Ohio | Department of Medicaid

2019 MyCare Ohio Program CAHPS® Member Experience Survey Methodology Report

July 22, 2020

IPRO
Better healthcare, realized.

Corporate Headquarters
1979 Marcus Avenue
Lake Success, NY 11042-1072
(516) 326-7767
ipro.org

Table of Contents

List of Tables

List of Figures

1. Introduction

Overview

This report provides the methodology for calculating the results for the 2019 Medicare Advantage and Prescription Drug Plan (MA & PDP) Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Survey for the Ohio Department of Medicaid's (ODM's or Ohio Medicaid's) MyCare Ohio program.

MyCare Ohio is a five-year financial alignment demonstration program aimed at coordinating health care delivery for Ohio residents served by both Medicare and Medicaid. The demonstration is a collaborative effort between Ohio Medicaid, the federal Centers for Medicare & Medicaid Services (CMS), and five managed care plans (MyCare Ohio Plans [MCOPs]). The MyCare Ohio program uses a managed care approach to provide the full continuum of benefits for Medicare-Medicaid members, including long-term services and supports, behavioral health services, and physical health services.

ODM requires a variety of quality assessment and improvement activities to ensure MCOP members have timely access to high quality health care services. These activities include annual member surveys, including the MA & PDP CAHPS Survey. The MA & PDP CAHPS Survey assesses topics such as quality of care, access to care, the communication skills of providers and administrative staff members, and overall experience with providers. Under the demonstration, the MCOPs are required to contract with a CMS-approved survey vendor to administer the MA & PDP CAHPS Survey and to submit their survey data to CMS annually. Five MCOPs participated in the 2019 survey, as listed in Table 1-1 below. The 2019 survey was conducted in the first half of 2019.

Table 1-1: MCOPs

MCOP Name	MCOP Abbreviation
Aetna Better Health of Ohio	Aetna
Buckeye Health Plan	Buckeye
CareSource	CareSource
Molina Healthcare of Ohio, Inc.	Molina
UnitedHealthcare Community Plan of Ohio, Inc.	UnitedHealthcare

Survey Instrument

The survey instrument selected was the 2019 MA & PDP CAHPS Survey. This is the version of the Medicare CAHPS Survey that Medicare-Medicaid Plans (MMPs) were required by CMS to use during reporting year 2019, which represents measurement year 2018. The Medicare CAHPS Survey is part of a group of surveys developed by a consortium of researchers from American Institutes for Research, Harvard Medical School, the RAND Corporation, and Research Triangle Institute (RTI) International under a cooperative agreement between CMS and the Agency for Healthcare Research and Quality (AHRQ), a component of the U.S. Public Health Service. The set of Medicare CAHPS Surveys is sponsored by CMS to collect information on Medicare beneficiaries as required by the Balanced Budget Act of 1997 and the Medicare Modernization Act of 2003.

The MA & PDP CAHPS survey instrument includes questions from the Medicare Advantage Plan (MA-only), Medicare Advantage Prescription Drug Plan (MA-PD), and Prescription Drug Plan (PDP-only) surveys. The MA & PDP CAHPS Survey includes 68 core questions that comprise 15 measures. These measures include five global rating questions, six composite measures, two individual item measures, and two other measures reported to MA-PD contracts, including MMPs. The global ratings are derived from individual, stand-alone survey questions and reflect overall satisfaction with the respondent's health plan, health care quality, drug plan, personal doctor, and specialist. The composite measures are sets of questions grouped together to address different aspects of care (e.g., "getting needed care" or "getting appointments and care quickly"). The individual item measures are individual questions that look at a specific area of care (i.e., "annual flu vaccine" and "pneumonia vaccine"). The other measures reported to contracts assess contact received from a doctor's office, pharmacy, or drug plan regarding medications. Table 1-2 lists the global ratings,

composite measures, individual items, and other measures reported to contracts included in the MA & PDP CAHPS Survey. Table 1-3 lists the items (i.e., questions) that compose the composite measures.

Table 1-2: MA & PDP CAHPS Survey Measures

Global Ratings	Composite Measures	Individual Items	Other Measures Reported to Contracts
Rating of Health Plan	Getting Needed Care	Annual Flu Vaccine	Contact from Doctor's Office, Pharmacy, or Drug Plan: Reminders to Fill Prescription
Rating of Health Care Quality	Getting Appointments and Care Quickly	Pneumonia Vaccine	Contact from Doctor's Office, Pharmacy, or Drug Plan: Reminders to Take Medications
Rating of Drug Plan	Doctors Who Communicate Well		
Rating of Personal Doctor	Customer Service		
Rating of Specialist	Getting Needed Prescription Drugs		
	Care Coordination		

Table 1-3: Items within Composite Measures

Getting Needed Care	Getting Appointments and Care Quickly	Doctors Who Communicate Well	Customer Service	Getting Needed Prescription Drugs	Care Coordination
Getting Needed Care, Tests, or Treatment	Getting Care Needed Right Away	Providing Clear Explanations	Give Information Needed	Ease of Getting Prescribed Medicines	Doctors Have Medical Records
Getting Appointments with Specialists	Getting Appointments	Listening Carefully	Courtesy and Respect	Ease of Filling Prescriptions[1] (combined item)	Doctors Communicate About Tests[2] (combined item)
	Getting Seen Within 15 Minutes of Your Appointment	Showing Respect for What Patients Have to Say	Forms Were Easy to Fill Out		Doctors Discuss Taking Medicines
		Spending Enough Time with Patients			Getting Help to Coordinate Care
					Doctors are Informed about Specialist Care

[1] The Ease of Filling Prescriptions composite measure item is scored based on responses to two survey questions related to the ease of filling a prescription. These individual survey question items include: "ease of filling prescriptions at a local pharmacy" (Question 44) and "ease of filling prescriptions by mail" (Question 46).

[2] The Doctors Communicate About Tests composite measure item is scored based on responses to two survey questions related to communication about blood test, x-ray, or other test results. These individual survey question items include: "how often personal doctor's office followed up to give results" (Question 20) and "how often results were received as soon as needed" (Question 21).

Sampling Procedures

Sample Frame

CMS required the MCOPs to administer the 2019 MA & PDP CAHPS Survey according to the MA & PDP Quality Assurance Protocols & Technical Specifications.[3] Members eligible for sampling met the following criteria at the time the sample was drawn:

- Were 18 years of age or older (i.e., January 3, 2019).
- Were MCOP members.
- Were continuously enrolled in the same MCOP for at least six months.
- Were living in the United States, Puerto Rico, or the United States Virgin Islands.
- Were not institutionalized at the time the sample was drawn.

Sample Selection

In January 2019, CMS selected a random sample of eligible members from the Integrated Data Repository (IDR) for each participating contract. CMS allowed oversampling at the contract level if there was sufficient eligible member volume to support additional sampling after the required MA & PDP CAHPS Survey sample was drawn. MCOPs were required to request an increase in sample size for their contract by December 1, 2018. Following MA & PDP Quality Assurance Protocols & Technical Specifications, CMS selected a random sample of at least 800 MyCare Ohio members from each MCOP.[4] Table 1-4 provides a breakout of the targeted sample size, oversample size, and total sample size for each MCOP.

Table 1-4: MCOP Sample Sizes

MCOP	Targeted Sample Size	Oversample Size	Total Sample Size
Aetna	800	480	1,280
Buckeye	800	800	1,600
CareSource	800	800	1,600
Molina	800	1,000	1,800
UnitedHealthcare	800	0	800

Survey Protocol

The MCOPs contracted with separate CMS-approved CAHPS survey vendors to perform the administration of the MA & PDP CAHPS Survey. The survey administration protocol employed by the MCOPs' vendors was the standardized mixed mode methodology, which allowed for two methods by which members could complete the surveys. The first phase, or mail phase, consisted of a pre-notification letter being mailed to all sampled members, alerting them of the first mailing of the questionnaire, and assuring the sampled members that the survey is sponsored by CMS. Following the pre-notification letter, all sampled members received the first survey mailing. A second survey mailing was sent out to all non-respondents. The second phase, or telephone phase, consisted of Computer Assisted Telephone Interviewing (CATI) for sampled members who had not mailed in a completed survey. A series of at least five CATI calls was made to each non-respondent.[5] It has been shown that the addition of the telephone phase aids in the reduction of non-response bias by increasing the number of respondents who are more demographically representative of a health plan's population.[6]

[3] Centers for Medicare & Medicaid Services. *MA & PDP Quality Assurance Protocols & Technical Specifications, V9.0*. November 2018.

[4] Per CMS' sampling protocol, the targeted sample size is based on the type of contract. For MA contracts, with or without a PDP component, a targeted random sample of 800 members was selected for surveying.

[5] Centers for Medicare & Medicaid Services. *MA & PDP Quality Assurance Protocols & Technical Specifications, V9.0*. November 2018.

[6] Fowler FJ Jr., Gallagher PM, Stringfellow VL, et al. "Using Telephone Interviews to Reduce Nonresponse Bias to Mail Surveys of Health Plan Members." *Medical Care*. 2002. 40(3): 190-200.

The survey protocol allowed sampled members the option to use a proxy (i.e., another individual's assistance with completing the survey) during both the mail and telephone phases of survey administration. Additionally, sampled members had the option to complete the survey in English, Spanish, Chinese, or Vietnamese.[7]

According to CMS' specifications for the MA & PDP CAHPS Survey, these surveys were completed using the time frames shown in Table 1-5.[8]

Table 1-5: MA & PDP CAHPS Survey Time Frames

Basic Tasks for Conducting the Surveys	Time Frames
Send first pre-notification letter to all sampled members one week before the first questionnaire mailing.	0 days
Send first questionnaire with cover letter within one week after mailing the pre-notification letter. Send a postcard reminder to non-respondents four to 10 days after mailing the first questionnaire.	7 – 9 days
Send a second questionnaire with cover letter to non-respondents within four weeks after mailing the first questionnaire.	30 – 32 days
Initiate CATI interviews for non-respondents approximately 21 days after mailing the second questionnaire.	50 – 57 days
Initiate systematic contact for all non-respondents such that at least five telephone calls are attempted at different times of the day, on different days of the week, and in different weeks.	58 – 92 days

2. Data Analysis

A number of different analyses were performed to generate the 2019 MyCare Ohio Program CAHPS Member Experience Survey results. This section provides a detailed discussion of each of the analyses used to generate the MyCare Ohio Program CAHPS Member Experience Survey reports.

Response Rates

The administration of the MA & PDP CAHPS Survey is comprehensive and designed to achieve the highest possible response rate. A high response rate facilitates the generalization of the survey responses to an MCOP's population. Per CMS specifications, the response rate is the total number of completed and partially completed surveys divided by all eligible members of the sample.[9] A survey was assigned a disposition code of "completed" if response items for at least one reportable measure (e.g., Getting Needed Care) were answered, and greater than or equal to 50 percent of the applicable to all (ALA) items were answered.[10,11] A survey was assigned a disposition code of "partially completed" if response items for at least one reportable measure (e.g., Getting Needed Care) were answered, and less than 50 percent of the ALA items were answered. Eligible members included the entire random sample minus ineligible members. Ineligible members of the sample met one or more of the following criteria: were deceased, were invalid (they did not meet criteria described on page 6 of this report), or were mentally or physically incapacitated.

$$Response\ Rate = \frac{Number\ of\ Completed\ Surveys + Number\ of\ Partially\ Completed\ Surveys}{Random\ Sample - Ineligibles}$$

[7] Survey vendors have the option to offer a Chinese or Vietnamese translation of the MA & PDP CAHPS Survey questionnaires.
[8] Centers for Medicare & Medicaid Services. *MA & PDP Quality Assurance Protocols & Technical Specifications, V9.0*. November 2018.
[9] Centers for Medicare & Medicaid Services. *MA & PDP Quality Assurance Protocols & Technical Specifications, V9.0*. November 2018.
[10] Please refer to Appendix K in the *MA & PDP Quality Assurance Protocols & Technical Specifications, V9.0* for the complete list of survey items applicable to all respondents.
[11] Please refer to Appendix L in the *MA & PDP Quality Assurance Protocols & Technical Specifications, V9.0* for the complete list of reportable measures.

It is important to note that CMS' specifications for the MA & PDP CAHPS Survey provide guidelines for calculating a response rate as a close approximation only. While IPRO followed CMS' specifications for calculating response rates, the response rates presented in the Executive Summary and Full reports may not match the response rates presented in the MCOP reports produced by CMS.

Demographics

Five separate demographic categories were evaluated using CMS administrative data or specific survey questions for the MyCare Ohio population. Table 2-1 depicts the data sources (either CMS administrative data or the MA & PDP CAHPS survey question items) used in calculating the demographic frequencies.

Table 2-1: MyCare Ohio Demographic Categories

Demographic Category	Data Source
Age	CMS Administrative Data
Gender	CMS Administrative Data
Education	Question 61
Race	Question 63
General Health Status	Question 48

Respondent/Non-Respondent Analysis

An analysis of the demographic characteristics of the respondents and non-respondents to the 2019 MyCare Ohio Program CAHPS Member Experience Survey was conducted. The demographic information analyzed was derived from CMS administrative data. Member age, gender, and race/ethnicity were broken into categories and analyzed for statistically significant differences between the respondent and non-respondent populations. Given the differences in the data sources used to determine race for the demographic analysis and respondent and non-respondent analysis (i.e., MA & PDP CAHPS Survey responses vs. CMS administrative data), the respondent results and values for the race demographic category will differ in each section.

Confidence Interval Overlap Test

A 95% confidence interval overlap test was performed to determine whether the percentage of respondents was statistically significantly different from the percentage of non-respondents within a particular demographic category.[12] The binomial proportion method (SAS®, Version 9.4, Cary, NC) was used to calculate 95% confidence intervals for the MyCare Ohio program and each MCOP. If intervals for the plan and the MyCare Ohio program did not overlap, then the difference between the groups was considered to be statistically significant.

$$lower\ limit\ of\ interval = p - z_{1-\frac{\alpha}{2}} * \sqrt{\frac{p*(1-p)}{n}} - \frac{1}{2n}$$

$$upper\ limit\ of\ interval = p + z_{1-\frac{\alpha}{2}} * \sqrt{\frac{p*(1-p)}{n}} + \frac{1}{2n}$$

$p = proportion\ of\ respondents$
$n = denominator\ for\ proportion$
$\alpha = 0.05$
$z = critical\ z\text{-}score = 1.96$

The above formula calculates a confidence interval ranging around the score (*p*) by the z-score times the standard error of *p*. Since we are approximating a discrete distribution with a continuous distribution, we use a correction of 0.5/*n* in the formula.

[12] This methodology differs from the methodology used for previous years' reports.

For each demographic category, arrows were assigned to each MCOP's and the MyCare Ohio program's respondent percentages to indicate whether there were statistically significant differences from the non-respondent percentages. An overlap in the confidence intervals between respondents and non-respondents is said to be statistically non-significant whereas a non-overlap in the confidence intervals between respondents and non-respondents is said to be statistically significant. MCOP-level and program-level percentages for the respondent population that were statistically significantly higher than the non-respondent population are noted with upward (↑) arrows. MCOP-level and program-level percentages for the respondent population that were statistically significantly lower than the non- respondent population are noted with downward (↓) arrows. MCOP-level and program-level percentages for the respondent population that were not statistically significantly different from the non- respondent population are not noted with arrows.

National Comparisons Analysis

The National Comparisons analysis was conducted using CMS benchmarks. The linear means were calculated in accordance with CMS' technical specifications for survey measures.[13] In order to assess the overall performance of the MyCare Ohio program and each MCOP, IPRO reported the linear means for the five global ratings (Rating of Health Plan, Rating of Health Care Quality, Rating of Drug Plan, Rating of Personal Doctor, and Rating of Specialist), six composite measures (Getting Needed Care, Getting Appointments and Care Quickly, Doctors Who Communicate Well, Customer Service, Getting Needed Prescription Drugs, and Care Coordination), and two other measures (Influenza Vaccination and Pneumonia Shot), but for those CMS-calculated scores reported as NA for select plans, IPRO calculated the missing scores using CMS' scoring methodology (for Rating of Personal Doctor, Rating of Specialist, and Doctors Who Communicate Well).

Calculation of the Means

For each global rating, composite measure, individual item, and other measures reported to contracts, an unadjusted mean was first calculated. For each of the global rating questions, scoring was based on a 10-point scale: a response value of 0 was given a score of 0, a response value of 10 was given a score of 10, etc. Table 2-2 illustrates how the global rating score values were determined. The 10-point global rating unadjusted mean was the sum of the response scores divided by the total number of responses to the global rating question.

$$Global\ Rating\ Mean\left(GRM\right)=\frac{\sum_{i=1}^{n}x_i}{n}$$

$i = 1,\ldots,n$ *members responding to question*
x_i = *score of member on question (from 0 to 10)*

Unadjusted means were calculated for the composite measures (Getting Needed Care, Getting Appointments and Care Quickly, Doctors Who Communicate Well, Customer Service, Getting Needed Prescription Drugs, and Care Coordination). In general, scoring was based on a four-point scale: responses of "Always" or "Yes, definitely" were given a score of 4; responses of "Usually" or "Yes, somewhat" were given a score of 3; responses of "Sometimes" or "No" were given a score of 2; and "Never" response was given a score of 1. Table 2-2 illustrates how the composite score values were determined.

The composite mean was the average of the unadjusted mean score for each question included in the composite measure. That is, each question contributed equally to the average, regardless of the number of respondents to the question. The unadjusted mean for each composite item was the sum of the response scores divided by the total number of responses to the item.

[13] Centers for Medicare & Medicaid Services. *MA & PDP Quality Assurance Protocols & Technical Specifications, V9.0*. November 2018.

$$Composite\ Measure\ Mean\ (CMM) = \frac{1}{m}\sum_{i=1}^{m}\left(\sum_{j=1}^{n_i}\frac{x_{ij}}{n_i}\right)$$

$i = 1,\ldots,m$ questions in a composite
$j = 1,\ldots,n_i$ members responsing to question i
x_{ij} = score of member j on question i (from 1 to 4)

For the individual item measures (Annual Flu Vaccine and Pneumonia Vaccine) and the other measures reported to contracts (Contact from Doctor's Office, Pharmacy, or Drug Plan: Reminders to Fill Prescription and Contact from Doctor's Office, Pharmacy, or Drug Plan: Reminders to Take Medications), unadjusted mean scores were calculated on a scale of 0 to 1.

Table 2-2 illustrates how score values were determined for these measures.

Table 2-2: Determining Unadjusted Mean Values

Response Category	Score Values
Global Ratings: 0–10 Format	
0	0
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
Composite Measures: Never/Sometimes/Usually/Always Format	
Never	1
Sometimes	2
Usually	3
Always	4
Care Coordination Composite Measure: No/Yes, somewhat/Yes, definitely Format	
No	2
Yes, somewhat	3
Yes, definitely	4
Individual Item and Other Measures Reported to Contracts: No/Yes Format	
No	0
Yes	1

Case-Mix Adjustment

Given that variances in respondents' demographics can result in differences in ratings between MCOPs that are not due to differences in quality, the data were adjusted to account for disparities in these characteristics. Case-mix refers to the characteristics of respondents used in adjusting the results for comparability among MCOPs. The unadjusted means were case-mix adjusted for age, education, self- reported general health status, self-reported mental health status, proxy assistance or completion of the survey form, Medicaid dual eligibility, low-income subsidy eligibility, and

completion of the survey in the Chinese or Vietnamese language.[14] IPRO followed the case-mix adjustment values provided in the MA & PDP Quality Assurance Protocols & Technical Specifications. In addition, case-mix adjusted scores are calculated using the following formula:

$$Adjusted\ Score = Raw\ Score - Net\ Adjustment$$

Where net adjustment is calculated using the following equation:

$$Net\ Adjustment = \left(Contract\ Adjuster's\ Mean - State\ Adjuster's\ Mean\right) \times Coefficient$$

The coefficient in the above equation was estimated using linear regression. It is important to note that CMS uses national data for case-mix adjustment; therefore, the results presented in the Executive Summary and Full reports will not match the results presented in the MCOP reports produced by CMS.

Linear Mean

Once means are adjusted, a linear mean can be determined by expanding the adjusted mean to a 0–100 scale. The linear mean is calculated using the following equation:

$$y = 100 \times \frac{\left(x - a\right)}{\left(b - a\right)}$$

$y = 0 - 100\ score$
$x = adjusted\ score$
$b = highest\ possible\ score\ on\ the\ original\ scale$
$a = lowest\ possible\ score\ on\ the\ original\ scale$

Assignment of Star Ratings

The National Comparisons analysis depicts member satisfaction using a one- to five-star rating system as displayed in Table 2-3. Star assignments were assigned based on a comparison of each measure's linear means to CMS' MMP national benchmarks, which were requested by ODM, for the MA & PDP CAHPS Survey results.

The MyCare Ohio program's and the MCOPs' scores were compared to national MMP percentile benchmarks provided by CMS to derive the overall star ratings for each measure. Based on this comparison, ratings of one (★) to five (★★★★★) stars were determined for each CAHPS measure, where one is the lowest possible rating (i.e., Poor) and five is the highest possible rating (i.e., Excellent), as shown in Table 2-3.[15] Please note that while CMS typically uses hundreds of MA-PD contracts for benchmarks, the national MMP benchmarks were produced using 42 MMPs, which may skew the distribution if there are outliers and result in more or less variation than the overall MA-PD contracts; therefore, caution should be exercised when interpreting these results.

[14] Age, Medicaid dual eligibility, and low-income subsidy eligibility were derived from CMS administrative data.
[15] IPRO used a different methodology to determine star ratings than is specified in the *MA & PDP Quality Assurance Protocols & Technical Specifications, V8.0*.

Table 2-3: Star Ratings

Stars	Percentiles
★ Poor	Below the 25th percentile
★★ Fair	At or between the 25th and 49th percentiles
★★★ Good	At or between the 50th and 74th percentiles
★★★★ Very Good	At or between the 75th and 89th percentiles
★★★★★ Excellent	At or above the 90th percentile

Statewide Comparisons Analysis

The Statewide Comparisons analysis presents results based on CMS' methodology. A description of the process used in calculating the mean values is presented on page 9 of this report. Per CMS specifications, results were case-mix adjusted. Once means were adjusted, the linear mean was calculated. A description of this process is presented beginning on page 11 of this report.

For purposes of reporting MyCare Ohio member experience with care results, CMS requires a minimum of 11 respondents per response category per measure (i.e., a minimum cell size of 11). If a cell size was fewer than 11, additional analyses were performed to determine the appropriate data suppression approach. If one or more of the response categories for a measure did not meet the minimum number of 11 responses, IPRO combined response categories to create aggregate categories that met or exceeded the minimum cell size requirement. In instances where aggregation of the data still resulted in cell sizes of fewer than 11, the measure's results were suppressed in full. Suppressed results are noted in the report figures as "Insufficient Data" (for the response category percentage) and "S" (for the mean).

Response Category Percentages

Response category percentages were calculated for each measure. For the global ratings, responses were classified into three categories:

- Satisfied—9 to 10
- Neutral—7 to 8
- Dissatisfied—0 to 6

For composite measures, responses were classified into three categories:

- Satisfied—Always/Yes, definitely
- Neutral—Usually/Yes, somewhat
- Dissatisfied—Never/Sometimes/No

For individual item measures and other measures reported to contracts, responses were classified into two categories:

- Satisfied—Yes
- Dissatisfied—No

For the global ratings, composite measures, composite items, individual items, and other measures reported to contracts, separate response category percentages were calculated for each of the response categories, as applicable.

Comparative Hypothesis Tests

MCOP-level scores were compared to the MyCare Ohio program (program average) scores to determine whether there were statistically significant differences between the 2019 scores for each MCOP and the 2019 MyCare Ohio program scores. Each of the response category percentages and the overall means were compared for statistically significant differences.

Two types of comparative hypothesis tests were applied. First, a global *F* test was calculated, which determined whether the difference between MCOP scores was significant. If the *F* test demonstrated MCOP-level differences (i.e., $p < 0.05$), then a *t* test was performed for each MCOP. The *t* test determined whether each MCOP's score was significantly different (i.e., $p < 0.05$) from the overall scores of the other MCOPs in the state.

Arrows were assigned to each MCOP's overall means to indicate whether there were statistically significant differences between MCOP-level means and the MyCare Ohio program average. The difference in MCOP performance from the MyCare Ohio program average was considered significant if the two-sided *p* value of the *t* test was less than 0.05. MCOP-level mean scores that were statistically significantly higher than the MyCare Ohio program average were noted with upward (↑) arrows. MCOP-level mean scores that were statistically significantly lower than the MyCare Ohio program average were noted with downward (↓) arrows. MCOP-level mean scores that were not statistically significantly different from the MyCare Ohio program average were not noted with arrows. Arrows noting statistically significant results are only displayed for the overall means in the figures. Statistically significant results for response category percentages and overall means are described in the text below the figures for each measure. National MMP benchmarks are also presented in the figures for comparison.

In some instances, the scores for two MCOPs may be the same, but one was statistically significantly different from the MyCare Ohio program average, and the other was not. In these instances, it is the difference in the number of respondents between the two MCOPs that explains the different statistical results. It is more likely that a statistically significant result will be found in an MCOP with a larger number of respondents.

Trending Hypothesis Test

For each MCOP and the MyCare Ohio program, scores in 2019 were compared to scores in 2018 to determine whether there were statistically significant differences. The MCOP's unadjusted scores were compared for trending given that minimal differences were expected between years in each MCOP's demographics. Each of the response category percentages and the overall means were compared for statistically significant differences. One type of hypothesis test was applied to the trend results. A *t* test was performed to determine whether the MCOP or MyCare Ohio program average means in 2019 were statistically significantly different from the MCOP or MyCare Ohio program average means in 2018.

Directional triangles were assigned to the overall means to indicate whether there were statistically significant differences between scores in 2019 and scores in 2018. The difference in performance from 2018 to 2019 was considered significant if the two-sided *p* value of the *t* test was less than 0.05. Scores that were statistically significantly higher in 2019 than in 2018 were noted with upward (▲) triangles. Scores that were statistically significantly lower in 2019 than in 2018 were noted with downward (▼) triangles. Scores in 2019 that were not statistically significantly different from scores in 2018 were not noted with triangles. Triangles noting statistically significant results are only displayed for the overall means in the figures. Statistically significant results for response category percentages and overall means are described in the text below the figures for each measure.

Priority Areas for Quality Improvement

To determine potential items for quality improvement (QI) efforts, a priority areas analysis was performed at the MyCare Ohio program and MCOP levels to help identify specific aspects of care that will benefit most from QI activities. The analysis provides information on:

- How well the MCOP/program is performing on the survey item.
- How important that item is to overall satisfaction.

The priority areas analysis focused on the following three global ratings: 1) Rating of Health Plan, 2) Rating of Health Care Quality, and 3) Rating of Drug Plan.

IPRO evaluated these global ratings to determine if particular MA & PDP CAHPS items (i.e., questions) have a high problem score (i.e., the MyCare Ohio program/MCOP has demonstrated poor performance) and are strongly correlated with one or more of these measures. These individual CAHPS items, which IPRO refers to as "priority areas," have the greatest potential to affect change in overall satisfaction with the global ratings and, therefore, are areas of focus for possible QI efforts.

Based on IPRO's evaluation of the survey response data, IPRO selected a list of survey questions to include in the priority areas analysis for each global rating. Table 2-4 presents the individual survey questions evaluated for the three global ratings (i.e., Rating of Health Plan, Rating of Health Care Quality, and Rating of Drug Plan) to determine priority areas for the MyCare Ohio program and each MCOP.

Table 2-4: Correlation Matrix

Question	Rating of Health Plan[1]	Rating of Health Care Quality	Rating of Drug Plan
Q4. Getting Appointments and Care Quickly—Getting Care Needed Right Away	✓	✓	
Q6. Getting Appointments and Care Quickly—Getting Appointments	✓	✓	
Q10. Getting Needed Care—Getting Needed Care, Tests, or Treatment	✓	✓	
Q18. Care Coordination—Doctors Have Medical Records	✓	✓	
Q23. Care Coordination—Doctors Discuss Taking Medicines	✓	✓	
Q26. Care Coordination—Getting Help to Coordinate Care	✓	✓	
Q29. Getting Needed Care—Getting Appointments with Specialists	✓	✓	
Q32. Care Coordination—Doctors are Informed about Specialist Care	✓	✓	
Q34. Customer Service—Give Information Needed	✓	✓	✓
Q35. Customer Service—Courtesy and Respect	✓	✓	✓
Q41a. Contact from Doctor's Office, Pharmacy, or Drug Plan—Reminders to Fill Prescription			✓
Q41b. Contact from Doctor's Office, Pharmacy, or Drug Plan— Reminders to Take Medications			✓
Q42. Getting Needed Prescription Drugs—Ease of Getting Prescribed Medicines		✓	✓
Q44. Getting Needed Prescription Drugs—Ease of Filling Prescriptions at a Pharmacy		✓	✓
Q46. Getting Needed Prescription Drugs—Ease of Filling Prescriptions by Mail			✓

[1]A checkmark (✓) indicates that the question was used in the priority areas analysis for the specified global rating.

Problem Scores

The perceived performance on an item is measured by calculating a problem score, in which a negative experience with care is defined as a problem and assigned a "1," and a non-negative experience is assigned a "0." The higher the problem score, the lower the member satisfaction with the aspect of service measured by that question. The problem score can range from 0 to 1. Table 2-5 depicts the problem score assignments for the different response categories.

Table 2-5: Problem Score Assignment

Never/Sometimes/Usually/Always Format		
Response Category	**Classification**	**Code**
Never/Sometimes	Problem	1
Usually	Not a problem	0
Always	Not a problem	0
No response	Not classified	Missing
No/Yes/Yes, somewhat/Yes, definitely Format		
Response Category	**Classification**	**Code**
No	Problem	1
Yes/Yes, somewhat/Yes, definitely	Not a problem	0
No response	Not classified	Missing

It should be noted that, since the priority areas analysis is based on data from individual MCOPs, the problem scores and correlations are not case-mix adjusted for differences among the populations.

Correlation Analysis

The relationship between the survey question's problem score and the global rating's 10-point mean was calculated using a Pearson product-moment correlation, which is defined as the covariance of the two scores divided by the product of their standard deviations.

The correlation can range from -1 to 1, with negative values indicating a negative relationship between the global rating and a particular item's problem score. However, the correlation analysis conducted is not focused on the direction of the correlation, but rather on the degree of correlation. Therefore, the absolute value of *r* is used in the analysis, and the range for *r* is 0 to 1. An *r* of zero indicates no relationship between the response to a question and satisfaction. As *r* increases, the importance of the question to the respondent's satisfaction increases.

A problem score above the median problem score is considered "high." A correlation above the median correlation is considered "high." Priority areas are those items for which the problem score and correlation are both at or above their respective medians. The median, rather than the mean, is used to ensure that extreme problem scores and correlations do not have disproportionate influence in prioritizing individual questions. The problem score mean was the sum of the problem scores (0 or 1) divided by the total number of responses to the survey question.

3. Reader's Guide

Understanding Statistical Significance

Statistical significance is the likelihood that a finding or result is caused by something other than chance. In statistical significance testing, the *p* value is the probability of obtaining a test statistic at least as extreme as the one that was actually observed. If a *p* value is less than 0.05, the result is considered statistically significant. Statistical tests enabled IPRO to determine if the results of the analyses were statistically significant. However, statistical significance does not necessarily equate to clinical significance and vice-versa. Statistical significance is influenced by the number of observations (i.e., the larger the number of observations, the more likely a statistically significant result will be found). Clinical significance depends on the magnitude of the effect being studied. While results may be statistically significant because the study was larger, small differences in rates may not be important from a clinical point of view.

Understanding Correlation Analysis

Correlations are statistical representations that are used to help understand how two different pieces of information are related to one another, and how one piece of variable information may increase or decrease as a second piece of variable information increases or decreases. In general, correlations may be either positive or negative.

- In a positive correlation, scores on two different variables increase and decrease together.
- In a negative correlation, as scores for one variable increase, they decrease for the other variable.

Calculating correlation statistics yields a number called the coefficient of correlation. The coefficient may vary from 0.00 to ±1.00. The strength of a correlation depends on its size, not its sign. For example, a correlation of −0.72 is stronger than a correlation of +0.53. As the correlation coefficient approaches 0.00, it can be inferred that there is no correlation between the two variables. For purposes of the priority areas analysis, the analysis was not focused on the direction of the correlation (positive or negative) but rather on the strength of the correlation; therefore, only the absolute values of the coefficients were used in the analysis, and the range is from 0.00 to 1.00.

It is important to understand that it is possible for two variables to be strongly related (i.e., correlated) but not have one variable cause another. For example, respondents may report a negative experience with ease of getting care, tests, or treatment and also a low overall rating of the health plan. This does not indicate that difficulty in getting care, tests, or treatment causes lower ratings of the health plan. The strength of the relationship between the two only helps to understand whether the difficulty of getting care, tests, or treatments should be a top priority or not.

Understanding Sampling Error

The interpretation of MA & PDP CAHPS results requires an understanding of sampling error, since it is generally not feasible to survey an MCOP's entire population. For this reason, surveys include only a sample from the population and use statistical techniques to maximize the probability that the sample results apply to the entire population.

In order for the results to be generalizable to the entire population, the sample selection process must give each person in the population an equal chance of being selected for inclusion in the study. For the MA & PDP CAHPS Survey, this is accomplished by drawing a systematic sample that selects eligible members for inclusion from the entire MCOP. This ensures that no single group of members in the sample is over-represented relative to the entire population. For example, if there were a larger number of members surveyed between the ages of 45 to 54, their views would have a disproportionate influence on the results compared with other age groups.

Since not every member in an MCOP's total population is surveyed, the actual experience of all members cannot be determined. Statistical techniques are used to ensure that the unknown actual experience of members lies within a given interval, called the confidence interval, 95 percent of the time. The 95 percent confidence interval has a characteristic sampling error (sometimes called "margin of error"). For example, if the sampling error of a survey is ±10 percent with a confidence interval of 95 percent, this indicates that if 100 samples were selected from the population of the same MCOP, the results of these samples would be within plus or minus 10 percentage points of the results from a single sample in 95 of the 100 samples. Table 7-3 depicts the sampling errors for various numbers of responses.[16]

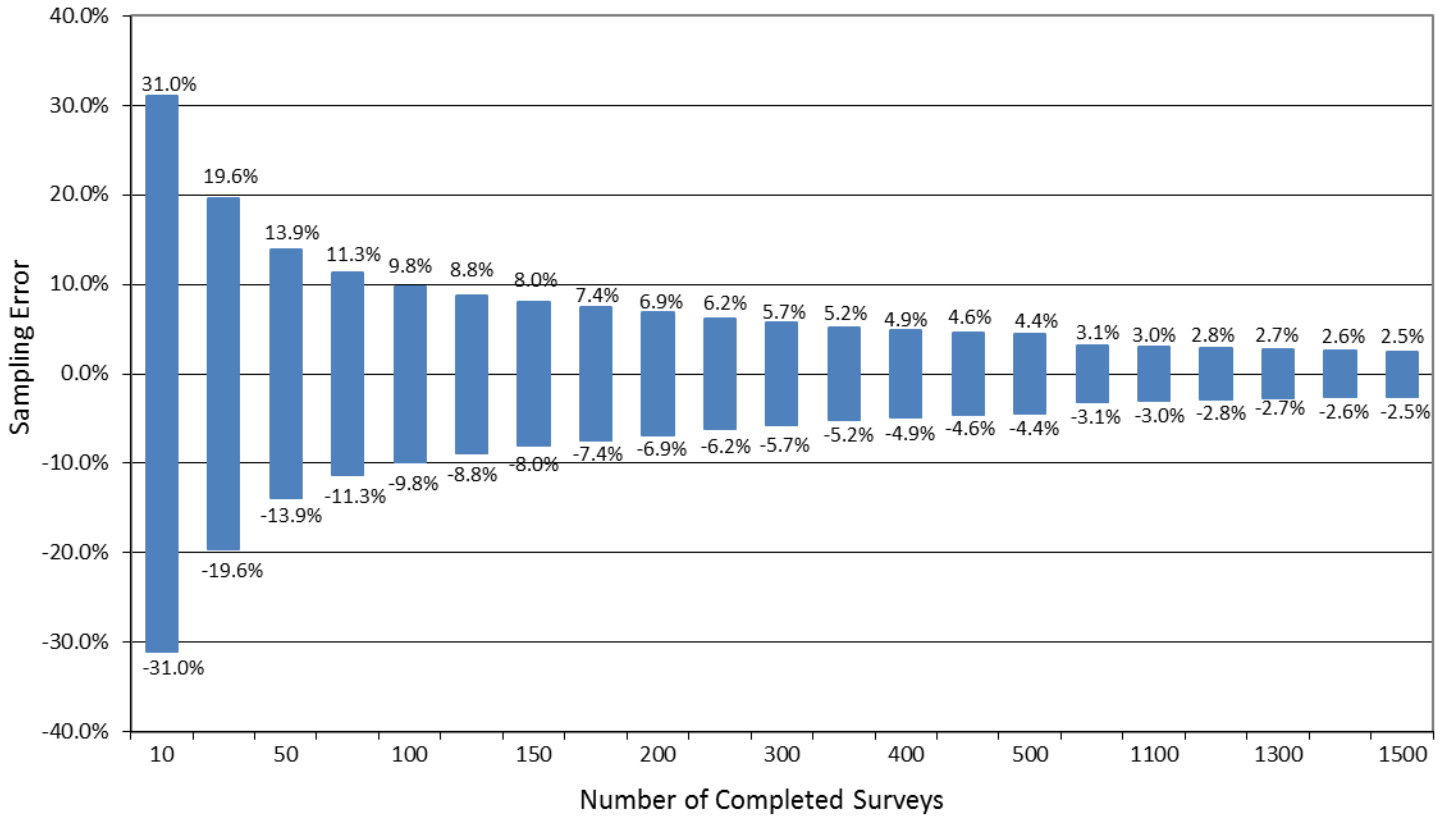
Table 3-1: Sampling Error and the Number of Survey Responses

Number of Responses	100	150	200	250	300	350	400	500
Approximate sampling error (%)	±9.8	±8.0	±6.9	±6.2	±5.7	±5.2	±4.9	±4.4

The size of the sampling error shown in Figure 3-1 is based on the number of completed surveys. Figure 3-1 indicates that if 400 MCOP members complete a survey, the margin of error is ±4.9 percent. Note that the calculations used in the graph assume that the size of the eligible population is greater than 2,000, as is the case with Ohio's MCOPs. As the number of members completing a survey decreases, the sampling error increases. Lower response rates may bias results because the proportion of members responding to a survey may not necessarily reflect the randomness of the entire sample.

[16] Ibid.

Figure 3-1: Sampling Error and the Number of Completed Surveys



As Figure 3-1 demonstrates, sampling error declines as the number of completed surveys increases.[17] Consequently, when the number of completed surveys is very large and sampling error is very small, almost any difference is statistically significant; however, this does not indicate that such differences are important. Likewise, even if the difference between two measured rates is not statistically significant, it may be important from an MCOP's perspective. The context in which the MCOP data are reviewed will influence the interpretation of results.

It is important to note that sampling error can impact the interpretation of MCOP results. For example, assume that 150 state MCOP respondents were 80 percent satisfied with their specialist. The sampling error associated with this number is plus or minus 8 percent. Therefore, the true rate ranges between 72 percent and 88 percent. If 100 of an MCOP's members completed the survey and 85 percent of those completing the survey reported being satisfied with their specialist, it is tempting to view this difference of 5 percentage points between the two rates as important. However, the true rate of the MCOP's respondents ranges between 75 percent and 95 percent, thereby overlapping the state average including sampling error. Whenever two measures fall within each other's sampling error, the difference may not be statistically significant. At the same time, lack of statistical significance is not the same as lack of importance. The significance of this 5 percentage-point difference is open to interpretation at both the individual MCOP level and the state level.

Cautions and Limitations

The findings presented in the 2019 MyCare Ohio MA & PDP CAHPS reports are subject to some limitations in the survey design, analysis, and interpretation. ODM should carefully consider these limitations when interpreting or generalizing the findings. The limitations are discussed below.

[17] Fink, A. *How to Sample in Surveys*. Thousand Oaks, CA: Sage Publications, Inc.; 1995.

Case-Mix Adjustment

While the data for the statewide comparisons analysis were case-mix adjusted for age, education, self- reported general health status, self-reported mental health status, proxy assistance, proxy completion of the survey form, Medicaid dual eligibility, low-income subsidy eligibility, and completion of the survey in the Chinese or Vietnamese language, it was not possible to adjust for differences in member and respondent characteristics that were not measured.[18] These characteristics include employment or any other characteristics that may not be under the MCOPs' control.

Non-Response Bias

The experiences of the survey respondent population may be different from those of non-respondents with respect to their health care services and may vary by MCOP. Therefore, the potential for non-response bias should be considered when interpreting the MA & PDP CAHPS Survey results.

Causal Inferences

Although the MA & PDP CAHPS reports examine whether members of various MCOPs report differences in experience with various aspects of their health care, these differences may not be attributed solely to the MCOP. The analyses described in the CAHPS reports identify whether members have different experiences with their MCOPs. The surveys by themselves do not reveal why the differences exist.

Survey Vendor Effects

The MA & PDP CAHPS Survey was administered by multiple survey vendors. CMS developed its Survey Vendor Certification Program to ensure standardization of data collection and the comparability of results across MCOPs. However, due to the different processes employed by the survey vendors, there is still potential for minor vendor effects. Therefore, survey vendor effects should be considered when interpreting the MA & PDP CAHPS Survey results.

Methods for Analysis

It is important to note that the MA & PDP CAHPS Survey results presented in the Executive Summary and Full reports for the MyCare Ohio program and all MCOPs represent the survey results calculated by IPRO. They are not official survey results and should be used for QI purposes only. To provide ODM with more information regarding MCOP and program performance, IPRO did not apply CMS' interunit reliability (IUR) threshold of "very low reliability" for reporting measure results.[19] For purposes of these reports, IPRO evaluated measure scores for small cell size criteria only (i.e., minimum of 11 responses); all MCOPs' results are reported for each item, regardless of the IUR reporting scoring.

Additionally, results were not weighted. Given these differences, the results presented in the Executive Summary and Full reports for MCOPs will not match the results presented in the MCOP reports produced by CMS. For the calculation of the national MMP benchmarks, 42 MMPs were used to produce these benchmarks; therefore, caution should be exercised when interpreting these results.

[18] Age, Medicaid dual eligibility, and low-income subsidy eligibility were derived from CMS administrative data.

[19] CMS defines "very low reliability" as measures scores with an IUR of less than 0.60. However, the specifications also indicate that no more than 12 percent of plans (those with the lowest IUR on the corresponding measure) are flagged as low reliability for a given measure, after excluding scores based on fewer than 11 responses.