

Implementation, Policy and Community Engagement Research Article

Cite this article: Welch LC, Tomoaia-Cotisel A, Noubary F, Chang H, Mendel P, Parajulee A, Fenwood-Hughes M, Etchegaray JM, Qureshi N, Chandler R, and Selker HP. Evaluation of initial progress to implement Common Metrics across the NIH Clinical and Translational Science Awards (CTSA) Consortium. *Journal of Clinical and Translational Science* 5: e25, 1–12. doi: [10.1017/cts.2020.517](https://doi.org/10.1017/cts.2020.517)

Received: 24 April 2020

Revised: 16 July 2020

Accepted: 17 July 2020

Keywords:

Performance improvement; Common Metrics; evaluation; clinical and translational science; CTSA

Address for correspondence:

L. C. Welch, PhD, Tufts Medical Center,
800 Washington Street, Box 63, Boston,
MA 02111, USA.
Email: LWelch2@TuftsMedicalCenter.org

© The Association for Clinical and Translational Science 2020. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.



Evaluation of initial progress to implement Common Metrics across the NIH Clinical and Translational Science Awards (CTSA) Consortium

Lisa C. Welch^{1,2}, Andrada Tomoaia-Cotisel³, Farzad Noubary^{1,2}, Hong Chang^{1,2}, Peter Mendel³, Anshu Parajulee¹, Marguerite Fenwood-Hughes¹, Jason M. Etchegaray³, Nabeel Qureshi³, Redonna Chandler⁴ and Harry P. Selker^{1,2}

¹Tufts Clinical and Translational Science Institute, Tufts University, Boston, MA, USA; ²Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA, USA; ³RAND Corporation, Santa Monica, CA, USA and ⁴The National Institute on Drug Abuse, National Institutes of Health, Bethesda, MD, USA

Abstract

Introduction: The Clinical and Translational Science Awards (CTSA) Consortium, about 60 National Institutes of Health (NIH)-supported CTSA hubs at academic health care institutions nationwide, is charged with improving the clinical and translational research enterprise. Together with the NIH National Center for Advancing Translational Sciences (NCATS), the Consortium implemented Common Metrics and a shared performance improvement framework. **Methods:** Initial implementation across hubs was assessed using quantitative and qualitative methods over a 19-month period. The primary outcome was implementation of three Common Metrics and the performance improvement framework. Challenges and facilitators were elicited. **Results:** Among 59 hubs with data, all began implementing Common Metrics, but about one-third had completed all activities for three metrics within the study period. The vast majority of hubs computed metric results and undertook activities to understand performance. Differences in completion appeared in developing and carrying out performance improvement plans. Seven key factors affected progress: hub size and resources, hub prior experience with performance management, alignment of local context with needs of the Common Metrics implementation, hub authority in the local institutional structure, hub engagement (including CTSA Principal Investigator involvement), stakeholder engagement, and attending training and coaching. **Conclusions:** Implementing Common Metrics and performance improvement in a large network of research-focused organizations proved feasible but required substantial time and resources. Considerable heterogeneity across hubs in data systems, existing processes and personnel, organizational structures, and local priorities of home institutions created disparate experiences across hubs. Future metric-based performance management initiatives across heterogeneous local contexts should anticipate and account for these types of differences.

Introduction

The National Institutes of Health (NIH) Clinical and Translational Science Awards (CTSA) Program, composed of about 60 CTSA hubs, is charged with growing and improving the nation's clinical and translational research enterprise. The CTSA Consortium is comprised of academic health care institutions that deliver research services, provide education and training, and innovate improved processes and technologies to support clinical and translational research. A 2013 Institute of Medicine (IOM, now the National Academy of Medicine) report on the CTSA Consortium [1] recommended the institution of “common metrics” to assess and continuously improve activities at each hub, and across the Consortium as a whole. In response, the NIH National Center for Advancing Translational Sciences (NCATS) and CTSA Consortium hubs implemented the Common Metrics Initiative, composed of establishing standardized metrics and using them for metric-based performance management.

Performance management, intended to identify and act on opportunities to improve, has been implemented in a variety of related settings, including clinical care [2,3], research hospitals [4], nonprofit organizations [5], governmental organizations [6], and academic institutions [7]. There are fewer examples of implementing performance management across a network of organizations, especially in biomedical research. Federal public health programs implemented by loosely integrated networks of local organizations face three challenges in measuring performance: complex problems with long-term outcomes, decentralized organization of program delivery, and lack of consistent data [8]. This is informative because the decentralized

organization of federal public health programs mirrors the CTSA Consortium. Although all CTSA hubs strive toward the same mission of catalyzing the clinical and translational research enterprise, each hub has autonomy to develop the approach and processes that are most effective in its local context. To our knowledge, the current paper reports the first evaluation of the implementation of shared metric-based performance management in a decentralized national network of health care research organizations.

Between June, 2016 and December, 2017, an implementation team from Tufts Clinical and Translational Science Institute (CTSI) led the rollout of three Common Metrics and the Results-Based Accountability performance improvement framework [9] across the CTSA Consortium in three waves of hubs, or implementation groups. As reported previously, implementation groups were used to manage training and coaching of a large volume of hubs and were assigned based on hubs' preferences [10]. The Common Metric topics focused on training scientists for careers in clinical and translational research, supporting efficiency by shortening Institutional Review Board (IRB) time, and ensuring results from CTSA Consortium pilot studies are disseminated (Supplemental Table 1). The Tufts Implementation Program entailed training on the metrics and performance improvement framework and seven every-other-week small group coaching sessions.

A separate Tufts CTSI team conducted a mixed methods evaluation to assess initial progress with Common Metrics. A 19-month follow-up period, ending in January, 2018, was intended to provide sufficient time for hubs to become oriented to the Common Metrics, incorporate the required activities into workflows, and implement performance improvement strategies. This report summarizes hubs' progress, and factors affecting that progress, in the follow-up period.

Methods

Research Design

We used an intervention mixed methods framework [11] to describe hubs' progress and experiences implementing the Common Metrics and performance improvement framework. The posttest design integrated quantitative measures, open-ended written responses, and qualitative interview data to describe *what* level of implementation hubs achieved in relation to the initial three Common Metrics and *why* full implementation was or was not achieved.

The primary evaluation outcome was implementation of the initial three Common Metrics and performance improvement framework for each metric. This outcome was measured quantitatively as the extent of completion of 13 activities, clustered into 5 distinct groups (Table 1). With input from the Tufts Common Metrics Implementation Team, we created a rubric with a point value for each activity. The sum of a hub's points indicated the extent of completion of activities, regardless of the order of completing them. The activities were not weighted for relative difficulty, effort, or time required because hub experiences varied.

To better understand lack of completion of each activity, we elicited reasons as open-text survey responses and conducted semi-structured interviews about contextual factors, challenges, and facilitators.

Data Collection

We collected data at various time points throughout the implementation period using four self-report surveys and a qualitative interview guide (Supplemental Table 2).

Table 1. Implementation of Common Metrics and performance improvement activities: definition and point assignments

Cluster and activities*	Points possible**
Creating the metric	
• Collected data	1.0
• Computed metric result according to operational guideline (self-report)	1.0
Understanding current performance	
• Forecasted future results or compared result to any other data	1.0
• Specified underlying reasons involving hub leadership/staff/faculty	0.5
• Specified underlying reasons involving any group outside hub leadership/staff/faculty	0.5
Developing a performance improvement plan	
• Involved hub leadership/staff/faculty when developing improvement plan	0.5
• Involved any group outside hub leadership/staff/faculty when developing improvement plan	0.5
• Specified actions for achieving desired outcome	1.0
• Prioritized actions	0.5
• When prioritizing actions, considered potential effectiveness of actions or feasibility	0.5
Implementing the performance improvement plan	
• Reached out to specific individuals or institutional partners for help in carrying out improvement plan	1.0
• Began to implement improvement plan	1.0
Documenting metric result and plan fully	
• Documented five elements in the Common Metric-specific Scorecard: metric result; underlying reasons; potential partners; potential actions; planned actions	1.0
<i>Total possible</i>	<i>10.0</i>

*Activities did not have to be conducted sequentially.

**Each distinct activity was assigned 1.0 points. For pairs of related activities (e.g., involving different types of stakeholders when specifying underlying reasons), each part of the pair received 0.5 points to equal 1.0.

Surveys

Before starting the Tufts Common Metrics Implementation Program, participating hubs completed a cross-sectional survey about hub prior experience with metric data collection and performance improvement activities in the previous calendar year. These data were used to construct a composite measure of each hub's prior experience with data-driven performance improvement.

Additionally, hubs completed a baseline and two follow-up surveys about progress on the 13 activities that composed the primary outcome (Supplemental Table 3). At the start (i.e., baseline), hubs were instructed to choose one of their local metrics that best exemplified how the hub had used metric data in the five months prior to starting the Common Metrics Implementation Program and to report on activities composing the primary outcome. We used these data to sample hubs for qualitative interviews (see below).

Two follow-up surveys collected data regarding hub progress on the Common Metrics. At the end of the implementation

program's coaching period, hubs were instructed to choose one Common Metric that best exemplified the hub's use of metric data and the performance improvement framework as of that time and report progress on completing the 13 activities for that metric. The second follow-up survey was conducted 19 months after Implementation Group 1 began, which was 17.5 and 15 months after Implementation Groups 2 and 3 began, respectively. This survey recorded any additional performance improvement activities completed for the Common Metric reported on during the first follow-up survey, activities completed for the remaining two Common Metrics and related performance improvement efforts, and additional information about hub experiences.

Semi-structured interviews

The interview guide included open-ended questions and probes to elicit an in-depth understanding of challenges, facilitators, and contextual factors for implementing Common Metrics (Supplemental Table 4). The Context Matters Framework [12] was applied to capture five domains that might have influenced hubs' experiences with Common Metrics implementation: (1) specific implementation setting, (2) wider organizational setting, (3) external environment, (4) implementation pathway, and (5) motivation for implementation.

The interview guide was adapted for three roles: the hub's Principal Investigator, the Administrator/Executive Director (or another individual filling the role of Common Metrics "champion"), and an "Implementer" staff member knowledgeable about day-to-day implementation. We piloted each version of the guide during mock interviews with personnel from Tufts CTSA. After each interview, three qualitative team members debriefed and revised the interview guide as needed to clarify content and improve the flow of the interview.

One of the two qualitative team members conducted each interview by telephone. Interviews lasted between 20 and 60 minutes. Each participant was emailed an information sheet describing the study prior to the interview and provided verbal consent. Interviews were audio recorded and transcribed verbatim.

Interviewer training entailed mock interviews and debriefing. To ensure consistency, the two study interviewers listened to and discussed audio recordings of early interviews and more difficult interviews. During weekly meetings, three qualitative team members discussed study participants' experiences with interview questions and, following procedures for qualitative interviewing, identified additional language to further facilitate future interviews.

Administrative data

Information on hub size and funding cohort was provided by NCATS and confirmed as current through publicly available sources when possible. Hub size was defined as total funding from NIH U, T, K, and/or R grant mechanisms for fiscal year 2015–2016. Hub funding cohort was calculated based on the year the hub was first funded.

Participants

Surveys

Sixty CTSA hubs were invited to participate in each survey by an invitation email to one principal investigator per hub. The email instructed the recipient to assign one person to complete the survey with input from others across the hub. To maximize response rate, reminder emails were sent to the principal investigator. All hubs responded to the survey about prior experience and the baseline survey, 57 hubs (95%) responded to the first follow-up, and 59 (98%)

responded to the second follow-up. Surveys were self-administered online using REDCap software [13].

Semi-structured interviews

Interviews were conducted with participants from a sample of 30 out of the 57 hubs that responded to both the baseline and the first follow-up surveys. The sampling plan sought balance primarily across hubs' experiences with metric-based performance improvement and, secondarily, across other key hub characteristics. First, to ensure hubs with a diversity of experiences on performance improvement, we created a matrix of hub scores on the study's primary outcome (implementation of the three Common Metrics) at two time points: baseline (i.e., prior experience on a local metric) and the first follow-up survey (i.e., early progress on a Common Metric). Hub scores for each time point were trichotomized into three levels (minimal, moderate, and significant), yielding nine cells representing combinations of baseline experience and early implementation progress (Supplemental Table 5).

After sorting the 57 hubs into the matrix, we targeted 3 or 4 hubs within each cell to achieve a sample of 30 hubs. For cells with fewer than four hubs, all hubs were designated for inclusion. For cells with more than four hubs, we randomly sampled four hubs. We then reviewed the resulting sample to ensure balance across a range of hub characteristics (years of funding, total funding amount, region, implementation group, and number of hub implementation team members reported). Selected hubs that declined or did not respond to invitations to participate were replaced by randomly selecting another hub from the same cell, when available. If no additional hubs were available in the same cell, we recruited a hub from another cell that represented a change in baseline experience and early implementation progress, with the goal of maximizing insight into challenges and facilitators for changing scores.

Recruitment for qualitative interviews began by seeking agreement for participation from the hub's principal investigator or designee administrator who nominated individuals in the other two roles addressed by the interview guide. If interviews for all three roles could not be scheduled, another hub was selected. A total of 90 interviews across 30 hubs were conducted.

Analytic Strategy

Quantitative and qualitative data were analyzed independently, and results were merged to develop a full description of hub experiences. Results from different data sources expanded our understanding by addressing different aspects of the experience (e.g., completion of activities vs. challenges and facilitators of that completion), and qualitative data provided insights to help explain associations identified in statistical analyses.

Statistical analyses

Hub characteristics were described overall and by implementation group using means and standard deviations for continuous variables and proportions for categorical variables. To assess differences in hub characteristics between implementation groups, we used *t*-tests for continuous data and chi-squared tests for categorical data. Similar numeric summaries were used to describe the frequencies of completion of activities. We also tested for differences in mean completion of activities for each Common Metric, using a linear mixed effects model with a hub-specific random intercept. Next, we fitted univariable (i.e., unadjusted) and multivariable (i.e., adjusted) linear regression models for the primary outcomes separately for each metric and for the overall sum. We included nine characteristics of hubs across three domains: hub basic attributes

(hub size and initial funding cohort), previous experience with metric-based performance improvement, and participation in the Tufts Implementation Program. For the multivariable linear regression model, a stepwise variable selection procedure using Akaike information criterion (AIC) was performed, starting with a full model including all covariates and proceeding with both backward and forward selection.

To construct the composite measure of a hub's prior experience with metric-based performance improvement, we conducted a factor analysis to create an experience factor score. The factor analysis used 10 survey items (Supplemental Table 6). Each response category was assigned a numerical value with a higher value indicating more experience. For questions with multiple parts, "yes" responses were summed to create a single score for that item. All 10 dimensions were used in an exploratory factor analysis, with results indicating a two-factor model based on the proportion of variance explained. After reviewing for meaningfulness, one factor was chosen. This single-factor score represented the "maturity of a performance management system" and was created using the weighted average of all dimensions involved. The resulting variable is a standardized normal score with a mean of zero and standard deviation of one. A higher score indicates a higher level of the underlying concept of maturity of systems.

Qualitative analyses

Semi-structured interview audio recordings were transcribed verbatim by a professional transcription company. Transcripts were uploaded into the NVivo qualitative data analysis software to facilitate coding and analysis [14].

The codebook was developed using a two-stage consensus-based process. First, the qualitative team developed an initial codebook using main topics of the interview protocol as preidentified categories. Then, analysts reviewed two transcripts, interview notes, and reflections to identify emergent concepts. The preidentified categories and emergent concepts were merged into a single initial codebook. This codebook was reviewed by the qualitative team for clarity and consistency.

Second, analysts applied the initial codebook to two small batches of transcripts (one transcript and then three more transcripts) with participants in different roles (Principal Investigator, Administrator, and Implementer) to ensure definitions were clear and codes were being used consistently. For each batch, the team met to compare the coding and resolve discrepancies, and the codebook was revised as needed.

Once consensus was reached on the codebook and coding was consistent between analysts, one team member coded the interviews using the codebook. To ensure consistency, another team member periodically reviewed a convenience sample of coded transcripts for fidelity to the codebook. The full qualitative team discussed all potential new themes or revisions before any changes were made to the codebook.

Over the course of coding transcripts, themes were grouped into four domains: metric design and content, stakeholder engagement, hub engagement, and perceived value of implementing Common Metrics. Once coding was completed, the four domains were divided among team members so that one analyst read all coded sections within one domain. Those analysts then categorized coded sections into facilitators and challenges, and summarized the range of themes, including illustrative quotations. Each analyst also identified intersections among themes that were discussed by

the full team and incorporated into the presentation of results. Subanalyses investigated whether hubs' engagement with the Common Metrics Implementation differed by participant role.

Open-ended survey responses followed similar consensus-based procedures. Two analysts independently developed initial codes and met to develop an initial codebook. Each analyst then applied the codebook to a subset of responses, met to discuss and resolve discrepancies, and modified the codebook as needed. After nine meetings, the analysts were applying the codebook consistently. At that point, one analyst coded the remaining responses and discussed questions with the other analyst. Given the straightforward nature of the responses, codes were summarized using frequencies and illustrative quotations.

Results

Description of Hubs

The primary quantitative analyses included the 59 hubs that responded to the second follow-up survey at the end of the evaluation study period (Supplemental Table 6). At the beginning of the Common Metrics Implementation Program, hubs ranged substantially in size of their annual budgets and year of initial CTSA funding across a 10-year time span. Across 10 indicators of experience with metric-based performance improvement, hubs generally reported average levels of experience in the middle of the possible response ranges for each indicator.

The composition of three implementation groups did not differ in size of annual budgets or experience with metric-based performance improvement, but did vary in composition based on initial year of funding. Compared to Implementation Groups 1 and 2, Implementation Group 3 was comprised of more hubs first funded in the earliest or latest cohorts of CTSA funding. As reported previously [10], Implementation Group 2 attended fewer training and coaching sessions than Implementation Groups 1 and 3 (average of 11.3, 12.6, and 11.9 sessions, respectively), and more hubs focused on the IRB Review Duration (38%) or Pilot Funding (39%) Metrics than the Careers Metric (23%) during coaching.

Completion of Metric and Performance Improvement Activities

After 19 months, all hubs reported that they had begun the work of implementing the Common Metrics and performance improvement for all of the first three metrics. However, less than one-third of hubs (17 of 59) had completed all 13 activities for each metric (score of 30; Fig. 1). About half of hubs (29 of 59) completed between 90% and 100% of activities (score of 27 or higher), one-quarter completed between 70% and 85% of activities (score of 21–25.5), and the remaining one-quarter completed between 27% and 65% of activities (score of 8–19.5).

On average, hubs completed almost all activities related to creating metric results, and the vast majority of activities related to understanding current performance (Table 2). However, variation was evident for activities related to developing performance improvement plans, which were completed less often for the IRB Metric compared to the Careers and Pilots Metrics. When a performance improvement plan was not developed, activities related to implementing it could not be completed. Additionally, not all hubs that developed a plan completed activities to implement the plan. Fully documenting a metric result and the four elements of the improvement plan was completed least often, on average.

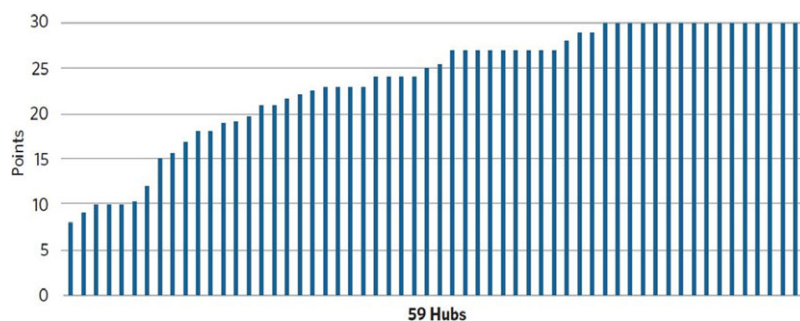


Fig. 1. Completion of Common Metrics and performance improvement activities per hub: three metrics combined (0–30 points possible).

Table 2. Completion of Common Metrics and performance improvement activities ($N = 59$ hubs*)

Mean (SD), range	Overall sum		By metric				
	Possible	Actual	Possible	Careers	IRB	Pilot	P-value
All activities	30	23.7 (6.6) 8–30	10	8.09 (2.6) 2.5–10	7.4 (2.9) 2–10	8.1 (2.5) 1–10	0.44
Clusters of activities**							
Creating metric result	6	5.9 (0.3) 4–6	2	2.0 (0.0) -	1.9 (0.3) 0–2	1.9 (0.1) 1–2	0.15
Understanding current performance	6	5.5 (0.8) 3–6	2	1.8 (0.4) 0.5–2	1.8 (0.4) 1–2	1.8 (0.4) 0–2	0.96
Developing improvement plan	9	6.4 (3.1) 0–9	3	2.3 (1.2) 0–3	1.9 (1.4) 0–3	2.3 (1.2) 0–3	0.05
Implementing improvement plan	6	4.1 (2.1) 0–6	2	1.4 (0.9) 0–2	1.2 (0.9) 0–2	1.4 (0.8) 0–2	0.17
Documenting metric result and plan fully	3	1.8 (1.2) 0–3	1	0.6 (0.5) 0–1	0.5 (0.5) 0–1	0.6 (0.5) 0–1	0.21

SD = Standard Deviation.

*One hub did not respond.

**Composition of clusters: (1) creating metric result entails data collection and computing metric according to operational guideline; (2) understanding metric result entails forecasting future performance or comparing results to any other data, and specifying underlying reasons with stakeholders; (3) developing improvement plan entails involving stakeholders, specifying actions, and prioritizing actions based on effectiveness or feasibility; (4) implementing the improvement plan entails reaching out to partners for help and starting implementation activities; (5) documenting includes entering metric result, describing underlying reasons, identifying partners, potential actions, and planned actions.

Factors Affecting Progress

Quantitative and qualitative results together identified seven key factors affecting hub progress. The characteristics that could be assessed quantitatively explained between 16% and 21% of the variation in completing improvement activities across hubs and metrics (Table 3). Qualitative results enhanced our understanding of these effects and identified additional factors.

Hub size and resources

Analysis of open-ended survey responses revealed that the most common reason hubs cited for not completing performance improvement activities was lack of time and resources. Hubs size, defined by funding level, varied greatly, and quantitative analysis showed that funding level appeared to have some effect, particularly for the Pilot Funding Metric. Compared to the smallest hubs, mid-size and large-size hubs consistently completed slightly more performance improvement activities on average. Yet, when considering activities completed across all metrics, the effect was largest for mid-size hubs, not the largest hubs (Table 3).

Qualitative results reveal that the size of a hub's funding award did not fully account for resource challenges. Investment from institutions within which hubs were situated, periods of interrupted funding, lack of data systems or lack of alignment of existing systems with the Common Metrics data requirements, and the availability of needed personnel and expertise all affected whether hubs could devote sufficient time and resources to fully implement Common Metrics and performance improvement activities (Table 4).

Hubs with available evaluation and other metric-related expertise, as well as institutional knowledge and general administrative support, reported that these greatly facilitated implementation. Hubs often formed a core team intended to provide an organized approach to implementation activities. Teams included mutually supporting roles such as site champions to engage stakeholders, keeping the principal investigator aware of activities, and conducting hands-on data collection and reporting. Participants identified three facilitators related to effective core teams: (1) one leader who is accountable for the work, (2) a "champion" or "real believer" on the team to encourage local ownership of the initiative, and

Table 3. Results of testing for effects of hub characteristics on completion of performance improvement activities (N = 59 hubs^c)

Characteristic	Univariable models				Multivariable models			
	Change in hub score				Change in hub score			
	Overall sum (0–30)	By metric			Overall sum (0–30)	By metric		
		Careers (0–10)	IRB (0–10)	Pilots (0–10)		Careers (0–10)	IRB (0–10)	Pilots (0–10)
Model N					55	55	55	55
Model Adjusted R ²					0.17	0.16	0.20	0.21
Basic attributes								
Size ^e at start of CMI program (tertiles)								
<\$4.56 million (Ref)	–	–	–	–	–	–	–	–
\$4.56–8.04 million	2.88	0.38	0.96	1.54*	–	–	–	1.27*
≥\$8.05 million	1.64	0.72	–0.20	1.12	–	–	–	1.42*
Initial funding cohort (tertiles)								
2010–2015	0.69	–0.14	0.63	0.20	0.89	–0.37	0.29	0.95
2008–2009	4.75**	1.41*	1.78*	1.56**	6.07***	1.61**	1.90**	2.05***
2007 or earlier (Ref)	–	–	–	–	–	–	–	–
Previous experience with metric-based performance improvement								
Maturity of performance management system	–0.31	–0.15	0.03	–0.19	–	–	–	–
Extent of automated data collection	–2.43	0.02	–2.76***	0.31	–	–	–2.16*	1.73*
Extent of data stored in centralized database	–1.57*	–0.52	–0.58	–0.47	–	–0.47	–	–0.63*
Participation in Tufts Implementation Program								
Attendance ^y								
Training (7 sessions)	1.21	0.22	0.35	0.64**	1.05	–	–	0.66**
Coaching (6 sessions)	2.25**	0.43	1.10**	0.72*	2.00	–	1.16**	–
Coaching metric								
Careers (ref)	–	–	–	–	–	–	–	–
IRB	–1.69	–1.89**	1.55	–1.35	–	–1.87**	0.77	–
Pilots	–2.46	–1.26	–0.29	–0.91	–	–0.72	–0.77	–
Primary coach								
Coach A (Ref)	–	–	–	–	–	–	–	–
Coach B	–0.49	0.04	–0.23	–0.30	–	–	–	–

Ref = reference group (indicated by dashes in cell); CMI = Common Metrics Implementation.

*≤0.10; **≤0.05; ***≤0.01.

^yOne hub did not respond.

^cCTSA size is defined as total funding from U, T, K, and/or R grants for fiscal year 2015–2016.

^yAttendance at a training or coaching session is defined as at least one person from the hub attended. Implementation Groups 1 and 2 were offered 7 coaching sessions; Implementation Group 3 was offered 6 coaching sessions.

Table 4. Challenges to hub progress, with illustrative quotation*

<p>Hub size and resources <i>In addition to size of hub's funding award, other resource-related factors contributed to an overall lack of available time and resources.</i></p>
<p>Lack of institutional investment[†]</p> <p>So a lot of the metrics, one would certainly hope could be facilitated by informatics systems, and our university, for example, has not invested in a citation index software, that would help a lot as we are trying to find investigator publications . . . Our . . . homegrown system works really well for the IRB, but any time anything needs to be added they have to contract with informatics people . . . , [who] are a scarce resource. So that's a challenge. – <i>Principal Investigator**</i></p>
<p>Interrupted funding</p> <p>. . . [G]iven our no-cost extension status, . . . we do not know yet if we are going to . . . turn the curve because we are not awarding, for example, . . . any more pilot awards . . . or K awards right now. – <i>Implementer</i></p>
<p>Lack of adequate staffing and expertise[†]</p> <p>Well, I can tell you the problem: we only pay a fraction of [his] time for evaluation because he does other functions for us, and our staff person who works with him does not have the capability to do this herself independently. . . . Nobody really thought about what impact it was going to have on the time allocation for the leadership that was responsible for evaluation . . . – <i>Principal Investigator</i></p> <p>Well, what I would like to change is to have an expert on hand, someone who has been trained in evaluation and metric design. And not so much just adding it on to people's job descriptions, but actually having someone who could truly represent us at the level of NCATS for Common Metrics. – <i>Administrator</i></p>
<p>Alignment with needs of Common Metrics Implementation <i>Lack of alignment of local data systems or institutional priorities created difficulties for metric data collection or local investment in the initiative.</i></p>
<p>Lack of data system or an existing system that was not aligned with the Common Metrics definitions created more effort for effective tracking[†]</p> <p>. . . our information systems were not automatically and easily aligned to collect information in the form that the initial set of metrics request demanded, and so we discovered . . . that there were various kinds of gaps and holes in the way various things are tracked. – <i>Principal Investigator</i></p>
<p>Lack of alignment with institutional priorities[†]</p> <p>We have tried to make sure that the deans and other leaders know about the Common Metrics. I don't know that those three Common Metrics have been exactly their highest priority. They look at it and they are happy with it. [But] it's not like they have said, "Oh yeah, we want to adopt that Common Metric for our university over time." But it's early in the process and they may. – <i>Principal Investigator</i></p>
<p>Hub authority <i>Lack of line authority over data, processes, or organizational components related to the metrics created challenges for implementation.</i></p>
<p>Lack of line authority over key drivers</p> <p>One issue with the CTSA's, particularly in a decentralized organization like ours, is we're responsible for outcomes but do not have authority over them. It is an exercise I am trying to lead from the middle. – <i>Principal Investigator</i></p> <p>There's thousands of IRB protocols submitted to the IRB every year. We only touch a small fraction of them, so how much control do we have over time to IRB approval? And so, the cynical answer is how can we affect the 90% of IRB submissions that we have nothing to do with? – <i>Principal Investigator</i></p>
<p>Hub engagement <i>Active hub engagement was important for completing implementation activities, but several factors undercut engagement.</i></p>
<p>Annual reporting cycle induced bursts of effort</p> <p>I think a limitation has been this idea that you can report [the metrics] once a year, which is good to report to NCATS, but it is not good as a management tool . . . – <i>Principal Investigator</i></p>
<p>Interrupted funding</p> <p>Given our no-cost extension status, we realized that we would not be able to implement all action plans that we proposed or we had outlined . . . – <i>Implementer</i></p>
<p>Reduced motivation due to lack of alignment with existing processes or unclear definitions</p> <p>. . . [W]hen I ask anybody on my staff to do something, I want to make sure it's not busy work and I want to make sure it's something that we're using. . . . And so when we did a change of operations to basically . . . [compute the metric] the other way [for the Common Metrics], . . . the report at the end wasn't useful to us. . . . – <i>Administrator</i></p>
<p>Stakeholder engagement <i>Engaging needed stakeholders external to the CTSA hub was crucial for performance improvement, but securing consistent participation was challenging.</i></p>
<p>Lack of a direct line of consistent communication with other units</p> <p>Unlike some institutions, we do not manage the IRB, and we don't manage contracting, so we are always the liaison working with those entities, to try and improve their performance. – <i>Principal Investigator**</i></p>
<p>Securing initial buy-in or sustained cooperation from key stakeholders</p> <p>Well, I think we have the same problems as everybody else. You give somebody a \$50,000 pilot grant, and then they forget to cite you on papers. We preach, we give seminars, we hand out mouse pads and mugs and do all kinds of things, and put it in our emails. But people still forget . . . So it is a constant struggle . . . – <i>Principal Investigator</i></p>

*Unless stated otherwise, themes manifest in more than one way; a quotation represents one manifestation.

**Participant is affiliated with a medical center that functions as a CTSA without current CTSA funding.

[†]Indicates that the challenge, under reverse conditions, becomes a facilitator.

Table 5. Facilitators for hub progress, with illustrative quotation*

Hub size and resources <i>In addition to the size of a hub's funding award, the presence of institutional resources, needed expertise, and effective teams facilitated progress.</i>
Availability of institutional resources [†] ... we use some IT [and other] resources that are institutionally supported to actually draw metrics for the Common Metrics. Because it's so highly integrated ... we don't necessarily separate out which effort is completely supported by NIH ... [versus] contributions to that task from non-NIH dollars. <i>-Principal Investigator</i>
Adequate evaluation and other specific expertise [†] We're fortunate in having a very experienced evaluator, and that's really made the difference. If we didn't have anyone who was so skilled in the metrics and assessment, some of these would have been more challenging. <i>-Principal Investigator</i>
Leveraging extended teams [†] Of all the possible factors that I could think of that might dictate whether or not we successfully implement the Common Metrics and whether it is beneficial to us, the structure of the team that was allocated to do the work has the greatest single effect. ... I am a department of one, so I need help doing evaluation activities. So, we have evaluation liaisons in every program. We also have a huge number of people on the Common Metrics team, ... and ... a parallel group of advisers, people who were interested in the Common Metrics. <i>-Implementer</i>
Effective core team And it did help to have one person willing to become the expert at the organization. Like, there isn't much she doesn't know about [the Common Metrics] at this point. So you have to have a go-to person who is immersed in it and can really get it done. <i>-Implementer</i> We have a pretty close-knit leadership team and our evaluator meets with us weekly. So I think there's the ability to address any of that quickly ... That's a facilitator that we're working on this together collaboratively. <i>-Administrator</i>
Alignment with needs of Common Metrics Implementation <i>Alignment with local data systems and institutional priorities facilitated metric data collection and local investment in the initiative</i>
Alignment of Common Metrics with and ability to use existing data collection tools [†] I can tell you that the IRB turnaround time was already being collected by both the IRBs. The pilot program, that was part of our ongoing evaluation to begin with, as was the KL2 ... <i>-Principal Investigator</i>
Alignment with institutional priorities [†] The institution is very interested in this. So, I think that this is something the institution is highly invested in doing well on. <i>-Principal Investigator</i>
Hub authority <i>Coupling hub leadership with institutional leadership positions helped to mitigate the problem of lack of direct authority over data or processes related to Common Metrics topics.</i>
Occupying institutional and integrated leadership roles I think reporting to the Provost helps, too ... Some of these data systems are not medical school-specific, so that helps getting access to big picture systems. <i>-Principal Investigator</i> So administratively ... we are a separate center even though I am in [a clinical department] ... , and it's kind of on purpose. We also have a lot of conflation of some of the personnel, so I am going to also hold a title of Associate Dean for Research, as did my predecessor, and that's by design. <i>-Principal Investigator</i>
Hub engagement—Principal Investigator (PI) <i>Hub PIs facilitated implementation in four key ways.</i>
Providing strategic guidance [The PI] doesn't do the day-to-day numbers, but he does the critical thinking of "how could we improve this number?" or "what could we do differently?". <i>-Administrator</i>
Serving as a champion I would say our PI, I think he has the role of champion on our Common Metrics team and he has definitely ... been that. So he welcomes ... those process improvement conversations and having a sort of data-driven context that we can use to make sure we're doing our work as best we can. <i>-Administrator</i>
Facilitating stakeholder engagement Our PI worked with a lot of the stakeholders to reengage them and to emphasize that this was going to be a process that we would have to comply with and that while it required more work up front, it was not only beneficial to the CTSA but it was going to be beneficial to them to have access to the data and the analyses in the long run. <i>-Administrator</i>
Providing hands-on oversight during start-up [The PI] was pretty directly involved with our Director of Evaluation to make sure that things were rolling out according to plan. I would say, compared to a lot of our sort of day-to-day initiatives and day-to-day work, he was more hands-on with the metrics than he is with some of the other things. <i>-Administrator</i>

(Continued)

Table 5. (Continued)

Stakeholder engagement <i>Successful engagement of stakeholders external to the CTSA hub facilitated performance improvement.</i>
Personal relationships and cooperative spirit [W]hen there would be meetings and conversations about getting data, and what mechanisms were in place, some of it was based on personal relationships that then needed to be shifted a little bit, with change in personnel. <i>–Principal Investigator</i>
Integration of Common Metrics with institutional priorities This has been embraced . . . as a barometer at the institution. . . . So, for us to have to . . . look at publication data or Pilot Award data, whatever we're instrumenting for the Common Metrics for the CTSA, we basically just extend across the institution. <i>–Principal Investigator</i>
CTSA location and hub size can strengthen relationships [O]ur primary research support activities . . . are all organized out of this independent laboratory, with the advantage being that it allows us very easy access to the other independent laboratories as well as . . . the schools and departments. <i>–Principal Investigator</i> We're very advantaged as a result of our small size. So, essentially, we have virtually all of our stakeholders around the table each week . . . <i>–Principal Investigator</i>

*Unless stated otherwise, themes manifest in more than one way; a quotation represents one manifestation.

[†]Indicates that the facilitator, under reverse conditions, becomes a challenge.

(3) a collaborative team climate with effective communication (Table 5).

Not all hubs, however, had available metric-related expertise, and for many hubs, the local team was relatively small. In smaller hubs, core teams may be particularly lean and exhibit less differentiation in roles related to Common Metrics implementation. To address this, some hubs leveraged other individuals and groups within their hub and academic institution to form extended teams that facilitated completion of data collection and performance improvement activities. In a number of cases, other stakeholders became part of extended teams to facilitate regular collaboration and sustained commitment. Directors of hub programs related to Common Metrics' specific topic areas, who played critical roles due to their ownership of the data and/or familiarity with the processes in their topic areas, were considered valuable for implementing improvement strategies.

Prior experience with performance improvement and alignment with needs of Common Metrics implementation

Although we anticipated that prior experience with metric-based performance improvement would facilitate completion of such activities for Common Metrics, the quantitative measure of prior experience (maturity of a hub's performance management system) appeared to have a small negative effect. This effect disappeared after accounting for other characteristics in multivariable statistical models, but similarly unexpected effects related to existing data collection and storage appeared more robust for the IRB and Pilot Funding Metric (Table 3).

Qualitative results revealed that alignment (or lack thereof) of the Common Metrics and performance improvement framework with a hub's prior experience, systems, and priorities affected implementation (Tables 4 and 5). As noted, one type of alignment was compatibility with technical needs of the Common Metrics, including local structures, processes, metrics, and experience. If systems and processes were aligned with the Common Metrics, prior experience with similar metrics or performance improvement frameworks facilitated implementation of the Common Metrics. When there was lack of alignment with existing systems and processes, more resources were required to conduct the work of the Common Metrics, and this hampered hubs' abilities to adapt to and engage in that work. Particularly for the IRB Metric, if existing institutional data systems were not aligned with the metric

definition, modifying existing systems to follow the metrics operational guidelines absorbed a great deal of time and resources.

A second type of alignment—compatibility of Common Metrics with existing institutional priorities—also shaped hubs' progress on the work of the Common Metrics. Alignment of the Common Metrics with local priorities (or the ability to create such alignment) made the Common Metrics more useful to hubs. This facilitated institutional investment in the work. In contrast, *lack of alignment* had the opposite effect on the perceived usefulness of, and investment in, the metrics.

Hub authority

Participating CTSAAs were diverse in how they were situated relative to their academic institutions. A hub leader's position in the institutional authority structure was important for accessing needed data, affecting improvements, and facilitating stakeholder engagement. Hubs with leaders that did not have line authority over the data, processes, or organizational components related to Common Metrics experienced challenges in implementing performance improvement (Table 4). The complexity of processes related to the Common Metrics, such as investigators' response times to IRB stipulations and the need to coordinate with multiple IRBs, exacerbated this challenge.

Although the problem of lack of direct authority could not be fully mitigated, some hubs noted that coupling the leadership role of the hub principal investigator with a leadership position at the school or institutional level, and integrating leadership relationships across the institution, facilitated the work of hubs generally and the work of the Common Metrics in particular. When direct lines of communication with relevant departments or leaders were not already existing, drawing on or creating personal relationships to build communication about the topics of the Common Metrics was a strategy to help gain buy-in of stakeholders (Table 5).

Hub engagement

A hub's type of engagement with the Common Metrics was associated with the degree to which it completed the performance improvement activities. Types of hub engagement, identified through qualitative interview analyses, included actively folding Common Metrics and the performance improvement framework into standard work processes (active engagement), complying with an external requirement (compliance-based approach), or some

Table 6. Results of testing for effects of hub engagement on completion of performance improvement activities (N = 30 hubs)

Engagement category Coded from qualitative interviews	N	Hub score (Mean, SE)			
		Overall sum (0–30)	By metric		
			Careers (0–10)	IRB (0–10)	Pilots (0–10)
All active engagement: All participants report active engagement	10	22.8 (2.28)	8.1 (0.88)	6.3 (0.93)	8.3 (0.91)*
Mix: Each participant reports both active engagement and compliance approach	4	22.8 (3.60)	8.5 (1.40)	6.0 (1.47)	8.2 (1.44)
Mix: Leader reports active engagement; Implementer reports compliance approach	12	23.1 (2.08)	7.5 (0.81)	8.0 (0.85)	7.7 (0.83)
All compliance-based engagement: All participants report compliance approach (ref)	4	17.0 (3.60)	6.4 (1.40)	5.3 (1.47)	5.4 (1.44)

Ref = reference group; SE = standard error.

* $p \leq 0.10$.

mixture of these approaches within the hub and/or its staff. Not surprisingly, hubs in which all participants reported only a compliance-based approach to the Common Metrics in qualitative interviews completed fewer activities related to Common Metrics and performance improvement (i.e., had lower scores on the primary quantitative outcome) than hubs in which one or more participants reported active engagement (Table 6).

Qualitative results revealed that engagement of a hub leader in particular appeared to affect completion of activities, particularly for the IRB Duration Metric which was often outside the hub's line authority. Principal investigators played four key facilitative roles: providing strategic and operational guidance, serving as a champion who kept Common Metrics work "on the agenda," facilitating stakeholder engagement, and providing hands-on oversight during start-up (Table 5).

Challenges for maintaining higher levels of engagement included periods of little Common Metrics-related efforts due to reporting on an annual basis, interruptions to hub funding, and reduced motivation due to perceptions of unclear metric definitions and lack of alignment with existing processes.

Variation in hub engagement revealed through qualitative interviews helped to explain the statistically significant effect of funding cohort. We expected that hubs funded earlier would have more established processes and stakeholder relationships to conduct the work of Common Metrics, but hubs in the middle CTSA-funded cohort completed an average of 15% more activities than the earliest funded hubs. This effect was the largest of all characteristics measured quantitatively, and it remained statistically significant when accounting for other hub characteristics in the multivariable models. Hubs funded in the earliest and latest cohorts completed about the same number of activities. Qualitative interview results explained why the potential benefit of more established processes was tempered.

Specifically, although all funding cohorts included hubs with multiple engagement approaches, a compliance-based approach was more common among hubs funded earlier while active engagement was more common among hubs funded later (Supplemental Figure 1). Qualitative results showed that this trend indicated that hub engagement, at least in part, reflected hubs' levels of willingness or ability to adjust processes to accommodate the requirements of the Common Metrics, which differed across funding cohorts.

Hubs funded in the latest cohort were less likely to have firmly established processes, which made the introduction of a performance improvement system more useful. Yet, these hubs sometimes had difficulties with resources or contextual issues (e.g., developing relationships with stakeholders). In contrast, hubs funded in the earliest cohort more likely had established processes. If these processes were aligned with the Common Metrics, then the work was more easily completed based on existing workflows. If

their processes were not aligned, then adaptation of existing processes presented difficulties. Many hubs in the middle cohort had fewer unresolved contextual issues than those funded later (e.g., they had already built relationships with home institutions and stakeholders), and their existing processes and systems appeared not quite as firmly established as those funded earlier, making it easier to adapt to Common Metrics.

Stakeholder engagement

Engaging stakeholders is a fundamental aspect of implementing Common Metrics using the shared performance improvement framework. Qualitative results showed that challenges for engaging stakeholders included lack of an existing line of consistent communication with other units in a hub's academic institution, difficulty securing initial buy-in, or sustaining cooperation over time (Table 4). Difficulty with initial buy-in resulted from resistance or "pushback" from stakeholders or from the hubs' hesitancy to involve stakeholders due to an expectation of resistance.

Facilitators included personal relationships (existing or new collaborations), a culture of cooperation in the academic institution, integration of the Common Metrics with institutional priorities, and structural features of hubs that supported access to institutional leaders and stakeholders (e.g., physical location and size). Line authority over the relevant domain also facilitated engaging stakeholders.

Hubs also identified proactive strategies for enhancing their abilities to successfully engage stakeholders in the Common Metrics (Supplemental Table 7). First, as relevant stakeholders varied by metric, persuading each set of stakeholders about the benefit to them from helping implement the Common Metrics and performance improvement plans was key. Second, creating avenues for discussion, dialogue, and feedback with stakeholders was important, including listening to stakeholders at the "ground level," not only leaders. Third, engagement of stakeholders may require persistence, both initially and over time. Fourth, positioning the CTSA hub as a "bridge" or "liaison" to engage stakeholders across the institution helped, even at times incorporating key stakeholders from other parts of the institution into roles within the hub to ensure engagement.

Training and coaching attendance

Hub attendance at training and coaching sessions provided by the Tufts Implementation Program appeared related to completion of activities according to statistical analyses (Table 3). As the number of training and coaching sessions attended by at least one hub team member increased, the average number of completed activities also increased. This trend was statistically significant for hubs that attended more coaching sessions. The benefit of attending more training and coaching sessions appeared to differ by metric.

Statistical results suggest that receiving coaching on a metric facilitated the completion of performance improvement activities for that metric. For the Careers and IRB Review Duration Metrics, receiving coaching while working on that metric was associated with completing more of the related performance improvement activities. For the Careers Metric, hubs that did not focus on this metric during coaching completed fewer of the related activities by the end of the evaluation. Although not statistically significant, hubs that focused on the IRB Metric during coaching completed about 1.55 more activities (out of 10) on the IRB Metric compared to hubs that focused on the Careers Metric during coaching.

Discussion

This mixed methods evaluation assessed progress in implementing Common Metrics and a shared performance improvement framework across the CTSA Consortium, a loosely integrated network of academic health care institutions, or hubs, charged with catalyzing clinical and translational research. After 19 months, the vast majority of hubs reported that they computed results for the initial set of three Common Metrics and undertook activities to understand current performance, but fewer hubs developed and carried out performance improvement plans for all metrics. Similar to performance management efforts in loosely integrated public health programs [8], heterogeneity across hubs' local contexts affected implementation of Common Metrics and performance improvement activities across hubs.

The most common reason cited for not completing an activity was limitation of available resources. Although the size of the hub's funding award played a limited role, other resource-related factors, such as investment from home institutions, periods of interrupted funding, availability of needed personnel and expertise, and effectiveness of core teams, varied across hubs and affected whether they could devote sufficient time and resources to fully implement Common Metrics and performance improvement activities.

Across hubs, alignment (or lack thereof) of the Common Metrics and performance improvement framework with a hub's local conditions and needs affected implementation. If existing local systems and processes were aligned with the needs of the Common Metrics Initiative, prior experience with similar metrics and/or performance improvement frameworks facilitated implementation. Without such alignment, more resources were required for implementation, and this hampered hubs' abilities to adapt to and engage in that work. Similarly, alignment of the Common Metrics with existing institutional priorities (or the ability to shape such alignment) made the initiative more useful to hubs and facilitated institutional investment in the work. In contrast, *lack* of this type of alignment had the opposite effect on the perceived usefulness of, and investment in, the metrics.

A hub leader's position in the institutional authority structure was important for accessing needed data, affecting improvements, and facilitating stakeholder engagement. Hubs with leaders who did not have line authority over the data or processes related to Common Metrics experienced challenges in implementation. Drawing on or creating personal relationships to build communication about the topics of the Common Metrics was a strategy to help gain buy-in of stakeholders.

Hubs also varied in their approach to engaging with Common Metrics work—including active engagement, a compliance-oriented approach or a mix—and this was associated with the degree to which hubs completed the performance improvement activities.

Not surprisingly, hubs in which all participants reported only a compliance-based approach completed fewer implementation activities than those in which one or more participants reported active engagement. The engagement of hub principal investigators was found to be important, particularly to provide strategic guidance and oversight, champion the project, and facilitate stakeholder engagement.

Attending training and coaching sessions, and opportunities for hubs to share experiences and best practices, were helpful for hubs. Although there was evidence of facilitation by these services, not completely clear is whether this related to the content of the training and coaching, the difficulty of the metric the hub was focusing on during coaching, or differences among hubs that chose to receive coaching on one metric rather than another.

Limitations

The design of the Common Metrics Implementation Program necessitated a descriptive evaluation study that focused on understanding hubs' progress and experiences. First, a controlled comparison group design was not compatible with the goal of having every hub implement the Common Metrics and a shared performance improvement framework to the fullest extent possible during the same time period. Second, without a control group, we considered a quasi-experimental pre-post design, but could not fully pursue this option because assessing change in Common Metrics' results was not feasible for two reasons. The metric definitions were newly released and not all hubs had retrospective data to compute the metric result for a prior time period, and even if hubs could collect retrospective data, the anticipated timeframe for achieving change in the metric results was longer than the study period. The resulting mixed methods approach yielded a multifaceted understanding of hubs' progress and related contextual factors, challenges, and facilitators.

Conclusion

Implementing Common Metrics and performance improvement in a large, loosely integrated network of research-focused organizations, the CTSA Consortium, proved feasible, but it required substantial time and resources. There was considerable contextual heterogeneity across hubs in their data systems, existing processes and personnel, organizational structures, and local priorities of home institutions, which created disparate experiences and approaches across hubs. To sustain engagement, future metric-based performance management initiatives should anticipate, and facilitate solutions to, barriers to implementation due to resources and authority and, for heterogeneous networks, account for local contexts. Future efforts should also consider the perceived value of the initiative, which is addressed for the CTSA Consortium's Common Metrics Initiative in a separate report.

Supplementary material. To view supplementary material for this article, please visit <https://doi.org/10.1017/cts.2020.517>.

Acknowledgements. The authors are grateful for the time and effort of many people who contributed to the Tufts Common Metrics Evaluation Study. CTSA Consortium hubs across the country invested resources, personnel, and time to implement the Common Metrics and provide data for the evaluation study. Debra Lerner, MS, PhD, provided expertise in study design and survey research. Annabel Greenhill provided valuable research assistance. Members of the Tufts CTSA Common Metrics Implementation Team, including Denise Daudelin, Laura Peterson, Mridu Pandey, Jacob Silberstein, Danisa Alejo, and Doris

Hernandez, designed and carried out the implementation program that is evaluated by this study.

The project described was supported by an administrative supplement award to Tufts CTSI from the National Center for Advancing Translational Sciences, National Institutes of Health (UL1TR002544 and UL1TR001064). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Center for Advancing Translational Sciences, the National Institute on Drug Abuse, or the National Institutes of Health.

Disclosure. The authors have no conflicts of interest to declare.

References

1. **Committee to Review the Clinical and Translational Science Awards Program at the National Center for Advancing Translational Sciences; Board on Health Sciences Policy; Institute of Medicine; Leshner AI, Terry SF, Schultz AM, Liveman CT**, editors. *The CTSA Program at NIH: Opportunities for Advancing Clinical and Translational Research*. Washington, DC: National Academies Press, 2013. doi: [10.17226/18323](https://doi.org/10.17226/18323).
2. **Pannick S, Sevdalis N, Athanasiou T**. Beyond clinical engagement: a pragmatic model for quality improvement interventions, aligning clinical and managerial priorities. *BMJ Quality & Safety* 2016; **25**: 716–725.
3. **Patrick M, Alba T**. Health care benchmarking: a team approach. *Quality Management in Health Care* 1994; **2**: 38–47.
4. **Catuogno S, et al.** Balanced performance measurement in research hospitals: the participative case study of a haematology department. *BMC Health Services Research* 2017; **17**: 522.
5. **Moxham C**. Understanding third sector performance measurement system design: a literature review. *International Journal of Productivity* 2014; **63**: 704–726.
6. **Northcott D, Taulapapa T**. Using the balanced scorecard to manage performance in public sector organizations. *International Journal of Public Sector Management* 2012; **25**: 166–191.
7. **Tari JJ, Dick G**. Trends in quality management research in higher education institutions. *Journal of Service Theory and Practice* 2016; **26**: 273–296.
8. **DeGroff A, et al.** Challenges and strategies in applying performance measurement to federal public health programs. *Evaluation and Program Planning* 2010; **33**: 365–372.
9. **Friedman M**. *Trying Hard Is Not Good Enough: How to Produce Measurable Improvements for Customers and Communities*. Victoria, BC: FPSI Publishing; 2005.
10. **Daudelin D, et al.** Implementing Common Metrics across the NIH Clinical and Translational Science Awards (CTSA) Consortium. *Journal of Clinical and Translational Science* 2020; **4**: 16–21.
11. **Fetters MD, Curry LA, Creswell JW**. Achieving integration in mixed methods designs—principles and practices. *Health Services Research* 2013; **48**(6): 2134–2156.
12. **Tomoaia-Cotisel A, et al.** Context matters: the experience of 14 research teams in systematically reporting contextual factors important for practice change. *Annals of Family Medicine* 2013; **11** (Suppl. 1): S115–S123.
13. **Harris PA, et al.** Research electronic data capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, 2009; **42**: 377–381.
14. **QSR**. *NVivo qualitative data analysis software, Version 10*. Melbourne, Australia: QSR International Pty Ltd, 2012.