Covered California CAHPS Ratings Fall 2014
Scoring Health Plan's Historical CAHPS Results

Summary of Key Elements

Covered California CAHPS Quality Rating System (QRS) is comprised of the following elements:

1. Report a single summary rating that is an aggregation of the available CAHPS measures
2. Top box 9+10 scoring for the global ratings measures and top box "always" for composite measures
3. PPO benchmarks
4. Regional benchmarks
5. 1-4-star performance classification

Figure 1. Distribution of Star Ratings for Covered California (9 Issuers)

	1 Star	2 Star	3 Star	4 Star
Health Plan Star Ratings Fall 2014	0	2	3	4
Health Plan Star Ratings Fall 2013	0	2	3	4

Purpose

Produce a health plan Quality Rating System (QRS) based on historical CAHPS survey results for all exchange plans with available data. For each Issuer produce a single summary "stars" performance rating using Reporting Year (RY) 2013 results.

Findings

The California plans span the 2-4 stars rating spectrum when the 7 CAHPS measures* are scored as a single summary rating after first aggregating the measures into 3 domains: Plan Service, Access, and Doctors and Care. Most plans scored well on Plan Service relative to benchmarks. Doctors and Care and Access domain scores varied; some health plans rated highly and other plans were low performers.

Geographic and Product Type Benchmarks

The NCQA San Francisco regional PPO benchmarks are used to classify plan performance. This region includes all PPO plans in California, Arizona, Hawaii, Nevada, and Pacific US Territories. The benchmark and performance data were both updated to RY2013 from RY2012.

Our previous report, based on RY2012 data and benchmarks, explored the impact of other geographic and product type options. Broadly speaking, the PPO regional benchmarks and PPO national benchmarks tend to be similar, perhaps because California is a large and diverse state compared to other states in the region and nationally. PPO benchmarks tended to be lower than most HMO benchmarks by 1-2 percent, but higher than Medicaid. The exchange plans are better fit to the PPO model in most cases given the higher cost sharing benefit designs. Other differences in data collection were judged to be minor and it was important to evaluate plans based on a common, level standard.

* Per Table 2, the 7 CAHPS measures are constructed from 10 survey questions

Top Box Scoring Alternatives

A top box scoring method is used -- defined as the sum of 9+10 responses for the global rating questions and the "always" response for the always-never response choice set. For the composite measures, we used the "always" responses as this method is used in several major accountability programs including Medicare Stars. For the global rating questions, the 9+10 formula yields higher reliability as plans are better differentiated compared to the 8+9+10 approach. In our RY 2012 work, the commercial plan ratings were identical for the global ratings' two top box options (9+10 vs. 8+9+10). The Medi-Cal plan ratings drop 1-2 stars when using the 8+9+10 compared to the 9+10.

Misclassification Error

No misclassification of error adjustment was used. In RY2012, we evaluated the probability of misclassification error -- that the true star rating differs from the assigned star rating -- in two ways.

Misclassification error occurs because the performance reported is not necessarily the same as the performance that a member in a particular plan might expect, due to limited or finite samples available. This can be distilled into two types of error: "plan error" in which the reported plan performance is lower than the correct value, and "consumer error" in which the reported plan performance is higher than the correct value.

First, we evaluated for each measure the probability that the reported result was two or more categories (across five categories) different from the correct result, assuming that the correct result mirrored the national PPO distribution. This test showed that, a priori of any data, most plans would be correctly classified. However, the error rates are higher for the CAHPS composites given these measures' lower reliabilities; in turn, we also evaluated the error potential using a probability of differences test. In this second test, for each measure and for the measure set combined, we evaluated the subjective probability that the correct result would be associated with each number of stars, from one to four, given the observed data. The results tell us that each plan probably was rated correctly but the possibility that the plan's true performance is above or below the assigned rating cannot be ruled out for certain plans.

Missing Data

Missing data was limited to 3 instances in which 3 plans' completed sample sizes for the customer service composite did not meet the CAHPS 100 respondents per measure minimum. We adopted the NCQA rule of weighting the non-missing components more highly where these customer service results were missing. The measures standardization step, in which each measure is converted to 1-5 points, was applied first so that no plan is unduly disadvantaged by the presence or absence of a particular measure. That is, each plan was scored by comparing the performance of its reportable measures to the performance of the benchmarks. Additionally, the customer service composite performance was comparable to that of other CAHPS composites -- with typical scores in the 80-85 range, it did not present a situation in which a missing measure was notably harder or easier to achieve than other CAHPS measures.

Eligible Plans

CAHPS scores are available for all of the Covered California 2014 participating plans except Chinese Community and Valley Health Plan (Table 2).

Methods

Measures

The QRS is comprised of 4 global rating questions and three composites, which are drawn from the Consumer Assessment of Health Plans and Providers Survey (CAHPS). The QRS uses three composites -- Getting Needed Care, Getting Care Quickly, and Customer Service -- as these are available for both the California commercial and Medi-Cal plans. The Claims Processing and Plan Information on Costs composites were not used as they were not included in the Medi-Cal plan

survey. All of the measures data is based on the Adult CAHPS survey. The commercial plan and Medi-Cal plan data represented RY2013, which was the most currently available data for the California Medi-Cal plans (Table 2).

Information Hierarchy

The quality information hierarchy consists of a set of CAHPS composites and questions that are organized into three domains (Plan Service, Doctors and Care, and Access to Care); in a subsequent step these three domains are combined to produce a summary rating.

Figure 2: CAHPS Information Hierarchy

Aggregate 3 Domains to Score Summary Indicator
Plan Service Rating of Health Plan Customer Service Composite
Doctors & Care Rating of Health Care Rating of Doctor Rating of Specialist
Access to Care Getting Care Quickly Composite Getting Needed Care Composite

Measure Scoring

For the global rating questions, the top box (9+10) responses were used. For composite measures, we used top box scoring to derive the average proportion of positive responses counting only the number of "always" responses for the "Never", "Sometimes", "Usually", and "Always" response choice set. As an example, to compute the average proportion score for the Getting Needed Care composite, we calculated the ratio of "always" responses to the total number of responses for each of the two questions in the composite. Then the two ratios of "always" responses were averaged by summing the ratios and dividing by 2.

Standardizing Measure Scores

Once the percent top box was computed, it was compared to the 0-100 score benchmarks provided by NCQA to create a standardized measure score in points. The purpose of this standardization is to give all the measures a common interpretation, to eliminate outlier problems, and to assure that no plan would be advantaged or disadvantaged because of missing data. Each health plan's scores were standardized on a 1-5 point scale by comparing the plan raw score with the percentile benchmark drawn from the distribution of regional commercial PPO plans. Each health plan's raw score was mapped to a place along this percentile distribution and, depending upon the performance interval of that distribution, it is assigned a 1-5 value.

We applied the commercial regional PPO benchmarks to create the QRS performance thresholds and assign plans to one of 5 rating categories. NCQA supplied these benchmarks for scoring the RY2013 data. The PPO benchmark data includes 100 values representing the regional performance on the 0-100 scale. Given the fewer number of plans in the regional benchmark data (compared to a national dataset) a benchmark score can occupy multiple percentile values (e.g., a score of "50" could represent the regional performance ranging from the 61st to the 66th percentiles). In instances in which a health plan score spanned a benchmark range, that score and benchmark were rounded to the smallest number of significant figures between the score and benchmark and then compared to assess whether the score met the benchmark. For example, a score and benchmark could be 55.1913 and 55.2, respectively. A score of 55.19, which is

rounded to 55.2, meets the 55.2 benchmark. We applied a 1-5 scoring rule using the regional PPO scores distribution such that if a plan scored less than the 25th percentile, its score = 1. Correspondingly, the 25th-49th = 2, 50th- 74th =3, 75th-89th = 4, =/> 90th = 5. For example, if the value fell between the 50th and 74th percentile of the measure, the standardized score was 3 points.

Domain and Summary Scoring

The standardized measure scores were averaged at the measure level to compute the domain scores and then the domain scores were averaged to compute the summary rating. In this computation, the component measures for each domain were equally weighted to score the domain and, in turn, the 3 domains are equally weighted to construct the summary rating. Next, the health plan domain and summary rating scores were converted into star ratings as described below.

Additional Scoring Rules for Selected Health Plans

Additional scoring steps were applied for several of the health plans. An "all-counties" rating was produced for Molina Health Plan by averaging the plan's Medi-Cal county-level scores. For each measure, the sum of the numerators was divided by the sum of the denominators to calculate the average measure rates across all counties. Then the all-county result was converted to a standard score using the national PPO benchmarks and aggregated into domain or summary indicator scores according to the QRS measure hierarchy in Figure 2. The aggregation steps were identical to those used for all QRS scores.

A single score was produced for the two commercial health plans that offered HMO and PPO products on the Exchange in 2014 (Anthem and Health Net). The historical product type-specific scores were combined to produce a single QRS score and the same result is reported for each of that plan's Exchange products. To combine the historical HMO and PPO results, weights were derived using the health plans' Covered California membership as of July 2014 per Figure 3. For QRS scoring, the health plan's HMO and PPO raw scores were calculated separately, then these scores were averaged weighed by the proportion of Covered California 2014 enrollees. The plan's weighted average score is assigned a star rating per the methodology described below and this single star rating is reported for all of the plan's Exchange product types. Note: no historical scores are available for commercial EPO plans. Plans offering EPO products on the Exchange in 2014 (Blue Shield and Anthem) were assigned the HMO/PPO blended rate as described above (Anthem) or the plan's PPO score (Blue Shield).

Conversion of Health Plan Scores to Star Ratings

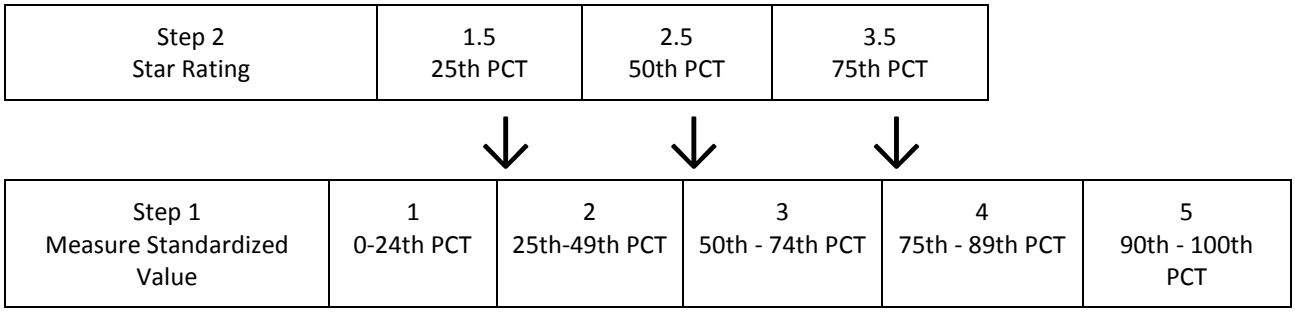
We applied a four-star category performance classification method to each eligible health plan. Performance ratings were produced at the Issuer level.

To create a four-star system, Covered California elected to use thresholds corresponding to average 25th, 50th and 75th percentile performance. The highest rating, four stars, is equivalent to saying that for all of the member-reported experience measures, the plan's average performance was at the 75th percentile or higher—a strong statement that implies that if we added another member-reported measure we could expect the plan to perform at the 75th percentile or higher on that measure, too.

The PPO performance benchmarks were used to assign point values at the 25th, 50th, and 75th percentiles -- to create the four-star ratings categories. In the first step each plan's measure scores were converted to 1-5 values and these values are averaged to produce domain and then a summary, single standardized score. In step two, these values were assigned a star-rating depending upon its position along the performance distribution (Figure 3). The 25th percentile equates to 1.5 points; the 50th percentile to 2.5 points and the 75th percentile to 3.5 points. That is, a plan with a 25th percentile rating falls between 1 and 2 so a midpoint, 1.5, is assigned as the stars threshold. As such, for a plan at the 25th percentile, half the measures would be less than the 25th percentile and score 1, and the other half would be somewhat above the 25th and score 2 and the average is 1.5. Similarly, for the 50th percentile a 2.5 is assigned as midpoint between 2 and 3 and at 75th percentile the 3.5 midpoint is assigned as the midpoint of 3 and 4. Given the very few number of measure-specific results in the top decile, we do not create a 90th percentile/5-star rating category. For those few

instances in which individual measure scores exceed the 90th percentile, the plan is assigned a standardized value of 5, which is an advantage toward achieving a 4-star rating.

Figure 3. Assignment of Points and Ratings Thresholds



Appendix

Table 1. Covered California Participating Plans

Commercial	Data Available
Anthem Blue Cross of California	√
Blue Shield of California	√
Health Net of California, Inc.	√
Kaiser Foundation Health Plan Inc. - Southern California	√
Kaiser Foundation Health Plan, Inc. - Northern California	√
Sharp Health Plan	√
Western Health Advantage	√
Chinese Community Health Plan	Not available
Valley Health Plan	Not available

Medi-Cal Plan	County	Data Available
LA Health Plan**	LA North, LA South	√
Molina HealthCare		
	El Dorado	Not available
	LA North, LA South	Not available
	Placer	Not available
	Sacramento	√
	Riverside, San Bernardino	√
	San Diego	√
	Yolo	Not available

** Single dataset for LA North and LA South

Table 2. CAHPS Measures*

Access to Care Domain

Getting Needed Care Composite

- In the last 12 months, how often was it easy to get appointments with specialists?
- In the last 12 months, how often was it easy to get the care, tests, or treatment you thought you needed through your health plan?

Getting Care Quickly Composite

- In the last 12 months, when you needed care right away, how often did you get care as soon as you thought you needed?
- In the last 12 months, not counting the times you needed care right away, how often did you get an appointment for your health care at a doctor's office or clinic as soon as you thought you needed?

Doctors and Care Domain

Rating of Health Care: Using any number from 0 to 10, where 0 is the worst health care possible and 10 is the best health care possible, what number would you use to rate all your health care in the last 12 months?

Rating of Personal Doctor: Using any number from 0 to 10, where 0 is the worst personal doctor possible and 10 is the best personal doctor possible, what number would you use to rate your personal doctor?

Rating of Specialist Seen Most Often: We want to know your rating of the specialist you saw most often in the last 12 months. Using any number from 0 to 10, where 0 is the worst specialist possible and 10 is the best specialist possible, what number would you use to rate that specialist?

Plan Service Domain

Rating of Health Plan: Using any number from 0 to 10, where 0 is the worst health plan possible and 10 is the best health plan possible, what number would you use to rate your health plan?

Customer Service Composite

- In the last 12 months, how often did your health plan's customer service give you the information or help you needed?
- In the last 12 months, how often did your health plan's customer service staff treat you with courtesy and respect?

*the Medi-Cal plan survey results are based on 6-month continuous enrollment using the "In the last 6 months…" question stem