HealthPartners®

HealthPartners Technical Documentation

Total Cost of Care
Intra-Class Correlation Analysis

Using Optum Symmetry Episode Risk Groups (ERG)

Purpose

Determine the Intra-Class Correlation reliability of the Total Cost of Care (TCI) measure using the Optum Symmetry Episode Risk Groups (ERG).

Table of Contents

HealthPartners®

Overview of Analysis

Total Cost of Care (TCI) is a measure of a provider's effectiveness of managing their primary care attributed population across the care continuum that can be used to profile and compare providers to one another. Effective provider profiling measures should be able to highlight real differences in performance while being minimally affected by random variation due to patient mix. The extent to which a measure accomplishes this can be quantified, and is termed its "reliability." The purpose of this paper is to measure the reliability of the TCI measure. One limitation to reliability is it distinguishes between measurement entities e.g., providers, but it does not quantify the accuracy or consistency of the measure. Rather, accuracy and consistency of the measure are addressed in the companion bootstrap and validity analyses.

Overall Conclusions

The HealthPartners' Total Cost of Care and Resource Use measures are reliable for clinic groups, clinics and physicians at the N size of 600. The results also indicate that the measures produce reliable results at lower N sizes, however the HealthPartners NQF endorsed Total Cost of Care and Resource Use measures using ACGs requires a minimum of 600 members. The 600 member threshold aligns with volumes necessary to measure quality and experience, which when combined with total cost of care supports triple aim reporting.

A measure can have good reliability for three reasons:

- There is little variation (error) within individual providers.
- There is comparatively high variability between providers.
- There are more members in the provider group.

Reliability Results – ERG TCI

The chart indicates the percentage of provider groups that have a reliability score above 0.7, the conventionally accepted threshold to be considered a reliable measure.

2011

	Clinic Groups			Clinic			Physician		
Person Count	Average Reliability	Group Count	Reliability > 0.7	Average Reliability	Clinic Count	Reliability > 0.7	Physician Reliability	Physician Count	Reliability > 0.7
2-10	0.271	62	2%	0.258	235	2%	0.261	2161	2%
10-20	0.474	26	15%	0.428	102	13%	0.450	837	13%
20-50	0.563	53	42%	0.562	161	32%	0.594	1336	40%
50-100	0.711	40	63%	0.702	106	58%	0.735	1205	68%
100-200	0.837	20	90%	0.764	100	73%	0.804	1052	84%
200-300	0.799	12	75%	0.865	50	92%	0.870	218	94%
300-400	0.920	7	100%	0.895	39	97%	0.915	81	99%
400-500	0.890	6	100%	0.924	23	100%	0.927	45	100%
500-600	0.941	5	100%	0.943	24	100%	0.920	22	95%
600-1000	0.943	11	100%	0.954	84	100%	0.962	12	100%
1000-2000	0.959	15	100%	0.970	58	100%	~	~	~
2000-5000	0.982	21	100%	0.983	45	100%	~	~	~
5000-10000	0.994	8	100%	0.994	9	100%	~	~	~
>10000	0.998	6	100%	~	~	~	~	~	~

HealthPartners®

2010

	Clinic Groups			Clinic			Physician		
Person Count	Average Reliability	Group Count	Reliability > 0.7	Average Reliability	Clinic Count	Reliability > 0.7	Physician Reliability	Physician Count	Reliability > 0.7
2-10	0.233	49	0%	0.221	197	1%	0.257	2020	2%
10-20	0.433	36	6%	0.396	101	2%	0.442	844	13%
20-50	0.591	51	37%	0.519	138	24%	0.585	1336	36%
50-100	0.693	30	67%	0.627	93	47%	0.714	1199	66%
100-200	0.771	23	78%	0.737	96	74%	0.804	977	85%
200-300	0.839	14	93%	0.856	61	97%	0.874	239	95%
300-400	0.814	7	86%	0.875	38	95%	0.908	66	97%
400-500	0.889	5	100%	0.917	24	100%	0.934	51	98%
500-600	0.941	2	100%	0.911	26	100%	0.942	24	100%
600-1000	0.932	11	100%	0.932	77	100%	0.958	23	100%
1000-2000	0.940	20	100%	0.954	68	100%	0.977	1	100%
2000-5000	0.983	18	100%	0.982	39	100%	0.989	1	100%
5000-10000	0.992	7	100%	0.993	10	100%	~	~	~
>10000	0.997	6	100%	~	~	~	~	~	~

2009

	Clinic Groups			Clinic			Physician		
Person Count	Average Reliability	Group Count	Reliability > 0.7	Average Reliability	Clinic Count	Reliability > 0.7	Physician Reliability	Physician Count	Reliability > 0.7
2-10	0.271	62	2%	0.258	235	2%	0.261	2161	2%
10-20	0.474	26	15%	0.428	102	13%	0.450	837	13%
20-50	0.563	53	42%	0.562	161	32%	0.594	1336	40%
50-100	0.711	40	63%	0.702	106	58%	0.735	1205	68%
100-200	0.837	20	90%	0.764	100	73%	0.804	1052	84%
200-300	0.799	12	75%	0.865	50	92%	0.870	218	94%
300-400	0.920	7	100%	0.895	39	97%	0.915	81	99%
400-500	0.890	6	100%	0.924	23	100%	0.927	45	100%
500-600	0.941	5	100%	0.943	24	100%	0.920	22	95%
600-1000	0.943	11	100%	0.954	84	100%	0.962	12	100%
1000-2000	0.959	15	100%	0.970	58	100%	~	~	~
2000-5000	0.982	21	100%	0.983	45	100%	~	~	~
5000-10000	0.994	8	100%	0.994	9	100%	~	~	~
>10000	0.998	6	100%	~	~	~	~	~	~

Methodology

The Reliability Measure of Provider Performance Scores

The reliability of a performance measure is its ability to consistently distinguish one provider from another. Providers' performance scores vary from one another due to two distinct types of variation. First, there is variation that reflects systemic differences between providers' costs, such as treatment models, the effectiveness of preventative treatments, administrative and overhead costs, etc. These are considered "real" differences, in the sense that this type of variation is controllable and what a performance measure is intended to capture. The second type of variation is the result of random fluctuations in the patient population in conjunction with their treatment needs. This is considered "error." Risk adjustment methods, such as the Optum Symmetry Episode Risk Group (ERG) methodology, have been developed to minimize error; however, there are no existing methods

that can entirely remove the effect of error on provider performance scores. Even after risk adjustment, there is still a degree of random error due to a provider's patient mix. The extent of this error affects the quality of a measure's results. Therefore, a metric such as reliability must be used as method to determine the potential usefulness of providers' performance score.

Reliability is a metric that measures the extent to which real differences in performance effect provider profile scores. It is developed individually for each provider group and is defined as the ratio of variability between providers and the total variability (variability between + variability due to error).

$$(1)\ \text{Reliability} = \frac{\text{Variability between providers}}{\text{Variability between providers} + \text{Variability due to error}}$$

The variability is estimated as the variance (σ^2), or the squared standard deviation (σ). The variability due to error (or average error variance) can be estimated as the average variance within the provider group. The between provider group variability was determined by calculating the variance the provider group level with the variability due to error factored out.

$$(2)\ \text{Reliability} = \frac{\sigma^2_{\text{between provider groups}}}{\sigma^2_{\text{between provider groups}} + \frac{\sigma^2_{\text{within provider group}}}{n}}$$

Where n is the number of people. Reliability scores range between 0 and 1. A performance measure that primarily captures error will be close to 0 and low reliability, while a measure that captures real differences between providers has high reliability and will approach 1. Statistical convention is that a measure is considered reliable if scores greater than 0.7.

The Intra-Class Correlation Reliability Applied to TCOC Data

In the case of TCI reliability, the between provider group variance would be the variance amongst all provider TCI's or the risk adjusted cost per member per month (PMPM) with the within group variance factored out. The $\frac{\sigma^2_{\text{within provider group}}}{n}$ component is the average error variance within the provider group, and is the variance of subgroups' scores divided by the number of subgroups within the provider group. This is also referred to as Mean Squared Error (MSE), which is an estimate of the provider group score variance that would be observable with new selections of data (resampling), or patient churn. Any subdivision within the provider group can be used to calculate the average error variance, such as episodes, patients, physicians, clinics, etc. The subgroups used for this study were individual patients. In mathematical notation, a TCI reliability score would be the following:

$$(3)\ \text{Reliability} = \frac{\sigma^2_{\text{provider TCI}}\ *}{\sigma^2_{\text{provider TCI}}\ * + \frac{\sigma^2_{\text{patient TCIs within the provider}}}{\text{number of patients within the provider}}}$$

* It is important to note that $\sigma^2_{\text{provider TCI}}$ is not simply the variance of the providers' TCIs. It is actually the variance between the providers with the error variance factored out. This was achieved using a model developed via SAS's mixed procedure.

HealthPartners®

Reliability Provider Group Definitions

A reliability score can be calculated for groups of any size, such as a single physician, clinics, or systems of clinics (i.e. provider groups). The reliability score is a function of provider group sample size, the sampling error, and the differences between other provider groups. A greater difference between providers would increase the reliability score as the reliability score measures the confidence an analyst can have in the contrast observed between providers. It then follows that the between group variance should be calculated with provider groups that are typically used as points of comparison for analysis, such as physician vs. all other physicians, clinics vs. all other clinics. Systems of clinics may vary greatly in size as some "systems" are little more than a single clinic and their reliability score will be impacted by differences amongst much larger systems. However, if small systems of clinics are typically compared to large systems of clinics, it is important to include them all in reliability calculation, because the reliability of that specific comparison is the end goal. One should be aware that if larger systems were broken down into smaller systems, the reliability score for all systems would be affected as there are now new points of comparison that would alter the between group variance.

For the purposes of this project we created reliability scores at three different provider group levels.

- Physicians
- Clinics
- Clinic Groups (groups of clinics defined by their business relationship with one another)

We chose these particular levels because they are either typical levels of analysis within the health care community or within the HealthPartners organization.

TCI Data Criteria

The data was limited down to claims that occurred between Jan 1, 2009 and December 31, 2011 with 3 months of run out. We only included members that were continuously enrolled for at least 9 months, between the ages of 1 and 65, enrolled in a commercial product, and attributed to a primary care provider. The attribution methodology is described in detail in the following paper: http://www.healthpartners.com/ucm/groups/public/@hp/@public/documents/documents/cntrb_031064.pdf.