by level of out-of-pocket costs.

In summary for cohort 1, both the practice-site and patient-level averages show a very small increase over time in out-of-pocket expenses. The patient-level analysis includes a comparison group of patients attributed to providers not associated with SIM practice sites, and this group shows a slightly larger increase in out-of-pocket expenses over time. The lack of a statistically significant difference in the comparison group outcome means that we cannot attribute the low level of cost increases to participation in SIM, since patients attributed to non-SIM providers also experienced a small increase over time—and the difference between the groups was not statistically significant.

For cohort 2, which began participation in SIM during Q3 of 2017, we have a full four quarters of outcome data for Medicaid and commercially insured patients. But for Medicare patients we only have prescription expenses through the end of 2017, which is half of cohort 2’s first year of participation. For this reason, we report results with and without Medicare claims and limit the analysis that includes Medicare to the last two quarters of 2017.

Table 82. Out-of-Pocket Costs by Quarter (Cohort 2)



Both series show very similar trends during the overlapping quarters. The comparison of first quarter values of the shifted evaluation year (Q3) shows a small increase over time in out-of-pocket expenditures in each series. The same results occur when we compare Q4 of each year and for each series. These increases are not statistically significant.

Table 83. Cohort 2 Out-of-Pocket Costs

Cohort 2	2016 Q3 Mean	2017 Q3 Mean	Mean Change 2016-2017	P-Value	2016 Q4 Mean	2017 Q4 Mean	Mean Change 2016-2017	P-Value
With Medicare	$544.26	$561.28	$17.02	0.263	$538.71	$559.01	$20.30	0.067
Without Medicare	$444.04	$468.08	$24.04	0.059	$447.20	$466.49	$19.28	0.095

The lack of a statistically significant increase in cohort 2's increases over time in out-of-pocket expenses is confirmed using the comparison group analysis.

Table 84. Patient-Level Analysis: Table of Comparison Group Results for Cohort 2

		2016/2017	2017/2018	Difference	Difference in Difference	After Controlling for Quarters of Data	P-Value
Out-of-Pocket Expenses	SIM	$571.04	$559.38	-$11.66			
	Comparison	$541.48	$509.85	-$31.63	$19.97	$19.93	0.314

Unlike the practice site averages, both the SIM and comparison groups experienced *decreases* in the average out-of-pocket expenses. The comparison group had larger decreases ($32 versus $12), but the difference was not statistically significant.

Summary of Out-of-Pocket Expenses

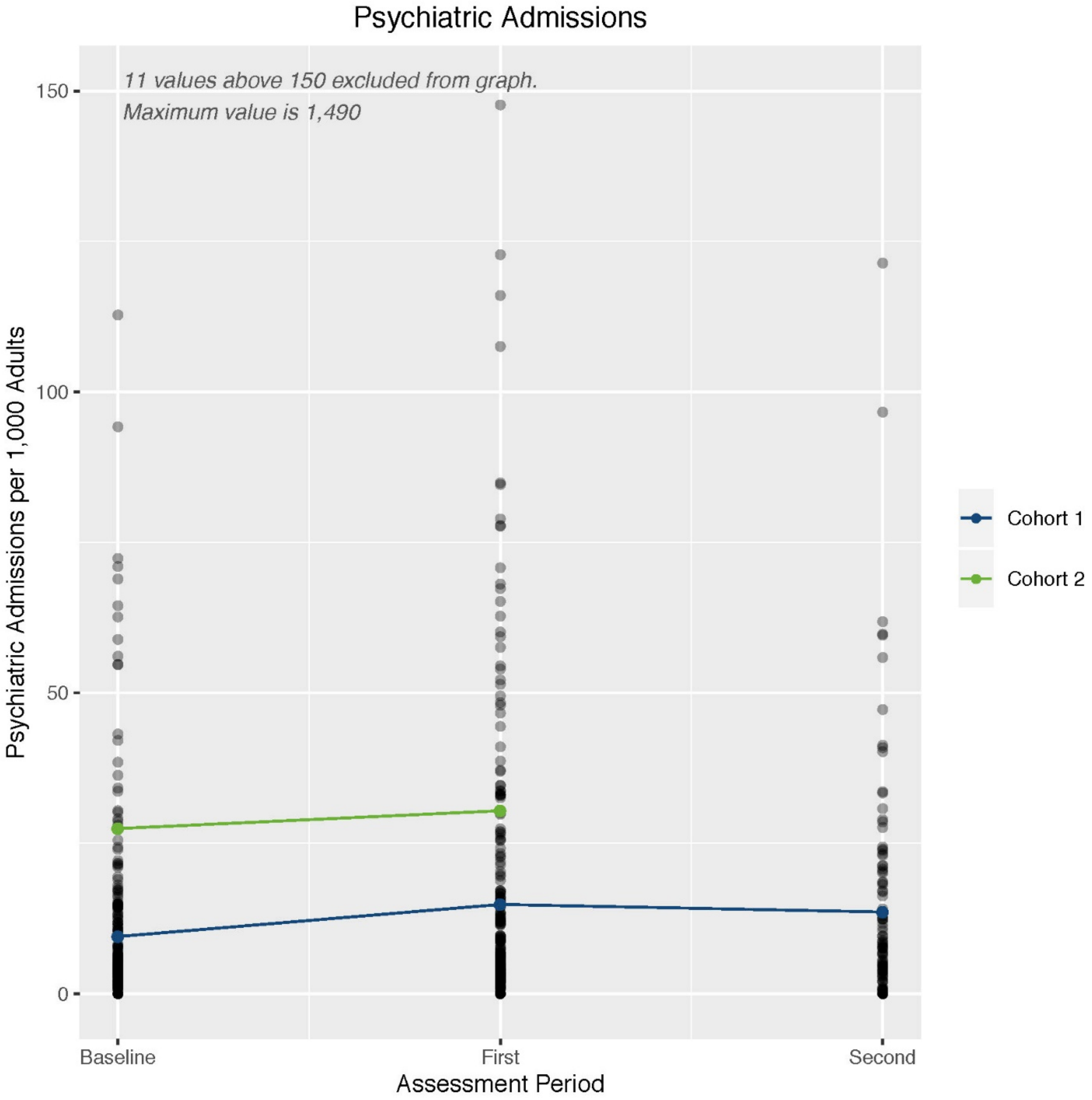
Out-of-pocket expenses were relatively stable during the period covered by the SIM program. Measured at the practice-site level, both cohorts had small increases, but only cohort 1's change was statistically significant. Measured at the patient level, increases were smaller (or negative). None of the comparison group results showed statistically significant changes, and we therefore conclude that the SIM program had no effect on out-of-pocket patient expenses.

The following items are behavioral health-related measures, many of which show improvement over time.

Psychiatric Admission Rate

The calculation of this rate is based on discharges for both mental illness and substance use disorders (SUD). Limits on SUD data within the APCD may result in undercounting SUD admissions. The population included in this calculation is limited to adults ages 18 years and older, and the rate is reported on a per 1,000 population basis.

Figure 41. Psychiatric Admissions



Both cohorts experienced increases in psychiatric admissions from baseline. The change in cohort 1 from baseline to the first assessment period was statistically significant. Some of this change vanished over the second assessment period, and a few outlying practice site values fell far from the mean, driving much of the change that did occur.

Table 85. Psychiatric Admissions

Psychiatric Admissions						
	Baseline		First		Second	
	Value	N	Value	N	Value	N
Cohort 1	**9.49**	**95**	**14.83**	**95**	**13.57**	**95**
Practice Sites Reporting Baseline and First Period	9.49	95	14.83	95	-	-
Difference / P-Value	-	-	5.34 < 0.001*	-	-	-
Practice Sites Reporting First and Second Periods	-	-	14.83	95	13.57	95
Difference / P-Value	-	-	-	-	-1.26 0.398	-
Cohort 2	**27.43**	**145**	**30.40**	**145**	**-**	**-**
Practice Sites Reporting Both Periods	27.43	145	30.40	145	-	-
Difference / P-Value	-	-	3.03 0.228	-	-	-

*P-Value ≤ 0.05

The comparison group analysis results varied by period and cohort. We can report the number of psychiatric inpatient hospital admissions per 1000 primary care patients, and also the cost on a per member per month basis of these admissions.

Table 86. Patient-Level Analysis: Cohort 1 One-Year Change

		2015	2016	Difference	Difference in Difference	After Controlling for Quarters of Data	P-Value
Psychiatric	SIM	6.86	11.37	4.51			

		2015	2016	Difference	Difference in Difference	After Controlling for Quarters of Data	P-Value
Inpatient Hospital Discharges Per 1000 Primary Care Patients	Comparison	29.67	43.50	13.83	-9.32	-9.31	<0.001*
Psychiatric Inpatient Hospital Costs	SIM	$3.07	$3.91	$0.84			
	Comparison	$7.69	$10.57	$2.89	-$2.05	-$2.05	<0.001*

*P-Value ≤ 0.05

Table 87. Patient-Level Analysis: Cohort 1 Two-Year Change

		2015	2017	Difference	Difference in Difference	After Controlling for Quarters of Data	P Value
Psychiatric Inpatient Hospital Discharges Per 1000 Primary Care Patients	SIM	6.86	11.78	4.93			
	Comparison	29.67	32.61	2.95	1.98	1.74	0.378
Psychiatric Inpatient Hospital Costs PMPM	SIM	$3.07	$3.86	$0.79			
	Comparison	$7.69	$7.67	-$0.02	$0.81	$0.75	0.101

Table 88. Patient-Level Analysis: Cohort 2 One-Year Change

		2016/2017	2017/2018	Difference	Difference in Difference	After Controlling for Quarters of Data	P-Value
Psychiatric Inpatient Hospital Discharges Per 1000 Primary Care Patients	SIM	23.81	25.08	1.27			
	Comparison	39.78	32.73	-7.06	8.33	8.46	<0.001*
Psychiatric Inpatient Hospital Costs	SIM	$5.29	$6.12	$0.83			
	Comparison	$8.29	$7.67	-$0.63	$1.46	$1.49	<0.001*

*P-Value ≤ 0.05

In contrast to the comparison group, cohort 1 experienced a statistically significant decline in admissions and costs from baseline to Year 1. But by Year 2 that result reversed to a non-significant increase. For both admissions and their costs, cohort 2 experienced an increase in psychiatric admissions while its comparison group had a decrease. These results were statistically significant.

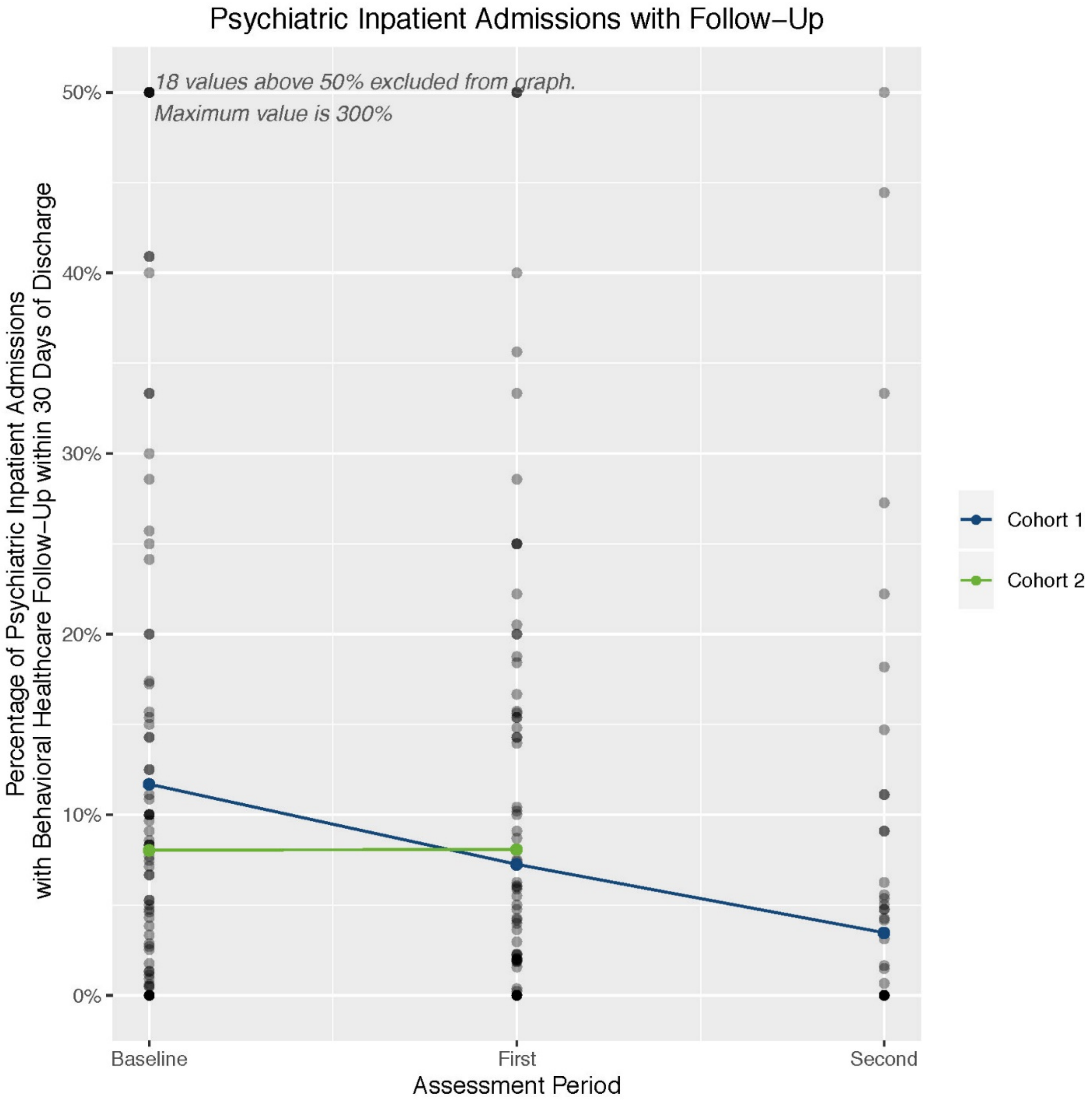
In summary, inpatient psychiatric admissions displayed differing patterns by cohort, leading to the absences of a clear conclusion on the effect of participation in SIM.

Percentage of Psychiatric Inpatient Admissions with Follow Up within 30 Days of discharge

This measure calculates the percentage of behavioral health inpatient admissions that have a follow-up visit with a mental health practitioner within 30 days of discharge. The metric includes all patients ages six years or older. This measure has the potential to be directly influenced by the provision of integrated care since primary care practices can coordinate community-based follow-up care for their patients who experience inpatient psychiatric admissions.

In calculating the percentage, Milliman counts for each practice and any given patient the number of behavioral health outpatient visits that occur within 30 days of any psychiatric inpatient admission and divides this by the total number of inpatient admissions. In that sense, a patient could have more than one outpatient visit per inpatient visit, or a rate of greater than 100%. The maximum reported for a practice site is 300%.

Figure 42. Psychiatric Inpatient Admissions with Follow Up



Cohort 1 experienced an average decrease from baseline to Year 1 and a larger and statistically significant decrease from Year 1 to Year 2.

While these results are contrary to our expectations, the large number of practice sites with very small numerators and denominators raises concerns about the applicability of this measure to integrated primary care practices. Either we were not capturing psychiatric admissions and follow up in our data or patients attributed to primary care practice sites did not tend to have behavioral health conditions that led to frequent inpatient admissions. This may be a better measure for the CMHCs.

Although the change over time was generally in the wrong direction, of greater concern is the very low level of follow up. As is presented in Table 88, an average of 11.7% of psychiatric inpatient admissions received an outpatient 30-day follow up for cohort 1 practice sites at baseline. This level declined to 7.3% at the first assessment period and 3.5% at the second assessment period. Once again, either the claims data were incomplete or patients with psychiatric admissions had been receiving follow-up care at very low rates.

Table 89. Psychiatric Inpatient Admissions with Follow Up

Psychiatric Inpatient Admissions with Follow Up						
	Baseline		First		Second	
	Value	N	Value	N	Value	N
Cohort 1	**11.7%**	**79**	**7.3%**	**85**	**3.5%**	**81**
Practice Sites Reporting Baseline and First Period	11.8%	78	7.9%	78	-	-
Difference / P-Value	-	-	-3.9% 0.161	-	-	-
Practice Sites Reporting First and Second Periods	-	-	8.1%	76	3.4%	76
Difference / P-Value	-	-	-	-	-4.7% 0.027*	-
Cohort 2	**8.0%**	**112**	**8.1%**	**103**	**-**	**-**
Practice Sites Reporting Both Periods	7.1%	95	7.2%	95	-	-
Difference / P-Value	-	-	0.0% 0.992	-	-	-

*P-Value ≤ 0.05

As with psychiatric and non-psychiatric inpatient readmissions reported below, the comparison group analysis of follow up after psychiatric admission requires a different approach than we use with most of the cost and utilization measures. Fewer than half of the patients with an inpatient psychiatric admission in the baseline period also had an admission in subsequent periods. Because having a follow-up outpatient behavioral health visit is predicated on first having an admission, we could not calculate rates of follow up for most of the patients in the SIM and comparison groups. Since our other analysis involved differencing each patient's outcomes over time, we require a different approach.

As a solution, we conducted difference in difference using a methodology appropriate for two independent samples, one in the baseline period and a second at Year 1 or Year 2. With this approach, we calculate the difference in the average follow-up rate in both the baseline and follow-up period as independent averages. The averages do not include the same patients in each period. And because we are not using all patients, the comparison patient weights are not correct, since the weights assume all comparison members are used. By using regression to calculate difference in difference, we can control for patient characteristics such as age, MARA risk score, and the other variables we used to develop the comparison group. We are therefore controlling for heterogeneity in the SIM and comparison group through multiple regression instead of CEM weights.

Table 90. Patient-Level Analysis: Cohort 1 One-Year Change

		2015	2016	Difference	Difference in Difference	After Controlling for Matching Variables	P-Value
Rate of Follow Up After Psychiatric Discharge	SIM	13.2%	9.6%	-3.6%			
	Comparison	7.6%	6.9%	-0.7%	-2.8%	-1.2%	0.504

Table 91. Patient-Level Analysis: Cohort 1 Two-Year Change

		2015	2017	Difference	Difference in Difference	After Controlling for Match Variables	P-Value
Rate of Follow Up After Psychiatric Discharge	SIM	13.2%	7.4%	-5.8%			
	Comparison	7.6%	4.1%	-3.5%	-2.4%	-1.5%	0.383

Table 92. Patient-Level Analysis: Cohort 2 One-Year Change

		2016/2017	2017/2018	Difference	Difference in Difference	After Controlling for Match Variables	P-Value
Rate of Follow	SIM	7.5%	6.2%	-1.3%			

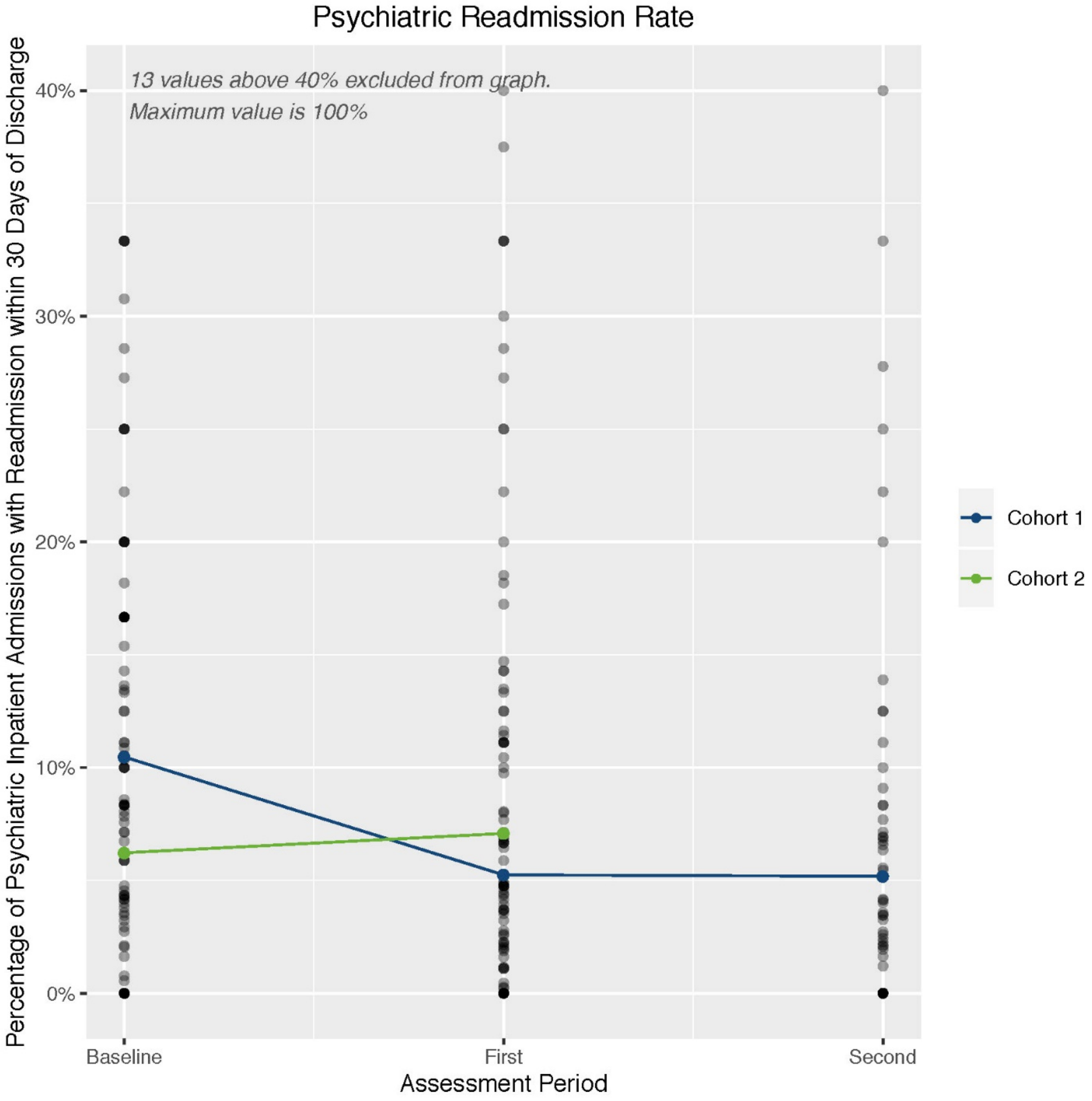
		2016/2017	2017/2018	Difference	Difference in Difference	After Controlling for Match Variables	P-Value
Up After Psychiatric Discharge	Comparison	4.8%	4.4%	-0.4%	-0.9%	-0.2%	0.895

Consistent with the practice-site-level analysis, follow-up rates declined for both cohorts during one- and two-year differences. Unlike the practice-site-level results, the difference-in-difference results were not statistically significant. This tells us that the declining rates observed by SIM practice sites were not the result of participating in SIM since the declines were also observed by the comparison groups.

Psychiatric Readmission Rates

Psychiatric readmissions have the potential for reduced rates when patients receive quality outpatient post-discharge care. The reported rate is the percentage of behavioral health inpatient discharges that have a readmission within 30 days of initial discharge.

Figure 43. Psychiatric Readmission Rate



Cohort 1 experienced a statistically significant decline in the practice site average from baseline to Year 1 of SIM. This was followed by a small increase that was not statistically significant. Cohort 2's results depended on the inclusion of practice sites reporting both periods. Excluding these data resulted in an increase in the rate, whereas including them resulted in a non-statistically significant decrease.

Although direct evidence of improvement in post-discharge follow up is lacking, the reduction in readmission rates offers some evidence that SIM patients have been receiving better care.

The change from baseline for patients attributed to cohort 1 practice sites is particularly impressive.

Table 93. Psychiatric Readmission Rate

Psychiatric Readmission Rate	Baseline		First		Second	
	Value	N	Value	N	Value	N
Cohort 1	**10.5%**	**66**	**5.2%**	**77**	**5.2%**	**77**
Practice Sites Reporting Baseline and First Period	9.8%	64	5.0%	64	-	-
Difference / P-Value	-	-	-4.7% 0.002*	-	-	-
Practice Sites Reporting First and Second Periods	-	-	5.2%	68	5.7%	68
Difference / P-Value	-	-	-	-	0.5% 0.792	-
Cohort 2	**6.2%**	**99**	**7.1%**	**96**	**-**	**-**
Practice Sites Reporting Both Periods	7.0%	85	6.8%	85	-	-
Difference / P-Value	-	-	-0.2% 0.902	-	-	-

*P-Value ≤ 0.05

As with psychiatric follow-up rates, readmission rates require an initial admission, which most patients did not have. We could therefore not calculate the difference in each patient's rate from baseline to the one- or two-year follow-up period since most patients were missing one or both values. Our solution was the same as psychiatric follow up; we used a two-sample regression technique, controlling heterogeneity in patient characteristics through multiple regression. We are also able to report on the cost of psychiatric inpatient hospital readmissions. Because these are not rates, but instead costs per person, we are able to use our first difference methodology.

Table 94. Patient-Level Analysis: Cohort 1 One-Year Change

		2015	2016	Difference	Difference in Difference	After Controlling for Matching Variables	P-Value
Psychiatric Hospitalization Readmission Rate	SIM	4.6%	3.0%	-1.5%			
	Comparison	7.3%	3.5%	-3.8%	2.3%	2.3%	0.004*
Cost of Psychiatric Readmissions PMPM	SIM	$0.20	$0.17	-$0.02			
	Comparison	$0.73	$0.21	-$0.52	$0.49	$0.49	<0.001*

*P-Value ≤ 0.05

Table 95. Patient-Level Analysis: Cohort 1 Two-Year Change

		2015	2017	Difference	Difference in Difference	After Controlling for Match Variables	P-Value
Psychiatric Hospitalization Readmission Rate	SIM	4.6%	5.4%	0.8%			
	Comparison	7.3%	4.0%	-3.4%	4.2%	4.1%	< 0.001*
Cost of Psychiatric Readmissions PMPM	SIM	$0.20	$0.19	$0.00			
	Comparison	$0.73	$0.16	-$0.56	$0.56	$0.56	< 0.001*

*P-Value ≤ 0.05

Table 96. Patient-Level Analysis: Cohort 2 One-Year Change

		2016/2017	2017/2018	Difference	Difference in Difference	After Controlling for Match Variables	P-Value
Psychiatric Hospitalization Readmission Rate	SIM	4.7%	4.4%	-0.3%			
	Comparison	3.5%	3.7%	0.2%	-0.5%	-0.2%	0.788
Cost of Psychiatric Readmissions PMPM	SIM	$0.21	$0.20	-$0.01			
	Comparison	$0.18	$0.19	$0.01	-$0.02	-$0.02	0.713

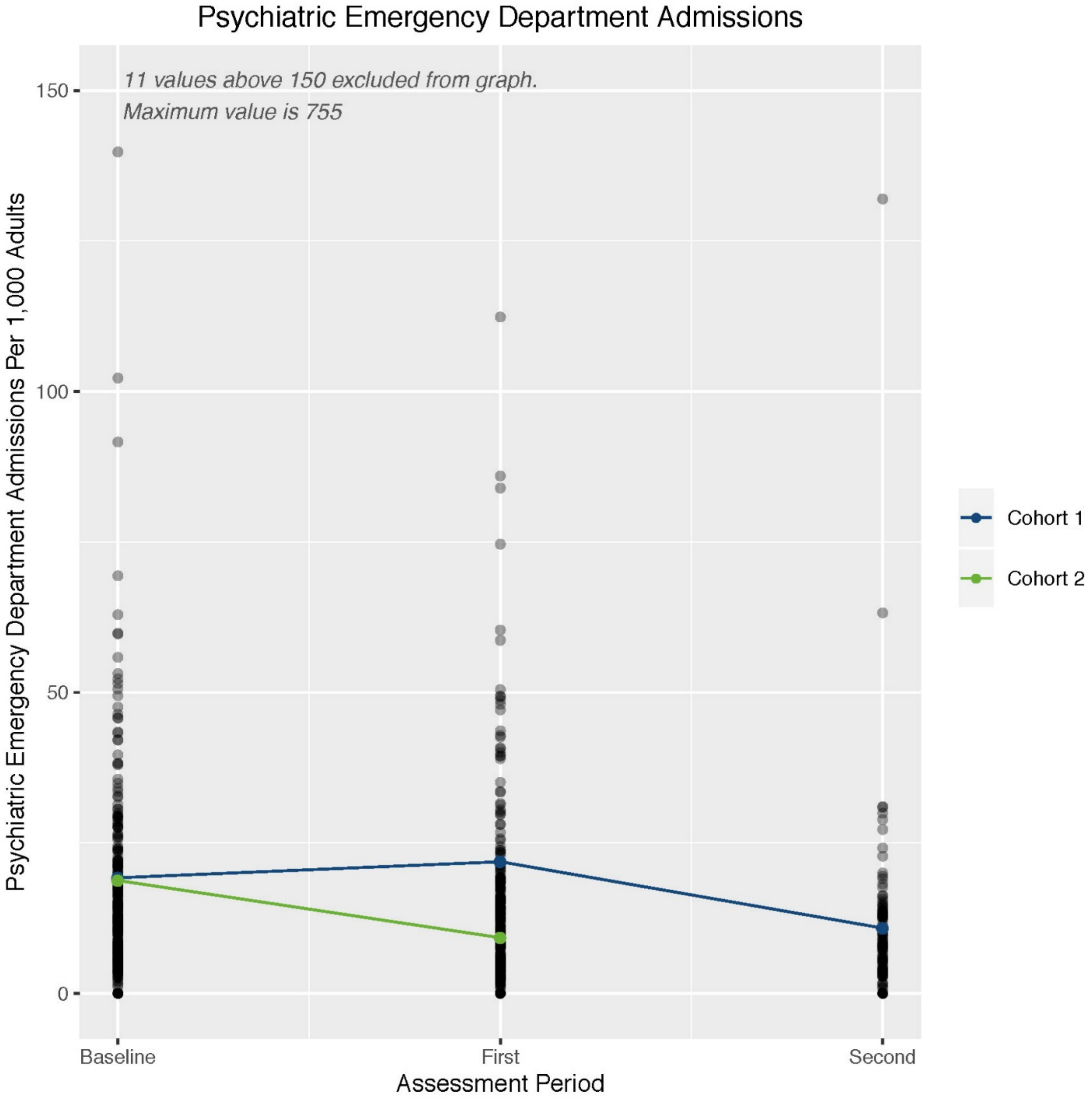
As with the practice site analysis, analysis of the SIM group showed a decline from 2015 to 2016. This decline was not as large as the decline in readmissions for the comparison group, and the difference was statistically significant. The same occurred for the 2015 to 2017 change for cohort 1. For both periods, the SIM group's readmission costs were greater than the comparison groups' costs, and the differences were statistically significant. Cohort 2 had a larger decline in both readmission rate and cost than its comparison group, but the difference was not statistically significant.

Based on these results, the decline in psychiatric readmission rates was not the result of SIM participation, since the SIM groups did not have statistically significant greater declines than the comparison groups.

Psychiatric Emergency Department Admissions

We report this measure in the same manner as we reported non-behavioral health ED visits: as a rate per 1,000 population. Patients with poor access to community-based behavioral healthcare are likely to use the emergency department at higher rates. Integrated care, therefore, has the potential to directly change the psychiatric ED rate.

Figure 44. Psychiatric Emergency Department



Cohort 1 experienced a large but not statistically significant increase from baseline to the first year of SIM, then a larger and statistically significant decrease. Cohort 2 practice data showed statistically significant decreases in the rate of emergency department utilization for behavioral health conditions. Excluding cohort 1's initial increase, these results are a promising indicator that behavioral health crisis ED visits may decline when patients receive integrated primary care.

Table 97. Psychiatric Emergency Department Admissions

Psychiatric Emergency Department Admissions						
	Baseline		First		Second	
	Value	N	Value	N	Value	N
Cohort 1	**19.19**	**95**	**21.86**	**95**	**10.83**	**95**
Practice Sites Reporting Baseline and First Period	19.19	95	21.86	95	-	-
Difference / P-Value	-	-	2.7 0.244	-	-	-
Practice Sites Reporting First and Second Periods	-	-	21.86	95	10.83	95
Difference / P-Value	-	-	-	-	-11 < 0.001*	-
Cohort 2	**18.74**	**145**	**9.24**	**145**	**-**	**-**
Practice Sites Reporting Both Periods	18.74	145	9.24	145	-	-
Difference / P-Value	-	-	-9.5 < 0.001*	-	-	-

*P-Value ≤ 0.05

Table 98. Patient-Level Analysis: Cohort 1 One-Year Change

		2015	2016	Difference	Difference in Difference	After Controlling for Quarters of Data	P-Value
Psychiatric ED Visits Per 1000 Primary Care Patients	SIM	18.60	28.71	10.11			
	Comparison	15.99	22.14	6.14	3.96	3.9	<0.001*
Cost of Psychiatric ED Visits PMPM	SIM	$0.66	$1.11	$0.46			
	Comparison	$0.58	$0.91	$0.33	$0.12	$0.12	0.010*

*P-Value ≤ 0.05

Table 99. Patient-Level Analysis: Cohort 1 Two-Year Change

		2015	2017	Difference	Difference in Difference	After Controlling for Quarters of Data	P-Value
Psychiatric ED Visits Per 1000 Primary Care Patients	SIM	18.60	10.60	-8.00			
	Comparison	15.99	9.27	-6.72	-1.28	-1.34	0.033*
Cost of Psychiatric ED Visits	SIM	$0.66	$0.44	-$0.22			
	Comparison	$0.58	$0.46	-$0.12	-$0.09	-$0.10	0.045

*P-Value ≤ 0.05

Table 100. Patient-Level Analysis: Cohort 2 One-Year Change

		2015	2017	Difference	Difference in Difference	After Controlling for Quarters of Data	P-Value
Psychiatric ED Visits Per 1000 Primary Care Patients	SIM	15.91	7.30	-8.62			
	Comparison	13.48	7.78	-5.70	-2.92	-2.90	< 0.001*
Cost of Psychiatric ED Visits PMPM	SIM	$0.70	$0.40	-$0.30			
	Comparison	$0.72	$0.52	-$0.20	-$0.11	-$0.10	0.114

*P-Value ≤ 0.05

Relative to the comparison groups, the cohort 1 group had a statistically significant increase in psychiatric ED visits at the one-year point, and costs also increased. But the two-year change was the opposite: a statistically significant decline. Cohort 2 had a statistically significant decline in visits and costs.

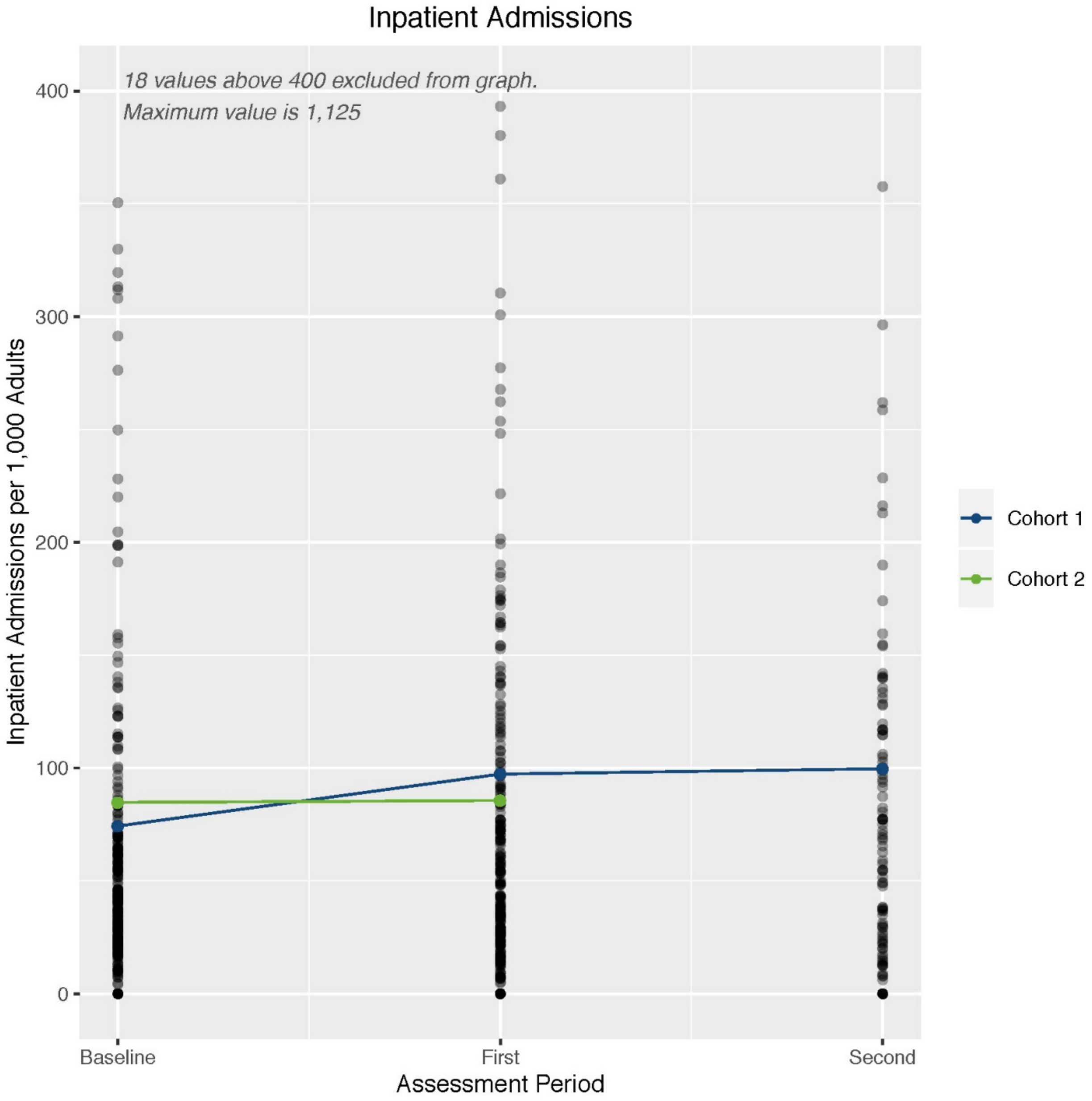
The effect of SIM participation on psychiatric ED use therefore varies by period and cohort. But for the longest periods available, both cohorts demonstrated a decline in psychiatric ED use and costs.

The remaining measures relate to non-behavioral health medical conditions.

Inpatient Admission Rates

We based calculations of the admission rate on hospital discharges for any cause for patients age 18 years and older. For ease of interpretation, we converted these to an annual per 1,000 population basis.

Figure 45. Inpatient Admissions



Cohort 1 practice sites had increases in admission from baseline to Year 1 and further increases from the first to second year of SIM participation. The second year of increases was smaller than the first, which may indicate delayed effects from participating in SIM. With a very wide distribution in each period, and only small changes in the average across practice sites, only change from baseline to the first assessment period was statistically significant. Cohort 2 had similar performances in both assessment periods leading to no statistically significant changes.

Table 101. Inpatient Admissions

Inpatient Admissions						
	Baseline		First		Second	
	Value	N	Value	N	Value	N
Cohort 1	**74.30**	**95**	**97.34**	**95**	**99.72**	**95**
Practice Sites Reporting Baseline and First Period	74.30	95	97.34	95	-	-
Difference / P-Value	-	-	23.04 0.006*	-	-	-
Practice Sites Reporting First and Second Periods	-	-	97.34	95	99.72	95
Difference / P-Value	-	-	-	-	2.4 0.663	-
Cohort 2	**84.77**	**145**	**85.65**	**145**	**-**	**-**
Practice Sites Reporting Both Periods	84.77	145	85.65	145	-	-
Difference / P-Value	-	-	0.88 0.869	-	-	-

*P-Value ≤ 0.05

Using the patient-level comparison groups, we can compare the SIM groups' changes in inpatient admissions and costs to those of the comparison groups.

Table 102. Patient-Level Analysis: Cohort 1 One-Year Change

		2015	2016	Difference	Difference in Difference	After Controlling for Quarters of Data	P-Value
Non-Psychiatric	SIM	48.18	63.30	15.12			

		2015	2016	Difference	Difference in Difference	After Controlling for Quarters of Data	P-Value
Inpatient Hospital Discharges Per 1000 Primary Care Patients	Comparison	83.06	97.64	14.57	0.55	0.62	0.780
Non-Psychiatric Inpatient Hospital Costs PMPM	SIM	$19.00	$21.56	$2.56			
	Comparison	$23.20	$28.93	$5.73	-$3.18	-$3.10	0.003*

Table 103. Patient-Level Analysis: Cohort 1 Two-Year Changes

		2015	2017	Difference	Difference In Difference	After Controlling for Quarters of Data	P-Value
Non-Psychiatric Inpatient Hospital Discharges Per 1000 Primary Care Patients	SIM	48.18	66.20	18.02			
	Comparison	83.06	92.95	9.89	8.13	7.54	0.010*
Non-Psychiatric Inpatient Hospital Costs PMPM	SIM	$19.00	$22.90	$3.90			
	Comparison	$23.20	$29.81	$6.62	-$2.72	-$2.86	0.012*

Table 104. Patient-Level Analysis: Cohort 2 One-Year Changes

		2016/2017	2017/2018	Difference	Difference In Difference	After Controlling for Quarters of Data	P-Value
Non-Psychiatric Inpatient Hospital Discharges Per 1000 Primary Care Patients	SIM	70.70	71.01	0.31			
	Comparison	95.45	90.01	-5.43	5.74	6.20	0.046*
Non-Psychiatric Inpatient Hospital Costs PMPM	SIM	$20.74	$24.68	$3.94			
	Comparison	$27.04	$29.49	$2.45	$1.49	$1.62	0.304

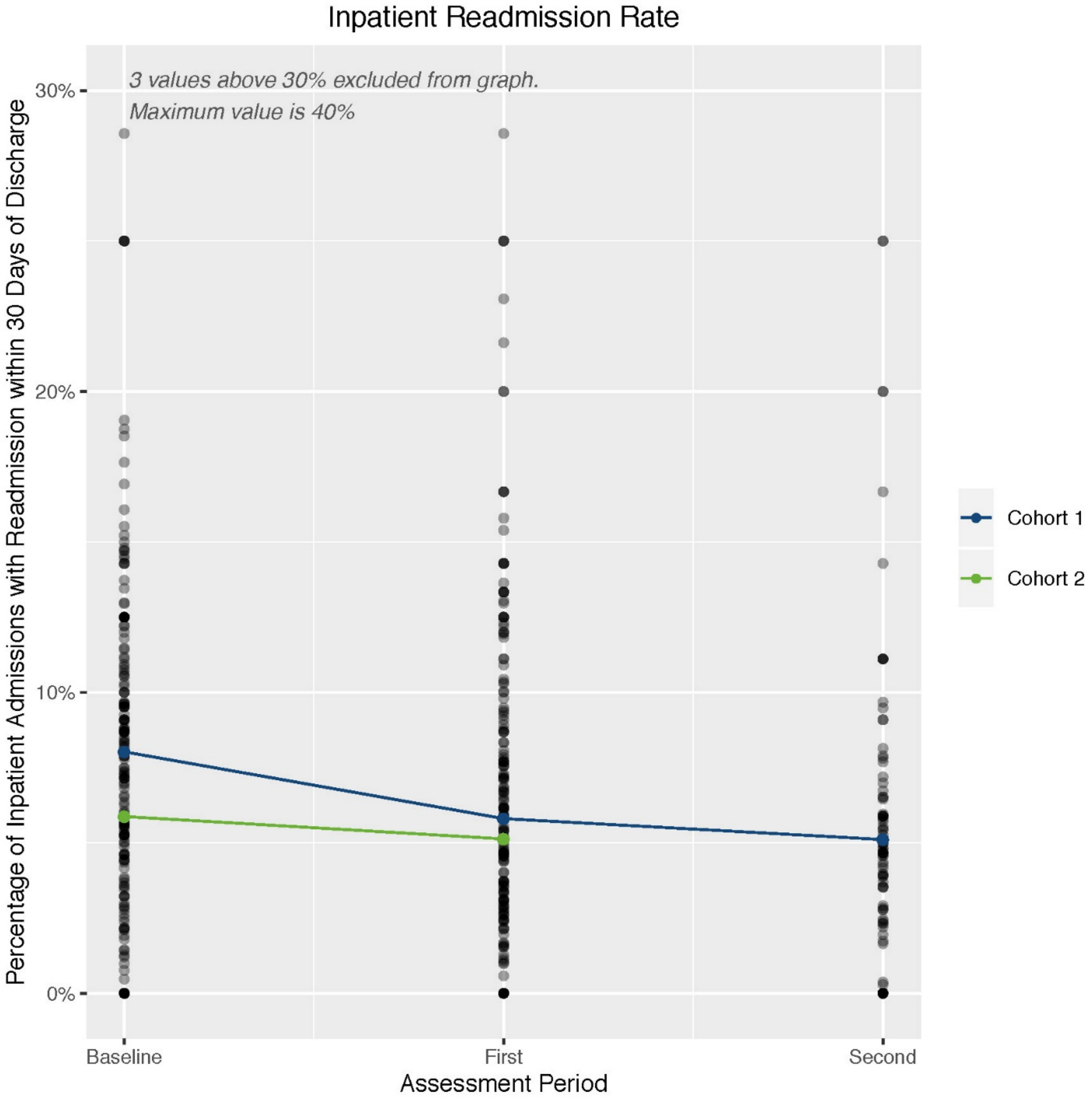
For cohort 1, the SIM group had larger increases in admissions but smaller increases in costs. These inpatient admission cost savings were statistically significant. Cohort 2 had statistically significant increases in admissions, but the difference in cost changes was not statistically significant.

Based on these results, participation in SIM did not reduce non-psychiatric inpatient admissions or costs uniformly.

Inpatient Readmission Rate

This rate uses the number of inpatient hospitalization discharges within 30 days of a previous discharge for the numerator. The denominator is the total number of inpatient discharges. Both initial and subsequent discharges are limited to non-psychiatric admissions and the metric is limited to adult patients ages 18 years and older.

Figure 46. Inpatient Readmission Rate



This measure showed declines in readmissions from baseline to Year 1 for both cohorts, and the declines were statistically significant for cohort 1. These declining rates are consistent with improvements in access to good follow-up care after discharge.

Table 105. Inpatient Readmission Rate

Inpatient Readmission Rate						
	Baseline		First		Second	
	Value	N	Value	N	Value	N
Cohort 1	**8.0%**	**89**	**5.8%**	**91**	**5.1%**	**89**
Practice Sites Reporting Baseline and First Period	8.0%	89	5.9%	89	-	-
Difference / P-Value	-	-	-2.1% 0.008*	-	-	-
Practice Sites Reporting First and Second Periods	-	-	5.9%	89	5.1%	89
Difference / P-Value	-	-	-	-	-0.8% 0.224	-
Cohort 2	**5.9%**	**128**	**5.1%**	**126**	**-**	**-**
Practice Sites Reporting Both Periods	6.1%	123	5.3%	123	-	-
Difference / P-Value	-	-	-0.9% 0.301	-	-	-

*P-Value ≤ 0.05

As we did with psychiatric inpatient readmission, we analyze non-psychiatric inpatient readmissions by comparing difference in difference based on two samples. Cost of non-psychiatric readmissions are based on each patient's change over time.

Table 106. Patient-Level Analysis: Cohort 1 One-Year Changes

		2015	2016	Difference	Difference in Difference	After Controlling for Matching Variables	P-Value
Non-Psychiatric Hospitalization Readmission Rate	SIM	4.2%	4.3%	0.1%			
	Comparison	5.7%	5.3%	-0.4%	0.5%	0.2%	0.550
Cost of non-	SIM	$1.93	$1.37	-$0.56			

		2015	2016	Difference	Difference in Difference	After Controlling for Matching Variables	P-Value
Psychiatric Readmissions PMPM	Comparison	$1.54	$1.22	-$0.32	-$0.24	-$0.23	0.490

Table 107. Patient-Level Analysis: Cohort 1 Two-Year Changes

		2015	2017	Difference	Difference In Difference	After Controlling For Match Variables	P-Value
Non-Psychiatric Hospitalization Readmission Rate	SIM	4.2%	3.7%	-0.5%			
	Comparison	5.7%	5.2%	-0.5%	0.0%	-0.1%	0.654
Cost of non-Psychiatric Readmissions PMPM	SIM	$1.93	$0.99	-$0.95			
	Comparison	$1.54	$1.61	$0.08	-$1.03	-$1.03	0.007*

Table 108. Patient-Level Analysis: Cohort 2 One-Year Change

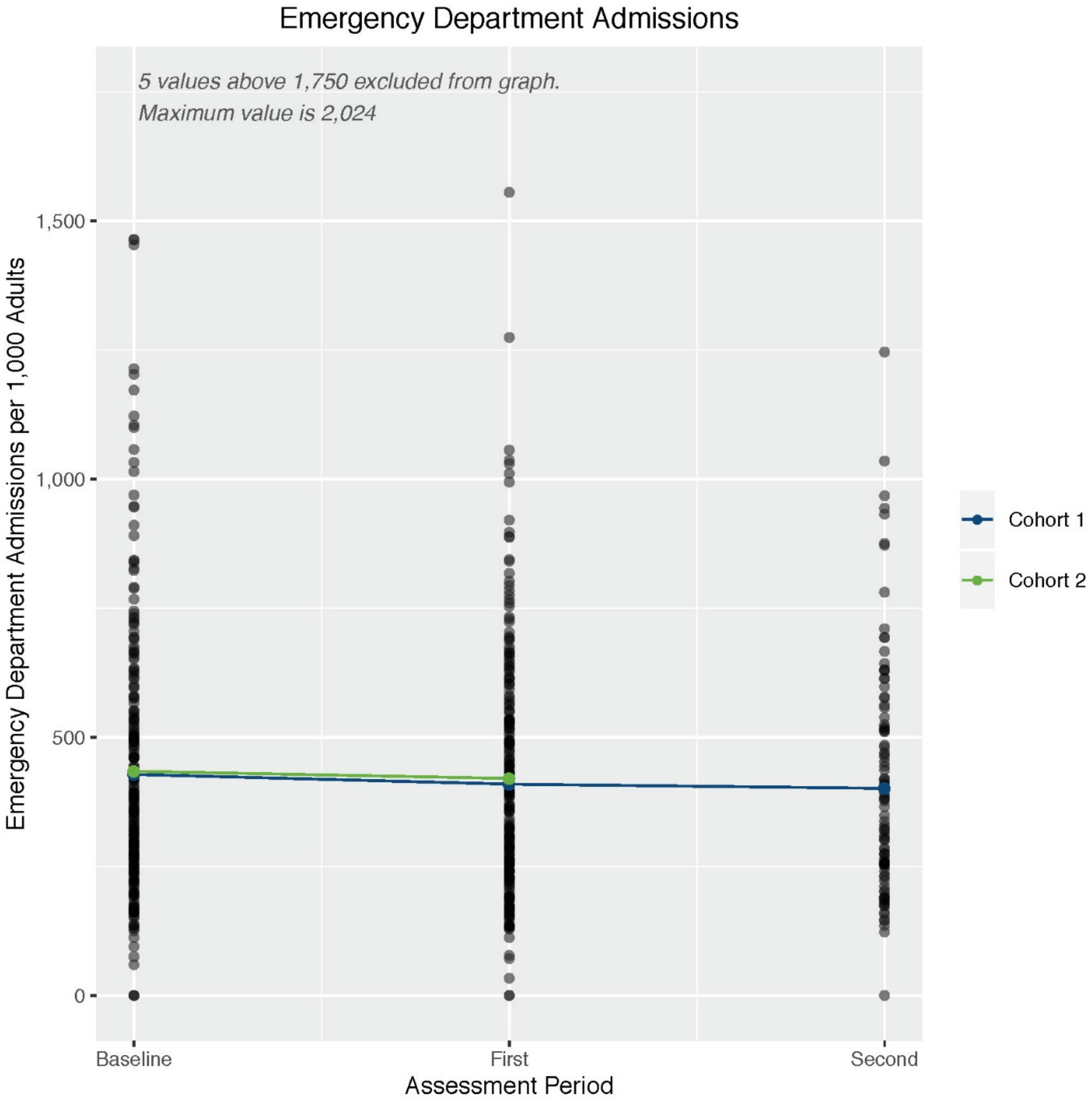
		2016/2017	2017/2018	Difference	Difference In Difference	After Controlling For Match Variables	P-Value
Non-Psychiatric Hospitalization Readmission Rate	SIM	3.4%	3.3%	-0.1%			
	Comparison	4.6%	4.9%	0.3%	-0.4%	-0.4%	0.395
Cost of non-Psychiatric Readmissions PMPM	SIM	$1.03	$0.87	-$0.16			
	Comparison	$1.13	$1.33	$0.20	-$0.36	-$0.35	0.330

The only statistically significant result from the patient-level analysis was a greater decline in two-year readmission costs for the cohort 1 group over its comparison group. Readmissions also had a greater decline for both cohorts, but results were not statistically significant. Readmission costs are therefore an example of an outcome in which SIM participation may have led to more positive outcomes.

Emergency Department Admissions

ED visits included ED observational units and excluded visits for behavioral health conditions. We report these on a per 1,000 population basis.

Figure 47. Emergency Department Admissions



Both cohorts experienced nearly identical declines in the emergency department utilization rate for all periods. With small changes and a wide distribution of practice site values, only the average change for cohort 1 between baseline and the first period was statistically significant.

Table 109. Emergency Department Admissions

Emergency Department Admissions	Baseline		First		Second	
	Value	N	Value	N	Value	N
Cohort 1	**428.14**	**95**	**409.28**	**95**	**401.14**	**95**
Practice Sites Reporting Baseline and First Period	428.14	95	409.28	95	-	-
Difference / P-Value	-	-	-18.86 0.012*	-	-	-
Practice Sites Reporting First and Second Periods	-	-	409.28	95	401.14	95
Difference / P-Value	-	-	-	-	-8.14 0.266	-
Cohort 2	**434.19**	**145**	**420.29**	**145**	**-**	**-**
Practice Sites Reporting Both Periods	434.19	145	420.29	145	-	-
Difference / P-Value	-	-	-13.90 0.139	-	-	-

*P-Value ≤ 0.05

Table 110. Patient-Level Analysis: Cohort 1 One-Year Changes

		2015	2016	Difference	Difference In Difference	After Controlling for Quarters of Data	P-Value
Non-Psychiatric ED Visits Per 1000 Primary Care Patients	SIM	462.93	441.58	-21.36			-
	Comparison	475.48	440.93	-34.55	13.19	11.34	<0.001*
Cost of non-Psychiatric ED Visits PMPM	SIM	$13.34	$13.96	$0.61	-	-	-
	Comparison	$14.15	$14.40	$0.25	$0.37	$0.31	0.125

*P-Value ≤ 0.05

Table 111. Patient-Level Analysis: Cohort 1 Two-Year Changes

		2015	2017	Difference	Difference in Difference	After Controlling for Quarters of Data	P-Value
Non-Psychiatric ED Visits Per 1000 Primary Care Patients	SIM	462.93	439.42	-23.51			
	Comparison	475.48	433.31	-42.17	18.66	16.73	<0.001*
Cost of non-Psychiatric ED Visits PMPM	SIM	$13.34	$12.75	-$0.60			
	Comparison	$14.15	$13.19	-$0.96	$0.36	$0.30	0.154

*P-Value ≤ 0.05

Table 112. Patient-Level Analysis: Cohort 2 One-Year Changes

		2016/2017	2017/2018	Difference	Difference in Difference	After Controlling for Quarters of Data	P-Value
Non-Psychiatric ED Visits Per 1000 Primary Care Patients	SIM	441.82	390.72	-51.09			
	Comparison	459.50	407.47	-52.03	0.94	2.34	0.498
Cost of non-Psychiatric ED Visits PMPM	SIM	$15.93	$13.64	-$2.29			
	Comparison	$15.77	$14.76	-$1.02	-$1.27	-$1.21	<0.001*

*P-Value ≤ 0.05

Based on the comparison group analysis, both cohorts had declining ED utilization, but the comparison groups had larger declines. These results are statistically significant. Given these results, it is surprising that cohort 2 experienced larger and statistically significant declines in ED costs. The results are therefore mixed; participation in SIM appears to increase ED utilization but may reduce ED costs.

Summary Table of Practice-Site-Level Outcomes

In the previous sections, we focused on change over time. In the case of practice-site-level data, we limited results to sites with data in each period. In the table that follows, we include summary statistics that include standard deviations for all practice sites for which we received cost and utilization data.

Table 113. Summary Statistics for Practice Site Cost and Utilization

Outcome	Cohort	Baseline			Assessment Period 1			Assessment Period 2		
		Mean	N	Standard Deviation	Mean	N	Standard Deviation	Mean	N	Standard Deviation
Total Cost of Care	1	$557.50	95	$361.00	$576.31	95	387.17	$604.36	95	$373.29
	2	$540.10	145	$417.64	$560.00	145	421.46	-	-	-
	2 Without Medicare	$444.57	143	$362.49	$480.15	143	383.53	-	-	-
Psychiatric Admissions	1	9.49	95	23.54	14.83	95	22.62	13.57	95	18.56
	2	27.43	145	143.66	30.4	145	127.98	-	-	-
Psychiatric Inpatient Admissions with Follow Up	1	11.69	79	23.96	7.25	85	17.27	3.47	81	9.28
	2	8.03	112	20.63	8.08	103	33.24	-	-	-
Psychiatric Readmission Rate	1	10.47	66	13.5	5.25	77	8.8	5.18	77	9.5
	2	6.22	99	13	7.08	96	16.31	-	-	-
Psychiatric Emergency Department Admissions	1	19.19	95	23.91	21.86	95	22.04	10.83	95	14.5
	2	18.74	145	23.66	9.24	145	10.53	-	-	-
Inpatient Admissions	1	74.3	95	106.28	97.34	95	144.21	99.72	95	135.89
	2	84.77	145	157.65	85.65	145	141.82	-	-	-
Inpatient Readmission Rate	1	8.02	89	6.61	5.8	91	4.97	5.11	89	5.2
	2	5.87	128	7	5.13	126	6.29	-	-	-
Inpatient Admissions	1	428.14	95	259.16	409.28	95	224.09	401.14	95	216.05
	2	434.19	145	248.09	420.29	145	231.13	-	-	-

CMHC Cost and Utilization

Because of the small number of CMHCs, we report all changes in a single table and forgo graphical analysis. All reported estimates are simple averages across the four sites. Because there were only four CMHCs, we were unable to conduct tests for statistical significance.

Table 114. CMHC Cost and Utilization

Metric	Baseline Mean	First Period Mean	Mean Change Baseline to First	Second Period Mean	Mean Change First to Second
Total Cost of Care PMPM	$1,151.00	$1,425.00	$273.78	$1,304.00	-$120.84
Psychiatric Admissions	76.03	147.92	71.89	86.88	-61.03
Psychiatric Inpatient Admissions with Follow Up	17.2%	15.0%	-2.2%	31.6%	16.6%
Psychiatric Readmission Rate	7.3%	6.6%	-0.7%	2.4%	-4.2%
Psychiatric Emergency Department Admissions per 1000 CMHC patients	236.88	426.34	189.46	110.76	-315.68
Inpatient Admissions per 1000 CMHC patients	71.50	150.00	78.50	156.02	6.00
Inpatient Readmission Rate	5.4%	6.7%	1.3%	3.0%	-3.7%
Emergency Department Admissions per 1000 CMHC patients	1,303.00	1,467.00	163.50	1,467	0.5

These rates tend to be significantly more volatile than the practice-level averages, in part because a maximum of four CMHCs reported in any given period. For example, psychiatric admissions increased from 76 to 148 admissions per 1,000 people from baseline to the first reporting period (a 71.9 admission, or 95%, increase). However, from periods one to two, readmission rates had a 61 admission decrease, making this drop larger by percentage than the previous increase. The psychiatric readmission rate, which is of concern to the CMHCs, decreased steadily and consistently, with a 0.7-percentage-point (or 9.6%) drop from baseline to the first assessment period, and then a 4.2-percentage-point (or 64%) drop from the first to second assessment periods. Overall, however, variation in CMHC rates was too high to deduce any meaningful trends.

We also developed a unique matched comparison group for the patients at each CMHC. For Jefferson Center, which had a bi-directional intervention with characteristics similar to a pediatric practice yet also served some patients with serious mental illness, we developed two comparison groups. The first used a methodology similar to the other CMHCs, including matching on behavioral health diagnosis. The second matched to patients attributed to pediatric NPIs. See the matching methodology section of this report for details.

In the table that follows, we provide the regression-based difference in difference for each outcome variable. This result is based on controlling for each patient's quarters of data used in the annual totals. The differences are based on 2015 to 2017 changes over time. The P-value for each DID is from the corresponding regression equation. We report all costs on a per patient per month basis, and utilization measures are generally reported per 1,000 CMHC patients. A few of the regressions had insufficient observations to calculate an effect for one or more of the CMHCs. For two measures, *Psychiatric Inpatient Hospitalization 30-Day Readmissions* and *Cost of Psychiatric Readmissions*, none of the CMHCs had sufficient observations.

We have not presented these results in any particular order, and the names of the CMHCs have been masked to protect confidentiality. A few of the regressions had insufficient observations to calculate an effect for one or more of the CMHCs, and for two measures none of the CMHCs had sufficient observations.

Table 115. Patient-Level Analysis: Difference in Difference for Outcome Variables

	DID	P-Value	DID	P-Value	DID	P-Value	DID	P-Value	DID	P-Value
Total Cost of Care	-$21	0.832	$278	0.097	-$513	0.062	$135	0.098	$11	0.138
Non-Psychiatric Outpatient Hospital Costs	$29	0.379	$117	0.064	-$24	0.783	-$67	0.081	$13	<.001*
Psychiatric Outpatient Hospital Costs	-$3	0.177	$2	0.680	-$1	0.617	$0	0.97	-	-
Primary Care Professional Services Costs	$1	0.835	-$5	0.404	-$10	0.107	$8	0.008*	-$2	0.067
Non-Psychiatric Professional Specialty Care Costs	-$3	0.510	$11	0.096	-$16	0.240	$6	0.205	$1	0.131
Psychiatric Professional Specialty Care Costs	-$15	0.019*	-$6	0.463	-$100	0.017*	$61	<.001*	-$2	0.051

	DID	P-Value	DID	P-Value	DID	P-Value	DID	P-Value	DID	P-Value
Non-Psychiatric Prescription Drug Costs	$114	0.023*	-$114	0.111	-$79	0.435	$31	0.18	$0	0.648
Psychiatric Prescription Drug Costs	-$5	0.799	$7	0.855	-$48	0.347	$1	0.961	-$2	0.163
Cost of Other Medical Services	-$106	0.075	$238	0.006*	-$136	0.346	$98	0.07	$2	0.704
Out-of-Pocket Expenses	$170	0.040*	$135	0.478	-$35	0.780	$189	0.039*	$7	0.517
Psychiatric Inpatient Hospital Discharges Per 1000 CMHC Patients	-97	0.168	-1,522	0.013*	-1,468	0.086	-22	0.686	-	-
Psychiatric Inpatient Hospital Costs	-$9	0.219	-$45	0.254	-$100	0.155	$2	0.872	-	-
Percentage of Psychiatric Inpatient Admissions with Follow Up Within 30 Days	-	-	-12%	0.804	37%	0.283	-26%	0.733	-	-
Psychiatric ED Visits Per 1000 CMHC Patients	-73	0.032*	-32	0.721	-57	0.576	-101	0.007*	-2	0.631
Cost of Psychiatric ED Visits	-$1	0.136	-$9	0.311	$0	0.990	-$1	0.794	$0	0.578
Non-Psychiatric Inpatient Hospital Discharges Per 1000 CMHC Patients	59	0.485	158	0.297	153	0.463	-31	0.577	2.204	0.587

	DID	P-Value	DID	P-Value	DID	P-Value	DID	P-Value	DID	P-Value
Non-Psychiatric Inpatient Hospital Costs	-$18	0.546	$84	0.128	$5	0.934	-$6	0.754	$1	0.425
Non-Psychiatric Inpatient Hospitalization 30-Day Readmissions per 1000 CMHC patients	0.015	0.886	-0.008	0.866	-0.052	0.637	0.469	0.014*	-	-
Cost of non-Psychiatric Readmissions	-$1	0.591	$4	0.298	-$22	0.669	$0	0.945	-	-
Non-Psychiatric ED Visits Per 1000 CMHC Patients	12	0.942	44	0.860	-81	0.850	-135	0.486	26	0.522
Cost of non-Psychiatric ED Visits	-$4	0.459	-$1	0.945	-$4	0.783	$1	0.91	$0	0.816

Although most outcome variables did not have statistically significant changes, a few did have significant results for multiple CMHCs. *Psychiatric Professional Specialty Care Costs* decreased in four of five comparison studies, and two of the decreases were statistically significant. A third decrease had a P-value of 0.051, and the fifth had a statistically significant increase. *Psychiatric Inpatient Hospital Discharges* had a very large and statistically significant decline but for only one CMHC. And *Psychiatric ED Visits* had statistically significant declines for two CMHCs.

Effect of Practice Transformation Factors on Cost and Utilization Outcomes

In addition to reporting the change over time in the included cost and utilization measures, we also attempted to identify which practice transformation factors drove any identified change. We can report on three potential factors: (1) the degree of integration at each practice, (2) staff satisfaction survey results, and (3) HIT survey results.

As with the CQMs, we analyzed the relationship between these transformation factors and cost and utilization using both cross-section (between practices) and across-time data. For each of the two assessment periods, we examined whether practice sites with greater integration clinician/staff job satisfaction or HIT also had improved results in the cost and utilization metrics. Because we report out-of-pocket results on a quarterly basis, and our transformation factors are annual, we exclude out-of-pocket measures from this portion of the analysis.

To determine which transformation factor most influences outcomes, we relied on the change over time analysis. Only cohort 1 had integration (Practice Monitor measure of overall integration) data for both baseline and at the end of SIM participation; we were limited to using this cohort for the change over time analysis.

In the tables that follow, the rows marked “Baseline” and “Final” refer to cross-sectional relationships for the corresponding assessment period. The row marked “Baseline to Final” includes the change over time analysis.

The Effect of Integration on Cost and Utilization

We note the following results in the summary. The only statistically significant relationship between Integration and cost or utilization is that an increase in integration correlates with a decrease in cost in cohort 2 when Medicare data are excluded. There does not appear to be any relationship between changes in integration over time and changes in the cost and utilization measures.

Table 116. Effect of Integration on Cost and Utilization

Assessment Period	Cohort 1			Cohort 2			Cohort 2 Without Medicare		
	Effect	P-Value	N	Effect	P-Value	N	Effect	P-Value	N
Total Cost of Care									
Baseline	-2.98	0.112	92	-4.48	0.005	131	-4.11	0.005*	130
Final	-3.79	0.090	92	-	-	-	-	-	-
Baseline to Final	-1.01	0.094	92	-	-	-	-	-	-
Psychiatric Admissions									
Baseline	0.04	0.714	92	-0.71	0.148	131	-	-	-
Final	0.07	0.533	92	-	-	-	-	-	-
Baseline to Final	-0.01	0.909	92	-	-	-	-	-	-
Psychiatric Inpatient Admissions with Follow Up									
Baseline	-0.12	0.178	83	0.21	0.190	94	-	-	-
Final	-0.07	0.256	80	-	-	-	-	-	-
Baseline to Final	-0.09	0.474	75	-	-	-	-	-	-

Assessment Period	Cohort 1			Cohort 2			Cohort 2 Without Medicare		
	Effect	P-Value	N	Effect	P-Value	N	Effect	P-Value	N
Psychiatric Readmission Rate									
Baseline	0.09	0.069	75	0.03	0.657	87	-	-	-
Final	-0.13	0.053	76	-	-	-	-	-	-
Baseline to Final	-0.05	0.658	67	-	-	-	-	-	-
Psychiatric Emergency Department Admissions									
Baseline	0.02	0.825	92	0.03	0.414	131	-	-	-
Final	-0.10	0.276	92	-	-	-	-	-	-
Baseline to Final	0.04	0.668	92	-	-	-	-	-	-
Inpatient Admissions									
Baseline	-0.26	0.719	92	-0.99	0.067	131	-	-	-
Final	-0.23	0.787	92	-	-	-	-	-	-
Baseline to Final	-0.11	0.728	92	-	-	-	-	-	-
Inpatient Readmission Rate									
Baseline	-0.04	0.099	89	-0.03	0.249	117	-	-	-
Final	-0.02	0.501	87	-	-	-	-	-	-
Baseline to Final	0.02	0.555	87	-	-	-	-	-	-
Emergency Department Admissions									
Baseline	1.51	0.178	92	0.74	0.361	131	-	-	-
Final	1.35	0.307	92	-	-	-	-	-	-
Baseline to Final	0.00	0.996	92	-	-	-	-	-	-

*P-Value ≤ 0.05

The lack of effect for the measure of integration is unexpected since this measure did show very large changes during the SIM implementation. We observed this same lack of effect in the comparison of the integration measure to changes in the CQM results. For this we need to compare change over time in staff satisfaction at each practice site with changes over time in the outcome measures.

The Effect of Staff Survey Results (Increased Clinician/Staff Satisfaction) on Cost and Utilization

Table 117. The Effect of Staff Survey Results on Cost and Utilization

Assessment Period	Cohort 1			Cohort 2			Cohort 2 without Medicare		
	Effect	P-Value	N	Effect	P-Value	N	Effect	P-Value	N
Total Cost of Care									

Assessment Period	Cohort 1			Cohort 2			Cohort 2 without Medicare		
	Effect	P-Value	N	Effect	P-Value	N	Effect	P-Value	N
Baseline	2.40	0.618	95	-9.50	0.067	144	-6.53	0.149	142
Final	-4.13	0.434	91	-6.13	0.227	135	-4.46	0.338	133
Baseline to Final	-1.19	0.256	91	-0.65	0.543	135	-0.02	0.983	133
Psychiatric Admissions									
Baseline	-0.03	0.922	95	-1.61	0.368	144	-	-	-
Final	-0.06	0.834	91	-0.66	0.671	135	-	-	-
Baseline to Final	-0.27	0.283	91	0.03	0.942	135	-	-	-
Psychiatric Inpatient Admissions with Follow Up									
Baseline	0.13	0.752	79	-0.01	0.968	112	-	-	-
Final	0.02	0.874	79	-0.04	0.937	94	-	-	-
Baseline to Final	-0.09	0.805	74	1.81	0.015*	87	-	-	-
Psychiatric Readmission Rate									
Baseline	-0.29	0.255	66	0.09	0.673	99	-	-	-
Final	-0.12	0.425	75	0.28	0.331	87	-	-	-
Baseline to Final	-0.55	0.055	60	-0.40	0.178	78	-	-	-
Psychiatric Emergency Department Admissions									
Baseline	-0.05	0.867	95	-0.26	0.383	144	-	-	-
Final	-0.38	0.071	91	-0.23	0.064	135	-	-	-
Baseline to Final	-0.33	0.271	91	0.54	0.054	135	-	-	-
Inpatient Admissions									
Baseline	-0.48	0.734	95	-4.07	0.037*	144	-	-	-
Final	1.04	0.602	91	-3.11	0.069	135	-	-	-
Baseline to Final	-0.39	0.757	91	-0.35	0.690	135	-	-	-
Inpatient Readmission Rate									
Baseline	0.13	0.194	89	0.01	0.958	127	-	-	-
Final	0.09	0.266	86	0.22	0.005*	118	-	-	-
Baseline to Final	0.17	0.147	85	0.41	0.002*	115	-	-	-
Emergency Department Admissions									
Baseline	0.25	0.942	95	-3.46	0.259	144	-	-	-
Final	-4.07	0.184	91	-5.55	0.035*	135	-	-	-
Baseline to Final	0.70	0.628	91	-2.70	0.089	135	-	-	-

*P-Value ≤ 0.05

In the change over time analysis, the only statistically significant relationship occurred in cohort 2. Increases in staff satisfaction are correlated with increases in 30-day readmission rates. This result is contrary to our expectation on the relationship between these two variables. Increases in staff satisfaction are also correlated with follow ups to psychiatric inpatient admissions, which is in line with our expectations.

In summary, changes in the staff survey do not appear to drive changes in these cost and utilization measures. Given that many of the cost and utilization measures did not show large changes over time—and the staff survey showed almost no change as well—this lack of effect is expected.

The Effect of HIT Improvements on Cost and Utilization

HIT improvements have the potential for driving change in several of the cost and utilization metrics. Most generally, to the extent that HIT systems assist primary care providers in meeting the needs of their patients, expensive inpatient services may be reduced or avoided, saving costs. More specifically, good management of patient information will help providers perform the type of screening measured with the CQMs and track other specific patient needs such as follow up after inpatient discharges.

Table 118. Effect of HIT Improvements on Cost and Utilization

Assessment Period	Cohort 1			Cohort 2			Cohort 2 without Medicare		
	Effect	P-Value	N	Effect	P-Value	N	Effect	P-Value	N
Total Cost of Care									
Baseline	-50.23	0.853	90	-80.68	0.668	144	-19.10	0.908	142
Final	359.14	0.209	91	-296.35	0.460	135	-156.27	0.672	133
Baseline to Final	22.73	0.605	89	26.72	0.485	135	36.13	0.397	133
Psychiatric Admissions									
Baseline	20.07	0.271	90	-63.78	0.324	144	-	-	-
Final	18.49	0.208	91	-365.25	0.003*	135	-	-	-
Baseline to Final	-10.29	0.347	89	3.11	0.833	135	-	-	-
Psychiatric Inpatient Admissions with Follow Up									
Baseline	-21.88	0.250	76	20.60	0.047*	112	-	-	-
Final	18.38	0.017*	79	6.52	0.859	94	-	-	-
Baseline to Final	-12.58	0.411	73	-42.40	0.073	87	-	-	-
Psychiatric Readmission Rate									
Baseline	4.52	0.704	63	0.86	0.900	99	-	-	-
Final	1.92	0.815	75	-8.38	0.656	87	-	-	-
Baseline to Final	1.92	0.895	59	4.19	0.673	78	-	-	-
Psychiatric Emergency Department Admissions									
Baseline	7.35	0.691	90	14.66	0.167	144	-	-	-
Final	2.58	0.823	91	0.58	0.953	135	-	-	-
Baseline to Final	9.89	0.439	89	4.49	0.656	135	-	-	-

Assessment Period	Cohort 1			Cohort 2			Cohort 2 without Medicare		
	Effect	P-Value	N	Effect	P-Value	N	Effect	P-Value	N
Inpatient Admissions									
Baseline	35.40	0.670	90	-32.36	0.649	144	-	-	-
Final	192.73	0.072	91	-233.70	0.083	135	-	-	-
Baseline to Final	48.84	0.362	89	-0.46	0.989	135	-	-	-
Inpatient Readmission Rate									
Baseline	7.49	0.146	85	-0.77	0.818	127	-	-	-
Final	2.17	0.603	86	2.57	0.665	118	-	-	-
Baseline to Final	10.52	0.028*	83	8.04	0.090	115	-	-	-
Emergency Department Admissions									
Baseline	346.80	0.078	90	-146.39	0.186	144	-	-	-
Final	32.05	0.849	91	-308.45	0.139	135	-	-	-
Baseline to Final	23.48	0.707	89	-48.58	0.399	135	-	-	-

*P-Value ≤ 0.05

In the cohort 1 cross-sectional analysis, higher scores on the HIT composite measure are associated with a higher percentage of *Psychiatric Inpatient Admissions with Follow Up*. For cohort 2, higher HIT scores are associated with lower psychiatric admissions, and a higher percentage of *Psychiatric Inpatient Admissions with Follow Up*.

In the change over time analysis, increases in HIT scores are correlated with increases in the *Inpatient Readmission Rate*, which counters our expectation that it would lead to reductions in readmissions.

No other regression analysis yields statistically significant results. Because the change over time analysis does not result in statistically significant results consistent with the model, we conclude that although there were large improvements in the HIT composite measure, these improvements did not drive improvements in cost and utilization.

Changes in Cost and Utilization by Practice Characteristics

As supporting tables, we provide analysis of change over time separately by practice characteristics. Because we anticipate observing greater changes over a longer period, these tables cover changes over a two-year period, from 2015 baseline to 2017 final year, and for cohort 1 only. We report data for only practice sites that report in both periods and include the number of sites reporting. We do not perform any type of statistical test for difference in difference.

Table 119. Cost and Utilization Metrics by Location

Metric	Urban			Rural		
	Baseline	Final	Difference/N	Baseline	Final	Difference/N
Total Cost of Care PMPM	**$553.57**	**$597.10**	**$43.53 (71)**	**$569.14**	**$625.84**	**$56.70 (24)**
Psychiatric Admissions	7.27	11.94	4.67 (71)	16.07	18.38	2.31 (24)
Psychiatric Inpatient Admissions with Follow Up	11.53%	4.38%	-7.15% (59)	11.38%	1.31%	-10.07% (17)
Psychiatric Readmission Rate	7.41%	6.10%	-1.31% (48)	21.59%	3.41%	-18.18% (14)
Psychiatric Emergency Department Admissions per 1000 Primary Care Patients	22.18	11.87	-10.31 (71)	10.36	7.76	-2.6 (24)
Inpatient Admissions per 1000 Primary Care Patients	56.96	81.15	24.19 (71)	125.61	154.65	29.04 (24)
Inpatient Readmission Rate	8.07%	5.03%	-3.04% (66)	8.25%	5.59%	-2.66% (22)
Emergency Department Admissions per 1000 Primary Care Patients	474.2	440.28	-33.92 (71)	291.88	285.37	-6.51 (24)

Overall, participation in SIM showed a greater positive and lesser negative impact for urban practice sites than for rural sites for cost and utilization metrics. Rates decreased more for urban practice sites in the areas *of Psychiatric Admissions to Emergency Departments, Inpatient Readmissions,* and *Emergency Department Admissions*. Urban practice sites also had a lower increase in *Total Cost of Care* and *Inpatient Admissions* as compared to rural practice sites. Follow ups to inpatient admissions declined less dramatically. For rural practice sites, *Psychiatric Admissions* increased less and *Psychiatric Readmissions* declined more than for urban practice sites. Rural practice sites reported less frequently than urban sites reported, resulting in small sample problems for rural practice sites.

Table 120. Cost and Utilization Metrics by Proportion of Underserved Patients

	Measure			
	Total Cost of Care PMPM	Psychiatric Admissions per 1000 Primary Care Patients	Psychiatric Inpatient Admissions with Follow Up	Psychiatric Readmission Rate

	Measure			
Low				
Baseline	$713.66	14.96	21.34%	13.24%
Final	$765.67	17.13	3.20%	3.77%
Difference(N)	$52.01 (25)	2.17 (25)	-18.14% (21)	-9.47% (20)
Medium				
Baseline	$617.39	11.1	12.88%	13.66%
Final	$670.76	12.8	6.80%	7.52%
Difference(N)	$53.37 (27)	1.7 (27)	-6.08% (22)	-6.14% (20)
High				
Baseline	$526.23	5.73	4.45%	10.70%
Final	$579.39	18.25	0.48%	4.90%
Difference(N)	$53.16 (13)	12.52 (13)	-3.97% (12)	-5.8% (6)
Very High				
Baseline	$361.46	5.42	4.44%	3.70%
Final	$392.16	9.46	2.90%	5.03%
Difference(N)	$30.7 (28)	4.04 (28)	-1.54% (20)	1.33% (15)
	Psychiatric Emergency Department Admissions per 1000 Primary Care Patients	**Inpatient Admissions per 1000 Primary Care Patients**	**Inpatient Readmission Rate**	**Emergency Department Admissions per 1000 Primary Care Patients**
Low				
Baseline	17.08	77.48	8.57%	309.78
Final	9.46	99.52	5.72%	291.82
Difference(N)	-7.62 (25)	22.04 (25)	-2.85% (25)	-17.96 (25)
Medium				
Baseline	13.43	93.81	9.89%	301.52
Final	10.42	130.63	5.63%	281.35
Difference(N)	-3.01 (27)	36.82 (27)	-4.26% (24)	-20.17 (27)
High				
Baseline	16.82	79.12	5.57%	435.45
Final	11.85	119.86	3.54%	420.32
Difference(N)	-4.97 (13)	40.74 (13)	-2.03% (12)	-15.13 (13)
Very High				
Baseline	28.65	51.82	6.65%	663.83

	Measure			
Final	12.43	61.05	4.94%	613.44
Difference(N)	-16.22 (28)	9.23 (28)	-1.71% (25)	-50.39 (28)

Overall, practice sites serving very high levels of traditionally underserved populations (i.e., practice sites with over 50% of patients using Medicaid or having no insurance) saw the most improvement with the implementation of SIM. *Psychiatric Emergency Department Admissions, Inpatient Admissions,* and *Emergency Department Admissions* all saw the greatest decline for practice sites serving these patients, *total cost of care* increased the least. And although *follow Ups after Psychiatric Inpatient Admissions* declined for this group, it declined the least of all for this subcategory. However, wherever very highly underserved populations did not gain the *most* from SIM (or lose the least), they benefitted the *least* of all groups from SIM implementation. Examples of this include *Psychiatric Admissions* and *Psychiatric Readmission Rates*—both of which increased when we would have anticipated a decline—and *Inpatient Readmission Rates*, which declined least for practice sites serving very high levels of underserved patients than for any other group.

Table 121. Cost and Utilization Metrics by Practice Size

Metric	Small			Medium			Large		
	Baseline	Final	Difference (N)	Baseline	Final	Difference (N)	Baseline	Final	Difference (N)
Total Cost of Care PMPM	$558.14	$590.53	$32.39 (21)	$637.93	$689.72	$51.79 (27)	$513.45	$564.12	$50.67 (46)
Psychiatric Admissions per 1000 Primary Care Patients	6.04	12.18	6.14 (21)	7.97	11.98	4.01 (27)	12.16	15.25	3.09 (46)
Psychiatric Inpatient Admissions with Follow Up	19.04%	2.67%	-16.37% (11)	14.83%	3.51%	-11.32% (20)	8.17%	4.02%	-4.15% (45)
Psychiatric Readmission Rate	20.52%	10.48%	-10.04% (9)	9.64%	3.01%	-6.63% (18)	8.56%	5.49%	-3.07% (35)

Metric	Small			Medium			Large		
	Baseline	Final	Difference (N)	Baseline	Final	Difference (N)	Baseline	Final	Difference (N)
Psychiatric Emergency Department Admissions per 1000 Primary Care Patients	19.61	6.65	-12.96 (21)	22.61	14.55	-8.06 (27)	17.29	10.8	-6.49 (46)
Inpatient Admissions per 1000 Primary Care Patients	52.4	79	26.6 (21)	84.32	126.6	42.28 (27)	79.57	93.46	13.89 (46)
Inpatient Readmission Rate	7.84%	7.38%	-0.46% (16)	9.07%	5.15%	-3.92% (26)	7.57%	4.44%	-3.13% (45)
Emergency Department Admissions per 1000 Primary Care Patients	392.29	334.38	-57.91 (21)	477.73	443.89	-33.84 (27)	415.91	406.23	-9.68 (46)

For small practices, SIM tended to have a larger magnitude of effect than for medium or large practices, whether positive or negative. For *Psychiatric Admissions and Follow Ups to Psychiatric Admissions,* there was a net negative impact for all practice sizes. Impact was greatest for small practices and became smaller for medium and large practices. On the other hand, for *Psychiatric Readmissions, Psychiatric Emergency Department Admissions,* and *Emergency Department Admissions*, for which SIM had a positive impact, the change was again greatest for small practice sites and decreased in magnitude for medium and large practice sites.

Table 122.Cost and Utilization Metrics by Practice Type

Metric	Pediatrics			Mixed Primary Care			Adult Primary Care		
	Baseline	Final	Difference (N)	Baseline	Final	Difference (N)	Baseline	Final	Difference (N)
Total Cost of Care PMPM	$210.54	$204.47	$-6.07 (21)	$601.01	$670.50	$69.49 (63)	$970.71	$988.99	$18.28 (11)

Metric	Pediatrics			Mixed Primary Care			Adult Primary Care		
	Baseline	Final	Difference (N)	Baseline	Final	Difference (N)	Baseline	Final	Difference (N)
Psychiatric Admissions per 1000 Primary Care Patients	1.79	3.94	2.15 (21)	12.07	15.54	3.47 (63)	9.42	20.66	11.24 (11)
Psychiatric Inpatient Admissions with Follow Up	0%	0%	0% (14)	11.73%	5.04%	-6.69% (53)	27.98%	1.53%	-26.45% (9)
Psychiatric Readmission Rate	1.15%	0.40%	-0.75% (3)	10.80%	4.96%	-5.84% (50)	12.72%	10.18%	-2.54% (9)
Psychiatric Emergency Department Admissions per 1000 Primary Care Patients	7.69	5.25	-2.44 (21)	20.24	13.01	-7.23 (63)	35.17	9.03	-26.14 (11)
Inpatient Admissions per 1000 Primary Care Patients	52.27	29.27	-23 (21)	78.74	113.13	34.39 (63)	90.95	157.41	66.46 (11)
Inpatient Readmission Rate	5.44%	4.90%	-0.54% (16)	8.58%	5.16%	-3.42% (62)	9.54%	5.64%	-3.9% (10)
Emergency Department Admissions per 1000 Primary Care Patients	413.75	410.54	-3.21 (21)	437.61	405.54	-32.07 (63)	401.44	358	-43.44 (11)

In some instances, SIM participation had a greater positive impact on practices serving adults than on those serving children. *Psychiatric Emergency Department Admissions, Inpatient Readmissions,* and *Emergency Department Admissions* all declined more for adult primary care practices than for mixed primary care facilities, and the least for pediatric practice sites. However, in other areas, such as in *Psychiatric Admissions and Inpatient Admissions*, rates increased, and more so for adult primary care facilities than for pediatric units. Otherwise, no major trends can be determined for cost and utilization metrics based on practice types. This is in part because practices of one type may not specialize in particular services utilized by patients. For example, 0% of pediatric practice sites followed up with patients after a psychiatric inpatient admission, in part perhaps because *Psychiatric Admissions* were already low for patients of those practice sites.

Summary and Conclusion

Because our evaluation questions addressed both practice site and patient changes in cost and utilization, we analyzed changes for each. Using SIM practice-site-level data, we observed some of the expected decreases in ED and inpatient utilization and costs associated with patients receiving better behavioral health and primary care in integrated primary care settings. These decreases include the rate of psychiatric and non-psychiatric emergency department admissions and 30-day psychiatric inpatient readmissions. These results were not uniform across cohorts and periods studied, but they were at least consistent with the underlying model of integrated primary care.

For patients attributed to SIM practice sites, we estimated the outcomes that would have occurred absent the SIM program by using a matched group of patients attributed to non-SIM primary care providers. By comparing changing outcomes over time between patients attributed to SIM (the “SIM” group) and non-SIM providers (the comparison group), we could estimate how much of the change experienced by the SIM group occurred because of the SIM program, and how much would have occurred absent SIM. Using this approach, we did observe for cohort 1 using 2017 versus baseline 2015 data, statistically significant larger declines in psychiatric ED utilization and cost, lower increases in the cost of non-psychiatric admissions and other medical services, and greater declines in the cost of readmissions. Some of these results also occur for patients attributed to cohort 2 SIM sites, including declines in psychiatric ED use.

But we also see some outcomes that had greater improvement in the comparison group. For cohort 1 2017 versus 2015, these include greater SIM group increases in non-psychiatric inpatient admissions, lower decreases for non-psychiatric ED visits, and greater increases in psychiatric hospital readmissions and readmission cost. Patients attributed to cohort 2 primary care practice sites had greater increases in psychiatric inpatient admissions and costs. They also had greater increases in non-psychiatric inpatient admissions.

Based on these inconsistent results, it is not surprising that the most aggregate measure, the total cost of care, did not robustly show cost savings. Limiting cohort 2’s data to equivalent six-month periods in the baseline and Year 1 of SIM participation, we did find statistically significant savings in the total cost of care. Using a full year of data for each period eliminated statistical significance but was also missing part of Medicare prescription drug costs for patients attributed to both SIM practice sites and the comparison group. Robust cost savings results may eventually occur as practice sites continue to refine the processes they use for providing integrated care. But based on full one- or two-year follow-up data, we do not observe statistically significant aggregate cost savings for either cohort 1, cohort 2, or the CMHCs.

The practice transformation factors of level of integration, staff satisfaction, and HIT improvements also have inconsistent relationships to the cost and utilization outcome measures. We are unable to conclude that any of them, as measured, drove systematic changes in cost and utilization of services. These results may be caused by limitations in the way we measured practice transformation or because some other factor drove changes in outcomes.

7 Population Health

Introduction

The SIM evaluation plan includes one population health outcome question.

PH4. To what extent did the selected behavioral and physical health related population health measures change over time?

One way to consider the overall impact of SIM efforts is to consider changes to the health of Coloradans. In other words, are Coloradans, as a statewide population, healthier today than they were before SIM?

To explore this question, the Evaluation and Population Health workgroups, SIM leadership, and CDPHE identified a set of population health measures to track over the course of the SIM effort. These population health measures correspond to clinical-level measures—clinical quality measures (CQMs)—that practice sites report on and use to monitor their patient populations. They also align with the goals set by LPHAs, BHTCs, and RHCs. These population health measures relate to health issues including depression, substance use, obesity, diabetes, maternal depression, suicide death rates, injuries, deaths from falls, and early childhood screenings for developmental concerns.

For example, the CQM related to maternal depression measures the percentage of mothers receiving screening for maternal depression by a clinician, at a SIM clinic, in the first six months following delivery. One of the related population health measures looks at the percentage of women ages 18–44 in Colorado who self-reported feeling down, depressed, or hopeless often or always following their most recent live birth. The CQM was measured and reported by SIM clinics whereas the population health measure, in the case of maternal depression, is measured by CDPHE using a representative statewide survey (Pregnancy Risk Assessment Monitoring Systems [PRAMS]) and reported as averages at state and Health Statistics Region (HSR) levels.

In addition to PRAMS, the other population health data sources used to address this evaluation question are Healthy Kids Colorado (also administered and reported by CDPHE), Vital Statistics (reported by the National Center for Health Statistics), the Behavioral Risk Factor Surveillance Survey (BRFSS, administered and reported by the Centers for Disease Control), Colorado Violent Death Reporting System (a subset of the National Violent Death Reporting System), the National Survey on Drug Use and Health (administered by the Substance Abuse and Mental Health Services Administration), and the Children’s Health Survey (administered and reported by the U.S. Health Resources and Services Administration).

This chapter includes a chart for each population health measure identified in the evaluation plan and for which we have data at the Health Statistics Region.[83] Each chart shows the statewide baseline (2015 in most cases), the most current statewide and national averages (usually 2017), and the most recent two-year averages for each of the 21 HSRs. For some measures we do not have comparable national averages. We note these instances in the respective graphs. For other measures, we do not have numbers for some or all HSRs. For measures for which we do not have any HSR-level data, we provide a table of baseline, state, and national averages and include this table at the end of the chapter (Table 123). For other measures for which we do not have rates all HSRs, we report the ones we do have.

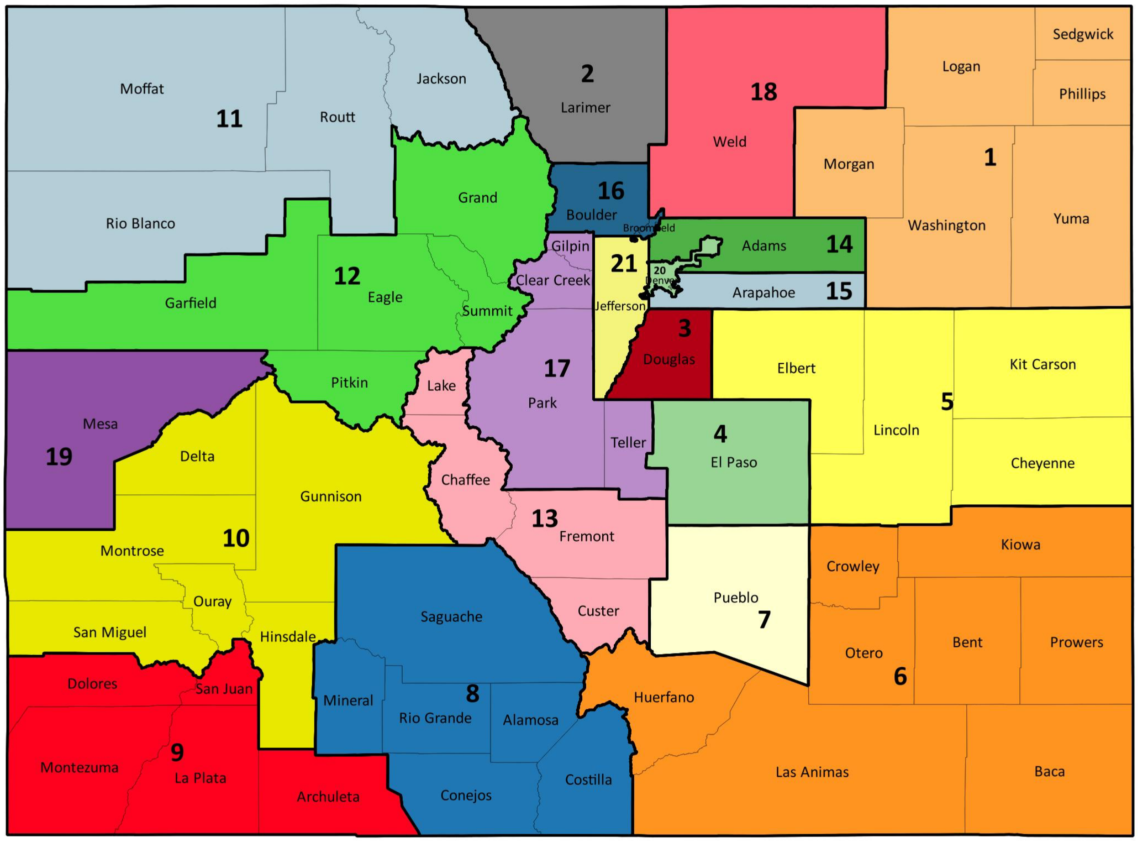
Dramatic change in state, national, or HRS-level population health measures over the four years of SIM implementation would be highly unlikely. These types of measures change slowly over decades. However, reporting these data in this report might provide a foundation for future population health policy makers to explore longer-term impacts of SIM.

The following map (Figure 48) shows the boundaries of the 21 Colorado Health Statistics Regions. We include this map here as a reference for the following charts, many of which include a break down by HSR.

---

[83] Measures for which we have at least statewide averages but not HSR-level data are reported in one table at the end of this chapter (see Table 123).

Figure 48. Colorado Health Statistics Regions.



Population Health Findings

We did not find much change in these broad population health measures over the four-year SIM implementation period. The Colorado average exceeds the national average on most measures, and the range of percentages for the HSRs varies widely on most measures.

Data received for the generation of graphical displays were provided to TriWest at an aggregate level (either in percentages or rates) and, with only the data from the Colorado Demography Office as the exception, did not include actual population sizes (N) that would allow for calculation of statistical significance. Additionally, the raw, individual-level survey data—from sources such as the BRFSS that are used to generate aggregate weighted percentages—were likewise not made available. This lack of availability limits our ability to accurately assess levels of statistical significance when comparing changes in outcomes over time.

Finally, the limited periods of time between baseline and “time-2” observations make generalizations of time-dependent effects problematic. However, in our view, these limitations still allow us to learn from these data. These results should be interpreted as largely descriptive. The variation at the aggregate HSR level should, likewise, be considered in the context of regional, community, and economic patterns that influence many of the outcomes we are observing. Still, the variation between HSRs, statewide, and national-level comparisons can provide a window into the status of important health indicators throughout Colorado.

All measures vary greatly across the state. Some HSRs have much higher rates, and others much lower rates than the statewide averages. So, the statewide average—although a true measure of the “overall” state as a whole—does not tell as a complete story of the many differences across regions. Original evaluation plans called for the comparison of the “saturation” of SIM to changes at the HSR region. However, stakeholders viewed these connects very tenuously and considered them poor comparisons. Therefore, we present the HSR data here to show the degree to which these measures fluctuate in various parts of the state.

There are several limitations to consider when interpreting these data. First, many of the surveys rely on self-reporting. Second, the most recent data are from 2017, which is only the second year of SIM implementation. Finally, higher rates may not be an indication of better health. For example, a higher rate of people reporting a diagnosis of depression may indicate more physicians are screening for and identifying depression rather than these rates signaling an actual increase in the number of people with depression.

Population Health Metrics Related to Depression and Suicide

The following set of graphs show population health metrics related to depression and suicide. There are no notable differences between the baseline (2013, 2014, or 2015), the most recent statewide average (2017), and national average (2017) for the depression measures. In each case, the differences are less than 2.2 percentage points. HSR percentages varied most on maternal depression (lower end 5.8% – higher end 17.5%), adults with frequent mental distress (lower end 6.6% – higher end 17.3%), and adults receiving mental health treatment (lower end 6.4% – higher end 17.3%). The high and low ranges were different HSRs in each case. However, for suicide attempts among high school students, Colorado rates (2.3 baseline in 2013 and 3.0 in 2015) are less than half the national average (7.4 2017). The data available to us do not allow any further analysis such as a test of significance.

Figure 49. Depression: Adults who Are Currently Depressed

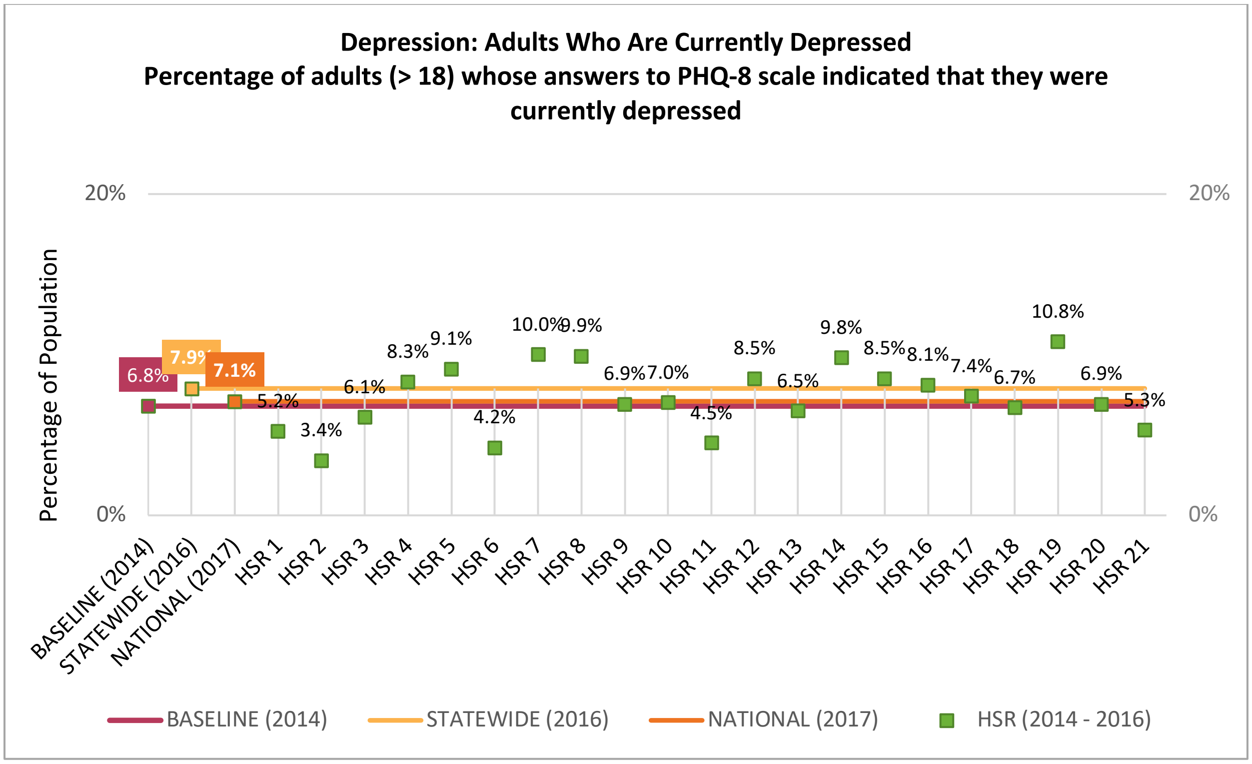


Figure 50. Depression: Adults with Frequent Mental Distress

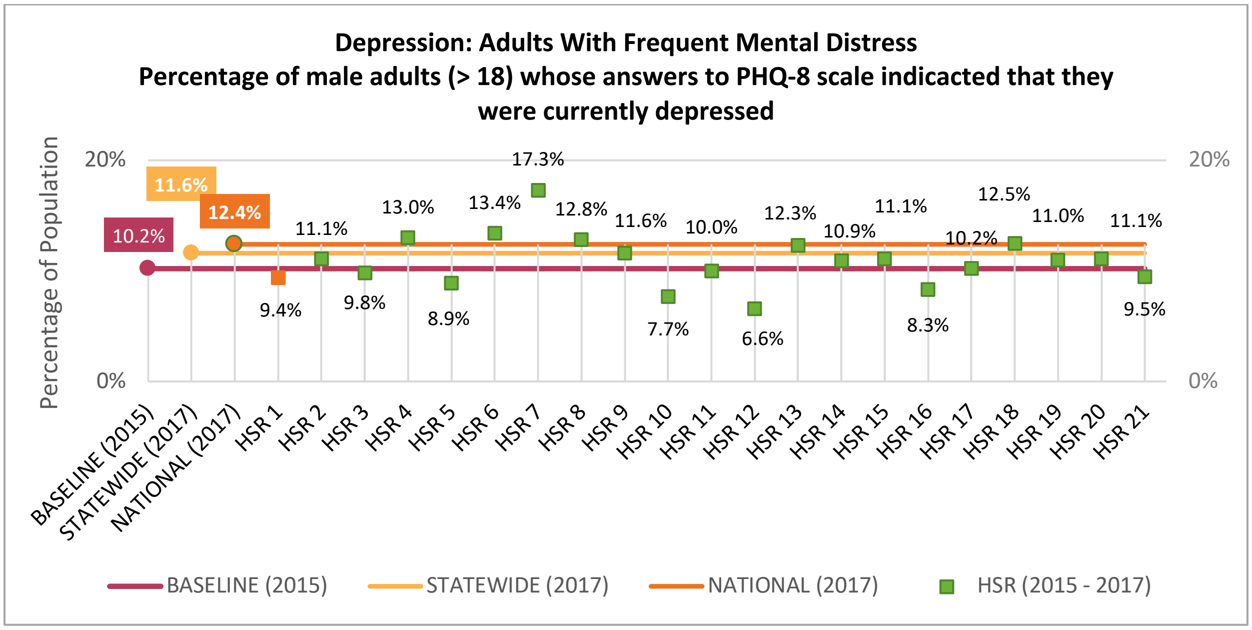


Figure 51. Depression: Adults Receiving MH Treatment

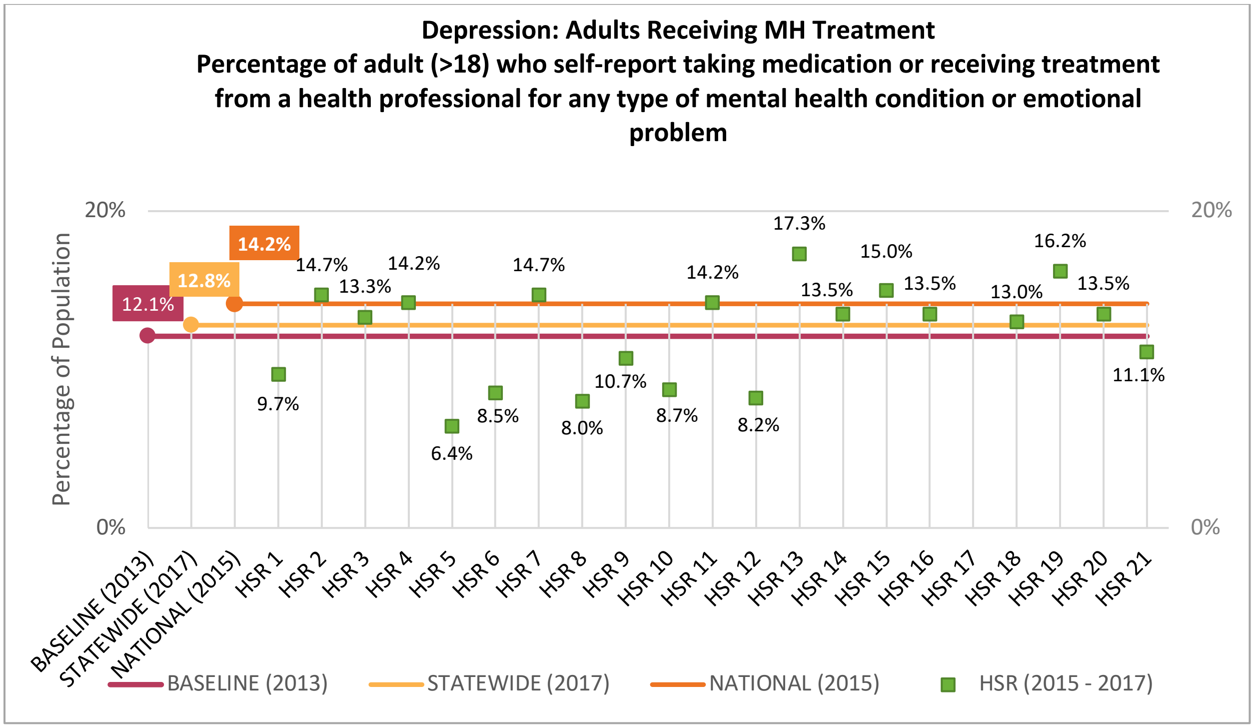


Figure 52. Depression: Symptoms Among High School Students

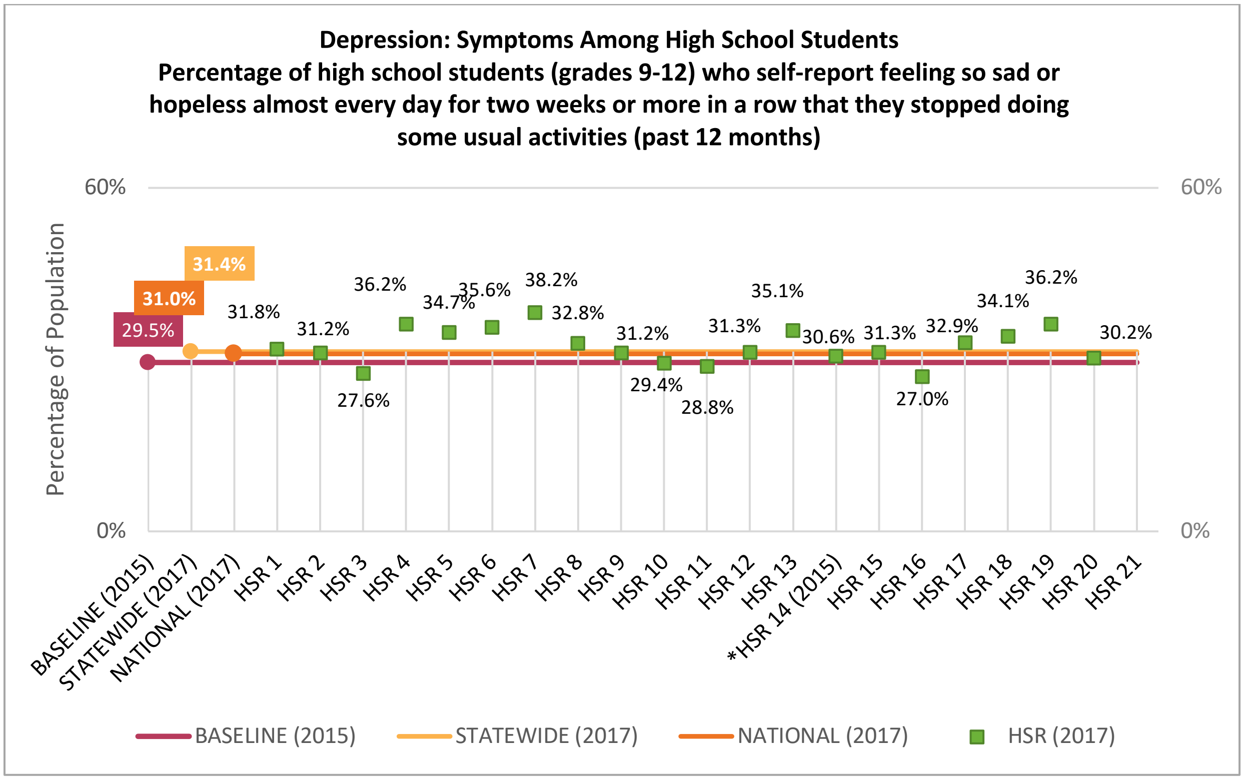


Figure 53. Suicide Attempts Among High School Students

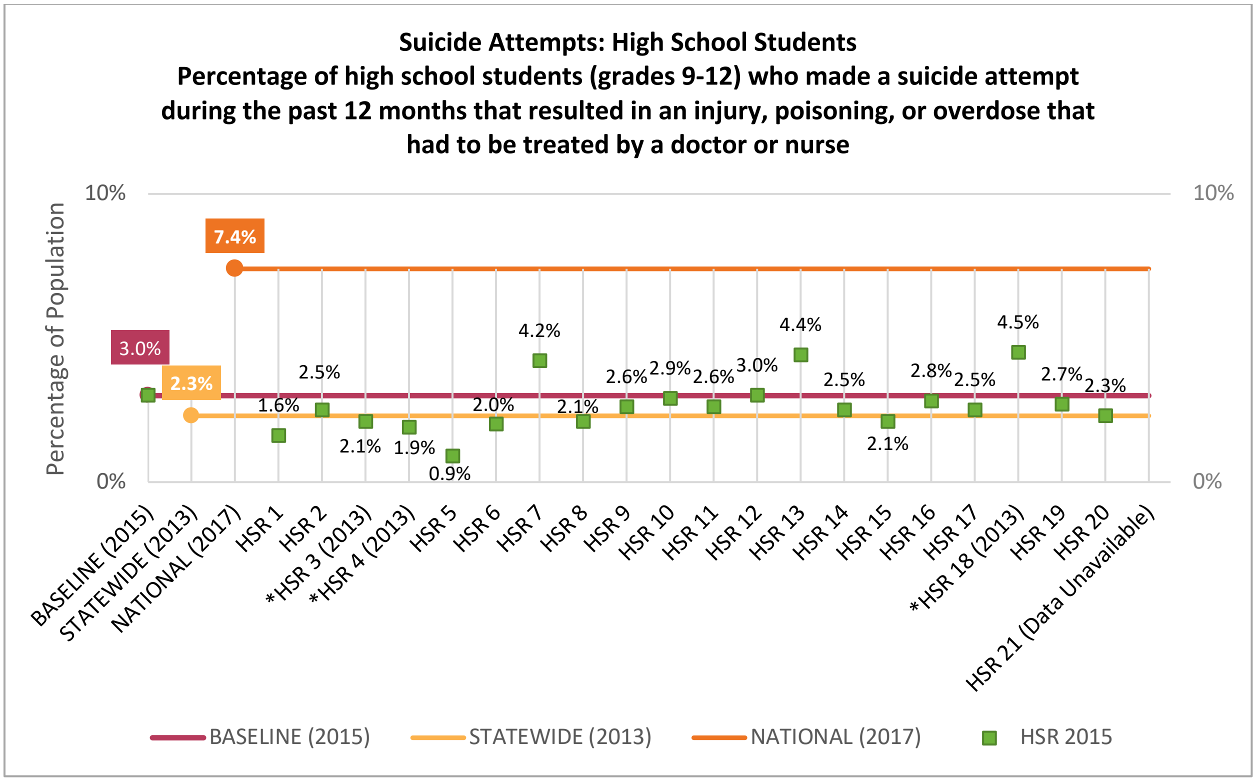


Figure 54. Maternal Depression

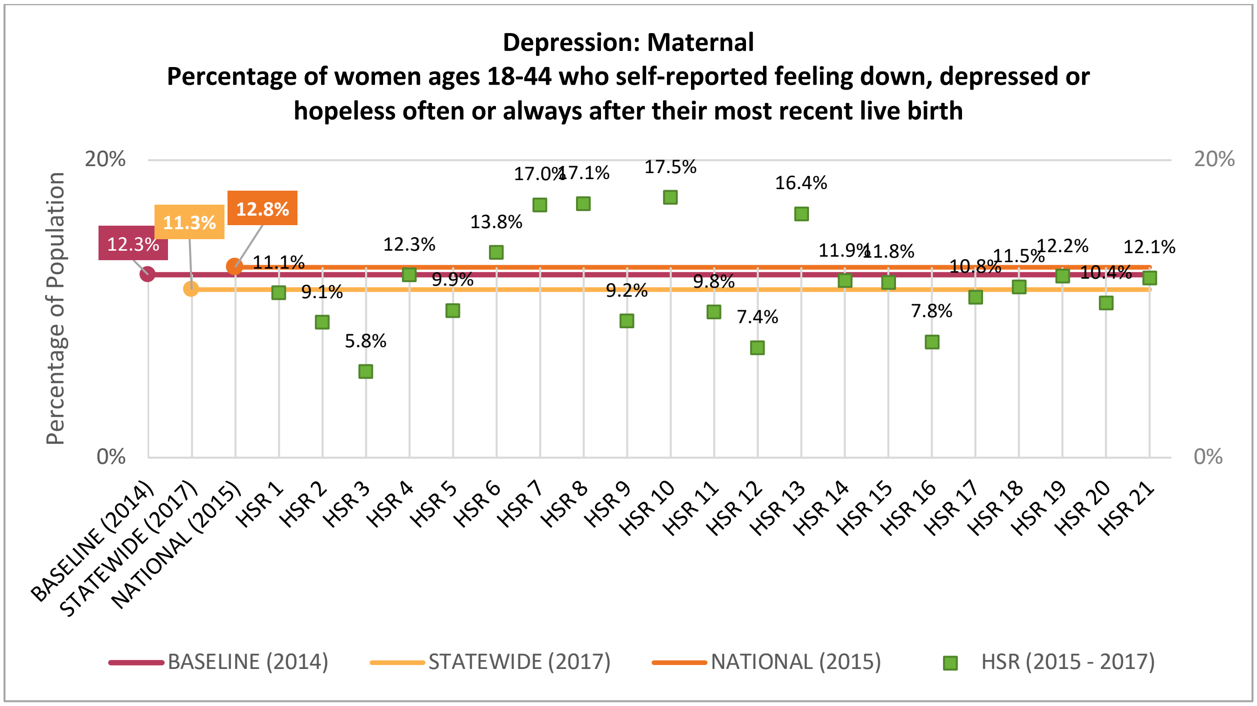
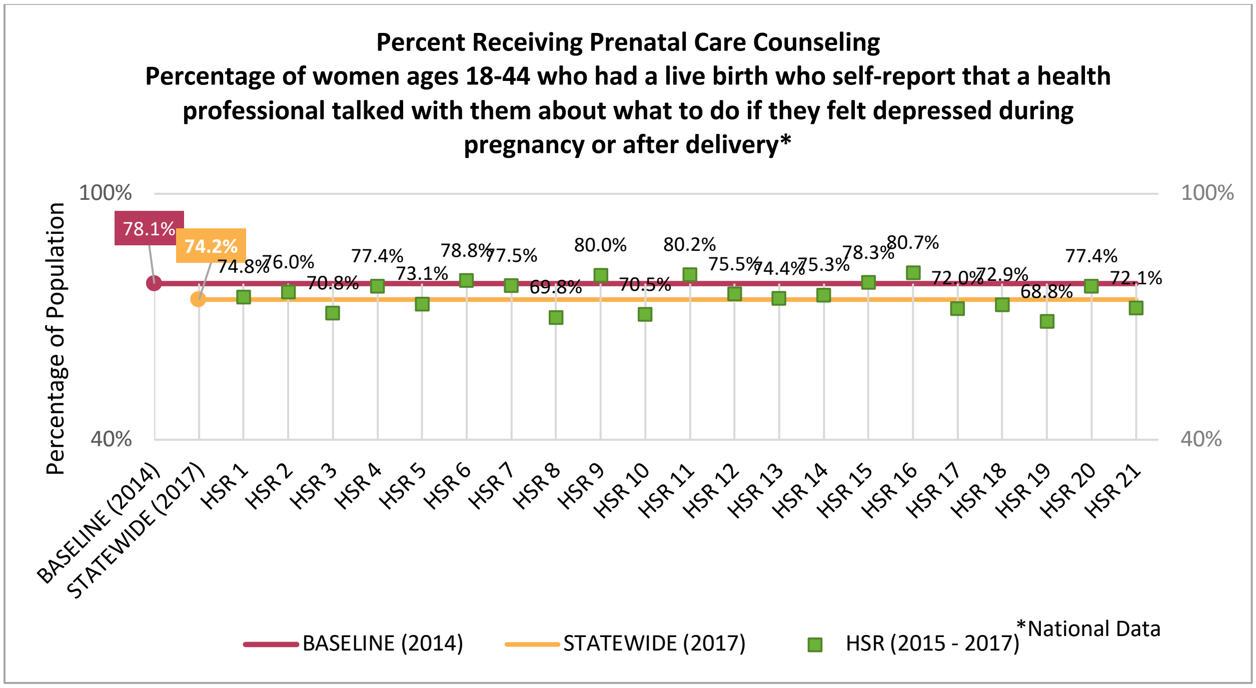


Figure 55. Prenatal Care

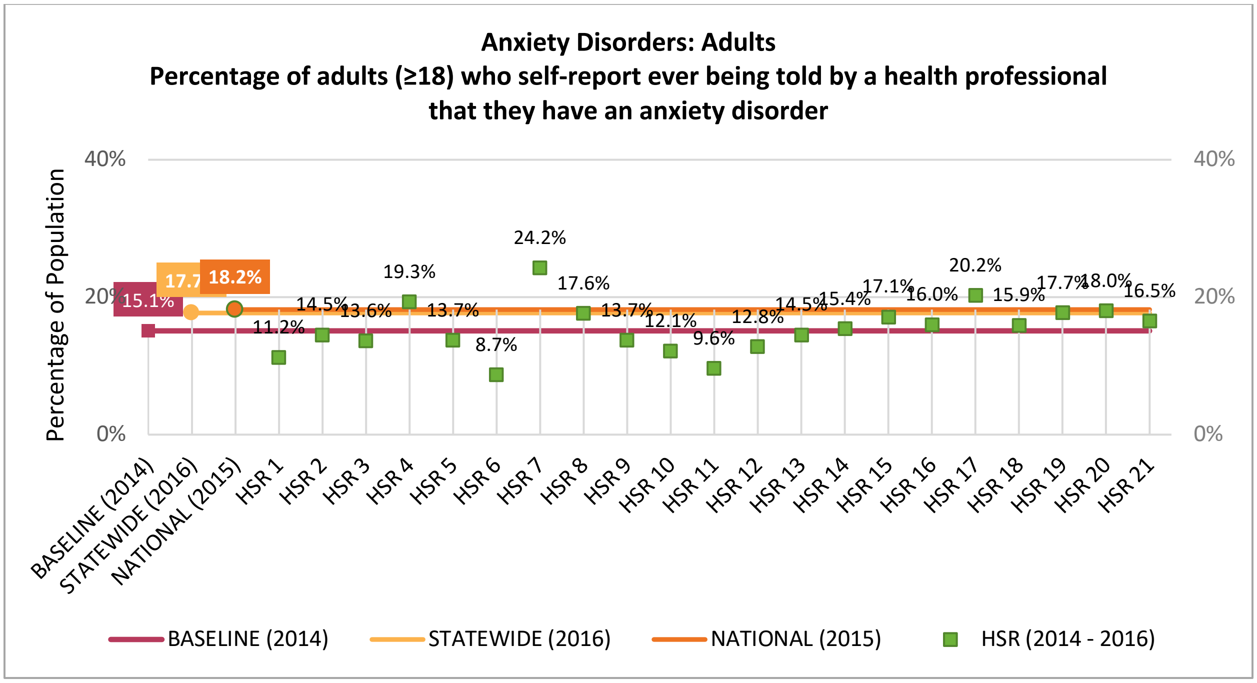


*National Data

Population Health Metrics Related to Anxiety

SIM tracked one population health measure related to anxiety disorders. The Colorado average for 2017 is within 0.5 percentage points of the 2017 national average and 2.6 percentage points higher than the 2014 baseline. This trend toward higher numbers of people reporting they have an anxiety disorder could result from either increased screening by primary care providers or an increase in the number of people with anxiety disorder. The data we received do not allow us to explore this distinction further. Percentages at the HSR level range from a high of 24.2 to a low of 8.7.

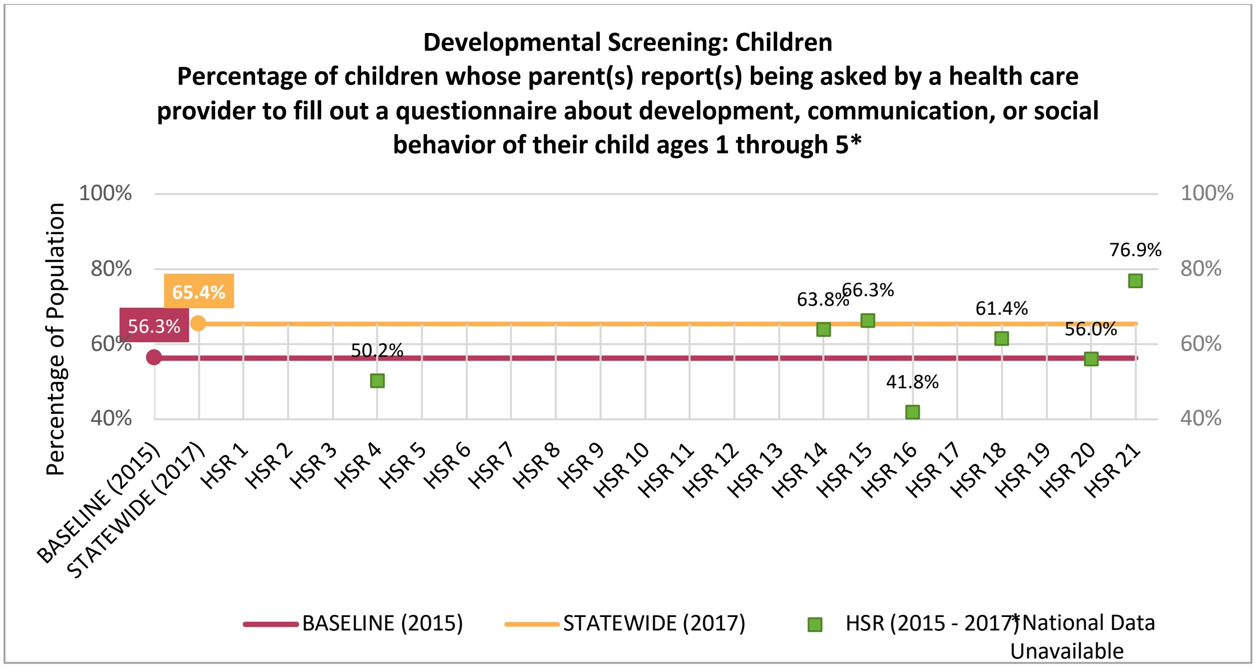
Figure 56. Anxiety Disorders in Adults



Population Health Metrics Related to Early Childhood Development

We received HSR-level data on one population health measure related to developmental screening for children. This measure comes from the Child Health Survey, and there is no comparable national average available. Parents are asked whether a healthcare provider requested they complete a questionnaire about their child's development. The 2017 Colorado average is up 9.1 percentage points over the 2015 baseline. Of the seven HSRs for which data are available, the range is (42.8–76.9). CDPHE is not able to report this measure for some HSRs because of low population numbers in more rural areas.

Figure 57. Developmental Screening (Children)



Population Health Metrics Related to Diabetes

There are two measures related to diabetes, both from the BRFSS. The first is the percentage of adults who report ever being told by a healthcare provider that they have diabetes. The 2017 Colorado average (7.4) remains close to the 2015 baseline (6.8%). Colorado remains below the national average by 3.4 % points. HSRs range from 3.9% to 15.2%. The second measure, percentage of adults with diabetes ever being told they have high blood pressure, follows the same pattern with Colorado averages from 2015 (65.7%) to 2017 (65.0%), remaining steady and below the national average for 2017 (73.6%) by 8.6 percentage points.

Figure 58. Diabetes (Adults)

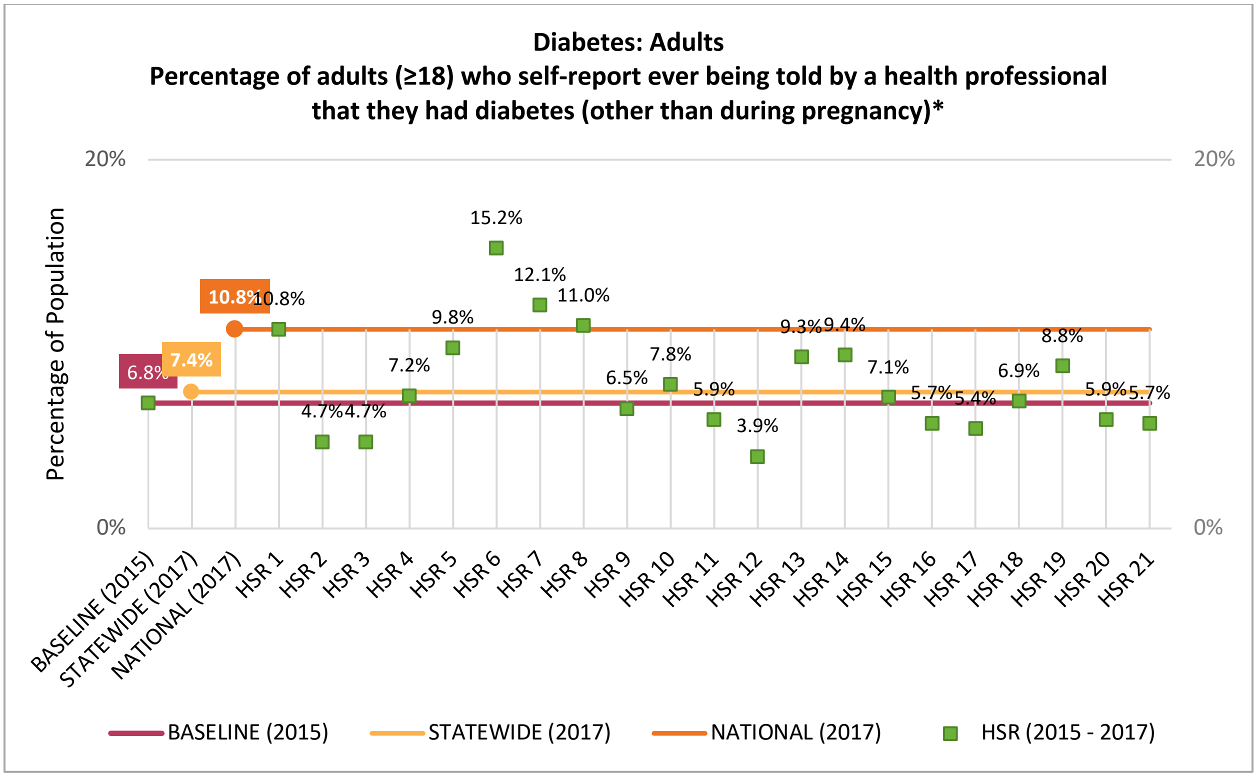
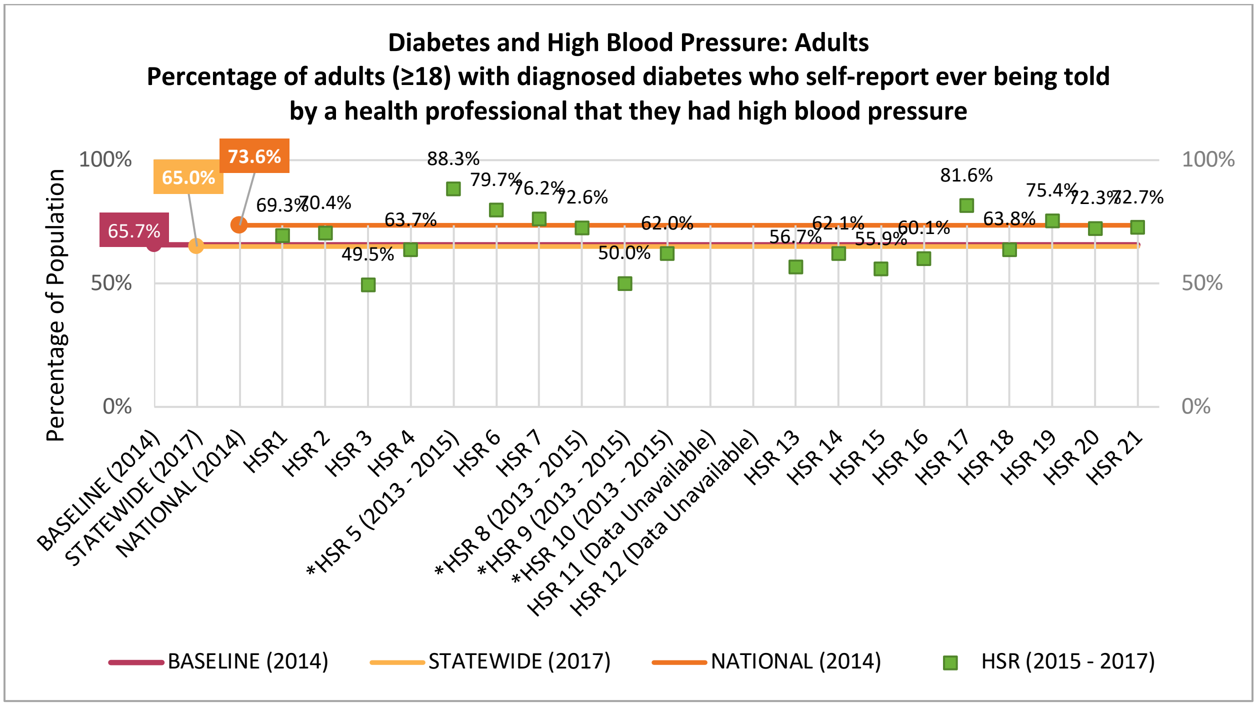


Figure 59. Diabetes and High Blood Pressure (Adults)



Population Health Metrics Related to Obesity

SIM selected one population health level obesity measure broken down into three age categories: adults ≥18, youth grades 9–12, and children ages 5–14. In each age category, Colorado remains below the national average, and the 2017 state averages for children and youth are slightly lower than baseline (2015), which may indicate a positive trend. The range of HSRs is 15.0%–32.0%.

Figure 60. Obesity (Adults)

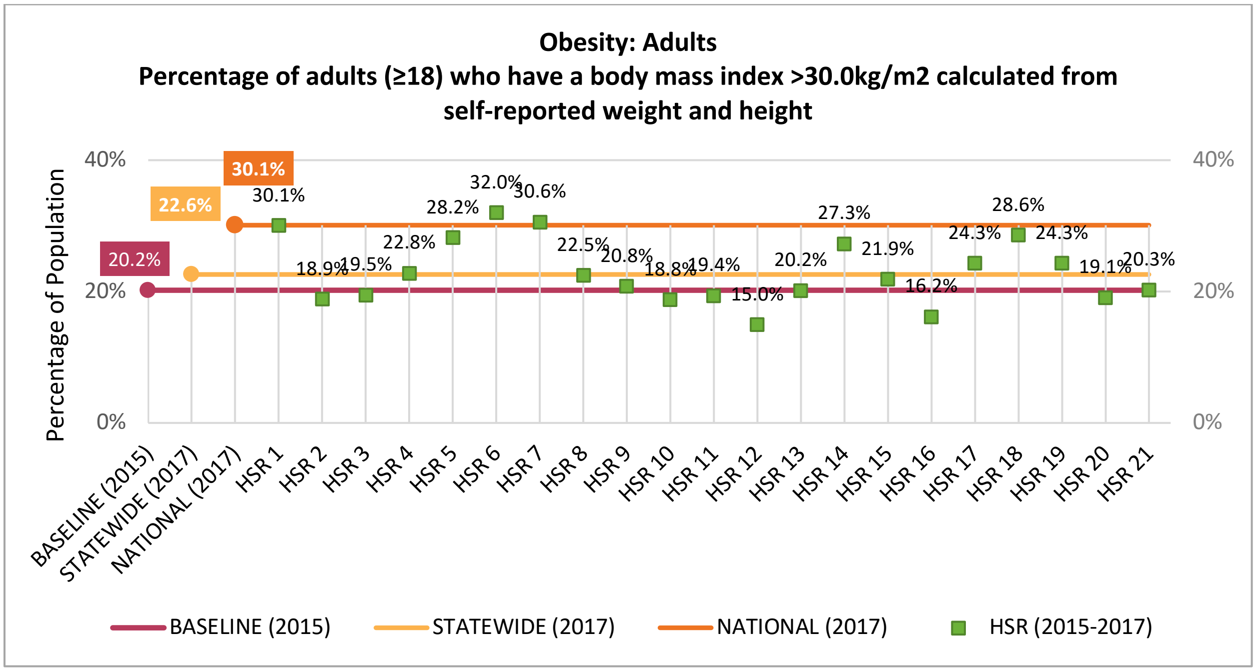


Figure 61. Obesity (Children and Youth)

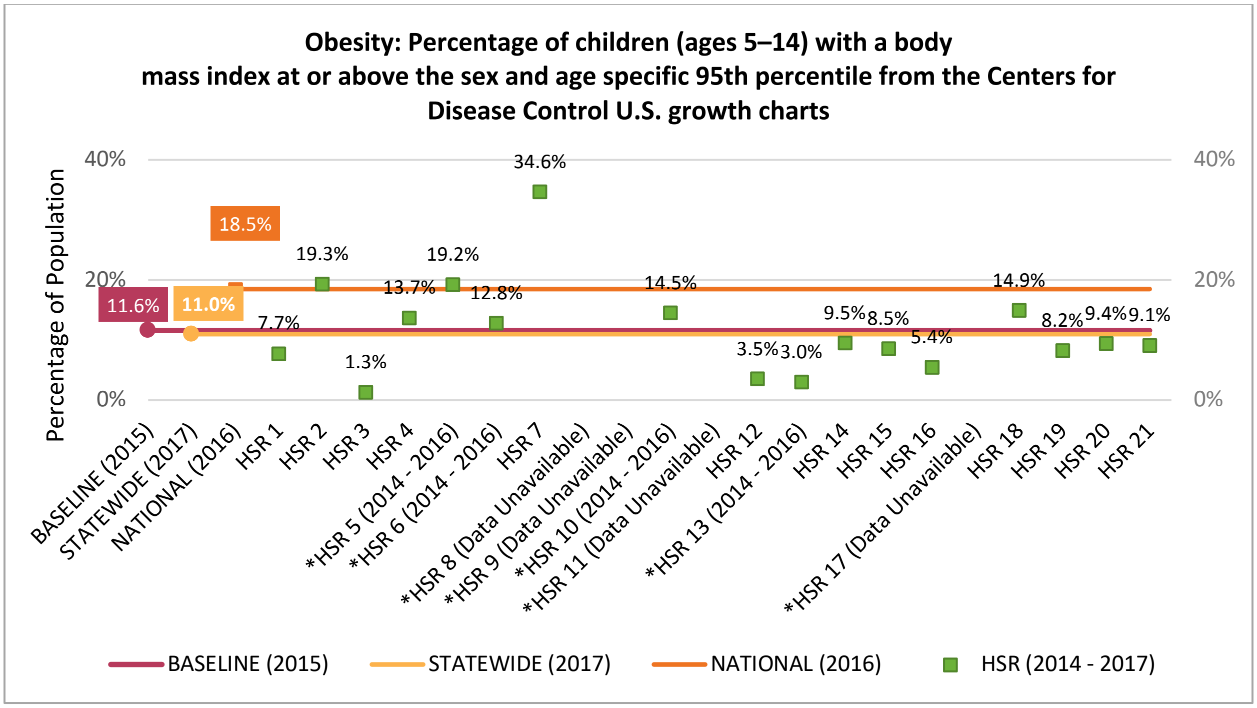
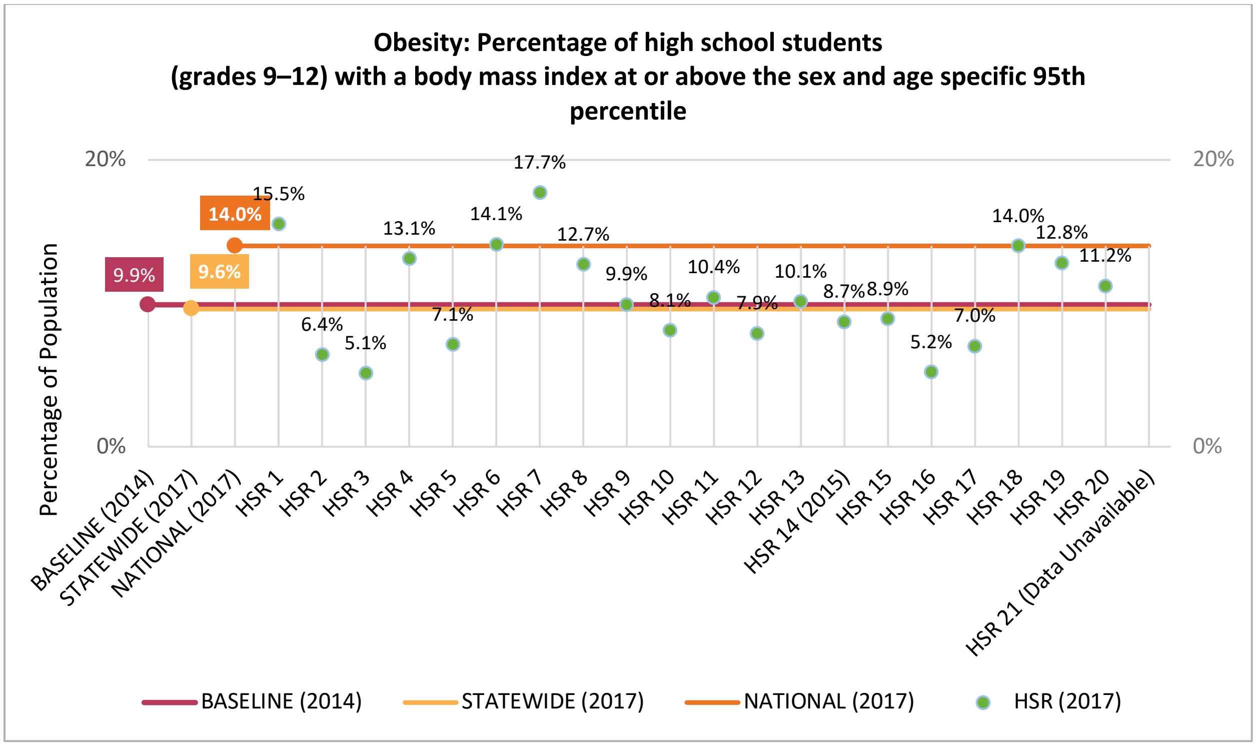


Figure 62. Obesity (High School Students)



Population Health Metrics Related to Asthma

As we did with the obesity measures, we tracked one population-health-level asthma measure broken down into adults ≥18 and youth grades 9–12. In both age categories the baseline, most current statewide and national averages are tightly clustered and within less than 1.5 percentage points. As noted before, this is not surprising or an indication that SIM has not had an impact. These type of population health measures change slowly over years and decades. The range of HSRs is within 10–12 percentage points, which is a tighter range than found in the other population health measures.

Figure 63. Asthma (Adults)

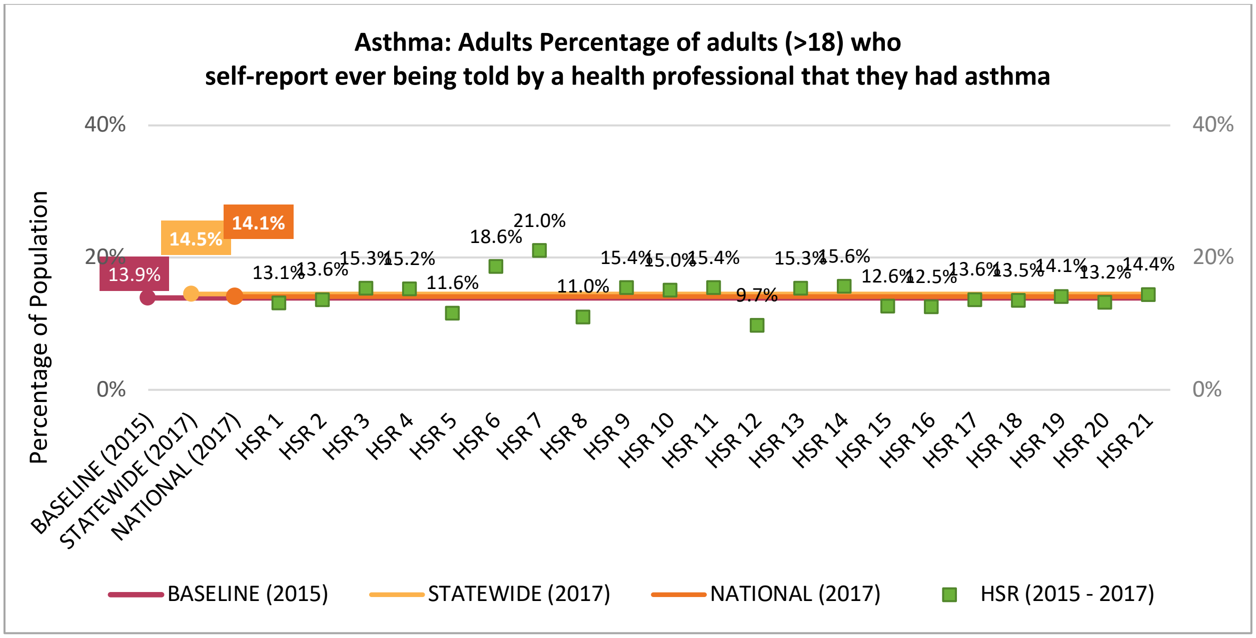
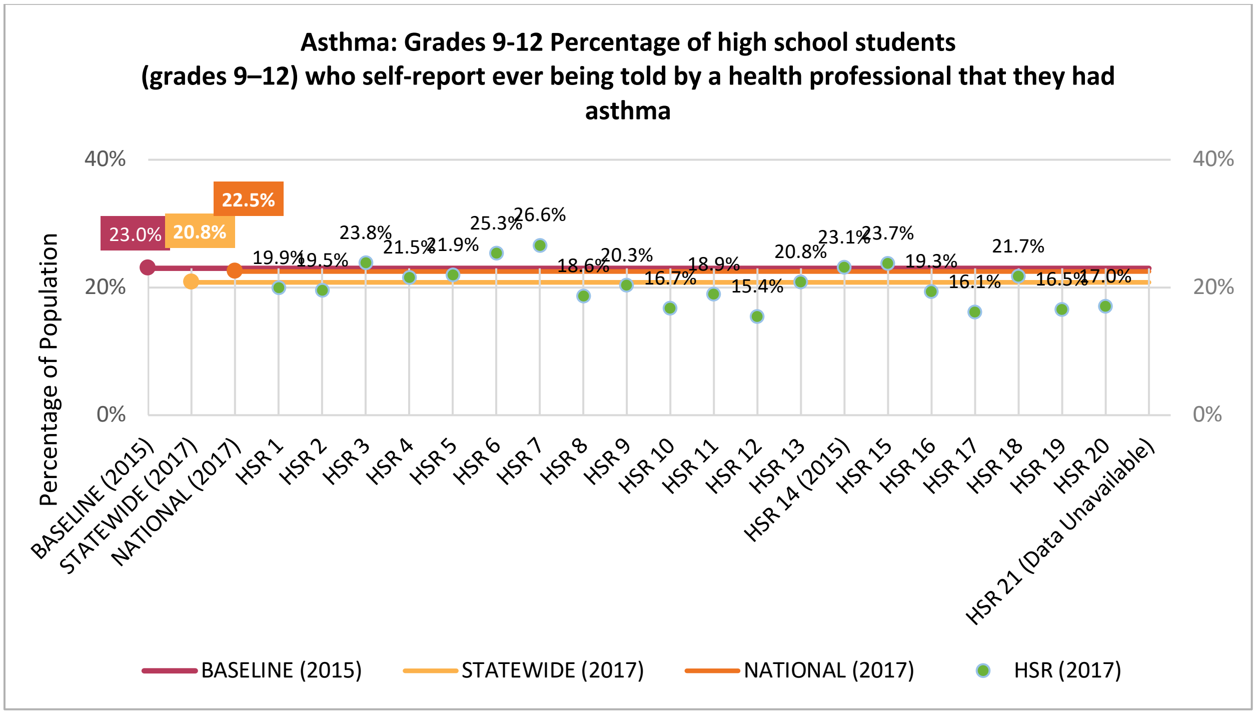


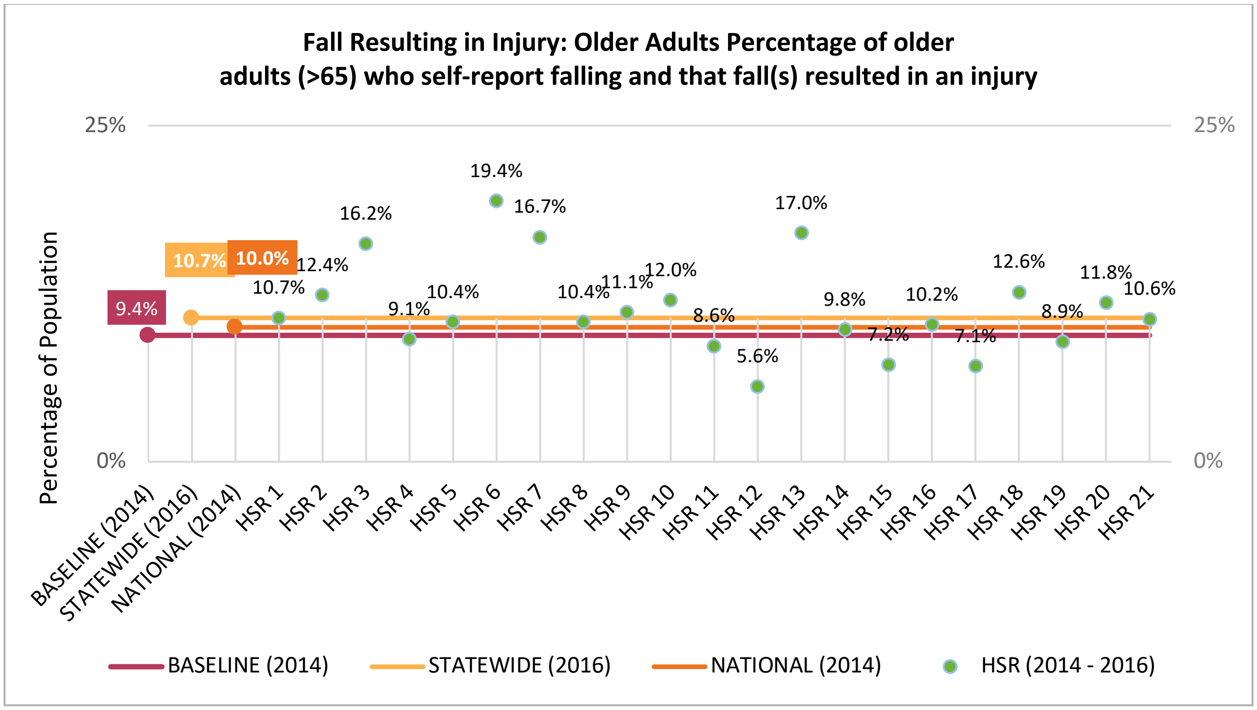
Figure 64. Asthma (Youth Grades 9–12)



Population Health Metrics Related to Safety

Both population health measures related to safety relate to falls in older adults. The first, measuring falls resulting in injury, shows very little difference between baseline and the most recent statewide and national averages. On this measure, HSRs ranged from 5.6% to 19.4%. The second safety-related measure looks at fall-related deaths. HSR-level data on this measure are not available. Therefore, we report this measure in the table at the end of this chapter (Table 123).

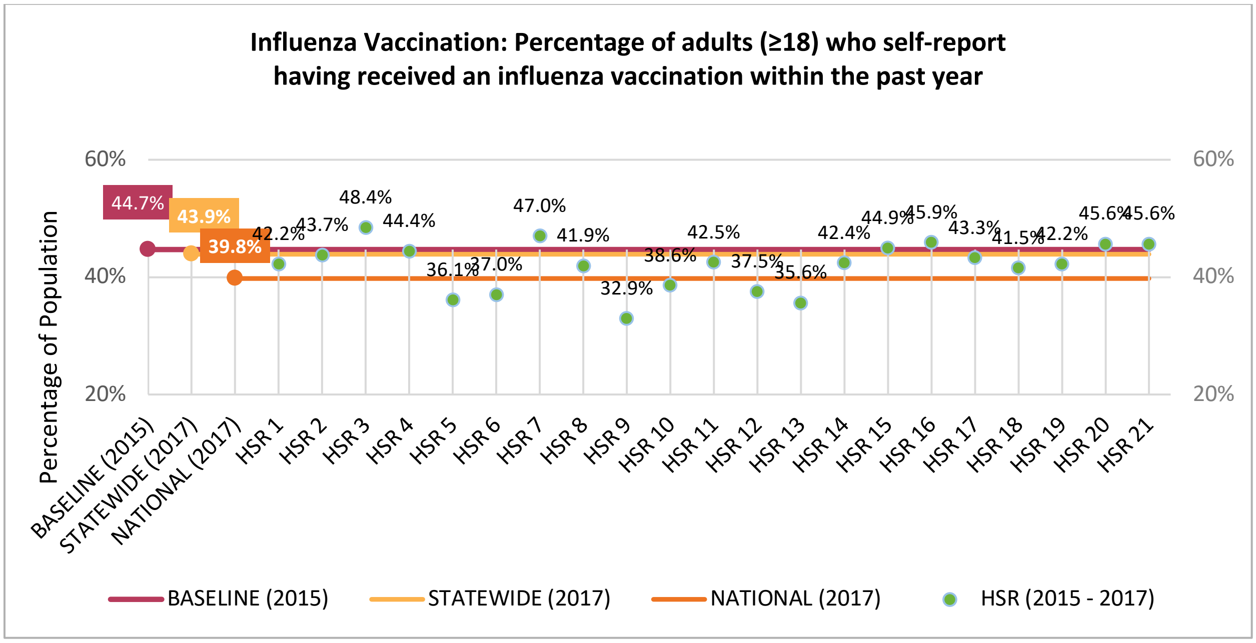
Figure 65. Falls Resulting in Injury



Population Health Metrics Related to Influenza

SIM selected one population health measure related to influenza. This measure asks adults whether they received a flu vaccine within the past year. The baseline (44.7%) and the most current Colorado average (43.9%) are within one percentage point and are higher than the national average (39.8%). HSR percentages range from 32.9% to 48.4%.

Figure 66. Influenza Vaccination (Adults)

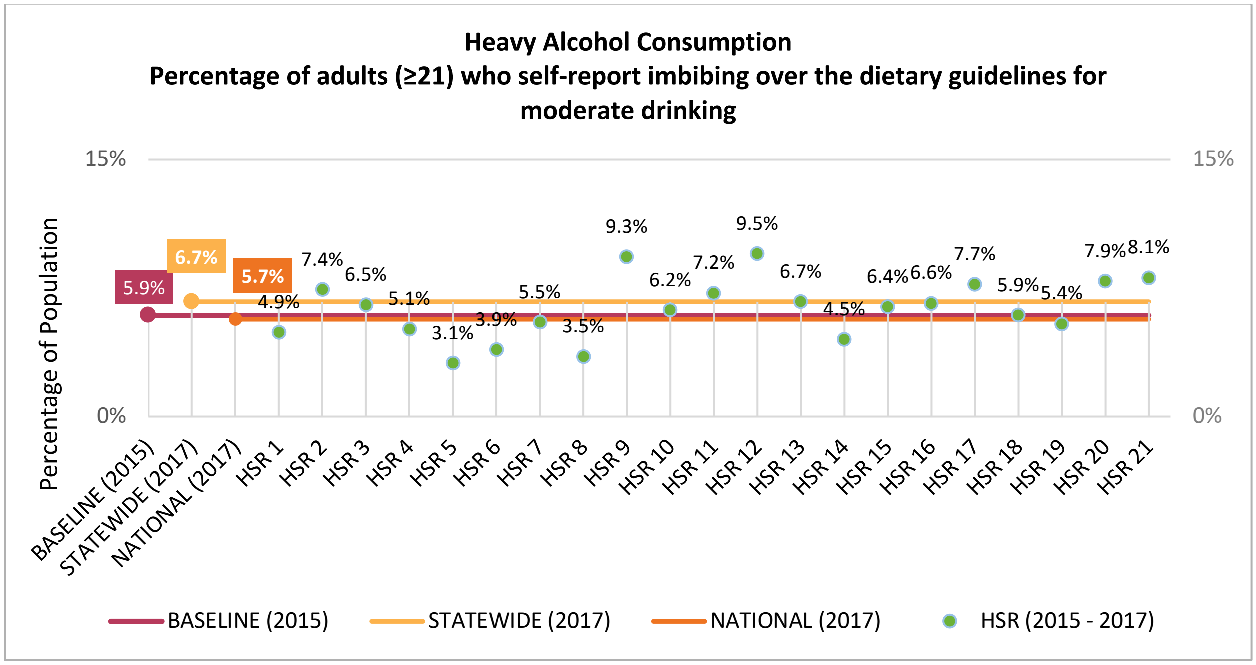


Population Health Metrics Related to Substance Use Disorders

SIM selected seven population health substance-use-related measures to follow: two each on alcohol and cigarette smoking, one on non-medical opioid use, one on risky prescription opioid dosage, and one on drug overdose deaths (not limited to opioids). We do not have HSR-level data for the two drug-related measures. We include these in Table 123 at the end of this chapter. Descriptions for the other substance use related measures immediately precede each graph in this section.

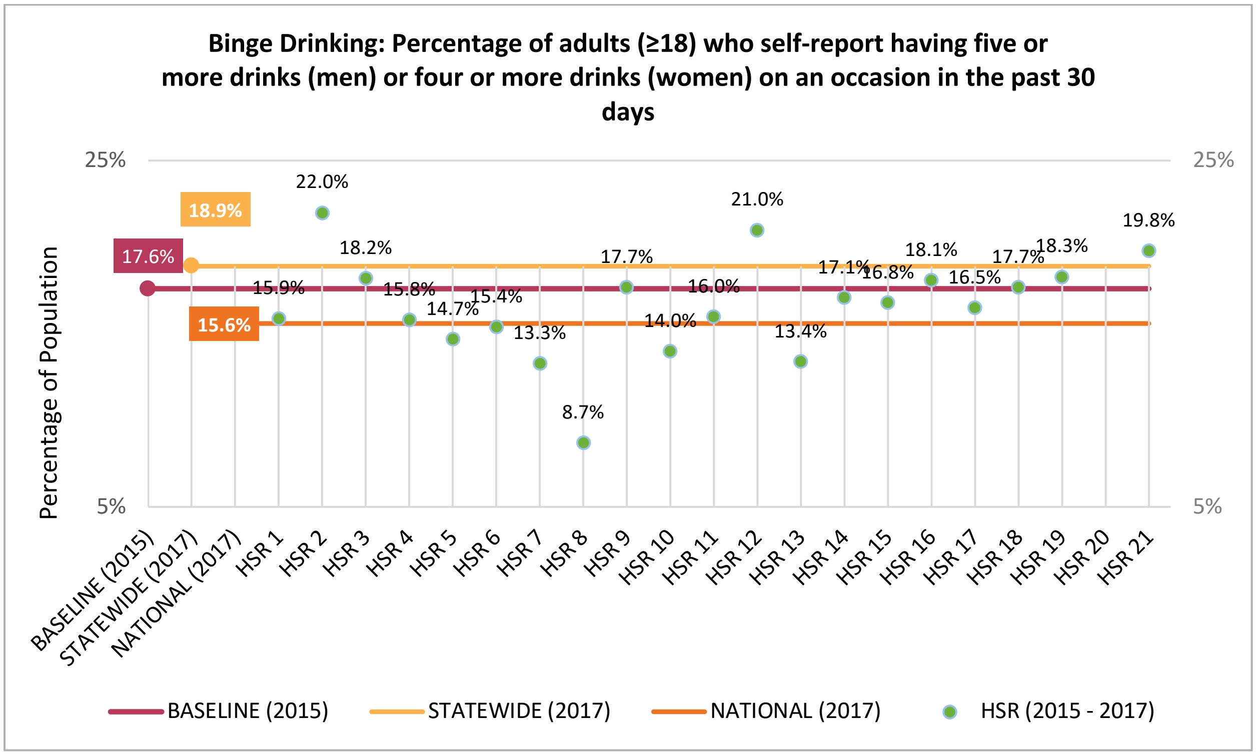
The first alcohol-related measure (heavy alcohol consumption) shows baseline (5.9%), the most recent statewide (6.7%), national averages (5.7%), and the HSR averages (3.1% - 9.5%) relatively close together.

Figure 67. Heavy Alcohol Consumption



Binge drinking among adults shows baseline (17.6%) and Colorado averages (18.9%) slightly higher than the national average (15.6%). HSRs range from 8.7% to 26.4%.

Figure 68. Binge Drinking (Adults)



The percentage of adults currently smoking at baseline (15.6%) and for the most recent statewide (14.6%) and national averages (15.5%) are within one percentage point. HSRs range from 8.4% to 20.7%.

Figure 69. Smoking (Adults)

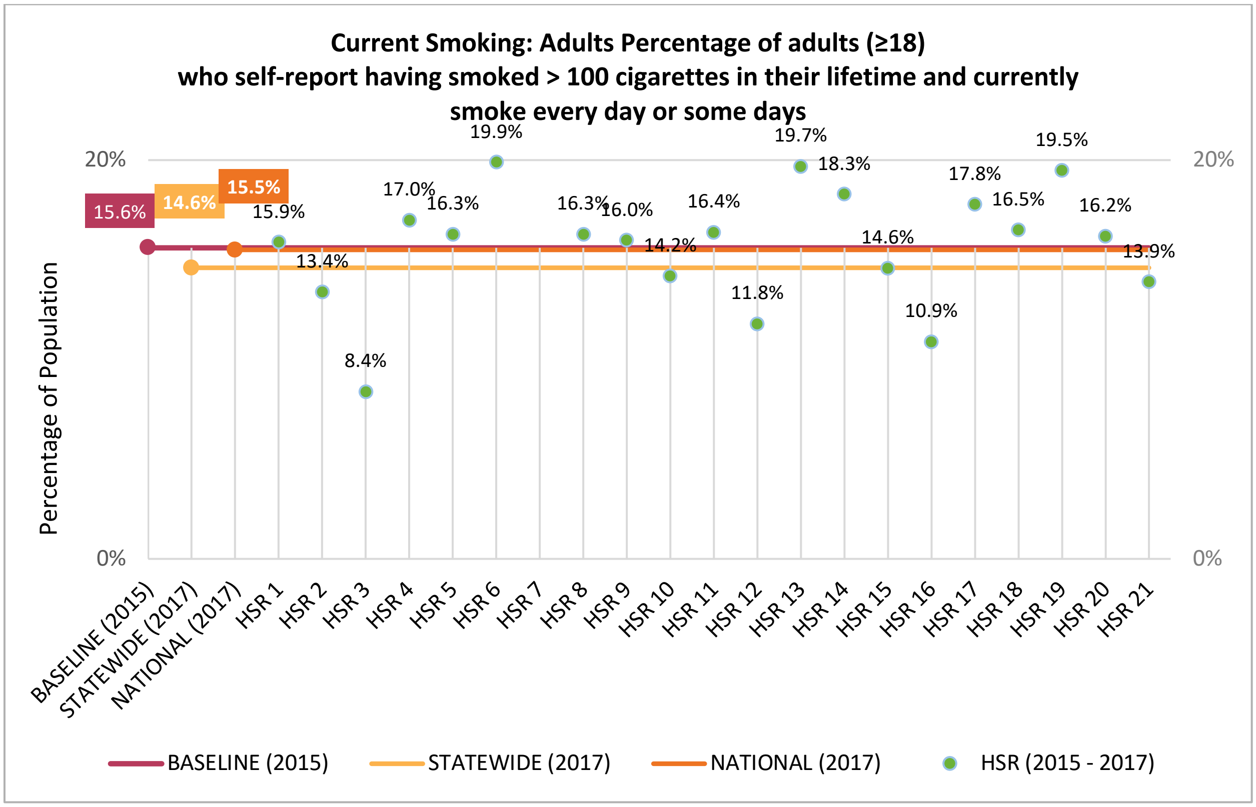
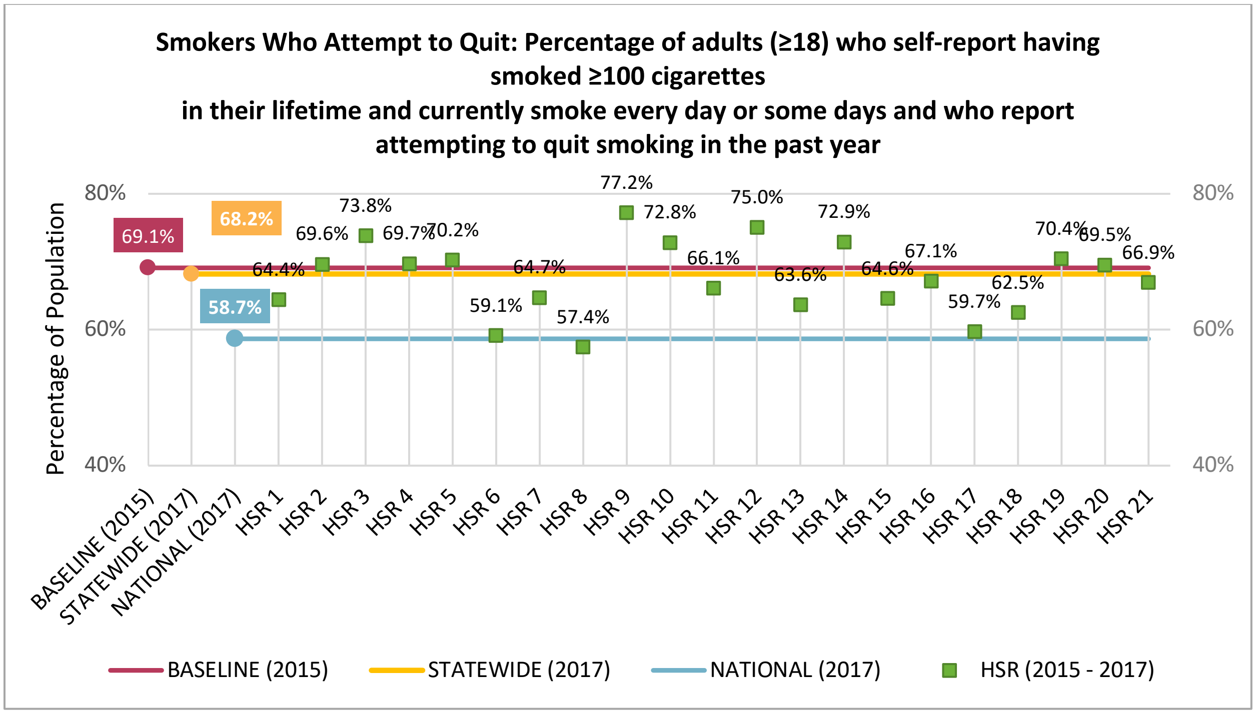


Figure 70. Smokers Attempting to Quit (Adults)



Measures Without HSR data

Five of the population health measures included in the SIM evaluation plan are not available at the HSR level at the time of this report. We report these measures in the following table.

Table 123. Measures Without HSR Data

Population Health Measures Without HSR Level Data						
Metric	**Baseline**	**Statewide**	**National % (2017)**			
Depression: Adult Males Who Are Currently Depressed. Percentage of male adults (≥18) whose answers to PHQ-8 scale indicated that they were currently depressed	5.7% (2014)	5.2% (2016)	5.3% (2017)			
Crude suicide death rate per 100,000	19.60/ 100,000 (2015)	20.40/ 100,000 (2017)	14.00/ 100,000 (2017)			

Population Health Measures Without HSR Level Data						
Non-Medical Opioid Use: Colorado. Percentage of those >12 years old who self-report ever having used pain relievers without a prescription, using in greater amounts, more often or longer than prescribed, or using it in any other way a doctor did not direct.	5% (2013-2014)	5% (2015-2016)	4% (2016)			
Rates per 100,000 deaths among older adults (≥65) with underlying cause of death related to falls within a calendar year.	102.8 per 100,000 residents 65 and older (2015)	96.36 per 100,000 (2018)	61.6 per 100,000 (2016)			
	2012 Female	**2012 Male**	**2014 Female**	**2014 Male**	**2016 Female**	**2016 Male**
Up to Date on Core Preventive Services: Older Adults Statewide. Percentage of older adults (≥65) (by male and female) who self-report receiving all of a core set of preventive services (influenza and pneumococcal vaccinations.)	48.4%	50.4%	47.1%	50.1%	34.4%	48.3%

8 Conclusion

Conclusions and Recommendations

The SIM implementation was a large-scale and wide-ranging effort, with activities that were expected to lead to outcomes at multiple levels: provider, patient, practice, statewide population, and overall systems (e.g., healthcare payment structures). To highlight some of the activities and results of such an immense project, we include this chapter, which summarizes findings from the summative evaluation of the SIM implementation, including outcomes achieved over the life of the 4.5-year-long project.

An implementation effort of this magnitude requires a similarly large evaluation effort. Yet, the evaluation does not necessarily cover every potential outcome. Furthermore, we do not cover all outcomes with the same level of comprehensiveness. Instead, we structured the summative evaluation around answering specific evaluation questions developed in partnership with SIM stakeholders. The Executive Summary of this report presents each of these questions and provides a brief summary for each. For this chapter, we organize findings by each of the subject chapters presented in this report. These chapters correspond to major outcome areas, which are derived from SIM primary drivers: practice transformation (including integration, access to care, and clinical outcomes), payment reform, cost and utilization, and population health.

We conclude the chapter with comments related to lessons learned based on the evaluation timing and data availability .

Practice Transformation: Integration Efforts

We measure change in integration with a composite indicator comprising 14 primary-care-reported and 13 CMHC-reported building blocks. Using this composite measure, we conclude that all three primary care cohorts and the four CMHCs experienced statistically significant improvements in the average amount of integration from baseline to the final assessment. The magnitudes of these changes are also large. The composite indicator reflects the extent that integration building blocks are fully implemented, and the average of baseline values varied between 52.6% and 71.2%. By the final assessment period, the cohort and CMHC averages varied between 77.4% and 90.4%.

In our analysis by practice type, we found that all three types (i.e., adult, mixed primary care, pediatric) experienced statistically significant improvements across all three primary care cohorts, with the exception of adult practices in cohort 1. This group did have improvements, but the small sample size (10 practice sites) led to a P-Value just above the 0.05 threshold (0.059).

Analyzing by practice size, location, and volume of underserved patients, we find that all categories had statistically significant improvements over time with the exception of small (P-value 0.096) and medium (p-value 0.145) size cohort 1 practice sites.

Additionally, when using the Clinician and Staff Experience Survey (CSES) composite measure, we found very stable provider satisfaction survey results over time. The only statistically significant change was an increase in cohort 2 satisfaction. All cohorts had large sample sizes, and the lack of statistical significance resulted from little change in the averages over time, not from limited sample sizes. Sub-analysis by staff rolls (i.e., behavioral or physical providers), practice type, or percentage of underserved patients did not yield any subgroups with statistically significant improvements over time.

For patient engagement, we had two data sources: one related to practice activities and the other a direct survey of patients. The Practice Monitor, which was our source of information on integration, also asked seven questions on the implementation of practice-level partnering with patients in such areas as collaboratively developing care plans. A composite measure made up of these seven items showed statistically significant improvement from baseline to final for all three primary care cohorts. The CMHCs also showed improvement over time, but we did not conduct statistical tests because of the small sample size.

Our second source of data was patient surveys. Although we had two sources of patient surveys, limits on data availability of both practice-conducted patient surveys and the HCPF-administered Child and Adult Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Patient-Centered Medical Home (PCMH) survey prevented us from drawing any conclusions about changing patient opinion.

Practice Transformation: Access to Care

The Agency for Healthcare Research and Quality (AHRQ) defines “access to care” as “the timely use of personal health services to achieve the best health outcomes."[84] In explaining this definition, AHRQ lists four components:

- **Coverage**: Facilitates entry into the healthcare system. Uninsured people are less likely to receive medical care and more likely to have poor health status.
- **Services**: Having a usual source of care is associated with adults receiving recommended screening and prevention services.
- **Timeliness**: Ability to provide healthcare when the need is recognized.
- **Workforce**: Capable, qualified, culturally competent providers.

---

[84] https://www.ahrq.gov/research/findings/nhqrdr/chartbooks/access/elements.html

In analyzing changing access to care for patients attributed to primary care practices participating in SIM, we used both direct and indirect approaches. For direct measurement, we examined changes in screening associated with the "services" component of access. These include statistically significant improvements in *Depression Screening*, *Adult Obesity Screening*, and *Fall Risk Screening* measures. Cohort 1 also has statistically significant improvements in *Hemoglobin A1c Control* and *Asthma Medication Management* measures. Cohort 2, likewise, showed an increase in *Adolescent Obesity Screening*. However, the other screening measures did not show statistically significant change.

In addition to these screening services, we also are able to report on changing use of preventative services. Milliman calculated one practice-site-level measure related to prevention services: the percentage of psychiatric admissions with a follow-up outpatient visit within 30 days. This is an outpatient service related to prevent re-hospitalization, and it is linked to access to behavioral healthcare. Unfortunately, it did not increase on average, either on average for SIM practice sites or compared to patients attributed to non-SIM providers.

We had limited capacity to measure other components of access. Coverage references health insurance. Patients without any type of coverage were not included in our APCD data, and we were unable to analyze changing insurance coverage. With hundreds of primary care practice sites participating in SIM (319 completed SIM), we were unable to gather systematic information on timeliness such as clinic hours, days until an appointment, and practices accepting new patients.

Our one quantitative measure for workforce is based on a survey of SIM practice staff and clinicians regarding their work satisfaction. Increased burnout, had it occurred, would have hindered access to care. Results of these surveys, presented in the Practice Transformation: Integration Efforts chapter show, however, that SIM work in integrating physical and behavioral healthcare did not increase provider burnout, thus not impeding access to care. The same chapter documents substantial programmatic activities undertaken through SIM to improve the availability of qualified behavioral health providers. Despite these efforts, there remain an insufficient number of qualified providers to meet the behavioral health needs of patients in primary care practices.

Our second approach to measuring changes in access to care is indirect. We use four AHRQ indirect indicators which sum the number of inpatient hospital admissions for conditions that could be treated successfully in primary care settings, such as dehydration or diabetes compilations. These measures were selected by the Evaluation Workgroup during the SIM planning phase and were not intended as targets for individual practices. Instead, they are used in this evaluation as a measure of the overall SIM impact on "access to care," one of the components of the overall SIM goals. Comparing changes in this measure between patients served by SIM practice sites and a matched comparison group of patients in the state allows for an isolation of effects that can be directly attributed to SIM.

A decrease in these admissions corresponds to improvements in access to outpatient care. Analysis of the SIM and comparison group time series shows higher levels of these admission for the SIM group, which does not support the hypothesis that the SIM intervention resulted in improved access for patients attributed to SIM-participating practices.

To measure statewide changes in access, we used the CHAS, but saw few changes over time. These findings are expected based on (1) the results above showing that these measures did not seem to improve significantly for the SIM population, specifically, and (2) the fact that statewide data for all measures were only available through 2017, which was very early in SIM implementation. Data points for 2019, and later will be better potential indicators of longer-term effects of full SIM implementation.

Despite these limitations, large majorities of Coloradans reported positive mental and general health, whereas much lower percentages indicated a lack of access to mental health counseling services or experiences of discrimination in medical settings. Finally, based on evidence from CHAS data, views towards the Colorado health system are improving slightly over time. For example, the number of individuals reporting concerns about mental health stigma (as a barrier to treatment) did decline significantly between 2015 and 2017.

In summary, patients attributed to primary care practices participating in SIM experienced increasing screening associated with improvements in access to care. Statewide survey results do not demonstrate improvements in access attributable to SIM, and comparison group analysis does not show improvement for patients attributed to SIM practices in either follow up after psychiatric hospitalization or the AHRQ indicators that indirectly measure access to quality outpatient care.

Clinical Outcomes

The Practice Transformation chapter focuses on reporting transformation efforts at SIM primary care and CMHC sites. The Clinical Outcomes chapter is related to outcomes from these efforts, reporting on statistically significant increases in the use of screenings measure that compose clinical quality measures (CQMs) and analyzing the effect of practice transformation factors on changes in the CQMs. Although we report average practice site outcomes for all cohorts, we focus results for cohorts 1 and 2, which have data allowing us to measure change over time and enough observations to perform statistical tests for difference in means.

Cohort 1 and 2 practice sites show statistically significant improvement in six of the 14 practice-site-reported CQMs during participation in SIM. Both cohorts had improvements in *Depression Screening* and *Fall Risk Screening* measures. Cohort 1 also had improvements in *Hemoglobin A1c Control* and *Asthma Medication Management* measures, and cohort 2 showed an increase in *Adolescent Obesity Screening* and *Adult Obesity Screening*.

Two other measures showed encouraging results for both cohorts. The average rates of *Maternal Depression Screening* increased over time, although differences were not significant. However, this may be caused by small sample sizes; likewise, both proxy and practice-site-reported measures had consistent increases over time. The same was true for the *Developmental Screening* measure.

Because of reporting issues with both the *Alcohol and Other Drug* and the *Alcohol Screening* measures, no conclusions can be made about changes over time.

We observed no consistent or statistically significant improvements in any of the other measures. Additionally, one measure showed declines over the reporting periods (*Tobacco Use Screening*), but this decline is likely related more to anomalies in data reporting than to a meaningful trend.

As measured by the reported CQMs, these results show some improvements in clinical quality over the course of SIM implementation. Although the lack of a comparison group means that we cannot definitively state that SIM alone contributed to these improvements, the results remain encouraging.

The proxy CQMs have the potential to offer insight into whether improvements in the SIM CQMs were caused by SIM involvement or resulted from statewide trends unrelated to SIM. A pattern of improvements in both the SIM CQMs and the SIM cohort proxy CQMs, with no corresponding improvement in statewide proxy CQMs, would support the hypothesis that SIM participation drove changes in CQM. However, we did not observe any examples with this pattern. Generally, the proxy CQMs for the SIM cohorts matched the statewide trends.

The relationship between practice transformation factors and CQMs is complex. As opposed to cross-sectional correlations, we focused our analysis on correlations between improvements in the transformation factors and improvements in the CQMs. The staff survey changes show little relationship to changes in CQMs, perhaps because the staff survey results had very little change over time. Changes over time in integration have a statistically significant relationship to changes in the *Hemoglobin A1c Control* and *Hypertension Management* CQMs. Both had negative relationships. And although we anticipated that improvements in integration would decrease the number of patients with hemoglobin above the recommended level (the outcome we hope to observe), improved integration should not have reduced the number of patients with correctly managed hypertension. There are many potential reasons to observe this pattern. For example, practices with the most improvement in integration may draw in those patients with disproportionately difficult to control cases of hypertension. Our data do not allow us to explore these possibilities.

Improvements in HIT over time correlated with improvements in the *Depression Screening* CQM. But they also correlated with worsening values for the *Adolescent Obesity Screening* measure. Based on these conflicting results, the transformation factor least associated with improvements in CQMs is the staff survey. HIT and integration improvements were associated with changes in the CQMs, albeit in inconsistent ways.

Cost and Utilization

Because cost and utilization outcomes could be measured for all attributed patients in the APCD, we were able to compare outcomes between patients attributed to cohort 1, cohort 2 and CMHC practice sites to outcomes for a matched comparison group of patient attributed to non-SIM providers. We calculate “difference-in-difference,” which measures the change over time for patients attributed to SIM-participating providers (the SIM group) and the change over time for the comparison group. Because both the SIM and comparison groups were subject to the same statewide factors that would change cost and utilization, but only the SIM group benefited from the SIM practice transformation efforts, the difference in difference is the estimated effect of SIM participation net of any statewide changes.

The most aggregate outcome measure, the total cost of care, did not *robustly* show cost savings. By limiting cohort 2’s data to equivalent six-month periods in the baseline and Year 1 of SIM participation, we were able to find a single example of statistically significant savings in the total cost of care. Using a full year of data for baseline and year 1 eliminated statistical significance for cohort 2. Based on full one- or two-year follow-up data, we do not observe statistically significant aggregate cost savings for either cohort 1, cohort 2, or the CMHCs.

While we did not find robust measures of savings in the total cost of care, analysis of individual cost categories did yield some statistically significant examples of cost or utilization savings. For cohort 1 using 2017 versus baseline 2015 data, we observed statistically significant larger declines in psychiatric ED utilization and cost, lower increases in the cost of non-psychiatric admissions and other medical services, and greater declines in the cost of readmissions. Some of these results also occurred for patients attributed to cohort 2 SIM sites, including declines in psychiatric ED use.

But we also see some outcomes that had greater improvement in the comparison group. For cohort 1 2017 versus 2015, these include greater SIM group increases in non-psychiatric inpatient admissions, lower decreases for non-psychiatric ED visits, and greater increases in psychiatric hospital readmissions and readmission cost. Patients attributed to cohort 2 primary care practice sites had greater increases in psychiatric inpatient admissions and costs. They also had greater increases in non-psychiatric inpatient admissions. These examples of higher cost or utilization increases for the SIM group offset some of the cost and utilization savings and are one reason for the lack of robust results for the total cost of care.

In addition to identifying statistically significant improvements that resulted from primary care practices participating in SIM, we also attempted to identify which practice transformation factors drove the observed changes. These include the level of integration, staff satisfaction, and HIT improvements. We compared changes in these factors to changes in outcomes, and were unable to conclude that any of them, as measured, drove systematic changes in cost and utilization of services. This may be because of limitations in the way we measured practice transformation, or because some other factor drove changes in cost and utilization outcomes.

Evaluation Summary, Lessons Learned, and Evaluation Design Recommendations

The SIM initiative was an ambitious and comprehensive effort, touching many aspects of healthcare in Colorado. Major activities took place in four areas: practice transformation, payment reform, health information technology, and population health. Each of these was supported by a series of stakeholder engagement workgroups, which were guided by a SIM Steering Committee, with an Advisory Board providing oversight.

Performing a large-scale program evaluation within this context lead to these five evaluation lessons learned:

Data availability for addressing evaluation questions was a challenge.

As mentioned in the process evaluation report, In the early planning stages of the initiative, stakeholders helped develop evaluation questions. For some of these, there were no readily available data sources to use in addressing the questions. As the evaluator, we attempted to accommodate the diverse and broad interests of the stakeholders across the workgroups. This approach may have been too ambitious. One factor that increased the difficulty of fully addressing all initial interests was SIM office staff changes, particularly changes in evaluation staff. With these changes came new considerations of the value and viability of some of the original evaluation questions.

In response to data limitations and new direction from the SIM office staff, we attempted to revise evaluation questions while adapting data sources to address the original evaluation questions. Ultimately, we recognized that the evaluation could have been improved if we, as evaluators, had worked with the SIM office to better refine initial evaluation questions to the limitations of existing data while acknowledging the diverse and significant interests of stakeholders.

The UCDFM SPLIT set of assessments provided data helpful in examining the implementation of SIM drivers at the individual primary care practice and CMHC levels. For the Process Evaluation report, these data provided useful information on practice successes and challenges that were used during the implementation for mid-course corrections and for annual/final process evaluation reports.

Although it was very helpful to have access to data gathered specifically for the SIM initiative and related to SIM drivers, the SPLIT data also had significant limitations. One is related to the timing of reporting. Initially SPLIT measures were intended to be reported quarterly, and initial evaluation designs were predicated on having two years of quarterly observations for each practice site. Timing of these assessments changed over time and made evaluation adjustments necessary. Cohort 1 practice sites found that the frequency of assessments was a burden, and, therefore, assessments were done less frequently and sites had greater reporting flexibility (e.g., in choosing which CQMs to report). This flexibility was important for the implementation effort but resulted in the ability to report only one change over time for most practice sites: from the last calendar quarter of the initial participation year to the last calendar quarter of the second participation year. This data structure makes detection of change driven by SIM participation less reliable.

A second limitation to the SPLIT data is that they were practice-reported. TriWest did not have any independent method of assessing data standardization and quality, and based recommendations of CHITAs we interviewed, we limited our analysis to the two periods mentioned and did not use other quarters of reported data.

A third limitation is that key data elements, especially measures of integration, were only available for SIM-participating practices. This prevented comparisons to non-SIM practice sites for any evaluation question involving the degree of integration and its impact on outcomes.

Claims data offer both benefits and challenges.

The Colorado SIM project was unique because it involved payers beyond just Medicaid and Medicare. The All Payers Claims Database (APCD) provides an enormous opportunity to include data claims from private payers in order to include a broader patient population in evaluation efforts. The APCD was only source of data on true baseline (pre-SIM participation) outcomes and for outcomes for patients not attributed to SIM providers.

However, the complexity of these data leads to challenges in merging disparate sources. Both CIVHC, the manager of the APCD, and the state Medicaid office changed data processing vendors during the course of SIM initiative. Reconciliation of Medicaid claims in the APCD versus in Medicaid's own data systems was not entirely completed, leading to concerns about Medicaid APCD data quality for this important payer.

In addition, claims data have limitations in their utility for measuring outcomes because they are tied to billing for services rather than provision of services. Some key services for this evaluation, such as depression screening, are not normally billed and therefore included in the APCD. This and other limitations are discussed extensively throughout this report.

Small grants offer concrete examples of progress.

Throughout this report and the Final SIM Process Evaluation report, examples provided from practice sites receiving small grants provide concrete examples of practice transformation activities that furthered integration and improved patient care. These examples could be very helpful for other practices interested in greater physical-behavioral healthcare integration. We were unable to directly measure outcomes from these grants. Doing so would be a useful next evaluation step.

With a short follow-up period to complete the evaluation, we are more likely to identify process improvements than large changes in health, utilization, or cost outcomes.

All cohorts saw gains in the level of integration and overall progress in reaching the milestones designed to indicate practice improvement. In addition, practice sites saw some significant increases in indicators of quality integrated practice, such as increased depression screening and better diabetes control. Further, as discussed in the process evaluation report, many of the foundations laid by SIM, in the efforts made to facilitate better communication and collaboration around VPBs, the population health Call-to-Action document, and many others, are likely to serve as conduits for ongoing healthcare reform in the state.

We had much less success in identifying resulting changes in cost and utilization outcomes of patients. This was driven in part by the reporting lag in the APCD and potentially by the need for practices to further refine integration procedures through experience.

The timing of the assessments for the final process evaluation was also contributor to the limited number of outcomes. We recommend for any similar future efforts that the evaluation timeframe have at least a six-month lag between end of all project activities and data gathering and the final evaluation report completion.