

## Psychometric Evaluation of the Care Transition Measure in TRACE-CORE: Do We Need a Better Measure?

Milena D. Anatchkova, PhD; Constance M. Barysaukas, MS; Rebecca L. Kinney, MPH; Catarina I. Kiefe, PhD, MD; Arlene S. Ash, PhD; Lisa Lombardini, BS; Jeroan J. Allison, MD

**Background**—The quality of transitional care is associated with important health outcomes such as rehospitalization and costs. The widely used Care Transitions Measure (CTM-15) was developed with a classic test theory approach; its short version (CTM-3) was included in the CAHPS Hospital Survey. We conducted a psychometric evaluation of both measures and explored whether item response theory (IRT) could produce a more precise measure.

**Methods and Results**—As part of the Transitions, Risks, and Actions in Coronary Events Center for Outcomes Research and Education, 1545 participants were interviewed during an acute coronary syndrome hospitalization, providing information on general health status (Short Form-36), CTM-15, health utilization, and care process questions at 1 month postdischarge. We used classic and IRT analyses and compared the measurement precision of CTM-15-, CTM-3-, and CTM-IRT-based score using relative validity. Participants were 79% non-Hispanic white and 67% male, with an average age of 62 years. The CTM-15 had good internal consistency (Cronbach's  $\alpha=0.95$ ) but demonstrated acquiescence bias (8.7% participants responded "Strongly agree" and 19% responded "Agree" to all items) and limited score variability. These problems were more pronounced for the CTM-3. The CTM-15 differentiated between patient groups defined by self-reported health status, health care utilization, and care transition process indicators. Differences between groups were small (2 to 3 points). There was no gain in measurement precision from IRT scoring. The CTM-3 was not significantly lower for patients reporting rehospitalization or emergency department visits.

**Conclusion**—We identified psychometric challenges of the CTM, which may limit its value in research and practice. These results are in line with emerging evidence of gaps in the validity of the measure. (*J Am Heart Assoc.* 2014;3:e001053 doi: 10.1161/JAHA.114.001053)

**Key Words:** acute coronary syndromes • care transitions measure • validity • IRT scoring

Transitional care refers to actions to ensure the coordination and continuity of care for patients as they transfer between locations of care. Care transitions are particularly important for the 145 million Americans who live

with chronic illness and receive care from multiple providers in various settings, often with noncommunicating medical record systems.<sup>1</sup> Poor transitions can fragment care, resulting in conflicting recommendations, increased medical error and duplication, and inadequate information to both patients and caregivers.<sup>2</sup> An effective measure of care transition quality is a key tool for reducing care fragmentation.

The most widely used measure of care transition quality is the Care Transitions Measure (CTM-15).<sup>2</sup> The 4 CTM domains, derived from patient focus groups, are (1) Information Transfer, (2) Patient and Caregiver Preparation, (3) Support for Self-Management, and (4) Empowerment to Assert Preferences. A 4-factor structure was retained for the final measure using confirmatory factor analysis. The labels for the retained domains differ a little from these: (1) critical understanding, (2) preferences important, (3) management preparation, and (4) care plan. All 15 items in the CTM-15 use a 4-point scale with responses ranging from "Strongly Disagree" to "Strongly Agree." The items are scored by

From the Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA (M.D.A., R.L.K., C.I.K., A.S.A., L.L., J.J.A.); Department of Biostatistics and Computational Biology, Dana Farber Cancer Institute, Harvard Medical School, Boston, MA (C.M.B.).

This work has been selected for presentation as a top-rated abstract during the annual Quality of Care and Outcomes Research Scientific Sessions, on June 4, 2014 in Baltimore, MD.

**Correspondence to:** Milena D. Anatchkova, PhD, Department of Quantitative Health Sciences, University of Massachusetts Medical School, 55 Lake Ave N, ACCES7, Worcester, MA 01655. E-mail: milena.Anatchkova@umassmed.edu  
Received April 25, 2014; accepted April 29, 2014.

© 2014 The Authors. Published on behalf of the American Heart Association, Inc., by Wiley Blackwell. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

summing the responses (between 1 and 4) followed by linear transformation to a 0-to-100 range. The CTM-15 has been shown to discriminate between patients who did and did not have a subsequent emergency department visit or rehospitalization for their index condition and between health care facilities with different levels of system integration.<sup>3</sup> A 3-item version of the CTM was found to explain 88% of the variance in the full measure; the CTM-3 has been demonstrated to have the same ability to detect group differences as the longer version in earlier studies.<sup>4</sup> The short form includes items from 2 of the 4 originally identified domains (critical understanding and preferences are important). The developers argue that the information lost by using the CTM-3 measure is small compared with the reduction in response burden. The CTM-3 measures the extent to which the hospital staff accomplished essential care processes in preparing the patient for discharge and participating in posthospital self-care activities. Therefore, it is advised that survey data collection be administered between 48 hours and 6 weeks postdischarge. The CTM-3 was endorsed by the National Quality Forum and included in the Consumer Assessment of Healthcare Providers and Systems (CAHPS) Hospital Survey in 2010.

As with most self-report questionnaires currently used in health research, the CTM is scored using a classic sum score transformed on a 0-to-100 scale. While this approach is widely used, more-sophisticated approaches to scale scoring based on item response theory (IRT) are gaining popularity, since the IRT method uses more of the available information in each item. Thus, theoretically, IRT can produce a more discriminating score based on the same underlying data than a sum score.<sup>5</sup> In previous studies, comparisons of sum-versus IRT-based scoring of patient-reported measures have produced mixed results; that is, the benefits of IRT scoring differ and are really important for some scales but not others.<sup>5</sup> To the best of our knowledge, the benefits of IRT-based scoring for the CTM-15 have not been explored.

The development and initial validation of the CTM-15 and CTM-3 are well documented<sup>3,6</sup>; however, these studies were conducted on relatively small samples of less than 250 people. Given that the CTM-3 is now included among the core questions designed “to provide a standardized survey instrument and data collection methodology for measuring patients’ perspectives on hospital care” as part of a national quality of care initiative, examining CTM measure performance in an independent large-scale study is warranted.

The aim of this report is to conduct a psychometric analysis of CTM items and evaluate the dimensionality, internal consistency, and construct validity of CTM-15 and CTM-3. We also explore the effect of IRT-based scale scoring on measurement ability and compare the performance of the simple summation and IRT-based scoring approaches head to head.

## Methods

### Sample

Transitions, Risks, and Actions in Coronary Events Center for Outcomes Research and Education (TRACE-CORE) is a large, multiracial cohort of adult patients hospitalized with acute coronary syndromes (ACS) from 6 hospitals in Massachusetts and Georgia.<sup>7</sup> A total of 2300 patients were enrolled in the study and were approached for computer-assisted telephone follow-up interviews at 1, 3, 6, and 12 months post ACS hospitalization. Interviews focus on individual patient quality of life, rehospitalization, behavioral, and psychosocial characteristics. This study relies on data from the 1545 patients who completed both baseline and 1-month follow-up interviews, including the CTM-15. The study has been approved by the University of Massachusetts Medical School Institutional Review Board and all participants gave informed consent for participation.

### Measures

We used CTM-15 data from the 1-month follow-up interview and demographic characteristics and a self-reported health status measure collected at the baseline interview. Other data collected at 1 month include self-reported questions on rehospitalization, emergency department visits and symptom development postdischarge along with 3 independent questions evaluating transition quality (access to medical records, prescheduled follow-up appointments, and knowing who to contact if symptoms worsen).

### Analytic Plan

#### *Classic Test Theory Psychometric Analyses*

Classic test theory psychometric analyses were performed to provide an overall understanding of item and scale quality. We computed item-level statistics, including mean scores, proportions of response categories, and item-total correlations. Items with a negative or low (<0.2) item-total correlation have undesirable discriminating power and should be flagged for further study.<sup>8,9</sup> We also examined scale-level score distribution, ceiling and floor effects, and internal consistency (Cronbach’s  $\alpha$ ).<sup>10</sup>

#### *Dimensionality Analyses*

Dimensionality analyses were conducted to examine the factor structure of the CTM items in the TRACE-CORE cohort. The evidence of factor structure—how many latent factors are needed in explaining the variance of item responses—can be used to both justify the use of a single scale score as currently reported in the CTM and lay the groundwork for

satisfying the unidimensionality assumption for further advanced psychometric modeling (eg, IRT).<sup>11,12</sup> Confirmatory factor analyses were conducted in the framework of structural equation modeling using the software Mplus.<sup>13</sup> The polychoric correlation matrix and weighted least squares with adjustments for mean and variance estimation ordered categorical data were used.<sup>14</sup> In addition to a model replicating the CTM's reported 4-factor structure, we tested a unidimensional model and a bifactor model (items loading on 1 general factor and 4 secondary factors). The model–data fit was evaluated by examining the magnitude and patterns of item factor loadings, as well as the commonly used fit statistics compared with established thresholds (eg, the comparative fit index (CFI)  $\geq 0.95$ , root mean square error of approximation (RMSEA)  $\leq 0.08$ ).<sup>15</sup> The residual matrix and modification indices were used to detect any local item dependency, which can potentially arise when a subset of items have the same phrasing or format and thus are correlated with each other after the primary factor is controlled for.<sup>16</sup> It was hypothesized that the CTM items present essential unidimensionality with a dominant factor, namely, that the general factor has uniformly high loadings on all items and explains the majority of total item covariance.

### IRT Modeling

Modern measurement theory assumes that each response option has a specific relationship with the underlying construct, so a unique response can be selected for a different interval of the scale. If this assumption is violated, individual response options do not provide unique information and therefore can be collapsed before fitting an IRT model.<sup>17</sup> We inspected item characteristic curves for out-of-order or nondiscriminating response options by using the TestGraf program. Good items should have response choice categories with unequivocal and unique relationships to the latent trait appearing in rank order. Visual inspection of these item characteristic curve graphs provides insight into whether an item's response choice categories overlap and require rescaling (or collapsing across response choice categories) to achieve distinct categories.<sup>18</sup> To conduct the IRT analyses with the CTM, we used the polytomous generalized partial credit model<sup>19–22</sup> in PARSCALE.<sup>23</sup> The model fit for all items was evaluated using IRTFIT macro for SAS (SAS Institute).<sup>24</sup>

### Validity Analyses

Validity analyses were conducted to evaluate the ability of CTM scores (CTM-15, CTM-3, CTM-IRT) to discriminate patients of different transitional care quality and health status using previously reported variables. Specifically, the known-groups method of construct validation was used to compare groups using 1-way ANOVA and *t* tests.<sup>25</sup> Groups compared

were formed on the basis of demographic characteristics and patients' responses to non-CTM questions assessing care transition quality processes (eg, access to medical records at discharge, scheduled follow-up visits with a health care provider), health care use at 1 month postdischarge (emergency department visits, rehospitalization), health status scores, and symptom trajectories postdischarge. We compared the 3 scores by using relative validity coefficients.<sup>26</sup>

## Results

Study participants were predominantly white (79%) and male (67%), reporting a mean age of 62 years. Relatively few subjects reported having a college education or higher (27%) or annual household income greater than \$75 000 (29%). Self-reported health status was normally distributed across the 5 response categories from poor to excellent (Table 1).

**Table 1.** Baseline Characteristics of TRACE-CORE Participants

	% (n)	% (n)	P Value
	CTM-15 Completers (n=1545)	CTM-15 Noncompleters (n=750)	
Age, mean (SD) y	61.9 (11)	62.7 (12)	NS
Female	33.5 (518)	33.6 (252)	NS
Hispanic or Latino	3.0 (47)	4.2 (30)	NS
White	79.0 (1219)	72.8 (518)	<0.001
Education			
Less than high school	13.2 (204)	26.4 (190)	<0.001
High school graduate/ some college	59.8 (924)	54.0 (389)	
College or higher	26.9 (416)	19.7 (142)	
Family income, \$			
<34 999	39.8 (472)	51.2 (285)	<0.001
35 000 to 74 999	31.2 (370)	28.9 (161)	
75+	28.9 (343)	19.9 (111)	
Self-reported health status			
Excellent	6.0 (93)	3.6 (26)	<0.001
Very good	20.1 (310)	14.6 (105)	
Good	41.1 (633)	35.1 (252)	
Fair	22.4 (346)	30.9 (222)	
Poor	10.3 (159)	15.9 (114)	

CTM indicates care transitions measure; TRACE-CORE, Transitions, Risks, and Actions in Coronary Events Center for Outcomes Research and Education.

**Table 2.** CTM-15 Item Means, SDs, and Response Frequencies

	N*	Mean	SD	Frequencies						
				n	Strongly Disagree	Disagree	Agree	Strongly Agree	Don't Know	Refused
Before I left the hospital, the staff and I agreed about clear health goals for me and how these would be reached.	1471	3.2	0.7	n	25	115	861	470	74	1
				%	1.6	7.4	55.7	30.4	4.8	0.1
The hospital staff took my preferences and those of my family or caregiver into account in deciding what my health care needs would be when I left the hospital.	1435	3.2	0.7	n	27	137	835	436	105	6
				%	1.8	8.9	54.0	28.2	6.8	0.4
The hospital staff took my preferences and those of my family or caregiver into account in deciding where my health care needs would be met when I left the hospital.	1426	3.2	0.7	n	21	134	838	433	118	2
				%	1.4	8.7	54.2	28.0	7.6	0.1
When I left the hospital, I had all the information I needed to be able to take care of myself.	1532	3.3	0.6	n	21	70	864	577	13	1
				%	1.4	4.5	55.9	37.3	0.8	0.1
When I left the hospital, I clearly understood how to manage my health.	1529	3.2	0.6	n	20	108	911	490	16	1
				%	1.3	7.0	58.9	31.7	1.0	0.1
When I left the hospital, I clearly understood the warning signs and symptoms I should watch for to monitor my health condition.	1523	3.3	0.6	n	16	86	854	567	23	0
				%	1.0	5.6	55.2	36.7	1.5	0.0
When I left the hospital, I had a readable and easily understood written plan that described how all of my health care needs were going to be met.	1495	3.2	0.7	n	34	145	855	461	50	1
				%	2.2	9.4	55.3	29.8	3.2	0.1
When I left the hospital, I had a good understanding of my health condition and what makes it better or worse.	1518	3.3	0.6	n	17	84	902	515	28	0
				%	1.1	5.4	58.3	33.3	1.8	0.0
When I left the hospital, I had a good understanding of the things I was responsible for in managing my health.	1532	3.3	0.6	n	11	70	926	525	13	1
				%	0.7	4.5	59.9	34.0	0.8	0.1
When I left the hospital, I was confident that I knew what to do to manage my health.	1527	3.2	0.6	n	11	94	942	480	18	1
				%	0.7	6.1	60.9	31.1	1.2	0.1
When I left the hospital, I was confident I could actually do the things I needed to do to take care of my health.	1529	3.2	0.6	n	9	104	965	451	16	1
				%	0.6	6.7	62.4	29.2	1.0	0.1
When I left the hospital, I had a readable and easily understood written list of the appointments or tests I needed to complete within the next several weeks.	1510	3.3	0.6	n	20	72	880	538	35	1
				%	1.3	4.7	56.9	34.8	2.3	0.1
When I left the hospital, I clearly understood the purpose for taking each of my medications.	1533	3.2	0.6	n	12	105	918	498	12	1
				%	0.8	6.8	59.4	32.2	0.8	0.1
When I left the hospital, I clearly understood how to take each of my medications, including how much I should take and when.	1536	3.3	0.6	n	7	45	900	584	9	1
				%	0.5	2.9	58.2	37.8	0.6	0.1
When I left the hospital, I clearly understood the possible side effects of each of my medications.	1495	2.9	0.7	n	48	320	801	326	49	2
				%	3.1	20.7	51.8	21.1	3.2	0.1

CTM indicates care transitions measure.

\*Response rates varied by items. CTM-15 allows scoring with missing item data.

## Classic Test Theory Psychometric Analyses

Item-level review (Table 2) revealed sample means that were slightly higher (average item mean 3.2, range 2.9 to 3.3) than item means reported in earlier development and validation reports (average item mean 3.1, range 2.9 to 3.2),<sup>3,4</sup> yet, as reported in Table 3, these means were lower

than those reported in recent publications (eg, average item mean 3.7, range 3.6 to 3.9). In our study, the CTM-15 scale demonstrated good internal consistency (Cronbach's  $\alpha$  0.95) that is comparable to previous studies (Cronbach's  $\alpha$  0.90 to 0.95); however, the CTM-15 had a ceiling effect of 8.7%, substantially higher than the 1.1% reported in the original reports and similar to the 10% ceiling effect levels reported

**Table 3.** CTM-15 Summary Data: Development and Validation Studies

Study	Average Item Score	Item Score Range	CTM Mean	CTM SD	CTM Median	% Patients With Max Score	Chronbach's $\alpha$
TRACE-CORE (n=1545)	3.2	2.9 to 3.3	73.9	16.17	66.6*	8.7	0.92
Coleman et al <sup>3</sup> (n=200)	3.0	2.9 to 3.2	67.34	14.67	66.7	1.1	0.93
Coleman et al <sup>27</sup> (n=242)	3.1	2.9 to 3.3	70.1				
Parry et al <sup>4</sup> (n=225)			71.21	16.48			0.93 to 0.95
Shadmi et al <sup>6</sup> (Hebrew, n=217)	3.2	3.0 to 3.4	73.1	19.7	71.1	10.3	0.94
Shadmi et al <sup>6</sup> (Arabic, n=100)	3.5	3.2 to 3.7	81.8	16.5	86.1	10.9	0.9
Ryvicker et al <sup>28</sup> (n=495)			63.7	12.1	66.7 <sup>†</sup>		

CTM indicates care transitions measure; TRACE-CORE, Transitions, Risks, and Actions in Coronary Events Center for Outcomes Research and Education.

\*Of the sample, 26% had a value of 66.7.

<sup>†</sup>Of the sample, 40% had a value of 66.7.

in some later studies. None of our participants scored at the floor (at the lowest possible score); indeed, very few (5.5%) scored in the lower half of the scale's theoretical range (<50).

Among the 15 CTM item responses (Table 2), the most frequently selected was "Agree," followed by "Strongly Agree" and "Disagree." Nineteen percent (N=288) of all participants selected "Agree" as a response to all 15 questions, while

"Strongly disagree," "Don't know," and "Refused" were least selected, across all items.

Dimensionality analyses of the CTM in our study population supports use of the single score assessment and provided evidence that the scale is sufficiently unidimensional for IRT analysis. The bifactor model fit the data best (CFI=0.981, Tucker-Levis Index (TLI)=0.995, RMSEA=0.10) with the general factor having higher loadings on most of the items and

**Table 4.** Confirmatory Factor Analyses Model Results

	Unidimensional Model	Bifactor Model				
		General	Preferences Important	Management Preparation	Critical Understanding	Care Plan
Item 1	0.754	0.729	0.292			
Item 2	0.756	0.685	0.694			
Item 3	0.741	0.685	0.428			
Item 4	0.850	0.838		0.379		
Item 5	0.881	0.873		0.265		
Item 6	0.821	0.825		0.055		
Item 7	0.790	0.797				0.200
Item 8	0.874	0.888		0.109		
Item 9	0.947	0.955			-0.099	
Item 10	0.922	0.927			-0.071	
Item 11	0.821	0.825			0.057	
Item 12	0.700	0.706				0.121
Item 13	0.785	0.755			0.610	
Item 14	0.796	0.771			0.340	
Item 15	0.657	0.642			0.246	
Fit statistics						
CFI	0.931		0.981			
TLI	0.980		0.995			
RMSEA	0.19		0.10			

CFI indicates Comparative Fit Index; RMSEA, Root Mean Square Error of Approximation, TLI, Tucker-Lewis Index.



**Table 5.** IRT Parameters and Item Fit for CTM Items

Item	a	b1	b2	X2	Prob_X2
Item 1	2.02	-1.56	0.63	87.73	0.00
Item 2	2.15	-1.40	0.68	131.79	0.00
Item 3	2.29	-1.39	0.66	108.33	0.00
Item 4	3.09	-1.60	0.28	37.51	0.11
Item 5	4.07	-1.32	0.41	47.00	0.01
Item 6	3.28	-1.51	0.29	42.36	0.02
Item 7	2.82	-1.23	0.58	128.29	0.00
Item 8	3.53	-1.49	0.36	42.19	0.02
Item 9	4.17	-1.53	0.32	34.55	0.12
Item 10	3.73	-1.43	0.44	29.95	0.23
Item 11	2.71	-1.56	0.64	37.80	0.08
Item 12	2.54	0.40		30.52	0.05
Item 13	2.71	0.52		18.77	0.54
Item 14	2.79	0.29		34.95	0.01
Item 15	1.53	-0.93	1.17	216.45	0.00

CTM indicates care transitions measure; IRT, item response theory.

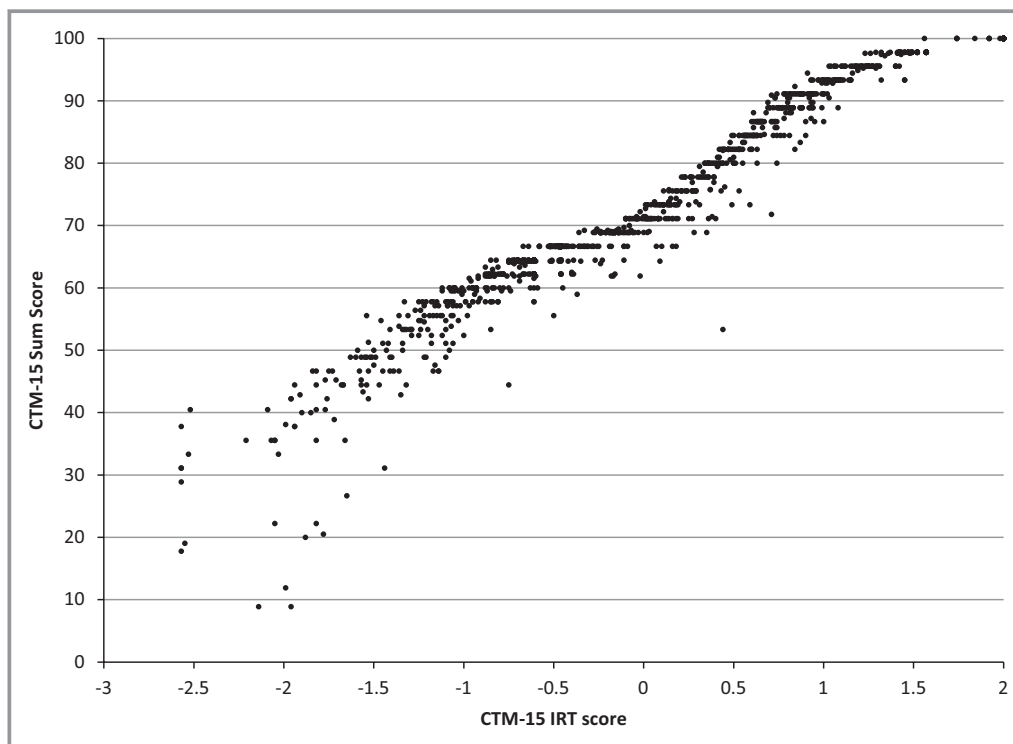
explaining most of the total item covariance (see Table 4). Review of the residual matrix did not detect any items with local dependency (had the lowest score), further supporting the essential unidimensionality of the data and the applicability of IRT methods.

## IRT Modeling

Review of the item characteristic curves revealed that all 15 items had some nondiscriminating response options that should be collapsed. For 3 items, we collapsed the 4 response categories into 2 (Agree versus Disagree), and for the rest of the items, the response options of “Disagree” and “Strongly Disagree” were collapsed into 1. The results of our IRT parameter estimates and item fit are presented in Table 5. Item slopes ranged between 1.53 and 4.17. The fit index suggested that 5 items did not fit the model well, but the violations were minor. These item parameters were used to calculate an IRT-based score for the 15 items of the CTM (denoted CTM-IRT). The correlation between the CTM-15 and CTM-IRT scores was very high ( $r=0.98$ ). A scatterplot of the 2 measures (Figure) revealed that the relationship followed the expected S-shaped curve for higher scores, but for scores below 40, the scatter is rather wide, reflecting the scarcity of information available for estimating the IRT parameters in this range.

## Validity Analyses

An ANOVA (Table 6) was used to evaluate the ability of CTM to differentiate scores among different patient groups. This analysis had mixed results. As in previous studies, the CTM-15 score was usually lower for patients who reported poor general health, problems with independent care transition



**Figure.** Scatter plot of CTM-15 sum scores and CTM-15 IRT scores ( $r=0.98$ ). IRT, item response theory.

**Table 6.** Known Group Validity Results for the Long (CTM-15), Short (CTM-3), and IRT-Based (CTM-15 IRT) Scores

Variable	Groups (n)	CTM-15				CTM-3				CTM-15 IRT**			
		Mean	SD	F	R <sup>2</sup>	Mean	SD	F	RV (95% CI)	Mean	SD	F	RV (95% CI)
Age	<64 (948)	74.8	16.5			75.2	17.1	4.5 <sup>†</sup>	1.03 (0.6 to 1.8)	74.8	15.5	5.7 <sup>†</sup>	1.31 (0.9 to 1.9)
	64 to 75 (412)	73.3	15.8	4.3 <sup>†</sup>	0.01	73.3	17.0			73.3	14.3		
	75+ (185)	71.2	14.8			71.7	15.0			71.0	13.0		
Sex	Male (1027)	74.9	15.3	10.7 <sup>†</sup>	0.006	75.2	16.1	9.2 <sup>†</sup>	0.86 (0.4 to 1.4)	74.7	14.4	7.7 <sup>†</sup>	0.71 (0.4 to 1.0)
	Female (518)	72.1	17.6			72.4	18.3			72.5	15.9		
Race	Asian/Pacific Islander (11)	76.5	18.1	0.4	NS	77.3	18.7	0.4	NS	75.9	17.5	0.5	NS
	Black (206)	73.2	15.6			73.3	16.5			73.3	15.3		
	Native American/Alaska (15)	70.3	11.1			71.5	12.7			70.6	9.0		
	White (1219)	74.1	16.3			74.5	17.0			74.1	15.0		
	More than one race (71)	74.6	16.9			74.7	18.4			75.2	15.1		
	Hispanic	Yes (47)	77.1	16.1	1.8	NS	76.8	16.7	1.1	NS	77.0	15.7	2.1
	No (1494)	73.9	16.2			74.2	16.9			73.9	14.9		
Education	<High school (204)	72.8	14.9	5.8 <sup>†</sup>	0.007	72.8	16.4	7.1	1.2 (0.8 to 2.1)	72.7	16.4	6.4 <sup>†</sup>	1.1 (0.8 to 1.7)
	High school/some college (924)	73.2	15.6			73.4	16.4			73.2	16.4		
Health status (baseline)	College graduate or higher (416)	76.3	17.6			76.9	17.9			76.1	17.9		
	Excellent (93)	77.6	15.8	3.5 <sup>†</sup>	0.009	78.6	16.0	2.5 <sup>†</sup>	0.72 (0.5 to 1.7)	77.3	15.2	3.4 <sup>†</sup>	0.97 (0.7 to 1.2)
	Very good (310)	75.7	16.0			75.3	17.1			75.5	15.2		
	Good (633)	74.1	15.8			74.2	16.3			74.0	14.5		
	Fair (346)	72.2	16.1			72.8	17.2			72.4	14.9		
	Poor (159)	72.2	17.5			73.3	18.3			72.3	15.9		
Self-report health utilization indicators													
ED visit since hospital discharge (1 m)	Yes (224)	72.0	17.5	3.8 <sup>†</sup>	0.002	72.7	17.6	2.3	0.61 (0.1 to 3.8)	72.7	15.6	1.8	0.48 (0.1 to 5.7)
	No (1317)	74.3	15.9			74.5	16.8			74.2	14.9		
Hospital readmission for any reason (1 m)	Yes (208)	71.9	17.8	3.9 <sup>†</sup>	0.002	72.7	18.0	2.1	0.54 (0.1 to 2.9)	72.4	14.8	2.7	0.69 (0.1 to 2.6)
	No (1337)	74.3	15.9			74.5	16.7			74.2	15.9		
Symptoms gotten better or worse after discharge (1 m)	A lot worse (30)	71.7	19.0	8.2 <sup>*</sup>	0.020	71.5	19.1	6.1 <sup>*</sup>	0.74 (0.5 to 0.9)	72.9	17.8	9.2 <sup>*</sup>	1.13 (0.9 to 1.3)
	A little worse (45)	66.7	16.1			68.8	16.4			67.4	14.4		
	Stayed the same (252)	72.1	16.3			72.5	17.8			71.9	14.8		

Continued

Table 6. Continued

Variable	Groups (n)	CTM-15			CTM-3			CTM-15 IRT**				
		Mean	SD	F	R <sup>2</sup>	Mean	SD	F	Mean	SD	F	RV (95% CI)
Self-report care transition indicators	A little better (317)	72.1	14.7			72.4	15.6		71.9	14.0		
	A lot better (867)	76.1	16.1			76.2	16.7		76.0	14.9		
Had access to medical records	Yes (930)	76.5	14.9	53.6*	0.04	76.7	15.7	50.5*	76.0	14.4	39.9*	0.74 (0.6 to 0.9)
	No (531)	70.1	17.7			70.3	18.4		70.9	15.6		
Prescheduled follow up visits	Yes (1178)	74.9	15.5	13.5 <sup>a</sup>	0.01	75.1	16.3	10.1 <sup>†</sup>	74.6	14.7	7.6 <sup>b</sup>	0.56 (0.2 to 0.8)
	No (338)	71.3	17.6			71.8	18.5		72.1	15.5		
Know who to contact if symptoms get worse	Yes (1478)	74.3	15.9	13.1*	0.01	74.5	16.7	8.1 <sup>†</sup>	74.2	14.8	10.9 <sup>a</sup>	0.83 (0.5 to 1.3)
	No (57)	66.4	20.5			68.0	20.1		67.5	17.2		

CTM indicates care transitions measure; ED, emergency department; IRT, item response theory; RV, relative validity.

\*\*CTM-IRT score linearly transformed to mean of 74 (SD 14).

\*P<0.001; <sup>†</sup>P<0.05.

indicators, and postdischarge rehospitalization and emergency department visits. In our study population, in contrast to the original reports, men and younger patients reported better CTM scores. For most indicators, the differences in the group mean scores were small (2 to 3 points; less than one-fifth of an SD).

Overall, the CTM-3 was able to detect differences in selected groups, as well as the CTM-15, for many patient characteristics, self-reported health, and self-report care transition indicator subgroups. However, the CTM-3 slightly inflated patient care transition scores, compared with the CTM-15; this difference was observed primarily for small subgroups. In this sample, the short measure failed to detect differences between patients with and without self-reported rehospitalization and emergency department visits. Relative validity coefficients suggest that this underperformance is not statistically significant. As expected, the SD of item responses was larger for the CTM-3 than for the CTM-15.

All relative validity coefficients for comparing the CTM-IRT with the traditional CTM-15 were close to, and not significantly different from, 1.0, suggesting that the IRT scoring approach did not improve measurement precision.

## Discussion

### Major Findings

The major findings of this work of this study are related to the psychometric characteristics of the CTM and the application of IRT methods for improvement of measurement precisions. Each of these areas is discussed separately in the context of previous evidence next. The basic psychometric characteristics of the CTM-15 in our data were similar to those previously reported; item-level means were comparable to those previously reported and the measure had good internal consistency and reliability. However, we also identified some undesirable characteristics of the CTM-15 scale score, which may influence the measurement of care transitions as a performance measure and future multivariable analyses. The distribution of CTM-15 summary score was severely left-skewed, due to a substantial ceiling effect and clustering of high summary transition scores. Moreover, the 4 response options do not provide unique information; 3 CTM items may be better assessed using a binary response. These problems can be partially explained by strong acquiescence bias (the tendency of respondents to agree with statements of opinion regardless of content),<sup>29,30</sup> which is common with the “agree/disagree” format used. Understanding the source does not fix the problem, however, since the highly skewed, clustered responses lead to scores with little variance, making it hard—even with rescaling—to discriminate levels of care transition quality. In our known-groups validity



analyses, the CTM-15 found statistically significant hypothesized differences supporting findings from earlier reports; however, the magnitude of differences (2 to 3 points on a 0-to-100 scale, with an SD of 16) observed for many of these tests, was small. In the absence of established guidelines for minimally important differences, we would typically only consider differences to be practically meaningful if they are  $\approx 0.5$  SD or greater.

The problems discussed here with the CTM-15 apply as well or more to its short form, the CTM-3. This is of particular concern given that the CTM-3 is included in the CAHPS Hospital Survey, used to judge hospital quality. Explorations of the validity of the CTM-3 based on the known-groups validity method had mixed results that raise questions about its ability to detect differences in care transition quality.

Our results complement those of a recent study of the CTM-15, which also identified some gaps in measurement performance for assessing the quality of care transitions in a complex population of older rehabilitation patients.<sup>31</sup> While the CTM was found to be reliable, the authors noted that the construct validity and utility of the measure could be improved. Qualitative data in their study revealed the arbitrary nature of the choice that some patients make when selecting a response category between “agree” and “strongly agree.” Comments included: “Either you agree or you don’t!” and “I don’t know why I keep saying ‘agree’ and not ‘strongly agree’—I guess I just don’t want you to think I’m not listening by choosing ‘strongly agree’ all the time!” In addition, focus-group respondents identified some aspects of transitional care that are not included in the CTM, such as, building a relationship, and effectively communicating with, one’s clinician, raising questions as to the CTM’s content validity. The authors also noted problems associated with the large proportion of participants responding in “agreement,”<sup>31</sup> which aligns with reports from the CTM’s developers that most patients agreed with each CTM question (range 69% to 94%).<sup>32</sup>

In this study, we also explored the possibility that the application of IRT scoring may improve the measurement properties of the CTM-15. Theoretically IRT scoring can often improve measurement precision<sup>5</sup>; however, as indicated by nonsignificant relative validity coefficients, IRT scoring did not improve the CTM-15. This is the first attempt to apply IRT scoring to the CTM of which we are aware. Several reports on the use of IRT or Rasch scoring approach to patient reported measures developed in the classical test theory framework have produced mixed results. When IRT was used to score the Short Form-36 Physical Function scale modest gains were observed, of these, the strongest gains were demonstrated in the most clinically dissimilar groups.<sup>33</sup> In sensitivity studies of the same measure, IRT scores were more sensitive for the general population across 7 countries, while results for the

Short Form-36 in patients with epilepsy were mixed depending on external criteria.<sup>34</sup> Considerable gains in precision were reported for the Rasch scoring of the Oxford Hip Score Questionnaire.<sup>35,36</sup> Precision was also improved using IRT-based scoring for the upper limb subscale of the Motor Assessment Scale, particularly in the scale’s extreme ranges.<sup>37</sup> An IRT-based scoring approach improved the sensitivity of the Visual Function Index, but produced no gains for cross-sectional comparisons.<sup>38</sup> No improvement in precision was observed for the Health Assessment Questionnaire<sup>39</sup> and the EORTC QLQ-C30 scales.<sup>5</sup>

Similarly, we could not improve the CTM-15’s measurement precision with IRT-based scoring. Possible causes for this have been discussed previously<sup>5</sup> and could be related to the misfit of some items to the IRT model, the fact that the measure was developed within the classic test theory framework and not specifically for the IRT, and the existence of several parallel items with highly similar content. Unique to the CTM was acquiescence bias, in which up to 30% of respondents selected the same response option (“agree” or “strongly agree”) for all 15 items, leading to extreme clustering of the scale. For these respondents, IRT scoring cannot improve over a traditional sum scored approach. With the tight clustering of scale scores in the upper end of the scale, it is possible that IRT scoring failed to capture any additional information for the respondents. These findings are also in line with previous reports that IRT gains are potentially largest at the extremes of a construct’s range.<sup>33,37</sup> Finally, it is worth noting that while the CTM-15 is a patient-reported measure, it aims to evaluate the patient’s experience with a care process; it does not relate to any of health, quality of life, or physical functioning.

## Strengths and Limitations

Our study has important strengths and several limitations. We used a large and diverse sample to conduct in-depth psychometric evaluation using both classic and modern analytic methods. To the best of our knowledge, this is the most comprehensive psychometric evaluation of the CTM. Our rich data set will also allow for future evaluation of relationship of the CTM with important clinical variables. While our sample is the largest one on which CTM-15 psychometric analyses have been reported, it is still limited to a particular patient population and geographic regions. Psychometric evaluation in a population with different characteristics may have different results. Good psychometric practice would require validation in samples that are significantly different from the ones previously evaluated. Our study participants who completed the measure reported on average higher level of education, better health, and higher income compared with all patients enrolled at baseline in the study.

These differences may have biased the average CTM-15 scores, leading to an underestimation of proportions of patients who report poor transitions and an overestimation of the ceiling effect in the measure. Given the diverse sample on which the analyses were completed, however, the general conclusions of the psychometric evaluation are most likely robust.

## Conclusions and Implications

Our findings have important implications for analyzing and interpreting CTM scores. To the extent that CTM score distributions are both clustered around certain values and skewed, it is important that appropriate analytic techniques are used in the analyses of CTM scores. If the linear scoring of the CTM-15 is not supported by available data, it would be more important to use analytic approaches for categorical data. Results from the CTM short form may have even a higher ceiling effect and conclusions, thus it may not always correspond to results of the CTM-15 as previously suggested. The strong acquiescence bias of the measure also suggests that CTM results may be representing an overly optimistic view of care transition quality. To increase the utility of the CTM, it will be important to determine whether observed magnitudes of difference (2 to 3 points on the CTM scale) between groups are clinically important using recommended triangulation approaches<sup>40–42</sup>. Future work should also aim to improve care transition quality assessments by constructing a measure that fully reflects what patients care about in the area of care transition, and that is less prone to acquiescence bias.

In summary, we identified some psychometric challenges in both the long and short forms of the CTM that could not be improved with IRT-based scoring. Combined with accumulating evidence of existing gaps in the CTM's content validity, this study suggests that CTM scores should be interpreted with caution and may not be a sufficiently sensitive tool for detecting meaningful improvements in the quality of transitional care. A new patient-reported measure of care transition quality that addresses these problems is needed.

## Sources of Funding

TRACE-CORE is supported by the National Institutes of Health, National Heart, Lung, and Blood Institute grant 1U01HL105268. Research also supported by UL1RR031982 from the National Center for Research Resources, NIH/NCRR, University of Massachusetts Center for Clinical and Translational Sciences, Biostatistics, Epidemiology & Research Design Component (BERD), and The Center for Health Equity Intervention Research (CHEIR), grant P60 MD006912. Partial salary support is additionally provided by the PhRMA Foundation Started Grant (Dr. Anatchkova).

## Disclosures

None.

## References

1. Anderson G. *Chronic Care: Making the Case for Ongoing Care*. Princeton, NJ: Robert Wood Johnson Foundation; 2010.
2. Coleman EA, Smith JD, Frank JC, Eilertsen TB, Thiare JN, Kramer AM. Development and testing of a measure designed to assess the quality of care transitions. *Int J Integr Care*. 2002;2:e02.
3. Coleman EA, Mahoney E, Parry C. Assessing the quality of preparation for posthospital care from the patient's perspective: the care transitions measure. *Med Care*. 2005;43:246–255.
4. Parry C, Mahoney E, Chalmers SA, Coleman EA. Assessing the quality of transitional care: further applications of the care transitions measure. *Med Care*. 2008;46:317–322.
5. Petersen MA, Groenvold M, Aaronson N, Brenne E, Fayers P, Nielsen JD, Sprangers M, Bjorner JB. European Organisation for Research and Treatment of Cancer Quality of Life Group. Scoring based on item response theory did not alter the measurement ability of EORTC QLQ-C30 scales. *J Clin Epidemiol*. 2005;58:902–908.
6. Shadmi E, Zisberg A, Coleman EA. Translation and validation of the care transition measure into Hebrew and Arabic. *Int J Qual Health Care*. 2009;21:97–102.
7. Waring ME, McManus RH, Saczynski JS, Anatchkova MD, McManus DD, Devereaux RS, Goldberg RJ, Allison JJ, Kiefe CI; TRACE-CORE Investigators. Transitions, risks, and actions in coronary events—center for outcomes research and education (TRACE-CORE): design and rationale. *Circ Cardiovasc Qual Outcomes*. 2012;5:e44–e50.
8. Nunnally JC, Bernstein I. *Psychometric Theory*. 3rd ed. New York, NY: McGraw-Hill Book Company; 1994.
9. Traub RE. *Reliability for the Social Sciences: Theory and Applications*. Oaks, CA: Sage; 1994.
10. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16:297–334.
11. Hambleton RK, Swaminathan H, Rogers J. *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Publications, Inc; 1991.
12. McDonald RP. The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*. 1981;34:100–117.
13. Muthen LK, Muthen BO. *Mplus User's Guide*. 3rd ed. Los Angeles, CA: Muthen & Muthen; 1998–2004.
14. Muthen B. A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*. 1984;49:115–132.
15. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model*. 1999;6:1–55.
16. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, Thissen D, Revicki DA, Weiss DJ, Hambleton RK, Liu H, Gershon R, Reise SP, Lai JS, Cella D. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the patient-reported outcomes measurement information system (PROMIS). *Med Care*. 2007;45:S22–S31.
17. Bjorner JB, Kosinski M, Ware JE Jr. The feasibility of applying item response theory to measures of migraine impact: a re-analysis of three clinical studies. *Qual Life Res*. 2003;12:887–902.
18. Bjorner JB, Chang CH, Thissen D, Reeve BB. Developing tailored instruments: item banking and computerized adaptive assessment. *Qual Life Res*. 2007;16:95–108.
19. Samejima F. Graded response model. In: Van der Linden W, Hambleton R, eds. *Handbook of Modern Item Response Theory*. New York, NY: Springer-Verlag; 1997:85–100.
20. Muraki E. Generalized partial credit model. In: Linden VD, Hambleton RK, eds. *Handbook of Item Response Theory*. New York, NY: Springer-Verlag; 1997:153–164.
21. Andrich D. *Rasch Models for Measurement*. Beverly Hills, CA: Sage Publications, Inc; 1998.
22. Andrich D. A rating formulation for ordered response categories. *Psychometrika*. 1978;43:561–573.
23. Muraki E, Bock R. *Parscale—IRT Based Test Scoring and Item Analysis for Graded Open-Ended Exercises and Performance Tasks*. Chicago, IL: Scientific Software, Inc; 1996.

24. Bjonner JB, Smith K, Stone C, Sun X. IRTFIT: a macro for item fit and local dependence tests under IRT models. 2007.
25. Kerlinger FN. *Foundations of Behavioral Research*. New York, NY: Holt, Rinehart, & Winston; 1973.
26. Deng N, Allison JJ, Fang HJ, Ash AS, Ware JE Jr. Using the bootstrap to establish statistical significance for relative validity comparisons among patient-reported outcome measures. *Health Qual Life Outcomes*. 2013;11:89.
27. Coleman EA, Parry C, Chalmers SA, Chugh A, Mahoney E. The central role of performance measurement in improving the quality of transitional care. *Home Health Care Serv Q*. 2007;26:93–104.
28. Ryvicker M, McDonald MV, Trachtenberg M, Peng TR, Sridharan S, Feldman PH. Can the care transitions measure predict rehospitalization risk or home health nursing use of home healthcare patients? *J Healthc Qual*. 2013;35:32–40.
29. Holbrook A. Acquiescence response bias. In: Lavrakas PJ, ed. *Encyclopedia of Survey Research Methods*. Thousand Oaks, CA: Sage Publications, Inc; 2008:4–5.
30. Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. *J Appl Psychol*. 2003;88:879–903.
31. McLeod J, Stolee P, Walker J, Heckman G. Measuring care transition quality for older patients with musculoskeletal disorders. *Musculoskeletal Care*. 2014;12:13–21.
32. Coleman EA, Smith JD, Frank JC, Min SJ, Parry C, Kramer AM. Preparing patients and caregivers to participate in care delivered across settings: the care transitions intervention. *J Am Geriatr Soc*. 2004;52:1817–1825.
33. McHorney CA, Haley SM, Ware JE Jr. Evaluation of the MOS SF-36 physical functioning scale (PF-10): II. Comparison of relative precision using likert and rasch scoring methods. *J Clin Epidemiol*. 1997;50:451–461.
34. Birbeck GL, Kim S, Hays RD, Vickrey BG. Quality of life measures in epilepsy: how well can they detect change over time? *Neurology*. 2000;54:1822–1827.
35. Fitzpatrick R, Norquist JM, Dawson J, Jenkinson C. Rasch scoring of outcomes of total hip replacement. *J Clin Epidemiol*. 2003;56:68–74.
36. Norquist JM, Fitzpatrick R, Jenkinson C. Rasch measurement in the assessment of amyotrophic lateral sclerosis patients. *J Appl Meas*. 2003;4:249–257.
37. Khan A, Chien CW, Brauer SG. Rasch-based scoring offered more precision in differentiating patient groups in measuring upper limb function. *J Clin Epidemiol*. 2013;66:681–687.
38. Las HAYAS C, Bilbao A, Quintana JM, Garcia S, Lafuente I; IRYSS Cataract Group. A comparison of standard scoring versus rasch scoring of the visual function index-14 in patients with cataracts. *Invest Ophthalmol Vis Sci*. 2011;52:4800–4807.
39. Wolfe F, Kong SX. Rasch analysis of the western Ontario MacMaster questionnaire (WOMAC) in 2205 patients with osteoarthritis, rheumatoid arthritis, and fibromyalgia. *Ann Rheum Dis*. 1999;58:563–568.
40. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61:102–109.
41. Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J*. 2007;7:541–546.
42. Turner D, Schunemann HJ, Griffith LE, Beaton DE, Griffiths AM, Critch JN, Guyatt GH. The minimal detectable change cannot reliably replace the minimal important difference. *J Clin Epidemiol*. 2010;63:28–36.