nrc HEALTH | Human understanding

1 800 388 4264
nrchealth.com

Calculations and Statistics in Catalyst Advanced Reporting

Scoring/Measure Type

The NRC Health standard method for displaying survey results is **top box scoring**. Catalyst Advanced Reporting offers four scoring options, labeled as Measure Type: Positive, Problem, Top Box and Top 2 Box. NRC Health recommends using Top Box when running reports in Catalyst Advanced Reporting and this calculation aligns with HCAHPS surveys.

Top Box scores are the proportion of respondents who answered the question in a certain predetermined manner. Take for example a question that uses a 0 – 10 response scale. When scores are displayed for this question, the top box score represents the percent of respondents who answered 9 or 10 out of the applicable responses. Generally speaking, the responses that are defined as a "top box response" are the desired behavior staff should be exhibiting. Only questions in which a desirable behavior can be identified in the responses are displayed in top box scoring.

If a question has 11 respondents distributed as: 8 answered "Always", 1 "Sometimes", 1 "Never", and 1 respondent skipping the question; the Top Box would be 80.0% (8/10 = 80%).

NRC Health routinely removes certain responses from its calculation. Examples include: 1) a respondent instructed not to answer the question by skip pattern, 2) the respondent skipping the question, and 3) the respondent indicating more than one response.

Since questions can have different response scales, reference the table below to determine which response options are considered Top Box and which are included in the other Measure Types available for reporting.

Positive	Never	Sometimes	Usually	**Always**
Problem	**Never**	**Sometimes**	**Usually**	Always
Topbox	Never	Sometimes	Usually	**Always**
Top2box	Never	Sometimes	**Usually**	**Always**

Positive	No	Yes, somewhat	Yes, mostly	**Yes, definitely**
Problem	**No**	**Yes, somewhat**	**Yes, mostly**	Yes, definitely
Topbox	No	Yes, somewhat	Yes, mostly	**Yes, definitely**
Top2box	No	Yes, somewhat	**Yes, mostly**	**Yes, definitely**

Positive	Definitely no	Probably no	Probably yes	**Definitely Yes**
Problem	**Definitely no**	**Probably no**	**Probably yes**	Definitely Yes
Topbox	Definitely no	Probably no	Probably yes	**Definitely Yes**
Top2box	Definitely no	Probably no	**Probably yes**	**Definitely Yes**

Positive	Poor	Average	**Good**	**Very Good**	**Excellent**
Problem	**Poor**	**Average**	Good	Very Good	Excellent
Topbox	Poor	Average	Good	Very Good	**Excellent**
Top2box	Poor	Average	Good	**Very Good**	**Excellent**

Positive	Poor	Fair	**Good**	**Excellent**
Problem	**Poor**	**Fair**	Good	Excellent
Topbox	Poor	Fair	Good	**Excellent**
Top2box	Poor	Fair	**Good**	**Excellent**

Positive	Strongly Disagree	Disagree	Agree	**Strongly Agree**
Problem	**Strongly Disagree**	**Disagree**	**Agree**	Strongly Agree
Topbox	Strongly Disagree	Disagree	Agree	**Strongly Agree**
Top2box	Strongly Disagree	Disagree	**Agree**	**Strongly Agree**

Positive	Strongly Disagree	Disagree	Neutral	**Agree**	**Strongly Agree**
Problem	**Strongly Disagree**	**Disagree**	**Neutral**	Agree	Strongly Agree
Topbox	Strongly Disagree	Disagree	Neutral	Agree	**Strongly Agree**
Top2box	Strongly Disagree	Disagree	Neutral	**Agree**	**Strongly Agree**

Positive	0	1	2	3	4	5	6	7	8	**9**	**10**
Problem	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	9	10
Topbox	0	1	2	3	4	5	6	7	8	**9**	**10**
Top2box	0	1	2	3	4	5	6	7	8	**9**	**10**

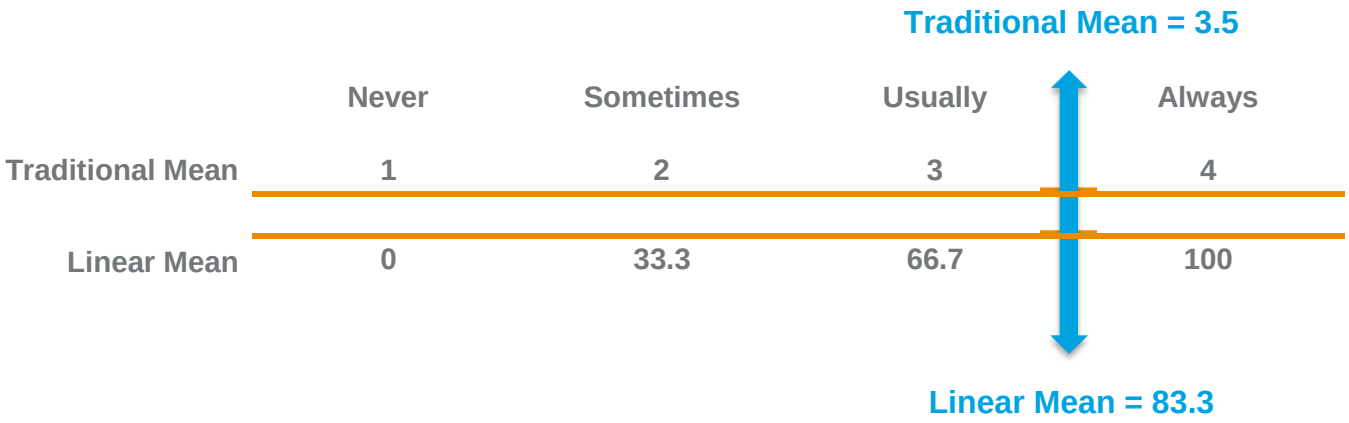
1 800 388 4264
nrchealth.com

Linear Mean

NRC Health offers Linear Mean Scoring as an enhancement to the existing Measure Types. Linear Means are calculated by converting response options to a 0-100 scale, in order to normalize questions with a differing number of response options, and reporting the score as an average of all the responses.

Benefits of Linear Mean over traditional mean scoring:

- **Provides context:** Linear mean scoring provides context around the highest and lowest possible scores. A traditional mean score of 3.5, for example, does not provide the user any information about whether that score is high or low; an average score of 3.5 on a 4-point scale is quite high while a 3.5 on an 11-point scale is quite low. By using a linear mean score approach, we can determine performance in relation to the lower bound of 0 and the upper bound of 100.
- **Standardizes:** Allows for easy comparison across items with different scales. For example, within a single report, users may want to evaluate items that are on a 4-point scale as well as items on an 11-point scale. By fitting both on 0 to 100 scale, users can compare scores across all items to gauge whether some are out performing others without having to keep upper and lower bounds in mind.



Low n-size and Masked Data

NRC Health strives to provide reliable and actionable data for improvement. Therefore, indicators appear on reports for data that does not meet standard thresholds. The mu symbol (μ) indicates data with an n-size less than 30. Questions or Dimensions with less than 30 responses could fluctuate and may not be ideal for establishing baseline performance or measuring improvement. The double dash (--) indicates data has been masked due to the n-size threshold set for that particular Survey Tool. Many CAHPS surveys have a minimum n-size threshold of 11 set by CMS (ICH, ACO, PQRS, Hospice, and Home Health CAHPS). NRC Health uses a minimum threshold of 5 for Employee and Physician surveys.

1 800 388 4264
nrchealth.com

Benchmarks

The optimum measure of improvement is to trend and compare against one's self over time, however, there can be value of comparing against other organizations as an important part of measuring performance. NRC Health strives to provide benchmarks that are reliable and informative. To do this, benchmarks are calculated at the question level, determined by the core number of the question. Different service lines (inpatient versus long-term care) and different patient populations (adult versus pediatrics) use different questionnaires with different core numbers to ensure that only like populations are using the same questions. By keeping the groupings broad we can compare similar experiences; thus ensuring benchmarks are against like populations, and have sufficient data for comparisons.

NRC Health employs several business rules to ensure that our comparative data is robust, reliable, and useable for clients. Below are the minimum requirements for Catalyst Advanced Reporting comparative data by question.

Benchmark Requirements:

- Made up of one year of data
- Questions must be used by at least five facilities
- Must have at least 1000 responses for the question
- For percentiles based comparisons* there must be at least 50 sample units using a question
- For percentile based comparisons* the 50 sample units must have a minimum of 30 responses

** Percentile based comparisons are those that use a percentile distribution to be determined, i.e. Top 10%, 75th Percentile, 90th Percentile*

NRC Health aims to keep survey content consistent from organization to organization, so that benchmarking is available, reliable, and relevant. Therefore, survey tools must be used as-is, without alternation; however clients can add a limited number of custom questions. In addition, there is an extensive set of standard question modules, which covers commonly requested topics that clients can choose to add to the core survey questionnaire. These initiatives will ensure that NRC Health clients are using the same survey tools and allow service teams to better predict what questions will have comparisons and better focus improvement initiatives. Ultimately, it will provide a better survey tool for all clients, which will translate into more reliable and actionable information.

Percentile Ranks

To calculate a **percentile rank**, NRC Health begins by sorting a group of scores from lowest to highest and dividing those scores into 100 equal portions. Where the score falls on that list determines the percentile rank. One way to interpret a percentile rank is if a score places a unit at the 55th percentile, than that unit's sample of respondents is better than 55 percent of the units NRC Health measures with that question.

Top 10% versus 90th Percentile

NRC Health provides both Top % and Percentile benchmarks. Top % scores are the Average Score of all the units that are at the corresponding percentile. The Percentile is the score one would need to achieve to be that percentile. For example, the Top 10% benchmark is calculated by determining the percentile rankings as described above, then averaging the scores of all the units at or above the 90th percentile, thus the score of the Top 10% of the benchmark group.

Top Box Scores vs. Percentile Ranks

NRC Health **recommends clients focus on top box scores** to track improvement initiatives. Percentile ranks and other benchmarks are useful reference points; however, top box scores measure specific behaviors and are best used to track measureable improvement. Moreover, the comparison database included in Catalyst Advanced Reporting is recalculated every quarter. Considering improvement initiatives almost always last longer than any given quarter; thus focusing on achieving a specific percentile rank creates a shifting target.

Scorecard Color Codes

Scorecard reports use **color codes** to indicate how a unit's score compares to selected comparison information.

Statistical significance is an expression used to describe the level of difference between two scores. If two scores are statistically significantly different that means the difference between the scores is likely not due to random chance.

On Scorecard reports, if a score is coded as Red it means that the score of the unit is statistically significantly lower than that of the benchmark group. If a score is coded as Green it means that the score of the unit is statistically significantly higher than that of the benchmark group.

Some questions may appear without any color coding, in these situations, there is not enough data to definitively say there is a statistically significant difference.

NRC Health uses a one sample 2-tailed *Z* test, adjusted for use with proportions, to test for statistically significant differences on Scorecard reports at the 95% confidence level. Below are the two equations NRC Health uses: one that calculates the specific *Z* score and the other is an additional equation that can be used to calculate whether a score should be coded as Red or Green.

To Calculate *Z* Score:

$$z = \left| \sqrt{p_s} \right|$$

Where:
p_u = population proportion (non-sample unit score the sample unit is being compared to)
p_s = sample population proportion
n = sample unit n size
Z = standard deviations (Z Score)

To determine if a Score should be coded Red or Green:

$$\text{Result} = |p_u - p_s| \perp 1.646 * \sqrt{p_u(1}$$

Where:
p_u = population proportion (non-sample unit score the sample unit is being compared to)
p_s = sample population proportion
n = sample unit n size

Interpretation:
If the result is a negative number, the difference is not statistically significant (no color coding). If the result is a positive number and the score is above the benchmark it's statistically significantly better than the benchmark (coded Green). If the result is a positive number and the score is below the benchmark it's statistically significantly lower than the benchmark (coded Red).

Key Drivers

On Scorecard reports, the **Key Driver** section identifies the questions that have the strongest relationship to the overall measure question selected for the report. These questions are the questions best suited for improvement initiatives because of their close relationship to the overall experience measures. NRC Health uses the **correlation coefficient** (r) to rank the question items in order from the strongest to weakest relationship with the overall measure question to create a comparative list.

For Scorecard and Priority Matrix reports, NRC Health uses the Pearson's Product Moment correlation coefficient:

$$r = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}}$$

Where:

$\overline{X}$ = Variable 1 mean
X_i = Variable 1 i^{th} sample response

$\overline{Y}$ = Variable 2 mean
Y_i = Variable 2 i^{th} sample response
n = Sample size
r = Correlation coefficient

The Key Driver section ranks the question items, in order of decreasing correlation coefficient creating a comparative list. Clients can view the correlation coefficient statistic by hovering the mouse cursor over the most recent time period on the Scorecard or by using the 'Export Data' function.
Correlations Coefficients can range from -1.000 to 1.000. A positive correlation coefficient indicates a relationship where a score increase for the question would predict a score increase for the overall measure. A negative correlation coefficient indicates a relationship where a score increase for a question would predict a score decrease for the overall measure.
Correlations do not indicate a causal relationship, in other words there is not a 1 to 1 association between questions and the overall measure. The overall measure is not expected to improve at the same rate as the Key Driver, however the closer the correlation coefficient gets to 1,the stronger the relationship.
There is much debate over what constitutes a moderate or strong correlation and there is no definitive answer. For the purposes of interpreting a Stoplight report NRC Health Corporation suggests adopting the following rule of thumb:

Interpretation Guide	
Correlation Coefficient (- or +)	Strength of Correlation
0.50 to 0.99 0.30 to 0.50 0.10 to 0.30 0.01 to 0.10	Strong correlation Moderate correlation Weak correlation None or very weak

Dimension Calculation

There are multiple ways to calculate dimension scores, and NRC Health offers two options in Advanced Reporting. A Dimension Calculation selector is available on Dashboard, Scorecard, Priority Matrix and Query Builder. The options for the selector are “NRC Standard – Respondent level” and “Modified – Question Level”, with the default being NRC Standard.

The NRC Health standard is to calculate a dimension score for each patient, and then average those scores to obtain the dimension score. In order to mitigate any potential biases related to unanswered questions, the “NRC Standard – Respondent Level” dimension calculation only includes patients who have valid responses to at least half (50%) of the questions in a given dimension.

The alternative method NRC Health offers calculates the positive score for each question in the dimension and averages those to obtain the dimension score. The “Modified – Question Level” gives the percent of all responses that were positive.

Neither calculation is more accurate than the other – they are simply different. Where the majority of respondents answered all of the questions in a dimension, these two approaches are very similar in their results. The benefits of the NRC Standard methodology becomes most evident where there are many respondents choosing not to answer or being directed to skip questions. The logic used in the NRC Standard protects from potential bias, by preventing patients with responses to less than half the questions, from contributing to dimension scores.

Things to keep in mind:

- All NRC benchmarks use the NRC Standard calculation. CMS Benchmarks use publically reported data.
- The n-size in both methods represents the number of respondents included in the dimension calculation.

Example Calculation:

NRC Standard - Respondent Level				
	Question 1	Question 2	Question 3	Respondent Dimension Score
Respondent 1	X	X		XX
Respondent 2	X		X	XX
Respondent 3	X	X	X	XXX
				Average
Dimension n-size = 3				

Modified - Question Level				
	Question 1	Question 2	Question 3	
Respondent 1	X	X		
Respondent 2	X		X	
Respondent 3	X	X	X	
Question Positive Score	XXX	XX	XX	Average
Dimension n-size = 3				