



HealthPartners®

HealthPartners Technical Documentation

Total Cost of Care and Total Resource Use Reliability Metric Analysis

Purpose

The HealthPartners Total Cost of Care (TCOC) measures, Total Cost Index (TCI) and Total Resource Use Index (RUI) are powerful analytical tools that can be used to evaluate the affordability of health care for a given population, such as health care providers, employer groups, geographic areas, etc. In this study, we will be focusing on the application of the TCOC measures to health care providers. Provider profile scores (TCI and RUI) are generated by risk adjusting the total costs and resource use of a provider's members. The scores can be influenced by either the actual performance of a provider or random variation due to a provider's patient mix and their treatment needs. The extent to which provider profile scores reflect real differences in performance is termed "reliability." More specifically, reliability is defined as the ratio of variability between providers and the total variability (variability between providers + variability within the provider). In this study, reliability scores were developed for both the TCI and RUI measures for individual provider groups at the physician, clinic, and clinic group levels.

The results indicate the HealthPartners TCOC measures are reliable for physicians, clinics, and clinic groups with more than 500 patients. The RUI reliability scores were higher than TCI reliability because cost scores include additional variation due to price. The highest reliability scores for both cost and resource use measures were found within clinics and physicians, with clinic groups demonstrating the smallest reliability scores. Within a given provider level, the number of patients attributed a provider is directly correlated with a higher reliability score. This study indicates that TCOC measures are able to highlight real differences in provider performance and would serve as an effective tool to improve health care affordability.

Introduction

TCOC Overview

The ballooning cost of health care in the United States has heightened the need to measure and compare health care affordability across multiple providers. While an industry standard has yet to be established, many organizations have begun to experiment with the measurement of costs and resource use. The various approaches to these measures have been dubbed "Total Cost of Care" (TCOC) measures because of their emphasis on measuring the aggregation of all costs associated with a patient across multiple places of service. HealthPartners has developed two population based indices, the Total Cost of Care Index (TCI) and the Total Resource Use Index (RUI), which were endorsed by the National Quality Forum (NQF) in both 2012 and 2017.

HealthPartners' TCI measures are created by summing the total costs for a provider's attributed members across all services (inpatient, outpatient, professional, and pharmaceutical) to create an average total cost per member per month (PMPM) for each provider. Each PMPM is then risk adjusted by dividing it by the provider specific Adjusted Clinical Group (ACG) retrospective risk score, which is weighted average of the member level ACG scores within a provider group. ACG risk scores, developed by Johns Hopkins University, reflect the relative expected resource use of a given member. Risk adjusted PMPMs are finally converted to TCIs by dividing them by the adjusted PMPM for the peer group (benchmark) as shown in equation (2).

$$(1) \text{ Adjusted Cost PMPM} = \frac{\frac{\text{Medical cost}}{\text{member months}} + \frac{\text{Pharmacy cost}}{\text{pharmacy member months}}}{\text{ACG Risk Score}}$$

$$(2) \text{ TCI} = \frac{\text{Provider Adjusted Cost PMPM}}{\text{Peer Group Adjusted Cost PMPM}} \quad (3) \text{ RUI} = \frac{\text{Provider ACG Adjusted Resource Use PMPM}}{\text{Peer Group ACG Adjusted Resource Use PMPM}}$$

The RUI measure is developed in a similar fashion. However, rather than comparing costs to a benchmark, the RUI compares Total Care Relative Resource Values (TCRRV) which measure resources consumed by medical procedures, services or products that are independent of price. Both the TCI and RUI measures include all services associated with treating patients including professional, facility inpatient and outpatient, pharmacy, lab, radiology, ancillary and behavioral health services. Further information on the Total Cost of Care methodology can be found on the HealthPartners website: www.healthpartners.com/tcoc.

Reliability Overview

The TCI and RUI measures are ways of profiling and comparing 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 and RUI measures.

Methods

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 Johns Hopkins ACG methodology used by HealthPartners, 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).

$$(4) \text{ Reliability} = \frac{\text{Variability between providers}}{\text{Variability between providers} + \text{Variability due to error}}$$

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.

$$(5) \text{ Reliability} = \frac{\sigma_{\text{between provider groups}}^2}{\sigma_{\text{between provider groups}}^2 + \frac{\sigma_{\text{within provider group}}^2}{n}}$$

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 Reliability Measure 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

$\frac{\sigma_{\text{within provider group}}^2}{n}$ 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:

$$(6) \text{ Reliability} = \frac{\sigma_{\text{provider TCI}}^2 *}{\sigma_{\text{provider TCI}}^2 * + \frac{\sigma_{\text{patient TCIs within the provider}}^2}{\text{number of patients within the provider}}}$$

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

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/RUI Data Criteria

The data was limited down to claims that occurred between January 1, 2015 and December 31, 2015 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.

Methodology

Reliability Scores by Level of Analysis

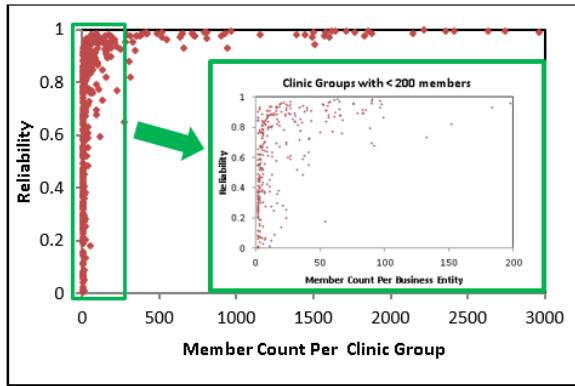
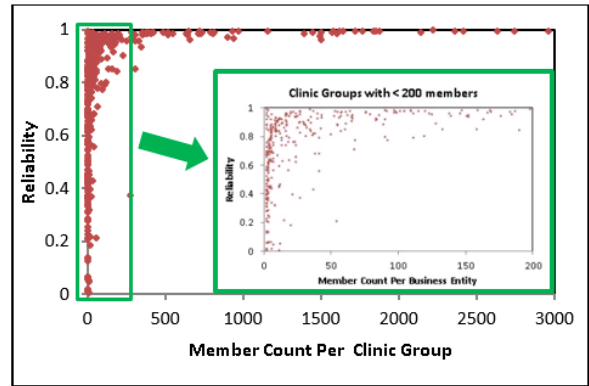
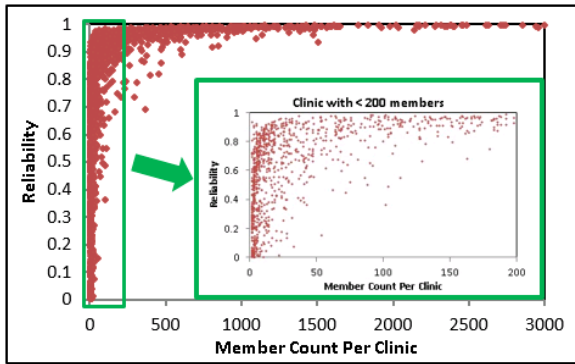
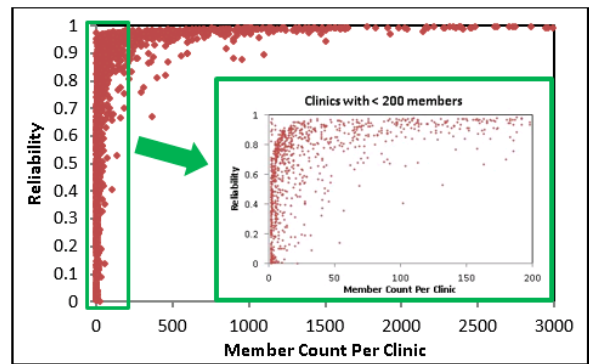
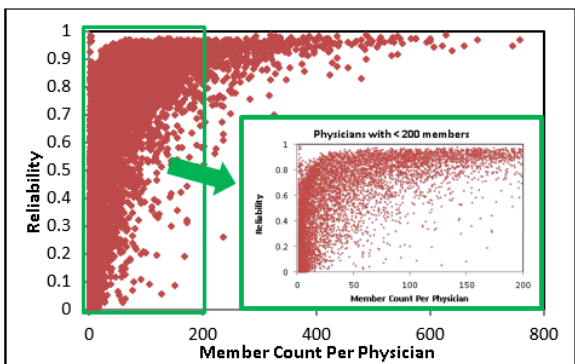
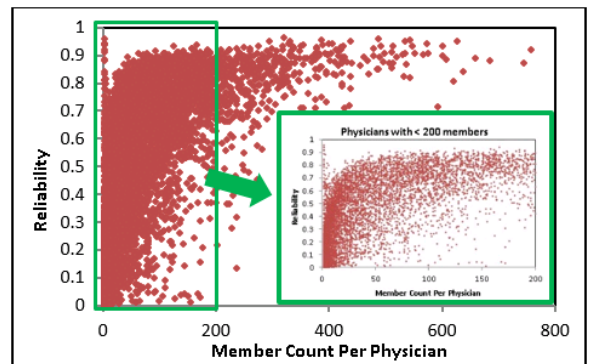
Table 1 summarizes the reliability results for three different levels of analysis: 1) physicians, 2) clinics, and 3) clinic groups. Each level was stratified by member count, and then the reliability scores for each level were averaged to produce a final reliability result for a given strata level. The leftmost column in the table indicates the strata (member count range) of the groups contributing to the average reliability scores. The three levels of analysis are listed across the top of the table, and nested beneath them are their reliability scores for a given strata and the number of providers that fall within the given strata. For example, there are 3328 physicians, 403 clinics, and 162 clinic groups with 2-10 attributed members. The "Reliability > 0.7" column indicates the percent of providers within a given strata that had reliability scores greater than the 0.7, a commonly used threshold. The red dotted line indicates the 600 person count level used in the TCI and RUI NQF endorsement, and the green shaded regions identify the stratum where at least 90% of the providers have a reliability score greater than the 0.7 threshold. Table 2 has the same organization as Table 1 but displays the reliability results for the RUI, rather than TCI.

Table 1: Reliability Scores for the TCI Measure

Person Count	Clinic Groups			Clinic			Physician		
	Average Reliability	Group Count	Reliability > 0.7	Average Reliability	Clinic Count	Reliability > 0.7	Average Reliability	Physician Count	Reliability > 0.7
2-10	0.501	162	27%	0.450	403	22%	0.290	3328	3%
10-20	0.678	44	59%	0.644	162	56%	0.504	984	29%
20-50	0.766	48	71%	0.702	156	60%	0.635	1430	47%
50-100	0.879	39	92%	0.838	110	86%	0.765	1355	74%
100-200	0.917	33	97%	0.903	118	96%	0.850	1244	91%
200-300	0.891	8	88%	0.929	55	100%	0.892	280	96%
300-400	0.948	11	100%	0.944	39	97%	0.934	105	99%
400-500	0.988	4	100%	0.957	40	100%	0.936	45	98%
500-600	0.975	3	100%	0.969	30	100%	0.961	19	100%
600-1000	0.976	15	100%	0.980	98	100%	0.958	10	100%
1000-5000	0.989	35	100%	0.991	127	100%	0.985	2	100%
5000-10000	0.998	9	100%	0.997	8	100%	~	~	~
>10000	0.999	10	100%	0.99878	1	100%	~	~	~

Table 2: Reliability Scores for the RUI Measure

Member Count	Clinic Groups			Clinic			Physician		
	Average Reliability	Group Count	Reliability > 0.7	Average Reliability	Clinic Count	Reliability > 0.7	Average Reliability	Physician Count	Reliability > 0.7
2-10	0.603	162	42%	0.480	403	27%	0.192	3328	1%
10-20	0.758	44	75%	0.697	162	64%	0.381	984	2%
20-50	0.828	48	81%	0.746	156	69%	0.493	1430	19%
50-100	0.903	39	97%	0.862	110	92%	0.612	1355	41%
100-200	0.942	33	100%	0.913	118	96%	0.709	1244	63%
200-300	0.888	8	88%	0.939	55	100%	0.776	280	80%
300-400	0.964	11	100%	0.946	39	97%	0.846	105	95%
400-500	0.990	4	100%	0.958	40	100%	0.848	45	91%
500-600	0.987	3	100%	0.966	30	100%	0.888	19	100%
600-1000	0.985	15	100%	0.974	98	99%	0.888	10	100%
1000-5000	0.993	35	100%	0.990	127	100%	0.961	2	100%
5000-10000	0.999	9	100%	0.997	8	100%	~	~	~
>10000	0.999	10	100%	0.99839	1	100%	~	~	~

Figure 1: Clinic Groups TCI Reliability

Figure 2: Clinic Groups RUI Reliability

Figure 3: Clinics TCI Reliability

Figure 4: Clinics RUI Reliability

Figure 5: Physician TCI Reliability

Figure 6: Physician RUI Reliability


Individual Provider Scores

Figures 1-6 show the relationship between reliability score and size of each level of analysis. Each data point represents an individual provider. The inset graph in each figure expands out the 0-200 member count range in order to provide a view with greater resolution of the data points.

Discussion

Average Reliability Scores by Level of Analysis

The results in Table 1 and 2 show the average reliability score given the member count within each level of analysis. A reliability estimate of at least 0.7 has been established as the industry standard threshold of what is considered a reliable measure for groups [3]. Both tables demonstrate that at all three levels of analysis, the TCI and RUI measures are more reliable as the membership count increases. At the NQF-endorsed level of 600 members, nearly all providers were above the 0.7 threshold. The results show that even below the 600 member level, the measure can still be reliable, though measuring small N-size groups will yield results that are more subject to natural variation. HealthPartners recommends using the NQF-endorsed 600-patient minimum, however, organizations that wish to support improvement efforts by measuring groups below 600 can do so, but must interpret results them with some discretion.

A measure can have strong reliability for three reasons: (1) there is little variation (error) within individual providers, (2) there is comparatively high variability between providers, or (3) there are more members in the provider group. The results in Table 1 confirm that as the member count increases so does the reliability of their measures. Mathematically, this is supported by equations (5) and (6), where one can see that as n increases towards infinity, the result rises to a perfect reliability score of 1.

Individual Provider Scores

Figures 1-6 show the individual reliability scores for each provider for both the TCI and RUI measures. It is clear that from these figures there are indeed many providers that fall below the 0.70, though those providers for whom the TCI and RUI measures are unreliable typically have less than 600 members. As the member counts increase, the reliability scores approach 1. There is a greater density of data points near the x axis for the TCI figures as compared to the RUI figures because the RUI measure is less variable between provider groups due to lack of price, thus lowering the reliability.

Conclusion

This study presents findings that demonstrate that the HealthPartners TCI and RUI measures are reliable methods to distinguish providers from one another at the physician, clinic, or clinic group levels. The results show that the measures are reliable above the NQF-endorsed 600 member level, which aligns with our other studies (using the bootstrapping methodology) for the NQF endorsement process. Additionally, the results demonstrate that the measures are reliable below this level, though it is important to keep in mind that when measuring at lower N-sizes, there will be a reduction in measure reliability. A TCOC and Resource Use Low Volume Measure Stabilization [module](#) is included in the TCOC grouper and shared on the HealthPartners' TCOC website in order to assist organizations with usability and reporting needs at smaller population sizes. HealthPartners recommends using the NQF-endorsed 600-patient minimum threshold for purposes of broad public reporting while certain business applications such as practice management and internal improvement efforts may need the option to report at smaller population sizes. In summary, the HealthPartners' TCI and RUI measures can be used to highlight real differences in provider performance and results can be used to identify opportunities to improve healthcare affordability.

Contact

Questions Contact: TCOCMeasurement@HealthPartners.com

Works Cited

Adams, John L., Ateev Mehrotra, J. William Thomas, and Elizabeth A. McGlynn. "Physician Cost Profiling — Reliability and Risk of Misclassification Technical Appendix." *New England Journal of Medicine* 362.11 (2010): 13-22. Print.

Adams, John L., Ph. D, Ateev Mehrotra, M.D., M.P.H, J. William Thomas, Ph. D, and Elizabeth A. McGlynn, Ph. D. "Physician Cost Profiling — Reliability and Risk of Misclassification." *New England Journal of Medicine* 362.11 (2010): 1014-021. Print.

Adams, John L., Ph.D. "The Reliability of Provider Profiling, A Tutorial." RAND Corporation (2009). Print.

"HealthPartners Total Cost of Care and Resource Use." HealthPartners Total Cost of Care and Resource Use White Paper (2012): 1-12. Web. 19 Aug. 2012.
https://www.healthpartners.com/ucm/groups/public/@hp/@public/documents/documents/dev_057649.pdf.

Singer, Judith D. "Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models." *Journal of Educational and Behavioral Statistics* 24.4 (1998): 323-55. Print.

Total Cost of Care Population Based PMPM Index Reliability Analysis. Tech. HealthPartners Health Informatics, July 2011. Web. 19 Sept. 2012. <https://www.healthpartners.com/public/tcoc/>.

Yang, Min. "A Review of Random Effects Modeling in SAS." Institute of Education, University of London 8.2 (2003). Print.