## **CTSA Consortium Common Metrics Initiative**

# Report by Tufts CTSI on Common Metrics Implementation Program and Evaluation Study



September 2018

## Contents

ACKNOWLEDGEMENTS iv ABBREVIATIONS vi LIST OF TABLES vii LIST OF FIGURES ix LIST OF APPENDICES x

## **Executive Summary**

Introduction 2 Selecting and Developing Metrics 2 Improving Performance 4 Addressing Barriers and Sustaining Engagement 6 Expanding Data-Driven Decision Making 7 Learnings about the First Three Common Metrics 8 Clinical Trial Accrual Metric Pilot Test 9 Conclusion 9

## Full Report of the Common Metrics Implementation Program and Evaluation Study Introduction 16

PART I.

## IMPLEMENTING COMMON METRICS AND ASTRATEGIC MANAGEMENT FRAMEWORK20

#### Selecting and Developing Metrics 20

#### Improving Performance 21

Effective Teams 21

Training, Coaching and Technical Assistance 22

Collecting Data and Developing Strategic

Management Plans 29

Disseminating Performance Drivers and Strategies for Improvement 31

1

15

#### Metric Data Quality 33

Data Checking 33

#### Learnings About The First Three Common Metrics 36

Metric: IRB Review Duration 36

Metric: Pilot Funding Publications 38

Metric: Careers in Clinical and Translational Science 40

#### Clinical Trial Accrual Metric Pilot Test 43

Background 43

Methods 43

Results 44

Transition to the Coordinating Center 48

## PART II. EVALUATION STUDY

**49** 

#### **Research Questions 50**

#### Methods 51

Study Design 51

Data and Analysis 53

#### Results 68

Description of Hubs 68

Hub Progress in Implementing Common Metrics and Performance Improvement Activities 70

Understanding Hub Progress and Experiences 79

Role of Tufts Implementation Program 110

Feedback on Results Based Accountability Framework and Scorecard Software 113

Perceived Value of Common Metrics Implementation 116

Additional Concerns and Hub Recommendations to NCATS 123

## Discussion 127

Hub Progress 127

Factors Affecting Implementation 127

Strategies to Facilitate Performance Improvement 129

Perceived Usefulness and Added Value 131

Limitations and Future Opportunities 133

## INTEGRATED SUMMARY AND CONCLUSIONS 135

Selecting and Developing Metrics 135 Improving Performance 136 Addressing Barriers and Sustaining Engagement 136 Expanding Data-Driven Decision Making 137 Metric-Specific Opportunities to Improve 137 First Three Common Metrics 137 Pilot Test: Clinical Trial Accrual Metric 138

References 140 Appendices 141

## ACKNOWLEDGEMENTS

We would like to acknowledge the significant investment of time and effort from many groups involved in the Common Metrics Initiative. Of particular mention, each CTSA devoted resources and personnel to implement Common Metrics. Members and staff of the Common Metrics Initiative Principal Investigator Working Groups and original Pilot Test hubs, the Median Accrual Metric Development Team and Accrual Metric Pilot Test hubs, the Vanderbilt University Medical Center CTSA Consortium Coordinating Center (C4), and the University of Rochester Center for Leading Innovation and Collaboration (CLIC) Coordinating Center provided expertise and feedback to enable the implementation process. Redonna Chandler, PhD, Meryl Sufian, PhD, Patricia Jones, DrPH, MPH and Samantha Jonson, MPS, of the National Center for Advancing Translational Science (NCATS) provided leadership and input for guiding implementation and evaluation. Members of the Tufts CTSI Common Metrics Implementation and Evaluation Teams, Quality Improvement Specialist Laura Peterson, RN, MS, and Results Based Accountability Consultant Phil Lee, JD, MPP, of Clear Impact, LLC, were integral to the design and implementation of the project. Finally, Andrada Tomoaia-Cotisel, Peter Mendel, Jason Etchegaray, and Nabeel Qureshi from RAND Health led the qualitative component of the evaluation study and provided key information for this report.

The final content of the report is solely the responsibility of the writing team (comprised of the Principal Investigator, Project Director, Quality Improvement Specialist, and Evaluation Study Co-Investigator). This team combined insights and findings from two components of the project (Common Metrics Implementation activities and an evaluation study) into an integrated report with unified recommendations that collectively represent the voices of hub participants, the experience of the Implementation Team, and analytic research results.

## **List of Participating Sites**

Institute	Ge
Case Western Reserve University Clinical and Translational Science Collaborative	Ha
Clinical and Translational Science Institute at Children's National	Ha
Colorado Clinical and Translational Sciences Institute	Inc
Columbia University Irving Institute for Clinical and Translational Research	Joł I
Dartmouth SYNERGY Clinical and Translational Science Institute	Ma
Duke Clinical and Translational Science Institute	Me
Emory University – Georgia Clinical and Translational Science Alliance	Ţ
Frontiers: University of Kansas Clinical and	Mi
mansiational science institute	INC

- Georgetown-Howard Universities Center for Clinical and Translational Science
- Harold and Muriel Block Institute for Clinical and Translational Research at Einstein and Montefiore\*
- Harvard Catalyst The Harvard Clinical and Translational Science Center
- Indiana Clinical and Translational Sciences Institute\*
- Johns Hopkins Institute for Clinical and Translational Research<sup>§</sup>
- Mayo Clinic Center for Clinical and Translational Science
- Medical College of Wisconsin Clinical and Translational Science Institute of Southeast Wisconsin
- Michigan Institute for Clinical and Health Research
- New York University Langone Health Clinical and Translational Science Institute

North Carolina Translational and Clinical Science Institute<sup>§</sup>

Northwestern University Clinical and Translational Science Institute

Oregon Clinical and Translational Research Institute<sup>§</sup>

Pennsylvania State Clinical and Translational Science Institute<sup>§</sup>

Scripps Research Translational Institute

South Carolina Clinical and Translational Research Institute<sup>§</sup>

Southern California Clinical and Translational Science Institute

Spectrum – The Stanford Center for Clinical and Translational Research and Education

The Institutes for Translational Sciences at Icahn School of Medicine at Mount Sinai

- The Ohio State University Center for Clinical and Translational Science
- The Rockefeller University Center for Clinical and Translational Science
- Tufts Clinical and Translational Science Institute\*§
- University at Buffalo Clinical and Translational Science Institute

University of Alabama Birmingham Center for Clinical and Translational Science

University of Arkansas for Medical Sciences Translational Research Institute

University of California Davis Clinical and Translational Science Center

University of California Irvine Institute for Clinical and Translational Science

University of California Los Angeles Clinical and Translational Science Institute

University of California San Diego Altman Clinical and Translational Research Institute

University of California San Francisco Clinical and Translational Science Institute

University of Chicago Institute for Translational Medicine

University of Cincinnati Center for Clinical and Translational Science and Training

University of Florida Clinical and Translational Science Institute

University of Illinois at Chicago Center for Clinical and Translational Science

University of Iowa Institute for Clinical and Translational Science

University of Kentucky Center for Clinical and Translational Science<sup>§</sup> University of Massachusetts Medical School Center for Clinical and Translational Science

University of Miami Clinical and Translational Science Institute

University of Minnesota Clinical and Translational Science Institute

University of New Mexico Health Sciences Clinical and Translational Science Center

- University of Pennsylvania
- University of Pittsburgh Clinical and Translational Science Institute\*
- University of Rochester Clinical and Translational Science Institute

University of Texas Health Science Center San Antonio – Texas Regional CTSA Consortium

University of Texas Health Sciences Center Houston – Texas Regional CTSA Consortium

University of Texas Medical Center Galveston – Texas Regional CTSA Consortium

University of Texas Southwestern Medical Center – Texas Regional CTSA Consortium

University of Utah Center for Clinical and Translational Science

University of Washington Institute of Translational Health Sciences

- University of Wisconsin Madison Institute for Clinical and Translational Research
- Vanderbilt University Medical Center Institute for Clinical and Translational Research
- Virginia Commonwealth University Center for Clinical and Translational Research
- Wake Forest School of Medicine Clinical and Translational Science Institute

Washington University in St. Louis Institute of Clinical and Translational Sciences

Weill Cornell Medicine Clinical and Translational Science Center

Yale Center for Clinical Investigation

University in St. Louis Institute of Clinical and Translational Sciences

\* Pilot sites for the first three metrics

§ Pilot sites for the Accrual Metric

## **ABBREVIATIONS**

CLIC	Center for Leading Innovation and Collaboration
CTMS	Clinical Trial Management System
CTR	Clinical and Translational Research
CTSA	Clinical and Translational Science Awards
CTSI	Clinical and Translational Science Institute
IRB	Institutional Review Board
NCATS	National Center for Advancing Translational Science
NIH	National Institutes of Health
QIS	Quality Improvement Specialist
PI	Principal Investigator
RBA	Results-Based Accountability
TTC	Turn the Curve Plan

## LIST OF TABLES

- 1. Summary and detailed recommendations 10
- 2. Number of hubs and self-selected coaching metric by Implementation Group 24
- 3. Implementation of Common Metrics and performance improvement activities: definitions and point assignments 54
- 4. Self-assessed ability to manage performance 55
- 5. Data collection instruments and chronology 56
- 6. Baseline and follow-up survey topics 57
- 7. Semi-structured interview topics 59
- 8. Semi-structured interviews: sampling frame 62
- 9. Semi-structured interviews: final sample of hubs 63
- 10. Hub characteristics used in linear regression modeling, by domain 65
- 11. Survey items used in factor analysis 66
- 12. Hub attributes and related experience, overall and by Implementation Group 69
- 13. Hub participation in Tufts Implementation Program, overall and by Implementation Group 70
- 14. Completion of Common Metrics and performance management activities 72
- Association of interim assessment and completion of Common Metrics and performance improvement activities by January 2018 73
- 16. Completion for individual activities, overall and by metric 74
- 17. Most common reasons for lack of completion, by activity (three metrics combined) 76
- 18. Reasons for lack of completion of activities, overall and by metric 78
- Results of testing for effects of hub characteristics on completion of performance improvement activities 80
- 20. Challenges and facilitators for implementation: funding 83
- 21. Challenges and facilitators for implementation: data system resources 84
- CTSA-funded personnel with important roles in most recent update (up to five) and approximate hours spent 85
- 23. Hub reasons that approximated hours are under- or over-estimated 86
- 24. Challenges and facilitators for implementation: personnel 88
- 25. Challenges for maintaining a higher level of engagement 92
- Results of testing for effects of hub engagement on completion of performance improvement activities 93
- 27. Challenges and facilitators for implementation: availability of accurate data 96
- 28. Challenge and facilitator for implementation: starting with "low-hanging fruit" 97
- 29. Challenges and facilitators for implementation: metric clarity and usefulness 98
- 30. Challenges and facilitators for implementation: engaging stakeholders 102
- 31. Strategies for engaging stakeholders 103

- 32. Challenges and facilitators for implementation: hub authority and control 104
- 33. Facilitators for implementation: core team 106
- 34. Facilitators for implementation: principal investigator involvement 107
- 35. Hub characterizations of main activities involved in most recent update 108
- 36. Hubs' anticipated changes to their approaches 109
- 37. Challenge and facilitator for implementation: Tufts Implementation Program 112
- 38. Benefits and limitations of Results Based Accountability framework and Scorecard software 114
- 39. Desired improvements for the Scorecard software 115
- 40. Hub concerns about the Common Metrics Initiative 124
- 41. Hub recommendations for Common Metrics Implementation 126
- 42. Summary recommendations 139

## LIST OF FIGURES

- 1. Timeline for Common Metrics implementation and evaluation activities 17
- 2. Common Metrics Implementation and Evaluation project teams, leadership 18
- 3. Common Metrics Implementation training and coaching timeline 22
- 4. Results Based Accountability framework 23
- 5. Evaluation of implementation training 26
- 6. Percent of hubs that produced a metric value by the end of the coaching period 29
- 7. Percent of hubs meeting criteria for meaningful application of RBA during the coaching period 31
- 8. Timeline for Common Metrics evaluation study 49
- 9. Conceptual framework for evaluation study 52
- 10. Flow of survey items related to implementation of Common Metrics and performance improvement activities, with allocated points 58
- 11. Data collection instruments and response rates 61
- 12. Completion of Common Metrics and performance management activities per hub: three metrics combined 71
- 13. Completion of Common Metrics and performance management activities by metric 71
- 14. Reasons for lack of completion: all metrics combined 77
- 15. Variation in personnel and hours 87
- 16. Variation in time and resources by metric 87
- 17. Hubs' self-described engagement 90
- 18. Association of hub engagement and funding cohort 93
- 19. Satisfaction with Tufts Implementation Program 110
- 20. Effectiveness of Tufts Implementation Program 110
- 21. Knowledge and proficiency gained from Tufts Implementation Program 111
- 22. Self-assessment over time 116

## LIST OF APPENDICES

- A. Hub experiences with specific elements of the Tufts Implementation Program 142
- B. Tufts Implementation Team's interim assessment tool for meaningful application of the Results Based Accountability framework 144
- C. Common Metrics IRB Review Duration change package 146
- D. Common Metrics Pilot Publications change package 157
- E. Common Metrics Careers in CTR change package 171
- F. Collaborative Learning Session agendas 187
- G. Tufts Implementation Team's decision rules for data quality checks of hub metric results 188
- H. Transition timeline 190
- I. Self-reported lack of completion, by activity 191
  - All metrics combined 191
  - Careers metric 192
  - IRB Review Duration metric 193
  - Pilot Funding Publications metric 194
- J. Reasons for lack of completion, by activity 195
- K. Reasons for lack of completion, by reason 200
- L. Combined challenges for implementing Common Metrics and performance improvement activities 202
- M. Combined facilitators for implementing Common Metrics and performance improvement activities 206
- N. Paired facilitators and challenges 211
- O. CTSA-funded personnel and hours spent 212
- P. Effect of hub engagement on completion of performance improvement activities, overall and by metric 214
- Q. Role of Tufts Implementation Program 216
- R. Hub self-assessment over time 218

# **Executive Summary**

Common Metrics Implementation Program and Evaluation Study



## INTRODUCTION

The Clinical and Translational Science Award (CTSA) Program Consortium has been charged with accelerating and improving clinical and translational research on a national scale (https://ncats.nih.gov/ctsa/about/hubs). In the Congressionally mandated 2013 Report on the CTSA Program, the Institute of Medicine (IOM) was very supportive of the program, but recommended that CTSAs become a "more tightly integrated network that works collectively." This included a call for standardized evaluation processes based on measurable strategic goals and uniform, actionable "common metrics" to enhance transparency and accountability in decision-making. This was seen as key for powering needed evolution of CTSAs and their research communities. In response, the National Institutes of Health (NIH) National Center for Advancing Translational Science (NCATS), the home of the CTSA Program, and the CTSA institutional hubs nationally, implemented the Common Metrics Initiative. Using standardized metrics and the principles of the Results Based Accountability (RBA) performance management framework,<sup>2</sup> this initiative aims to improve the strategic management of individual CTSA hubs and of the national CTSA Consortium. This initiative also was seen as an opportunity to develop, demonstrate, and disseminate methods of improving "the science of doing science."

Having a focus of research process improvement, Tufts Clinical and Translational Science Institute (CTSI) was asked by NCATS to implement the Common Metrics and the RBA framework across the CTSA Consortium, and to run an evaluation study of this implementation, for three years, starting September 2015. This report synthesizes results and recommendations both from the Tufts Common Metrics Implementation Program and from the Tufts Common Metrics Evaluation Study. Summary recommendations are included below in the text; more specific recommendations related to each are in Table 1 at the end of this Executive Summary.

The primary goal of this report is to inform decision-making on future directions of the CTSA Program Common Metrics Initiative. However, it also may be of interest to other research groups or networks implementing standardized metrics and performance improvement processes, potentially including other NIH institutes and centers. Additionally, as a perturbation of the complex operations of the CTSA Consortium and its hubs, the responses to, and successes with, the Common Metrics Initiative may provide insights into the characteristics and operation of such clinical and translational science enterprises, and how they may be enhanced.

## SELECTING AND DEVELOPING METRICS

The initial development of the metrics was outside the scope of the Tufts projects. However, the Tufts Implementation Program and Evaluation Study highlighted the importance of pilot testing and continuously reviewing metrics, considering local usefulness, and ensuring the validity of aggregating results across hubs.

### **Metric Development and Ongoing Review**

In collaboration with NCATS and CTSA hubs, the Tufts Implementation Team conducted two types of pilot testing. The first three common metrics (Careers in Clinical and Translational Research, Pilot Funding Publications, and Institutional Review Board [IRB] Review Duration) were pilot-tested for data collection feasibility. A fourth metric (Clinical Trial Accrual) was tested for data collection feasibility and for usefulness for strategic management. Although both types of pilot testing identified gaps in metric definitions, the more robust approach also uncovered confusion about metric calculations and challenges to the usefulness of the metrics for strategic management that were important to address before widespread implementation.

## **Summary Recommendation 1:**

Develop metrics using robust pilot testing, and engage stakeholders in ongoing review.

## **Metric Usefulness**

### **Usefulness to Local CTSA Hubs**

Hubs experienced value in implementing Common Metrics, but continued to have concerns about whether the metrics provided enough benefit to justify their effort. Hubs founds value in three ways: 1) implementing a formal structured process if one did not exist previously, 2) using the performance improvement process to enable strategic conversations, and 3) making improvements in processes and immediate outcomes. They also provided external requirements that helped hubs justify recommendations for institutional changes and/or targeted funding.

However, participants in qualitative interviews at more than two-thirds of CTSA hubs expressed concerns about the usefulness of the first three metrics and their overall value relative to the effort expended. Many hubs found it difficult to use the metrics for local improvement, particularly when the metrics did not align with local institutional priorities or addressed topics on which the hub was already performing well.

### Usefulness to the National CTSA Consortium

Hubs reported continuing concerns about between-hub variation in how metric data would be collected and computed. To be used for benchmarking, comparison, or aggregate reporting, metric data must be comparable. However, during the period in which Tufts led the implementation process, two indicators raised questions about comparability across hubs. First, for calendar year 2015 metric results, 80% of hubs revised data for one or more metrics after originally entered. This suggested a changing understanding or approach to calculating metric results. Second, there was limited ability to assess data quality or the extent that hubs correctly followed Operational Guidelines because the submission of raw data or contextual information was not required. In addition, meaningful interpretation of the range of metric values across hubs would require understanding the institutional, hub, and program-level characteristics that shape metric values.

## **Summary Recommendations 2 and 3:**

Maximize usefulness to hubs by selecting metrics that align with local needs. Maximize usefulness to the National CTSA Consortium by ensuring validity of aggregation and comparison reporting.

## **IMPROVING PERFORMANCE**

## **Training, Coaching and Technical Assistance**

Members of each hub received interactive webinar-based training and participated in small group coaching sessions during implementation of the initial Common Metric of their choice. Additional support included ad hoc individualized coaching, a Help Desk, worksheets to assist in calculating metric results, and exemplar strategic management plans ("Turn the Curve plans"). Small group coaching sessions and regular coach assessments of Turn the Curve plans were effective in gauging team progress, tailoring coaching sessions, and sharing experiences across hubs. Participants appreciated the opportunity coaching sessions provided for peer-to-peer learning and some valued a structured approach to meeting project milestones.

Eighty percent of hubs reported satisfaction with the training, coaching, and technical assistance received, and hubs that used these offerings indicated they facilitated their work. The vast majority of hubs reported that they gained the knowledge and proficiency they needed, or more, to carry out the work of the Common Metrics. However, almost one-third of hubs reported gaining more knowledge and proficiency than was needed, suggesting that hubs perceived different levels of need.

## **Summary Recommendation 4:**

Equip hubs to fully implement each metric and performance management by providing peer-to-peer learning and training, coaching, and assistance for varying levels of experience.

## **Support for Implementation**

#### **Effective Teams**

Each CTSA hub formed a core Common Metrics team to oversee data collection and the use of the metric for strategic management. Although team composition was often in line with the Implementation Team's guidance on team membership (e.g., included leadership, project management, and data system and subject matter experts), some hubs appeared to delegate responsibility for Common Metrics implementation to a small number of individuals and/or did not include subject matter experts with specialized knowledge of, or influence over, metric topic(s).

#### **Collecting Data and Developing Strategic Management Plans**

Implementing Common Metrics entailed collecting metric data and developing metricspecific strategic management plans. By the end of the evaluation period, the vast majority of hubs self-reported that they had computed metric results according to the Operational Guidelines and had completed activities to understand "the story behind" current performance. However, hubs experienced challenges in completing each element of the RBA performance improvement framework, some specific to the particular element and others more general across the improvement process. On average, hubs developed performance improvement plans for 77% of the metrics, with fewer plans for the IRB Review Duration metric than for the Careers or Pilot Funding Publications metrics.

### **Identifying Targets for Performance**

A number of hubs were interested in identifying benchmarks to help understand their performance and progress, and to prioritize areas for improvement. However, some participants were concerned about lack of comparability of metric results across hubs that would undermine their use for comparing. Some were also concerned about results being used by NCATS to judge hubs' performance.

#### **Disseminating Performance Drivers and Strategies for Improvement**

Driver Diagrams, Change Packages, Collaborative Learning Sessions, and publication of hub *Success Stories* were used to disseminate best and promising practices for metric strategic management. Hub teams expressed considerable interest in continuing to share challenges and learning from each other about proven and potentially successful strategies to improve performance for Common Metric topics.

#### Software Support

Hubs used Clear Impact's proprietary Scorecard software to report their metric values and to document, manage, and communicate strategic management plans. Hubs recognized the value of a common software platform but reported technical limitations that hindered its benefit. The Scorecard software platform was not configured to support data collection, data management, or quality checks for metric results. Also, reporting and analytic features were limited, and hubs indicated the desire for enhanced user experience and visualization capabilities.

## **Summary Recommendations 5 and 6:**

Support implementation by promoting metric-specific teams, allowing for capacity-building periods, providing accurate benchmarks, and updating performance drivers and best practices.

Maximize usefulness of the reporting platform by enhancing functionality, visualization options, and user experience.

## ADDRESSING BARRIERS AND SUSTAINING ENGAGEMENT

Not surprisingly, those hubs reporting active engagement in implementing Common Metrics completed more performance improvement activities than those reporting only a compliance-based approach. Many factors could affect a hub's level of engagement.

## **Resources and Personnel**

Availability of resources was the most common reason hubs cited for not completing Common Metrics activities. Of note, the size of a hub's funding award did not fully account for this. Challenges related to investment from home institutions, interrupted funding, alignment of existing systems with the needs of the Common Metrics, and 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.

## **Local CTSA Program Hub Context**

CTSA organizations are heterogeneous in their structures, organizational processes, and experience with metric-based performance improvement. Alignment with technical needs of the Common Metrics Implementation, especially compatibility with local structures, processes, metrics, and experience, facilitated completion of the work. When there was lack of alignment in these areas, more resources were required to conduct the work of the Common Metrics, and this hampered hubs' abilities to adapt and engage in that work. A second type of alignment, compatibility of Common Metrics with existing institutional priorities, also shaped hubs' progress on the work of the Common Metrics.

## **Local Authority**

A CTSA hub leader's position in their home institutional authority structure was important for accessing needed data, affecting improvements, and facilitating stakeholder engagement. Hubs whose leaders did not have line authority over data or processes related to Common Metrics experienced challenges in implementing performance improvement. Drawing on or creating personal relationships to build communication about the topics of the Common Metrics was helpful for gaining buy-in by stakeholders. However, this did not fully compensate for lack of direct authority.

## **Effective Communication**

Communication strategies included website postings, e-newsletter updates, presentations, conference panels, and roundtable discussions, and these appeared to provide an appropriate level of communication for initial implementation efforts. Communication challenges included ensuring that newly participating and existing staff were aware of how to access project resources and received initiative communications. Some hubs also indicated a need for more information about future plans for the Common Metrics Initiative, particularly when weighing the overall value of the initiative.

## **Summary Recommendation 7:**

Sustain engagement by facilitating solutions to barriers due to resources and authority, accounting for hub heterogeneity, and ensuring effective communication.

## **EXPANDING DATA-DRIVEN DECISION MAKING**

The Tufts Implementation and Evaluation efforts revealed many structural and cultural aspects of CTSA organizations that affected hubs' abilities to engage with and complete Common Metrics and performance improvement activities. These findings suggest two opportunities to expand data-driven decision making in the CTSA Program. First, several hubs indicated a desire to learn from other available data and research results in order to inform and enhance the Common Metrics Initiative. These hubs spoke of using clinical and operational data beyond that needed to calculate the metric result to inform the choice of Common Metrics and elucidate drivers of performance.

Second and more broadly, insights about structural and cultural factors affecting implementation of Common Metrics likely apply to other CTSA-wide initiatives as well. Expanding "the science of doing science" approach to governing the CTSA Program would address additional questions more deeply. For example, are there "best practices" for CTSA organizational structures and ways of linking to home institutions that expedite performance improvement or other types of initiatives? How did implementation of Common Metrics affect CTSAs and their relationships with home institutions? Are there metrics or processes that are commonly used across hubs that could inform selection of future Common Metrics or other priority initiatives?

## **Summary Recommendation 8:**

Expand use of data to inform future directions of the Common Metrics Initiative and the CTSA Program.

## LEARNINGS ABOUT THE FIRST THREE COMMON METRICS

Implementing the first three metrics and input from hubs during the evaluation study revealed questions about the usefulness of these metrics. As described above, there is an important opportunity to review the metrics with hubs to ensure they are useful relative to required effort and to address any needed modifications to the Operational Guidelines.

## **Metric: IRB Review Duration**

Hubs continued to have questions about, and differences in, how they applied the Operational Guideline definitions. They also were challenged in developing strategic management plans in conjunction with the IRB, which often required crossing organizational boundaries. Usefulness of the IRB metric at the local level varied depending on the number of IRBs, the types of protocols reviewed, the extent of reliance on central IRB models, and the work process of reviews for ethics, feasibility, and budgets/contracts.

## **Metric: Pilot Funding Publications**

Hubs identified two challenges with interpreting and using this metric. First, the metric is cumulative and some hubs reported that this made interpreting metric results at the hub level difficult and not useful for measuring improvement, particularly for hubs with large numbers of pilot awards. Regarding local usefulness, some hubs considered the Pilot metric to be too narrow in scope to capture the goals of their local programs. Specifically, there was disagreement as to whether publication was a primary desired outcome of pilot awards.

### **Metric: Careers in Clinical and Translational Science**

The Operational Guideline provided examples of what it means to be "engaged in research" rather than a required definition, and there was disagreement with, or confusion about, several of the metric exclusion criteria. As a result, hubs used a range of definitions in their data collection, which raised questions about comparability of metric results across hubs. Additionally, a number of hubs reported that the cumulative nature of these metrics made interpretation of metric results difficult at the hub level, and some hubs considered the Careers metric to be too narrow in scope to capture the goals of their local programs.

## **Summary Recommendation 9:**

Make improvements to the first three Common Metrics by clarifying Operational Guidelines and assessing usefulness with hubs.

## **CLINICAL TRIAL ACCRUAL METRIC PILOT TEST**

Both hubs with and without clinical trial management systems (CTMSs) faced numerous challenges collecting the metric data and developing strategic management plans. Only one of eight pilot hubs for this metric was able to assess the accrual ratio for all its eligible trials, and all hubs had difficulty creating a central list of trials at their institution to use for a sampling frame. At many hubs, existing data sources did not align with the inclusion/exclusion criteria and variable definitions in the Operational Guideline, or the hubs did not have the needed variables. The exclusion of trials with fewer than 10 targeted participants removed many otherwise potentially-eligible clinical trials from the sampling frame. Concerns about data quality limited the usefulness of the metric for strategic management.

## **Summary Recommendation 10:**

Use the results and recommendations detailed in the full Accrual Metric Pilot report to determine the future direction of metric implementation.

## CONCLUSION

The Tufts Common Metrics Implementation and Evaluation Study generated insights and evidence to assess and reflect on the Common Metrics Initiative. Findings and conclusions speak most directly to future directions of the Common Metrics Initiative, but they can also inform other CTSA Program initiatives and similar networks that plan to embark on implementing shared metrics and performance improvement frameworks.

## Table 1. Summary and detailed recommendations

1	Develop metrics using robust pilot testing, and engage stakeholders in ongoing review.
1a	For each new Common Metric, conduct a robust pilot test that equally emphasizes feasibility of data collection and usefulness of the metric for local and Consortium-wide strategic management.
1b	Consider phased pilot testing (i.e., test data collection first, then strategic management) for metrics for which data quality or feasibility issues are likely.
1c	Include requirements for collection and reporting of additional data points in Operational Guidelines to confirm that comparisons across hubs are valid. Ideally, reporting of all underlying data would allow for data audits. Short of that, reporting all data elements used to calculate metric values would support oversight of data quality.
1d	Periodically engage hubs in a review of each metric for completeness, clarity, usefulness, and required effort.
2	Maximize usefulness to hubs by selecting metrics that align with local needs.
2a	<ul> <li>Select metrics that better align with local CTSA and home institution needs and priorities. For example:</li> <li>i. Consider clustering similar CTSAs to address selected metric topics rather than creating Consortium-wide requirements.</li> <li>ii. Acknowledge and communicate to hubs that local priorities can influence performance targets.</li> </ul>
3	Maximize usefulness to the National CTSA Consortium by ensuring validity of aggregation and comparison reporting.
3a	Regularly review metric results for those missing, clearly incorrect or inconsistent with Operational Guidelines and follow-up with hubs.
Зb	If aggregation or comparisons of hubs' metric results are pursued, ensure results are comparable across hubs. Consider implementing a data coordinating center function with formal data cleaning or auditing processes.
3c	Until a full and thorough review of metric results can be performed, acknowledge inconsistencies in data collection when interpreting aggregated reports.
3d	To allow meaningful interpretation of metric results and comparison across diverse hubs, collect and report factors relevant to performance, including hub and institutional characteristics

4 Equip hubs to fully implement each metric and performance management by providing peer-to-peer learning and training, coaching, and assistance for varying levels of experience. **4a** Provide training and coaching that meets the needs of adult learners with different learning styles and various levels of prior experience in performance management. **4b** Add training and coaching on more advanced strategic management concepts and relevant examples as the Common Metrics Initiative matures and participants become more proficient in implementing metrics and creating performance management plans. **4c** Provide small group coaching when implementing each new Common Metric, and provide mechanisms to promote peer-to-peer learning and accountability for meeting implementation milestones. Consider extending coaching beyond initial metric implementation for hubs wanting or needing additional support. 4d Provide concrete examples of how to calculate each metric (e.g., metric calculation worksheets) and exemplar strategic management plans to assist hubs to conduct and document their planning. 5 Support implementation by promoting metric-specific teams, allowing for capacity-building periods, providing accurate benchmarks, and updating performance drivers and best practices. Encourage the use of metric-specific teams with active subject matter experts who are 5a able to address data issues and strategic management specific to the metric topic. **5b** Promote hub-identified facilitators for building effective teams, including identifying one team member who takes ownership of the project and a local champion on the team, and attending to team climate and interactions. **5c** Encourage involvement of the CTSA Principal Investigator to provide strategic guidance and oversight, to champion the project, and to facilitate stakeholder engagement. **5d** Allow for a capacity-building period prior to mandating collection and reporting of metric data to support revising existing data sources, developing data sources and systems, and training personnel. 5e Encourage hubs to engage partners and subject matter experts outside of the core team to gain a deeper understanding of underlying causes of existing hub performance, and to assist in selecting and implementing improvement strategies. **5f** Provide useful, accurate benchmarking data to help hubs better target areas for improvement. 5g If using aggregated hubs' metric results to identify performance benchmarks, acknowledge and describe local reasons for variation in metric results. **5h** Provide hubs a repository of best and promising practices, including newly developed and updated Driver Diagrams and Change Packages, to speed and focus development of strategic management plans. 5i Promote peer-to-peer learning and disseminate best and promising practices. i. Consider continuing and establishing additional mechanisms for shared learning (e.g., Collaborative Learning sessions) and disseminating best and promising strategies (e.g., publishing hub Success Stories). ii. Highlight successful adoption and application of RBA and CTSA Consortium achievements.

6	Maximize usefulness of the reporting platform by enhancing functionality, visualization options, and user experience.
6a	Expand software features and functionality to support data collection, storage, and quality checks.
6b	Enhance user experience (e.g., speed, intuitiveness, number of clicks required to navigate), and improve visualization capability (e.g., create more display options, display multiple metrics simultaneously).
7	Sustain engagement by facilitating solutions to barriers due to resources and authority, accounting for hub heterogeneity, and ensuring effective communication.
7a	Facilitate solutions to limited resources and personnel and use multiple strategies to account for heterogeneity across hubs. For example:
	<ul> <li>Consider aligning Common Metrics reporting with other required reporting (e.g., annual reporting).</li> </ul>
	<ul> <li>Consider an explicit process to weigh the value of a metric with the effort to obtain data.</li> </ul>
	iii. Consider a designated budget allocation to support Common Metrics work.
	iv. Use a software platform that does not limit the number of users due to fees.
7b	Account for heterogeneity of hub data, processes, and local priorities. For example:
	<ul> <li>Consider clustering similar CTSAs to address selected metric topics rather than creating Consortium-wide requirements.</li> </ul>
	ii. Offer expanded flexibility in choice of performance improvement framework
7c	Maintain realistic expectations about the amount of improvement that can be achieved and the pace of change, particularly when the CTSA leader does not have line authority over the target processes.
7d	Promote peer-to-peer learning about successful strategies for affecting change in the home institution.
7e	Develop and maintain effective ongoing communication strategies for hub leadership and staff, and particularly new staff.
7f	Inform hubs of future directions for the Common Metrics Initiative.
8	Expand use of data to inform future directions of the Common Metrics Initiative and the CTSA Program.
8a	Use hub data beyond what is needed to implement the Common Metrics (e.g., other clinical and operational data) to inform the selection of metrics and to identify potential drivers of outcomes.
8b	Use discussion of the Common Metrics Evaluation results to catalyze a broader conversation about other high impact research projects to drive data-driven decisions related to the structure of CTSAs and the CTSA Program.

9 Make improvements to the first three Common Metrics by clarifying Operational Guidelines and assessing usefulness with hubs. **Metric: IRB Review Duration** 9a Consider clarifying Operational Guideline definitions about inclusion of multiple institutional IRBs. 9b Collect additional data about the number of IRBs included in the metric calculation and the hub's ability to exclude pre-review activities to inform appropriate metric comparisons across the CTSA Consortium. **9c** Assess with hubs the usefulness of this metric to hubs and the CTSA Consortium given disparate local IRB processes, variation in types of clinical protocols, and concerns about comparability of metric values across hubs. **Metric: Pilot Funding Publications 9d** Consider modifying the metric only to include pilots that have had sufficient time to publish (e.g., one year after pilot conclusion). **9e** Assess with hubs the usefulness of this metric for local improvement, particularly the extent to which the metric reflects local priorities. **Metric: Careers in Clinical and Translational Science 9f** Modify the Operational Guideline to further define and clarify exclusion criteria: i. clarify whether hubs may use additional definitions of "engaged in research," ii. add exclusion of solely institutionally-funded scholars, iii. clarify definition of "still in training," iv. add criteria for "lost to follow-up." 9g Consider modifying the metric to be annual (e.g., percent of 2015 graduates who are in CTR; percent of 2016 graduates who are in CTR, etc.). **9h** Assess with hubs the usefulness of this metric for local improvement, particularly the extent to which the metric definitions reflect local priorities.

10	Use the results and recommendations detailed in the full Accrual Metric Pilot report to determine the future direction of metric implementation.
10a	Consider providing an infrastructure-building period prior to mandated collection of metric data to allow hubs time to devise and/or revise data sources and systems and data collection and data quality procedures, and train personnel.
10b	Modify the metric to be collected prospectively rather than retrospectively to increase its potential usefulness for strategic management, including the ability to identify and intervene in individual trials as needed.
10c	Revise the Operational Guideline to address certain multi-site clinical trials (e.g., those of competitive enrollment design) in which key accrual Metric variables are not known.
10d	Re-evaluate the exclusion criterion for trials with fewer than 10 targeted participants; consider lowering the cut-off (e.g., to trials with less than five targeted participants).
10e	Do not exclude clinical trials of dose-to-toxicity design.
10f	Collect and report additional information, including information about the mix of clinical trials at the primary institution or included in the Median Accrual Ratio, to understand how representative the median is of the intended sample.
10g	Provide a template of tested survey qeustions and survey considerations.
10h	Provide hubs with best or promising practices and strategies for implementing a CTMS to produce metrics.

# **Full Report**

Common Metrics Implementation Program and Evaluation Study

## INTRODUCTION

The Clinical and Translational Science Award (CTSA) Program Consortium has been charged with accelerating and improving clinical and translational research in the United States (<u>https://ncats.nih.gov/ctsa/about/hubs</u>). This has required multiple efforts and changes in the way research is done. In its 2013 Report on the CTSA Program,<sup>1</sup> the Institute of Medicine (IOM) set forth a vision for transformative change. Among its recommendations was a call for standardized evaluation processes based on measurable strategic goals and uniform, actionable "common metrics" to enhance transparency and accountability in decision-making. This was seen as key for powering needed evolution of CTSAs and their research communities.

Stimulated by the IOM recommendations, the NIH National Center for Advancing Translational Science (NCATS), home of the CTSA Program, and the CTSA hubs – a national network of research organizations – have implemented a number of efforts to become a "more tightly integrated network that works collectively."<sup>2</sup> Among them, the Common Metrics Initiative marks a fundamental shift in the management of research and the associated culture throughout the CTSA Consortium. In contrast to continuing as a loosely organized set of institutional academic homes, each with its own system for performance improvement, the Common Metrics Initiative creates a path for moving the CTSA Consortium toward a network that operates by sharing information and best practices. Using standardized metrics and the principles of the Results Based Accountability (RBA) performance management framework,<sup>3</sup> this initiative aims to continually transform management both within each CTSA hub and at the level of the CTSA Consortium by focusing stakeholders on common outcomes and on strategies to improve performance.

This far-reaching initiative provides a unique opportunity to expand the knowledge base on the effective management of science. Recognizing this, between September, 2015, and September, 2018, NCATS provided funding support to Tufts Clinical and Translational Science Institute (CTSI) to disseminate and evaluate the use of Common Metrics and the RBA framework throughout the CTSA Consortium.

## **OVERVIEW OF ACTIVITIES AND REPORT**

With input from NCATS, the Tufts Implementation Team developed and delivered a program of activities intended to provide the diverse CTSA organizations with the necessary tools to support full adoption of collaborative performance measurement and management. These activities included implementing a hub training program on collecting Common Metrics and using the RBA framework for performance management, coaching hubs through the implementation process, providing technical assistance, disseminating promising improvement strategies, and supporting hub use of a software reporting system.

Concurrently, the Tufts CTSI Common Metrics Evaluation Team conducted a mixedmethod evaluation to describe the CTSA Consortium's progress and experiences. Topics included hub progress with implementation; challenges, facilitators, and contextual factors that affected progress; hub perspectives on the Tufts Implementation Program; ways in which hubs integrated Common Metrics work and the personnel resources expended; and perceived benefits and concerns.

This report synthesizes results and recommendations from the Tufts Common Metrics Implementation Program and Evaluation. An overview of activities is described in the following timeline.



Figure 1. Timeline for Common Metrics implementation and evaluation activities

The primary goal of this report is to provide insights and evidence for decision-making on future directions of the CTSA Program Common Metrics Initiative. However, we believe that the results also will be of interest to other groups or networks that are poised to implement standardized metrics and performance improvement processes, potentially including other NIH institutes and centers. Additionally, the Common Metrics Initiative, as conducted across the many CTSA organizations and their home institutions, revealed characteristics of CTSAs and the ways they operate that enhance our understanding of the CTSA Program more generally.

## **PROJECT ORGANIZATION AND OVERSIGHT**

The overall organizational structure of the Common Metrics Initiative is depicted in Figure 2. Throughout the Common Metrics Implementation, project collaboration was promoted by maintaining clear and consistent communication between these groups, and with hub participants, leadership and other stakeholders. Standing meetings were held with NCATS, who also received a monthly status report. Sixty-four CTSAs and interested medical centers participated (see Acknowledgements for a list of participating sites).

CTSA Common Metrics Executive Team Harry Shamoon, MD Anantha Shekar, MD, PhD Harry Selker, MD, MSPH Bradley Evanoff, MD, MPH Rebecca Moen, MBA William Trochim, PhD Phil Lee, JD, MPP	Principal Investigator: Harry Selker, MD, MSPH Implementation: Project Director: Denise Daudelin, RN, MPH Quality Improvement Specialist: Laura E. Peterson, BSN, MS Evaluation Research: Co-Investigator: Lisa Welch, PhD Organizational Evaluation Advisor: Debra Lerner, MS, PhD	ConsultantsResults Based Accountability Consultant:Phil Lee, JD, MPP, Clear ImpactBurness CommunicationsAccrual Metric Pilot: Ken Getz, MBA Edward Kuczynski, MARAND Health Qualitative Evaluation Team: Andrada Tomoaia-Cotisel, PhD, MPH, MHA Peter Mendel, PhD Jason Etchegaray, PhD Nabeel Qureshi, MPH
Denise Daudelin, RN, MPH Michelle Kipke, PhD Redonna Chandler, MD Patricia Jones, DrPH, MPH	Common Metrics Evaluation Research Team Research Associate/ Qualitative Analyst: Anshu Parajulee, MPH Lead Statistician:	Common Metrics Implementation Team Communication Manager: Ramana- than Mahalingam, MAS Project Manager: Mridu Pandey, MPH

### Figure 2. Common Metrics Implementation and Evaluation project teams, leadership

## **CTSA Common Metrics Executive Team**

The Executive Team is comprised of CTSA Principal Investigators (PIs), Administrators, evaluation leads, NCATS staff, and an RBA expert. This group provides important input and guidance to the Common Metrics Initiative, ensuring that activities are relevant and useful for all CTSAs.

## **Tufts CTSI Common Metrics Leadership Team**

The Administrative Team at Tufts CTSI included three essential areas of the project: multi-center studies, process improvement, and evaluation research. Harry Selker, MD, MSPH, Dean and Principal Investigator of Tufts CTSI, was the overall project Principal Investigator. The implementation was led by Denise Daudelin, RN, MPH, Director of the Tufts CTSI Research Process Improvement Program. Evaluation research was led by Debra Lerner, MS, PhD, and Lisa Welch, PhD. Laura E. Peterson, BSN, MS, a Quality Improvement Specialist, supported metric and RBA implementation and dissemination of metric-related best and promising practices.

### Consultants

Management, communications, qualitative research and organizational improvement consultants were engaged as needed. Phil Lee, JD, MPP, President of Clear Impact, assisted in the development of the Common Metrics and the development and implementation of the RBA training sessions.

### **Implementation Team**

Under the leadership of the Principal Investigator and Project Director, the Implementation Team included staff members who carried out the day-to-day implementation and dissemination plan as well as conducted new metric pilot testing. These activities included maintaining clear and consistent communication across all hub sites and with NCATS, metric training, database development, quantitative and qualitative data collection, and data synthesis and reporting.

### **Evaluation Research Team**

Under the leadership of the Organizational Evaluation Advisor and Co-Investigator, the Evaluation Research Team included expertise in quantitative and qualitative research. Activities included study design, development of surveys and semi-structured interview guides, quantitative and qualitative data collection, data management, statistical analysis, qualitative analysis, and project management. With input and assistance from Tufts CTSI personnel, a four-member team from RAND Health led the qualitative component.

## PART I. IMPLEMENTING COMMON METRICS AND A STRATEGIC MANAGEMENT FRAMEWORK

Common Metrics Implementation included testing developed metrics; providing training on the metrics, the RBA improvement framework, and the software used by hubs to enter data and strategic management information; and providing coaching and opportunities for collaborative learning.

Each of the following sections describe Implementation activities, related results, insights gained, and recommendations for future activities of the Common Metrics Initiative.

## SELECTING AND DEVELOPING METRICS

The first three common metrics (Careers in CTR, Pilot Funding Publications, and Institutional Review Board [IRB] Review Duration) and their respective Operational Guidelines were developed by a workgroup of topic experts, evaluators, Principal Investigators (PIs), and NCATS staff. The process included review of relevant literature, discussions with other academic investigators and across NIH, and group meetings. Each of the three metrics were pilot-tested in four volunteer hubs for data collection feasibility. Pilot test results informed important changes to the metric Operational Guidelines and plans for metric implementation by identifying gaps in metric definitions, confusion about metric calculations, and challenges to the usefulness of the metrics for strategic management. A revised Operational Guideline template was developed to add numerator and denominator statements, additional definitions for terms, and more detailed inclusion and exclusion criteria. Implementation with the remaining hubs used the post-pilot versions of the metrics and Operational Guidelines.

## **Summary Recommendation 1:**

Develop metrics using robust pilot testing, and engage stakeholders in ongoing review.

- 1a For each new Common Metric, conduct a robust pilot test that equally emphasizes feasibility of data collection and usefulness of the metric for local and Consortium-wide strategic management.
- 1b Consider phased pilot testing (i.e., test data collection first, then strategic management) for metrics for which data quality or feasibility issues are likely.
- 1c Include requirements for collection and reporting of additional data points in Operational Guidelines to confirm that comparisons across hubs are valid. Ideally, reporting of all underlying data would allow for data audits. Short of that, reporting all data elements used to calculate metric values would support oversight of data quality.
- 1d Periodically engage hubs in a review of each metric for completeness, clarity, usefulness, and required effort.

## **IMPROVING PERFORMANCE**

## **Effective Teams**

Each CTSA Program hub was asked to form a "core" Common Metrics team, based on five responsibilities felt to be most important to success of the initiative:

- A **Project Champion** to ensure everyone at the hub was "on board" and committed to the ultimate success of the project. This role was often assumed by the Principal Investigator.
- A Project Leader, responsible for overall planning and execution of the project.
- An **RBA framework Lead**, responsible for helping their team learn and implement the RBA framework.
- A Scorecard Software Lead, responsible for helping others at the hub learn the Scorecard software.
- A Metrics Topic Lead, responsible for overseeing the collection of metric data. Based on staffing capabilities, hubs could elect to have a different metric expert for each of the common metrics.

Each hub was encouraged to configure their team as was best suited to their current staffing pattern and levels. Some hubs had the same person assume more than one role (such as project leader or RBA expert) or two people sharing the same role (e.g., project co-leaders). The resulting teams were quite variable in size (range: 2-11 named members) and some teams were operationalized by one person who had been assigned the role of "Common Metrics person" by senior leadership at their hub.

The core team was also instructed to identify additional subject matter experts to assist them and address specific metrics (e.g., a member of the IRB for the IRB Review Duration metric) but, teams often lacked participation of members with specialized knowledge of metric topic(s) or influence over topic areas. Greater inclusion of subject matter experts would improve the ability of teams to determine underlying causes for metric performance and to implement effective strategies.

## Summary Recommendation 5:

# Support implementation by promoting metric-specific teams, allowing for capacity-building periods, providing accurate benchmarks, and updating performance drivers and best practices.

5a Encourage the use of metric-specific teams with active subject matter experts who are able to address data issues and strategic management specific to the metric topic.

## **Training, Coaching and Technical Assistance**

Hub teams participated in training and coaching in one of three implementation "waves" or groups between July and October of 2016 (Figure 3); group assignment was based on hub preference. Each Implementation Group received training on the metrics and Operational Guidelines, Scorecard software, and RBA strategic management concepts, participated in small group coaching sessions, and received individualized coaching and ongoing technical assistance as needed.





## Training

The training was delivered via live and pre-recorded interactive webinars and conference calls, sometimes with "pre-work" assignments before sessions. It included:

- An "onboarding" call: A conference call reviewing the Common Metrics project activities, timeline, providing guidance on finalizing the hub's team, and answering questions.
- Common Metrics Training: A webinar reviewing each Operational Guideline.
- **Principal Investigator Training Session**: A review of RBA and Scorecard software specifically for hub Principal Investigators.
- **RBA and Scorecard Training:** A series of three sessions for hub team members using a mix of pre-recorded lectures, training videos and live interactive webinars to build knowledge and skills in the RBA framework, and in setting up and using the Scorecard software. The RBA framework is depicted in Figure 4.
- Kickoff Call: Designed for the full hub team, this conference call featured updates on the project timeline, examples of completed Scorecards, and guidance on next steps.

The training was modified after the first Implementation Group and again after the second, based on participant feedback. Interactivity was increased, and additional CTSA-specific examples were added. A set of annotated training slides was also developed to allow core team members to train additional staff at their hub regarding the Initiative, RBA, and Scorecard.



## Figure 4. Results Based Accountability framework

Tufts CTSI Common Metrics Report - Part I: Implementation Program

## Coaching

Following training, each hub selected one of the three Common Metrics for data collection and conduct of strategic management activities during a subsequent 14-week coaching period (Table 2). Each Implementation Group was divided into smaller, four to six hub, coaching groups. During seven every-other-week small group coaching calls, the Quality Improvement Specialist or Project Director facilitated a one-hour webinar during which hubs discussed their progress in meeting implementation milestones, and accomplishments, challenges and barriers in applying RBA and collecting Common Metric data according to the Operational Guidelines. Assessments of hub progress based on implementation criteria were used to direct webinar content and discussion. A "pre-work" assignment (e.g., come prepared to discuss at least one barrier you have overcome) was emailed before each webinar. During the webinars, hubs had the opportunity to ask questions about defining terms, sources of data, how to apply the Operational Guidelines to hub-specific situations and how other hubs were overcoming common challenges. Some hub questions led to clarifications in the Operational Guidelines themselves.

		First Common Metric		
Implementation Group	n	Median IRB Review Duration n (%)	Careers in CTR n (%)	Pilot Funding Publications n (%)
Pilots	4	0 (0%)	2 (50%)	2 (50%)
Implementation Group 1	20	8 (40%)	5 (25%)	7 (35%)
Implementation Group 2	18	7 (39%)	5 (28%)	6 (33%)
Implementation Group 3	22	9 (41%)	3 (13%)	10 (46%)
Total	64	24 (38%)	15 (23%)	25 (39%)

## Table 2. Number of hubs and self-selected coaching metric by Implementation Group

## **Technical Assistance**

Technical assistance was provided by the Implementation Team to help hubs to: a) interpret and apply the Operational Guidelines, b) identify and overcome barriers to data collection and strategic management, and c) complete strategic management plans. Assistance included:

## Help Desk

Questions were managed by a Help Desk process with telephone and email availability. Query volume varied depending on the stage of the initiative. In addition, as hub team members became familiar with Tufts Implementation Team members and their respective roles, they often "self-triaged" requests directly to relevant staff (e.g., strategic management inquiries to the QIC or Project Director).

## **Metric Worksheets**

Although the Operational Guidelines were updated with additional detail after the conclusion of pilot testing for the first three metrics, hubs continued to have numerous questions and had misunderstandings about how to calculate metric score values. Therefore, the Implementation Team developed worksheets for both the Pilot Publications and the Careers Common Metrics that provided examples of how to calculate each metric score, as well as a place for a hub to enter their own data and complete the calculation. These worksheets were posted on the Common Metrics Initiative website. Implementation staff also emailed worksheets directly to hubs when the data entered into Scorecard appeared to be inconsistent with the Operational Guideline (as identified during data checking).

## **Exemplar Turn the Curve Plan**

Teams had numerous questions about the level of detail that should be included in an initial Turn the Curve plan, as well as where various components of the plan should be documented (i.e., in what step of the process). Therefore, the Implementation Team developed an example for the Careers in Clinical and Translational Science Metric. Made available on the website and reviewed with hub teams on coaching calls, the plan contained context that can be applied to any Turn the Curve plan for a Common Metric, as well as example data and text for a hypothetical hub.

### **Scorecard Software**

Hub teams used Scorecard software to report common metric data and the various components of their Turn the Curve plan. The Implementation Team answered basic hub Scorecard functionality questions, and collaborated with personnel at Clear Impact about use of the system for the Common Metrics Initiative as a whole.

## Evaluation of Training, Coaching, and Technical Assistance

### Training

Hub responses to questions about the amount of time spent training on RBA and the metric Operational Guidelines, and ratings of the didactic training overall, are summarized in Figure 5. There was no statistically significant difference by Implementation Group (Appendix A). Hubs found the didactic webinars to be professionally presented but session content was felt to be redundant and hubs would have preferred that the training have taken less time and included more relevant examples.

I think in the very beginning some of the trainings with the Scorecard framework and like about the RBA framework was a little redundant. I think we could have only had one training on each of those versus four different calls about RBA framework or about Scorecard and how to use it. - Implementer


Figure 5. Evaluation of implementation training

## Coaching

Hubs generally found the coaching sessions supported implementing the Common Metrics as they facilitated sharing experiences with and successful strategies for implementation across hubs.

I thought the coaching calls were very useful; they kept us all on track, we had little mini homework assignments after each call, I feel like, which helped us get to the next step, in terms of developing our Turn the Curve plan. – Implementer

Hubs pointed to a mechanism to share experiences with each other as central to their ability to engage with the Common Metrics Initiative. Some hubs indicated they wished that these calls had continued or requested that they be started again.

It was really helpful to hear what everyone else is doing and how they're overcoming their challenges and what interventions they're considering for their plans...because once those calls ended, there is no communication on what people are doing with their Turn the Curve plans. We were kind of left on our own...– Implementer

Others indicated that the intensity and frequency of calls was crucial when they were initially learning the work of the Common Metrics Initiative, but that this same intensity was not needed thereafter.

Hub responses to questions about the amount of time allocated for the coaching sessions, as well as a rating of the amount of discussion and homework, are summarized in Figure 5. There was no statistically significant difference by Implementation Group (Appendix A).

#### **Technical Assistance**

Hubs found tools such as the metric worksheets helpful for informing their understanding of operationalizing the Common Metrics Initiative at their hub – and some mentioned wanting these resources earlier in the process.

I would advise [a new hub] to, obviously, do the training but hold out for some of these worksheets that made things a lot easier for us to do it right, quickly, rather than trying to figure things out for ourselves. – Implementer

Hubs that reached out for individual support from the help desk or relevant staff found it to be useful.

[A]ll the calls, the group calls as well as individual [calls] have been very helpful in getting those questions answered...the degree of feedback and responsiveness that we've gotten from the Tufts team has really helped with the implementation. –Principal Investigator

[Y]ou could kind of walk through that [metric] together with Tufts, which was helpful because it helps us to work on the other two [metrics] a little more quickly...when you have questions having someone to send an email to or even get on the phone and talk to is important. –Administrator

Hubs were glad to have resources on the website for accessing the latest information and for training new staff members.

I, for one, really appreciated everything they put up on the website. I refer to those resources a lot, the Change Packages and the Operational Guidelines, just sort of knowing where to go for the latest versions, because I know those have been evolving as well. – Administrator

## **Summary Recommendation 4:**

Equip hubs to fully implement each metric and performance management by providing peer-to-peer learning and training, coaching, and assistance for varying levels of experience.

- 4b Add training and coaching on more advanced strategic management concepts and relevant examples as the Common Metrics Initiative matures and participants become more proficient in implementing the metrics and performance management.
- 4c Provide small group coaching when implementing each new Common Metric, and mechanisms to promote peer-to-peer learning and accountability for meeting implementation milestones. Consider extending coaching beyond initial metric implementation for hubs wanting or needing additional support.
- 4d Provide concrete examples of how to calculate each metric (e.g., metric calculation worksheets) and exemplar strategic management plans to assist hubs to conduct and document their planning.

## **Effective Communication**

The Implementation Team employed numerous strategies to provide project-related resources and communicate initiative progress. Implementation challenges included ensuring that newly participating and existing staff were aware of how to access project resources and that they received initiative communications.

#### Website

During the implementation, a website for the Initiative was housed at the Tufts CTSI website and provided Common Metrics FAQs and training and dissemination materials. Upon the completion of the initial Common Metrics Implementation, the primary Common Metrics website was transitioned to CLIC as a part of their new Coordinating Center activities.

## **CTSA Program Update E-Newsletter**

During the implementation period, Common Metrics updates, milestones, deliverables and resources were provided in the NCATS e-newsletter. Hub Principal Investigators, Administrators and members of Common Metrics teams were placed on the distribution list. Significant hub team member turnover required frequent updating of the list.

## Presentations

Project-related presentations and round table discussions were used by the Implementation Team to inform NCATS staff, Common Metrics Initiative leadership and other stakeholders about the project's progress. Presentations and discussions were held with the following audiences:

- NCATS Program Officer. Program Officers began discussing Common Metrics results and hub-specific strategic management activities with their hubs in December 2017. Presentations to this group addressed the progress of the initiative, successes and challenges hubs were facing in data collection and strategic management, and new resources and tools available to hubs.
- *CTSA Common Metrics Executive Committee*. The Executive Committee was consulted for counsel and feedback at a number of junctions during implementation of the first three metrics, and planning for the pilot test of the fourth Common Metric regarding accrual.
- NCATS CTSA Steering Committee. Updates on the accomplishments of the initiative were provided to the NCATS Steering Committee, who provides leadership and guidance on opportunities, impediments, and joint agreement on broad issues affecting the Common Metrics Initiative.
- ACTS Translational Science Meetings. 2017/2018. Roundtable and panel discussions were held to provide hubs the opportunity to discuss data collection and strategic management challenges, lessons learned, and effective and innovative improvement strategies.
- *CTSA Evaluators' Group*. Regular updates on the status of the project and the use of performance improvement tools and methods were provided to Evaluators who were usually active participants on their hub's Common Metrics team.

• Common Metrics Informatics Development Team Meeting. An overview of the Common Metrics Initiative, the role of strategic management, and lessons learned from implementation of the first three metrics was provided to the Informatics Metrics Development team during the kick-off of their development process.

## **Summary Recommendation 7:**

Sustain engagement by facilitating solutions to barriers due to resources and authority, accounting for hub heterogeneity, and ensuring effective communication.

7e Develop and maintain effective ongoing communication strategies for hub leadership and staff, and particularly new staff.

#### **Collecting Data and Developing Strategic Management Plans**

Hubs collected metric data and developed metric-specific strategic management plans for each of the first three Common Metrics. Many hubs were collecting the same or similar data prior to metric implementation. For some hubs and some metrics, existing data sources did not align with Operational Guidelines or the hub's ability to access the data was limited. Developing new data sources, revising existing sources, or gaining access to data led to delays in collecting and reporting metric values. By the conclusion of the coaching period, the majority of hubs had produced a value for their first Common Metric (Figure 6).



Figure 6. Percent of hubs that produced a metric value by the end of the coaching period

#### \*Optional metric

IRB=Institutional Review Board, Pubs=Publications, Fund=Funding, KL2= Mentored Career Development Award, CTR=Clinical Translational Research, URP=Underrepresented Persons, TL1=Linked Training Award

Hubs also experienced challenges in completing each element of the RBA framework, some specific to particular elements and others more general across the improvement process. During the coaching period, hubs indicated that multiple factors affected the pace and quality of strategic plan development, including previous improvement efforts on the topic; access to data to compute the metric or analyze the result; availability and engagement of project team members, partners and subject matter experts; and the extent to which Common Metrics activities were competing or synergistic with other hub priorities. Some hubs were reluctant to develop plans until metric values were available or concentrated their plan on implementing or revising data sources. When hubs believed their current level of performance did not require improvement, their strategic plans primarily described activities they had previously undertaken relative to the metric topic.

The Implementation Team regularly assessed hubs' progress in meaningfully applying the RBA framework for their selected metric. Following each coaching webinar, the Turn the Curve plan for each hub was assessed based on 13 RBA implementation criteria developed in conjunction with the Clear Impact RBA consultant (Appendix B). The assessment scale included:

0 - Unable to assess
1 - Not Meeting criteria
2 - Approaching criteria
3 - Meeting criteria
4 - Exceeding criteria

The rating of "Exceeding criteria" was included to identify hubs whose Turn the Curve plan demonstrated best practices in applying the RBA framework. Hubs progressed at variable rates in implementing the framework and documenting their Turn the Curve plan. Several criteria related to the development and implementation of the Turn the Curve plan were unable to be fully assessed during the time interval of the coaching period (e.g., development of action plans) and it was expected that hubs would make continued progress following the conclusion of the period. Figure 7 illustrates the results of assessment of meaningful application of RBA for all hubs as of their respective final coaching session. These assessments also helped identify needs for additional support and tools (e.g., metric worksheets). Aggregate group progress was shared regularly with NCATS throughout implementation.



Figure 7. Percent of hubs meeting criteria for meaningful application of RBA during the coaching period

## **Summary Recommendation 5:**

Support implementation by promoting metric-specific teams, allowing for capacity-building periods, providing accurate benchmarks, and updating performance drivers and best practices.

- 5d Allow for a capacity-building period prior to mandating collection and reporting of metric data to support revising existing data sources, developing data sources and systems, and training personnel.
- 5e Encourage hubs to engage partners and subject matter experts outside of the core team to gain a deeper understanding of underlying causes of existing hub performance, and to assist in selecting and implementing improvement strategies.

#### **Disseminating Performance Drivers and Strategies for Improvement**

Several activities were undertaken to "harvest" and disseminate best and promising practices from the hubs' work.

#### **Driver Diagrams and Change Packages**

For each of the first three metrics, a "version 1.0" Driver Diagram and Change Package was developed (Appendices C-E). An evidence-based resource used in improvement efforts, Change Packages are meant to be concise and practical. They can provide ideas and inspiration for the work of improvement teams, but are expressly not proscriptive or requirements. Each Change Package is designed around an aim, or what teams are trying to accomplish (for the IRB Review Duration metric, for example, the aim is *Improve the median number of calendar days from the official IRB application receipt date to the official IRB final approval date for fully reviewed protocols*). The Implementation Team reviewed the positive or facilitating factors from the Story Behind the Curve and the What

Works sections of hub Turn the Curve plans, as well as any published evidence available. The results were grouped into Drivers; factors that, if they are present, can help achieve the aim. Drivers for each of the first three metrics are provided in the Learnings section below. For each of the identified drivers, other areas of the Turn the Curve plans, and the published evidence, were reviewed for potential Strategies that could turn the curve. The metric aim, the drivers and the strategies – and their relationship to each other – were depicted in a Driver Diagram.

In the Change Packages, for each driver, hotlinks were provided to example strategies from hubs, including a brief rationale and a description of the example. Since these tools were generated early in the Initiative, they can be updated in the future with additional drivers and strategies that are likely present in more recent plans. The exemplars are primarily strategies that hubs had already been doing prior to the Common Metrics Initiative and were not developed specifically as a part of Common Metrics strategic management. With additional time and focus on improvement for these topics, additional examples and potential best and promising practices should emerge from the Initiative.

#### **Collaborative Learning Sessions**

As the Initiative proceeded, and hubs began strategic management for common metric topics, the Implementation Team convened a series of collaborative Learning Sessions to showcase and share some of the best and promising improvement strategies that were being identified. Scheduled as one-hour webinars with a combination of didactic presentation and group interaction, all hub team members, including Principal Investigators and Administrators, were encouraged to attend. Slides and recordings of each session were made available on the website. Agendas for the Learning Sessions are provided in Appendix F.

#### **Success Stories**

Implementation Success Stories presented some of the ways the diverse institutions in the CTSA Consortium collected Common Metrics data and used the RBA framework to improve hub operations. These snapshots were meant to highlight innovations and provide an accessible source of practical ideas for supporting continuous improvement at any hub.

#### **Evaluation of Dissemination Activities**

Hubs largely found the Driver Diagrams and Collaborative Learning Sessions helpful and many wished that the latter had continued past early 2017. A portion of hubs felt there could have been more dialogue during the Learning Sessions; others felt they did not have much to present, or did not always feel fully comfortable in doing so.

[The] Driver Diagrams were by far the most helpful resource because they were options that others had used, which reinforced that we were on the right track and/or provided suggestions that would help us with our effort. -Administrator

...those [Learning Sessions] were somewhat helpful and...people could

volunteer to discuss, say, "Our IRB turnaround time is 40 days, and here's what we're doing," you know, that kind of thing can actually volunteer to share your information. Since then [2017], we haven't had any calls and I don't know how we are expected to share best practices, results, things we're doing because we don't have any mechanism for doing that. –Implementer

#### **Summary Recommendation 5:**

Support implementation by promoting metric-specific teams, allowing for capacity-building periods, providing accurate benchmarks, and updating performance drivers and best practices.

- 5h Provide hubs a repository of best and promising practices, including newly developed and updated Driver Diagrams and Change Packages, to speed and focus development of strategic management plans.
- 5i Promote peer-to-peer learning and disseminate best and promising practices.
  - i. Consider continuing and establishing additional mechanisms for shared learning (e.g., Collaborative Learning sessions) and disseminating best and promising strategies (e.g., publishing hub Success Stories).
  - ii. Highlight successful adoption and application of RBA and CTSA Consortium achievements.

## **METRIC DATA QUALITY**

#### **Data Checking**

In order to be used for benchmarking, comparison, or aggregate reporting, metrics must be collected, computed and reported in a comparable way. In addition, meaningful interpretation of the range of metric values reported across hubs requires understanding the institutional, hub, and program-level characteristics that shape metric values. During the Implementation, Operational Guidelines required hubs to enter Common Metric results without the underlying raw data or contextual information and with limited information on data elements used for metric calculations. Data checking of initial Common Metric results revealed many instances where metric values were not consistent with the Operational Guidelines. Although the Tufts Implementation Team developed data checking rules (e.g., range checks) and attempted to work with hubs to correct observable errors, there was limited ability to assess data quality or the extent to which hubs correctly followed Operational Guidelines.

A series of steps to check the 2015 data were undertaken to identify and attempt to remediate instances where hub metric data were either missing or obviously not in conformance with the Operational Guidelines. However, hubs do not enter their raw

data into Scorecard, and there was no audit function possible with the Common Metrics Implementation. Further, hubs were only required to provide 2015 data (although some provided data for some metrics for 2012-2014).

#### Timeline

Data checking of 2015 Common Metrics values began after the due date for hubs to enter their metric values into their Scorecard. Metric data for the first Common Metric selected by the hub was due at the end of the coaching period for their respective Implementation Group (IG1, IG2 or IG3). Data for the second and third metrics were due January 9, 2017 (IG1), February 13 (IG2), or April 3 (IG3), respectively. Data checking was accomplished for all 2015 metric values that were entered by hubs to Scorecard as of July 20, 2017.

#### **Missing Data**

An initial review identified missing metric results (no data value entered) for all hubs. Some metric results would be expected to be missing. For example, hubs whose KL2 or TL1 program did not begin until late 2014 or 2015 may have had no program graduates by 2015 and therefore no results would be expected for the respective KL2 or TL1 metrics. In each instance of a missing metric result, the hub's associated Turn the Curve plan, if there was one, was reviewed for an explanation; e.g., "We had not been tracking pilot publications but will start doing so prospectively." However, in the majority of cases of missing results, it was necessary to contact the hub for the explanation. Over time, hubs gradually populated many of these missing results. Since the system does not provide a notification when values are entered, successive reviews by project staff were required to determine the completeness of the data.

#### Identify Data Inconsistent with the Common Metrics Operational Guidelines

Decision rules (Appendix G) were developed to help identify instances in which the metric results were clearly inconsistent with the relevant metric Operational Guideline. Staff reviewed the results for each metric against these rules. Since the metric results themselves are not the sole source of information as to whether the data are inconsistent with the Operational Guidelines, Implementation Team members also reviewed hub Turn the Curve plans.

#### **Communication with Hubs**

Throughout the data checking process, there was frequent communication with members of hub Common Metric teams. Inquiries about the status of data, or questions about entered values, were made through serial emails and conference calls from the Implementation Team. Hubs were generally responsive to these inquiries.

Although all hubs had participated in coaching calls for a first Common Metric, and had access to the Operational Guidelines, Frequently Asked Questions, and other online resources for the subsequent metrics, many did not initially collect or enter the metric results correctly and required additional technical assistance in order to do so. Copies of the metric calculation worksheets were provided again for those who needed them. Frequent metric result errors were:

- Entering Pilot and Careers metric results annually rather than cumulatively
- Mathematical errors in calculating percent of TL1 or KL2 scholars in CTR
- Including data for years prior to 2012

Despite training, coaching, and technical assistance, some hubs had not entered all 2015 metric scores by July of 2017, or had entered scores that were still obviously inconsistent with the Operational Guidelines. Additionally, in a sample of hubs assessed in April, 2018, approximately 80% had changed one or more 2015 metric values from that which had been present in July, 2017. The new results may have updated earlier incomplete or missing values; some of which were obviously incorrect. Given the ongoing revisions to hub metric values, CTSA Consortium-wide metric results for calendar year 2015 were not reported as it was not possible to fully assess the degree to which the data were consistent with the Operational Guidelines.

## **Summary Recommendation 3:**

# Maximize usefulness to the National CTSA Consortium by ensuring validity of aggregation and comparison reporting.

- 3a Regularly review metric results for those missing, clearly incorrect or inconsistent with Operational Guidelines and follow-up with hubs.
- 3b If aggregation or comparison of hubs' metric results is pursued, ensure results are comparable across hubs. Consider implementing a data coordinating center function with formal data cleaning or auditing processes.
- 3c Until a full and thorough review of metric results can be performed, acknowledge inconsistencies in data collection when interpreting aggregated reports.
- 3d To allow meaningful interpretation of metric results and comparison across diverse hubs, collect and report factors relevant to performance, including hub and institutional characteristics.

## Summary Recommendation 5:

Support implementation by promoting metric-specific teams, allowing for capacity-building periods, providing accurate benchmarks, and updating performance drivers and best practices.

5g If using aggregated hubs' metric results to identify performance benchmarks, acknowledge and describe local reasons for variation in metric results.

# LEARNINGS ABOUT THE FIRST THREE COMMON METRICS

A number of learnings were generated during implementation of the first three metrics and in working with the hubs as they collected data, calculated metric results for 2015, and developed and implemented their strategic management plans.

## **Metric: IRB Review Duration**

Summary: This metric assesses the median number of calendar days from the IRB application receipt date to the IRB final approval date for fully reviewed protocols.

## **Collecting Data and Calculating Metric Values**

Hubs used IRB data from a variety of electronic (e.g., eIRB systems, databases and spreadsheets) and paper (e.g., copies of submitted protocols) sources to calculate the metric. The ease with which hubs could gain access to IRB-related data was variable. The degree to which they could was facilitated by the presence of a pre-existing organizational relationship with the IRB or the ability to develop a collaborative relationship. IRBs that already had or were in the process of applying for accreditation from the Association for the Accreditation of Human Research Protection Programs (AHRPP) already collected and reported a similar metric.

## **Errors and Areas of Confusion**

The IRB metric Operational Guideline indicates protocols "submitted to the institutional or 'local' IRBs at the CTSA Program primary institution (hub)" should be included in the metric. However, hubs had questions about this instruction, including which IRBs they should use if there was more than one at the primary institution or if there were multiple institutions. Hubs with more than one candidate IRB selected one or more to include in their IRB Review Duration metric based on considerations including availability of data, volume of protocols reviewed, and the extent to which the IRB was amenable to participating in strategic management. Considerations informing their choice of IRB(s) were variably described in Turn the Curve plans.

Some reported metric values used business days rather than calendar days as required by the Operational Guideline. The Operational Guideline also instructs to calculate the IRB Review Duration metric excluding the time period a protocol spends in triage or pre-review. However, hubs reported extensive variability in whether they use pre-review, and what activities were included when they do so. Most who reported using pre-review did not capture the duration of those activities. Therefore, the wide range of values observed for this metric for 2015 (8-109 days) is likely related to variable use by IRBs of pre-review activities and variable ability of hubs to exclude pre-review time from the metric value they reported.

#### **Developing and Implementing Strategic Management Plans**

Hub reports of usefulness of the IRB Review Duration metric for strategic management at the local level varied depending on the number of IRBs, the types of protocols reviewed, and the work process of reviews for ethics, feasibility, and budgets/contracts. Many IRBs

had previously addressed IRB review duration and felt that opportunities for further improvement were lacking. Others were challenged in forming collaborative relationships with IRB subject matter experts who would partner to create actionable strategic management plans.

#### Metric-specific Contextual Factors

Meaningful interpretation of the range of IRB Review Duration metric values reported across hubs would require understanding the institutional, hub, and program-level characteristics that shape performance for the IRB Review Duration metric. A number of such characteristics were identified including:

- IRB size / number of protocols (which may or may not correspond to hub size)
- Uptake and use of eIRB systems
- Use of external IRBs (and which types of studies are sent out for external review)
- Participation in centralized IRB models
- Configuration and meeting frequency of IRB review committees
- The extent of IRB application assistance provided by the hub to investigators and their staffs

#### **Drivers of Metric Performance**

A synthesis of information provided by hubs in their strategic management plans and from a review of the literature suggests that IRB review duration can be reduced when the following drivers of high performance are present:

- 1. Engaged and supported investigators create high quality applications and respond to inquiries in a timely manner
- 2. IRB staff and review committees are sufficiently staffed and have appropriate training with optimized workloads
- 3. Waste and redundancy in the review process are identified and eliminated
- 4. Use of appropriate technology is optimized
- 5. Review processes are improved based on feedback from researchers and system metrics

The Driver Diagram and related Change Package distributed to hubs for the IRB Review Duration metric is provided in Appendix C.

## Summary Recommendation 9:

Make improvements to the first three Common Metrics by clarifying Operational Guidelines and assessing usefulness with hubs.

#### **Metric: IRB Review Duration**

- 9a Consider clarifying Operational Guideline definitions about inclusion of multiple institutional IRBs.
- 9b Collect additional data about the number of IRBs included in the metric calculation and the hub's ability to exclude pre-review activities to inform appropriate metric comparisons across the CTSA Consortium.
- 9c Assess with hubs the usefulness of this metric to hubs and the CTSA Consortium given disparate local IRB processes, variation in types of clinical protocols, and concerns about comparability of metric values across hubs.

## **Metric: Pilot Funding Publications**

Summary: This metric assesses the number and percent of research projects that expended hub pilot funding that resulted in at least one publication.

#### **Collecting Data and Calculating Metric Values**

Most hubs already were collecting at least some information about the outcomes of pilot funded projects including publications. However, the frequency of data collection and the sources used varied. Hubs that conducted surveys did not always do so annually (range one to three years); some hubs used electronic searches (e.g., of publications indexed in PubMed), or combined survey responses and online data.

## **Errors and Areas of Confusion**

The Pilot Funding Publications metric is cumulative; the hubs reported that this made interpreting metric results at the hub level difficult or even meaningless. This concern was summarized by a hub lead: "[This metric] does not accurately reflect the success of the program and does not provide meaningful insights into our longitudinal performance as publications related to the program are produced over time. In cumulative reporting, the denominator increases each year while the relative amount of time to produce a publication (and therefore be included in the numerator) decreases, which inevitably results a downward curve".

From the perspective of checking for comparability across sites, variability across hubs in how pilot funding is used and the number of pilots funded made it challenging to identify hubs that were calculating the scores incorrectly (e.g., determining appropriate range checks). Some hubs don't fund pilots every year; other hubs fund over 100 pilots in a single year.

#### **Developing and Implementing Strategic Management Plans**

This metric did not align with the goals of some pilot programs; they disagreed that publication is a desired outcome of pilot grants. A substantial number of hubs indicated that publication is not a focus or necessarily a metric of success of the pilot award program for senior leadership at their hub.

#### Metric-specific Contextual Factors

Meaningful interpretation of the range of metric values reported across hubs would require understanding the institutional, hub, and program-level characteristics that shape performance for the Pilot Funding Publications metric. A number of such characteristics were identified, including:

- Extent and type of CTSI-sponsored resources and services
- Extent (if any) of assistance provided in manuscript or proposal-writing for pilot awardees
- Provision of mentorship for pilot awardees and/or their teams
- Variability in the size of funded pilot awards

#### **Drivers of Metric Performance**

A synthesis of information provided by hubs in their strategic management plans and from a review of the literature suggests the pilot funded publication rates may be improved when the following drivers of high performance are present:

- 1. Effective pilot-funded teams complete projects in a timely manner
- 2. Pilot-funded awardees receive high-quality mentorship
- 3. Awardees access CTSI sponsored resources and services
- 4. Pilot awards are made for projects with attributes that are associated with higher rates of publication
- 5. Awardees are aware of the need to cite pilot funding source and the method for citing pilot support in publications
- 6. Publications with pilot funding citation are identified and tracked (Tracking will not change the underlying rate of publication but is necessary to compute the metric and understand the Story Behind the Curve.)

The Driver Diagram and related Change Package distributed to hubs for the Pilot Funding Publications metric are provided in Appendix D.

## Summary Recommendation 9:

Make improvements to the first three Common Metrics by clarifying Operational Guidelines and assessing usefulness with hubs.

#### **Metric: Pilot Funding Publications**

- 9d Consider modifying the metric only to include pilots that have had sufficient time to publish (e.g., one year after pilot conclusion).
- 9e Assess with hubs the usefulness of the metric for local improvement, particularly the extent to which the metric reflects local priorities.

## **Metric: Careers in Clinical and Translational Science**

Summary: These metrics assess the number and percent of institutional scholars and trainees who completed the KL2 and TL1 programs, respectively, who are currently engaged in CTR, and, of those who are currently engaged in CTR, the number and percent of underrepresented persons and women.

## **Collecting Data and Calculating Metric Values**

Hubs varied greatly in the extent to which they had been tracking graduates and their career statuses over time, if they had, and by what method and frequency. A few hubs had not tracked this information at all. Some hubs surveyed former trainees but at frequencies less than annually (e.g., every other year). The Operational Guideline gives examples of what it means to be "engaged in research" but some hubs made their own determinations when trainees were engaged in activities not specifically addressed by the guideline. Also, hubs had been using a range of definitions in their pre-existing tracking systems (and some did not define it for respondents at all).

## **Errors and Areas of Confusion**

There was disagreement with or confusion about several of the metric exclusion criteria; including the following issues.

- Institutionally-funded scholars: In response to hub questions, NCATS clarified that scholars who are only institutionally-funded should be excluded from the Careers metric scores. A substantial number of hubs indicated that they don't distinguish scholars or graduates by funding stream source in their programs. Thus, providing incomplete metric results (those for only NIH-funded scholars), was confusing to program staff and other partners during their strategic management discussions. Despite the additional guidance, some hubs that administered programs with institutionally funded scholars continued to include those scholars in the metric.
- Scholars "still in training": The Operational Guideline states, "Trainees and scholars who are still in training...should not be included in this metric." In response to hub questions, NCATS clarified that if a trainee has completed the TL1/KL2 training program and is no longer on the TL1/KL2 grant, they can be assessed for whether they are engaged in research (added to the denominator of the metric).

If they have completed the TL1/KL2 training program and are participating in additional training (e.g., residency or PhD) that has dedicated time for research, they are considered "engaged in research" (added to the numerator of the metric).

Lost to follow-up: The Operational Guideline states that graduates for whom data cannot be obtained are not included in the numerator (number of graduates engaged in research) or in the denominator (total number of graduates). However, it does not give any additional guidance as to when a graduate should be considered lost to follow-up, how many and what types of efforts to locate graduates should be made, or acceptable response rates from graduate surveys. Further, hubs are not required to report the number or percent of their graduates who have been lost to follow-up and are therefore excluded from the Career metric scores. Given the variety of hub methods and frequency of graduate tracking, it is highly likely that hubs are not applying this exclusion uniformly and that lost to follow-up rates are quite variable across hubs. This creates the possibility for bias of results and gaming the metric.

Finally, this metric is cumulative rather than annual, and numbers of hubs reported that this made interpreting metric results difficult at the hub level.

## Developing and Implementing Strategic Management Plans

Further, many hubs asserted in their Turn the Curve plan that they were limited by the pool of applicants from which they can recruit and did not identify specific strategies that could change the characteristics of the pool because this is largely an institutional issue over which they had limited control.

## Metric-specific Contextual Factors

Meaningful interpretation of the range of metric values reported across hubs would require understanding the institutional, hub, and program-level characteristics that shape performance for the Careers in CTR metric values. A number of such characteristics were identified including:

- The extent of resources provided specifically for career development or networking
- Provision of exposure to or training in team science to enhance scholar ability to work effectively in research teams
- The extent to which hubs believe that they can improve participation of underrepresented persons in CTR

## **Drivers of Metric Performance**

A synthesis of information provided by hubs in their strategic management plans, and from a review of the literature, suggests careers in CTR may be facilitated when the following drivers are present:

- 1. Scholars receive high quality mentorship
- 2. Resources specifically for career development are provided
- 3. Networking is facilitated between current scholars, alumni, and other successful CTR researchers

- 4. Scholars develop research skills (e.g., grant writing, study management)
- 5. Scholars receive exposure to and training in team science
- 6. Recruitment, marketing, and applicant review strategies target more diverse applicants
- 7. Graduates and their career statuses are tracked over time (Tracking will not change the underlying rate at which they are engaged in CTR, but is necessary in order to collect the data for the Careers metric and understand the Story Behind the Curve)

The Driver Diagram and related Change Package distributed to hubs for the Careers in CTR metric is provided in Appendix E.

#### **Summary Recommendation 9:**

Make improvements to the first three Common Metrics by clarifying Operational Guidelines and assessing usefulness with hubs.

**Metrics: Careers in Clinical and Translational Science** 

- 9f Modify the Operational Guideline to further define and clarify exclusion criteria:
  - i. clarify whether hubs may use additional definitions of "engaged in research,"
  - ii. add exclusion of solely institutionally-funded scholars,
  - iii. clarify definition of "still in training,"
  - iv. add criteria for "lost to follow-up."
- 9g Consider modifying the metric to be annual (e.g., percent of 2015 graduates who are in CTR; percent of 2016 graduates who are in CTR, etc.).
- 9h Assess with hubs the usefulness of this metric for local improvement, particularly the extent to which the metric definitions reflect local priorities.

# **CLINICAL TRIAL ACCRUAL METRIC PILOT TEST**

## Background

Subsequent to the first three Common Metrics, an Accrual Metric Development Team developed a Median Accrual Ratio Metric and associated Operational Guideline. The Median Accrual Ratio is the median with-in trial ratio across a set of clinical trials:



The Common Metrics Implementation Team conducted a pilot test of the metric to determine the feasibility of collecting metric data, the quality of the data that could be collected, and the usefulness of the metric result for conducting strategic management of accrual at CTSA Consortium hubs and nationally.

## Methods

Eight hubs were selected in consultation with NCATS to participate in the four-month pilot, based on their expressed interest and site characteristics (institution size, use of a clinical trial management system [CTMS], and volume and types of clinical trials). Hubs with and without a CTMS were selected to represent a range of data collection capabilities across the CTSA Consortium. Each hub formed a team to participate in the pilot that included the Principal Investigator, subject matter experts, data system experts, and data analysts.

Participants received training in the Operational Guideline and RBA. During the 15-week pilot, hubs teams collected metric data, developed a Turn the Curve plan and participated in every-other week small group webinars to discuss their experiences, problem solve, and report on their progress. Operational Guideline clarifications and responses to hub questions were provided in conjunction with a subgroup of the Accrual Metric Development Team.

Data sources for the pilot included information collected by the Implementation Team from webinar notes, documentation in the Scorecard system, a post-pilot survey of hub teams, and key informant interviews following the data collection period.

## Results

## Accrual Metric Data Collection Feasibility and Quality

#### **Determining the Sampling Frame**

Of three possible sampling frames specified in the Operational Guideline, five of eight (62%) of hubs conducted a non-random sample of eligible clinical trials, and two of eight (25%) conducted a random sample. Only one hub (12.5%) was able to assess all eligible trials. Assessing all eligible trials, or a random sample of all eligible trials, requires identifying all such clinical trials at a hub's primary institution. However, hubs without a CTMS, or with a CTMS used by only a subset of trials, did not have a central list of trials at their institution to use for their sampling frame. The IRB was a potential source of such a list for some hubs, but IRB data were sometimes difficult to extract from existing electronic or paper systems, or the systems did not include the data elements needed to determine if trials were eligible for inclusion in the metric.

## **Applying Inclusion/Exclusion Criteria**

The Operational Guideline specified inclusion and exclusion criteria for trials to be used in calculating the median accrual ratio. Although most hubs had limited their sampling frame to a group of trials for which they believed these data were present, they were still not able to determine inclusion/exclusion criteria for all of the trials in the sample.

The Operational Guideline excluded trials with less than10 targeted participants to avoid potentially skewing the metric value by including very small trials, for which accrual is expected to be very low. Among the seven hubs that could estimate the effect of this exclusion on their sample of eligible trials, an average of 38.7% of clinical trials (median 32%; range 9%-74%) were excluded from hub sampling frames because they had less than10 targeted participants or it could not be determined if they had less than10 targeted participants. Hubs expressed concerns about the significant number of trials this criterion excluded and suggested lowering the minimum number of targeted participants.

## **Data Sources for Metric Variables**

Prior to pilot testing, it was assumed that CTMS use would improve the feasibility of creating the metric value. However, some hubs using a CTMS also needed to use additional information systems or collect primary data. Of the five hubs that used a CTMS to create the metric value, three used only the CTMS, and two used additional data sources (range three to six data sources). The three hubs that did not use a CTMS used three to five data sources. Data sources included non-CTMS electronic systems (e.g., eIRB), IRB progress reports, study protocols, surveys, emails or calls to study teams, and other methods. All hubs reported data quality issues.

## **Barriers to Collecting Key Metric Variables**

Data collection barriers were reported for all four variables in the accrual metric ratio. Barriers varied considerably by hub, data source(s) used, and the variable being collected.

#### Number of Participants Accrued

Hubs with a CTMS varied in their ability to determine the number of participants

accrued for a particular timeframe depending on the extent of CTMS use at the primary institution, and the system's reporting capabilities. Hubs without a CTMS collected this value via a research team survey or electronic data system. Survey respondents were not always clear about which participants to report, and multiple electronic data sources sometimes provided different values for the number of participants accrued to the same clinical trial.

## Number of Participants Targeted

At half of the pilot hubs, determining the number of targeted participants was a barrier for some or all clinical trials in the sampling frame. In several instances, the variable was not a field in the hub's CTMS or it was present but specified differently than in the Operational Guideline. Hubs without a CTMS typically used an eIRB database or a research team survey to collect the number of participants targeted. One hub that conducted a survey noted, "Only 50% of studies used a power calculation to obtain a number of targeted participants." Multiple hubs had considerable skepticism about the accuracy of data from IRB sources on the number of participants targeted, given investigator practice of overestimating to avoid IRB amendments.

## Date Open to Recruitment

For half of the pilot hubs, determining the date a trial opened to recruitment was a barrier for some or all clinical trials in their sample. Availability of data quality in the CTMS varied and non-CTMS hubs relied on research team survey data for this variable.

## Number of Days Planned to be Open to Recruitment

This was the most problematic of the variables in the accrual metric. Seven out of eight pilot hubs reported barriers to collecting this variable for some or all of the clinical trials in their sample. For hubs with a CTMS, the variable was not present, not specified as per the Operational Guideline, or not credible to the hub team. For hubs without a CTMS, the number was not available in pre-existing databases.

## Number of Trials in the Median Accrual Ratio

After applying sampling frames and inclusion/exclusion criteria and removing trials with missing or incomplete data (including survey non-response), the mean number of clinical trials included in the Median Accrual Ratio across pilot hubs was 76.1 (median=57.5; range 6-212). In three out of eight hubs (37.5%), the Median Accrual Ratio was calculated based on fewer than 20 eligible clinical trials.

## CTMS, Data System, and Survey Considerations

Whether using a CTMS or another electronic data system, many of the same challenges presented: system data entry requirements were not aligned with the Operational Guideline, systems were missing fields, and study team data entry was inconsistent. To address this, sites used proxy variables (e.g., projected length of study vs. projected days of recruitment), removed trials from the sample, or collected new primary data. Using a proxy variable compromised the accuracy of the metric result, and removing trials from the sample undermined the representativeness of their reported sample. Fifty percent of all pilot hubs conducted a REDCap survey to collect and/or verify at least one data element or inclusion/exclusion criterion. Considerable time was spent planning and fielding surveys, limiting the time the surveys could remain open, and survey response rates were therefore low. Hubs described logistical survey challenges and suggested that some study Principal Investigators may have concerns about sharing study data.

## Level of Personnel Effort and Feasibility of Expanding the Sample

Most hubs reported that their pilot data collection approach was not sustainable due to the high level of effort required, the low confidence in the data collected, and diversion of effort from other important CTSA areas and strategic management for accrual. Many respondents worried that expanding data collection to all eligible trials would be a "big mandate" that would not be feasible for at least two to three years, and some believed it would require a long, complex planning process. Even respondents from the site that used a CTMS more broadly expressed concern because not all trials use the CTMS and the CTSA would have limited influence when promoting mandatory entry of accrual data and conducting strategic management.

Five of the six hubs without broad CTMS implementation reported advocating for or planning to implement an institution-wide CTMS and expected feasibility of expanding their sample to additional clinical trials to increase in the future.

## Accrual Metric Usefulness for Strategic Management

All hubs believed in the importance of measuring and improving clinical trial accrual outcomes at their hub. Although only one hub produced an accrual metric value prior to the end of the pilot test, and no hubs used the value to inform their Turn the Curve plans, hubs expressed opinions and reported their observations about potential usefulness of the metric for accrual strategic management.

## Turn the Curve Plan Development

All hubs developed a Turn the Curve plan, and most described their data collection feasibility and quality issues, their current trial accrual barriers, and their potential improvement strategies. One hub solely addressed improving their metric data collection rather than clinical trial accrual. One hub did not provide a metric value because data were available for only a small number of trials and they felt the metric value would not be representative of the intended sample. Improvement strategies from across the pilot hubs were compiled in a resource document and provided to the accrual metric pilot participants.

## **Data Issues Affecting Strategic Management**

Hubs reported that data issues would be a barrier to usefulness of the metric for strategic management at a hub and/or CTSA Consortium level. This was seen as being because of low confidence in the quality of data collected at one's hub, concern that different hubs were not collecting data in the same way, and not yet having the ability to see trends with multiple data points. Given the small number of clinical trials (relative to all clinical trials) included in their Median Accrual Ratio, several hubs questioned the representativeness of their median relative to their intended sampling frame. Some respondents were concerned

that the effort required to collect data for the accrual metric would be so high that there would be little or no resources available for strategic management. Non-response bias and low response rates decrease the generalizability and usefulness of metric data collected through a survey.

#### Usefulness of the Median Value and Trial Level Accrual Ratio for Strategic Management

Over three-quarters (78%) of respondents believed that the median value would be useful for strategic management, but they also described limitations or concerns. The most common barrier identified was difficulty interpreting the value because it is difficult to understand the meaning of a single number that represents a broad spectrum of trials. Others found it difficult to use a median value to judge their hub's performance without comparator data or a benchmark. Moreover, even were comparison information available, a few respondents cautioned that a high median value relative to other hubs does not necessarily mean there are no areas for improvement. Some respondents noted that a median value may not change over time even when there are effective improvement efforts targeted toward a subset of trials. As a result, some respondents concluded that their CTSAs would have limited ability to influence change to improve the median ratio.

Most respondents thought the accrual ratio would be useful for strategic management at the trial level, with some describing limitations or concerns. The strategy most commonly described was to identify and provide support to poorly accruing trials. Many respondents had difficulty understanding the ratio and predicted that stakeholders, including Principal Investigators and institutional leadership, would have this same challenge. Contextual factors also were cited as important for interpreting accrual metric results. Many respondents noted that accrual is influenced by a variety of variables, such as study size, patient population, and type of treatment. In addition, several respondents noted that accrual does not necessarily occur in a linear fashion and that this is not accounted for in the metric Operational Guideline.

#### **Aggregating Accrual Metric Results**

The majority (63%) of respondents had mixed opinions on whether accrual metric results should be aggregated across the CTSA Consortium, and 28% described only limitations or concerns. The most common potential benefits cited were to help a hub assess its own accrual performance and to set hub-specific accrual goals. Many respondents cautioned against aggregating data without considering context and data quality to prevent comparisons between dissimilar studies or hubs. Some respondents guessed that other hubs also would have low confidence in their own data, and would conclude that an aggregate dataset with low-quality data points would not be useful.

#### Summary Recommendation 10:

# Use the results and recommendations detailed in the full Accrual Metric Pilot report to determine the future direction of metric implementation.

- 10a Consider providing an infrastructure-building period prior to mandated collection of metric data to allow hubs time to devise and/or revise data sources and systems and data collection and data quality procedures, and train personnel.
- 10b Modify the metric to be collected prospectively rather than retrospectively to increase its potential usefulness for strategic management, including the ability to identify and intervene in individual trials as needed.
- 10c Revise the Operational Guideline to address certain multi-site clinical trials (e.g., those of competitive enrollment design) in which key accrual metric variables are not known.
- 10d Re-evaluate the exclusion criterion for trials with fewer than 10 targeted participants; consider lowering the cut-off (e.g., to trials with less than five targeted participants).
- 10e Do not exclude clinical trials of dose-to-toxicity design.
- 10f Collect and report additional information, including information about the mix of clinical trials at the primary institution or included in the Median Accrual Ratio, to understand how representative the median is of the intended sample.
- 10g Provide a template of tested survey questions and survey considerations.
- 10h Provide hubs with best or promising practices and strategies for implementing a CTMS to produce metrics.

## TRANSITION TO THE COORDINATING CENTER

In June, 2017, NCATS designated CLIC as the new coordinating center to support activities of the Common Metrics Initiative. After joint development of a transition timeline by CLIC and Tufts CTSI (Appendix H), CLIC personnel joined the standing meetings with NCATS to assure smooth conduct of transition activities. The Common Metrics Initiative website was transitioned from Tufts CTSI to one hosted by CLIC in December, 2017. The Implementation Team provided training to CLIC personnel in providing technical assistance in applying the Operational Guidelines, and in data checking. A number of resources were also developed and shared, including a Questions Inventory database and a detailed Operations Manual. Implementation Team members also "shadowed" new CLIC Common Metrics staff for one month prior to completion of the transition.

# PART II. EVALUATION STUDY

Between March, 2016, and June, 2018, the Tufts CTSI Common Metrics Evaluation Team conducted a formal evaluation to assess hubs' experiences implementing the initial three Common Metrics and RBA management framework. This 28-month study period 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 (Figure 8). The multi-faceted evaluation used quantitative and qualitative methods to address:

- hub progress on implementing Common Metrics and performance improvement activities and related prior experience,
- challenges, facilitators, and contextual factors that affected that progress,
- hub perspectives on the role of the Tufts Implementation Program,
- ways in which hubs integrated Common Metrics work and personnel resources expended, and
- perceived benefit and concerns.

Figure 8. Timeline for Common Metrics evaluation study



## **RESEARCH QUESTIONS**

The goal of the evaluation was to provide a multi-faceted, rich understanding of hubs' progress and experiences. We assessed four specific research questions.

- 1. To what extent did hubs achieve meaningful implementation of performance management in relation to the initial three Common Metrics? The success of the Common Metrics Initiative depended on the extent to which hubs implemented the Common Metrics and performance management activities and whether that effort produced a benefit. Implementation was defined as completing activities related to computing the metric result and applying the RBA performance management framework. Although the timeframe of the evaluation did not allow for a full assessment of whether hubs improved their metric scores, "meaningfulness" was defined as perceived benefit to hubs in their abilities to manage performance.
- 2. What were the challenges and facilitators for achieving meaningful application of performance management? Although prior models existed for implementing metrics and metric-based management in other settings, doing so in a research-oriented environment and among 60 diverse organizations across the country was novel and entailed many unknowns. Understanding the challenges encountered and factors that facilitated meaningful application is essential not only for understanding hubs' progress but also for guiding decisions about future directions and potential expansion.
- 3. What was the role of the strategies employed by the Tufts Implementation Program? The Tufts Implementation Program sought to provide the necessary knowledge and tools to support implementation of the Common Metrics and a shared performance management framework. The evaluation assessed the role of the Tufts Implementation Program by eliciting hubs' feedback on the program's usefulness for the work of the Common Metrics.
- 4. How did CTSAs approach the work of implementing the initial three Common Metrics? Introducing new metrics and a shared performance management framework necessarily entailed changes to organizational processes and, potentially, structures and priorities. We anticipated that the heterogeneity among organizations within the national CTSA Consortium would generate wide variation in team structures and ways of organizing activities. In addition, personnel effort and resources expended to complete the work is an important factor for understanding and determining for the future what is necessary for successful Common Metrics implementation.

# **METHODS**

## **Study Design**

We used an intervention mixed methods framework<sup>3</sup> to describe in detail hubs' progress and experiences implementing the Common Metrics and the shared performance improvement framework. The post-test design integrated quantitative measures, openended textual responses, and nested qualitative interviews to describe *what* hubs achieved in relation to the initial three Common Metrics, *how* hubs incorporated this work into their existing structures, and *why* full implementation was or was not achieved.

The conceptual framework for the evaluation identified factors that may have affected hubs' abilities to implement Common Metrics, the strategies used by the Tufts Implementation Program in order to facilitate implementation, and expected outcomes by time period (Figure 9):

- immediate
- short term, and
- medium- to long-term.

Environmental factors that catalyzed the Common Metrics Initiative are also represented. The IOM report on the CTSA Program and the resulting charge for CTSAs to function and show impact as a network are central to any representation of the Common Metrics implementation. By definition, these factors did not vary between hubs and, therefore, were not explicitly measured in the evaluation.

In contrast, organizational factors varied by hub and could influence outcomes. These factors included expectations that hubs with larger funding awards would have more resources to implement the Common Metrics and, therefore, might be better positioned to complete that work. Also, we expected that earlier funded hubs would have more established work processes and relationships that might facilitate the Common Metrics work. Finally, we anticipated that hubs having previous experience with metric-based performance management might be better equipped for implementing Common Metrics.

Additionally, we anticipated that participation in the Tufts Common Metrics Implementation Program would directly facilitate data collection, metric computation, and the conduct of performance improvement activities. The Tufts program was based in a continuous learning approach that was modified for later Implementation Groups, which we expected to enhance Common Metrics implementation for those groups.

#### Figure 9. Conceptual framework for evaluation study



Considering the overall aim of organizational change nationally, the timeframe of the evaluation was relatively short. Within this constraint, the design focused on immediate and short-term outcomes:

- knowledge and proficiency gained from the Tufts Implementation Program,
- approaches to integrating the work of the Common Metrics into existing organizational structures,
- implementation of the initial three Common Metrics and performance management,
- self-assessed benefit to hubs, and
- personnel resources expended.

Aspects of the implementation of Common Metrics necessitated a descriptive study that focused on understanding hubs' progress and experiences. A controlled comparison group design was not compatible with the goal of having every hub implement the Common Metrics and a common performance management framework in the same time period. In consultation with NCATS, the Tufts Implementation Team prioritized the fullest possible adoption of the metrics and management framework by every hub. As a result, all hubs were offered the most complete program of training, coaching, and support that was feasible. Without a control group, we considered a quasi-experimental pre-post design but could not fully pursue this option because assessing change in performance on the Common Metrics' results was not feasible. First, the metric definitions were newly released and not all hubs had retrospective data to compute

the metric result for a prior time period. Second, even if hubs could collect retrospective data, achieving change in performance on the metric results was not expected within the relatively short time period of the study. The resulting mixed method approach yielded a multi-faceted understanding of hubs' progress and related contextual factors, challenges, and facilitators.

## **Data and Analysis**

To generate a full picture of the Implementation, we integrated quantitative and qualitative methods at multiple levels. First, to collect data from hubs, we used surveys, including both forced choice and open-text questions, and semi-structured qualitative interviews. Second, participants in the quantitative and qualitative components of the study were linked. Specifically, the sampling frame for the qualitative interviews was composed of hubs that were asked to complete the surveys. Third, at the level of analysis, results from independent analyses were merged to develop a fuller description and understanding.<sup>3</sup>

#### Outcomes

## **Primary Outcome**

The primary evaluation outcome was implementation of the initial three Common Metrics and performance management framework within the study timeframe. Implementation was measured quantitatively as the extent of completion of a range of related activities. In addition, to better understand the extent of completion, we elicited reasons for not completing each activity, when applicable, and conducted semi-structured interviews about contextual factors, challenges, and facilitators.

Implementation of Common Metrics and the performance management framework was defined as completion of 13 specific activities (Table 3). Following the RBA framework, activities ranged from collecting metric data to implementing a performance improvement plan. For analytic purposes, we further clustered the activities into five distinct groups:

- creating the metric result,
- understanding current performance,
- developing a performance improvement plan,
- implementing the performance improvement plan, and
- documenting the metric result and plan fully.

With input from the Tufts Implementation Team, we assigned a point value to each activity. The sum of a hub's points indicated the extent to which it completed the activities of the performance management framework. The activities were not weighted for relative difficulty, effort, or time required because hub experiences varied in this regard.

Table 3. Implementation of Common Metrics and performance improvement activities:
definition and point assignments

Cluster and Activities	Points Possible
Creating the Metric	
Collected data	1.0
<ul> <li>Computed metric result according to Operational Guideline (self-report)</li> </ul>	1.0
Understanding Current Performance	
<ul> <li>Forecasted future results -or- compared result to any other data</li> </ul>	1.0
<ul> <li>Specified underlying reasons involving hub leadership/staff/faculty</li> </ul>	0.5
<ul> <li>Specified underlying reasons involving any group outside hub leadership/staff/ faculty</li> </ul>	0.5
Developing a Performance Improvement Plan	
<ul> <li>Involved hub leadership/staff/faculty when developing improvement plan</li> </ul>	0.5
<ul> <li>Involved any group outside hub leadership/staff/faculty when developing im- provement plan</li> </ul>	0.5
<ul> <li>Specified actions for achieving desired outcome</li> </ul>	1.0
Prioritized actions	0.5
<ul> <li>When prioritizing actions, considered potential effectiveness of actions -or- feasibility</li> </ul>	0.5
Implementing the Performance Improvement Plan	
<ul> <li>Reached out to specific individuals or institutional partners for help in carrying out improvement plan</li> </ul>	1.0
<ul> <li>Began to implement improvement plan</li> </ul>	1.0
Documenting Metric Result and Plan Fully	
<ul> <li>Documented 5 elements in the Common Metric-specific Scorecard - metric result; underlying reasons (Story Behind the Curve); potential partners; potential actions (What Works); planned actions (Strategies)</li> </ul>	1.0
Total possible	10.0

#### Secondary Outcomes

Following the conceptual framework, two additional short-range outcomes were: i) perceived benefit and ii) personnel resources expended. Perceived benefit was defined as positive change in hubs' self-assessed ability to manage performance (Table 4), measured on a four-point Likert-type scale (not at all, a little, some, a lot). These results were merged with themes resulting from qualitative interviews regarding hubs' perceived value of the Common Metrics Initiative.

#### Table 4. Self-assessed ability to manage performance

#### Criterion

To what extent is your hub able to ...

- Assess whether its current performance is on track to meet its goals, aims, and objectives
- Assess whether future performance is likely to be on track to meet its goals, aims, and objectives
- · Engage hub leaders, faculty, and staff in discussions about operational or strategic issues
- · Engage stakeholders outside the hub in discussions about operational or strategic issues
- · Identify actions that have potential to influence/improve performance
- Efficiently address performance issues
- Effectively address performance issues
- Advance clinical and translational science

Personnel resources required to maintain implementation of the initial three Common Metrics and performance management activities were assessed mainly through open-text descriptions of the personnel positions involved, the approximate number of personnel hours used, and other important topics for understanding the time and effort involved. To further probe the relative degree of effort to implement each metric, multiple choice questions were used to compare the metrics.

In addition to short-range outcomes, there were two immediate process-related outcomes that might help hubs achieve full implementation of the performance improvement framework. First, we assessed the amount of knowledge and proficiency that hubs gained from the Tufts Implementation Program using quantitative measures with a five-point scale ranging from "much less than needed" to "much more than needed." These results were supplemented by qualitative themes related to the Tufts Implementation Program.

Second, we determined how hubs organized and integrated their work to implement the Common Metrics and performance management. To encourage a broad view of activities, we used a project management framework and provided the following definition:

Main activities are broad groups of related tasks that were important to implementing RBA in your hub. Activities may be related to the conception, planning, execution, documenting, quality control and closing of an RBA cycle, as well as training required. Feel free to label activities using RBA terminology (e.g., developing turn the curve plans) or other descriptors (e.g., data collection, data management, project management, communications, stakeholder engagement, etc.).

We also elicited open-text responses about changes in approach to implementing the Common Metrics that were being planned for the future.

## **Data Collection Instruments**

We collected data at various time points throughout the implementation period using five self-report surveys and a qualitative interview guide that was adapted for three participant roles (Table 5).

Instrument	Data collection period
Cohort assignment survey*	2016 March
Baseline survey	2016 July - September (by Implementation Group)
Follow-up survey 1	2016 October - 2017 March (by Implementation Group)
Qualitative interview guide	2017 November – 2018 March
Follow-up survey 2	2018 January - February
Organizational resources survey	2018 February - April

#### Table 5. Data collection instruments and chronology

\*The term to describe a group of hubs that began training at the same time initially was "cohort" and later changed to "Implementation Group".

## **Cohort Assignment Survey**

Prior to the start of the Tufts Implementation Program, participating hubs completed a cross-sectional survey designed to understand hubs' prior experiences with metric-based performance management and their preferences for when to begin the training. Survey topics included:

- Preference for participating in the first, second, or third cohort (i.e., Implementation Group)
- Anticipated difficulties implementing each of the initial three Common Metrics
- Hub experience in 2015 with:
  - metric data collection,
  - performance management activities and framework(s).

In addition to informing the formation of Implementation Groups, the Evaluation Team used these data to construct a composite measure of each hub's prior experience with data-driven performance improvement.

## **Baseline and Follow-up Surveys**

Hubs completed three surveys to understand performance improvement. At baseline, we asked hubs to choose the one local metric that best exemplified how the hub used metric data in the prior five months and to report on activities of performance management, as defined by the primary outcome. These data were used to achieve a sample for qualitative interviews that was balanced according to prior experience (see Data Collection and Hub Participation).

Data regarding hub progress on the Common Metrics were collected with two followup surveys. The first follow-up survey asked hubs to report progress on completing performance management activities for the Common Metric that best exemplified the hub's use of metric data and RBA at the end of the small-group coaching period. Hubs could report on the metric on which they focused during small-group coaching sessions but were not required to do so. The second follow-up survey recorded any additional performance management activities completed for the metric reported on during the first follow-up survey, activities completed for the remaining two Common Metrics, and additional information about hub experiences (Table 6). Questions about metric data collection and performance management followed the same logical skip patterns across surveys (Figure 10).

#### Table 6. Baseline and follow-up survey topics

Торіс	Baseline	Follow-up 1	Follow-up 2
Performance management activities	One Local Metric <sup>*</sup>	Any Common Metric	All Common Metrics
Completion of each activity per metric	$\checkmark$	$\checkmark$	$\checkmark$
Date of completion of activity, if applicable	$\checkmark$	$\checkmark$	$\checkmark$
<ul> <li>Reason for lack of completion of each activity, if applicable</li> </ul>			$\checkmark$
<ul> <li>Documentation of activities in software platform</li> </ul>			$\checkmark$
<ul> <li>Open-ended comments on status of data collection and performance management</li> </ul>		optional	optional
Self-assessment			
Overall ability to manage performance	$\checkmark$	$\checkmark$	$\checkmark$
Assessment of Tufts Implementation Program			
Amount and time allocated to components			✓
<ul> <li>Knowledge and proficiency gained</li> </ul>			$\checkmark$
Overall effectiveness and satisfaction			$\checkmark$
Other			
General open-ended comments			optional

\*For the baseline survey, hubs were instructed to choose any one metric that best exemplified how the hub used metric data in the prior five months.

Timing of the baseline and first follow-up surveys followed the Tufts Implementation Program's group-based approach. At baseline, Implementation Group 1 received the survey after the training period but before the small group coaching; Implementation Groups 2 and 3 received this survey on the initial day of the Tufts Implementation Program. Each cohort received the first follow-up survey between one and two weeks after the end of the respective small group coaching periods. In this survey, hubs reported progress on completing performance management activities for one of the Common Metrics. Near the end of the study, all hubs concurrently received the second follow-up survey.

# Figure 10. Flow of survey items related to implementation of Common Metrics and performance improvement activities, with allocated points



#### **Qualitative Interview Guide**

A semi-structured interview protocol included open-ended questions and probes to ensure interviews would elicit an in-depth understanding of experiences and viewpoints on the challenges, facilitators, and contextual factors for implementing Common Metrics (Table 7). The Context Matters Framework<sup>4</sup> was used to consider the different layers of context to probe. This framework identifies and sorts contextual factors into five domains: 1) the specific implementation setting, 2) wider organizational setting, 3) external environment, 4) implementation pathway, and 5) motivation for implementation. We applied this framework to ensure that we fully captured the various factors that might have influenced hubs' experiences with Common Metrics Implementation.

#### **Table 7. Semi-structured interview topics**

Торіс
Background information on the participant and the hub
Initial perceptions of the Common Metrics and the performance improvement framework
Preparation for implementation and order in which hubs addressed the metrics (if not simultaneously)
Experience, facilitators, and challenges for implementing Common Metrics
<ul> <li>Probing for contextual characteristics that may have influenced the experience</li> </ul>
Hub characteristics
Local institutional characteristics
<ul> <li>Role of Tufts Implementation Program and National Center for Advancing Translational Sciences (NCATS)</li> </ul>
<ul> <li>Environmental characteristics external to the local institution</li> </ul>
Probing for type of experience:
Evolution of implementation over time
Meaningfulness and added value of implementing Common Metrics
Recommendations
Common Metrics Initiative
<ul> <li>Advice for a new hub starting to implement the Common Metrics</li> </ul>

We adapted the guide as needed for three participant roles per hub: the hub's Principal Investigator, the Administrator/Executive Director (or another individual filling the role of 'champion' at the hub), and a staff member knowledgeable about day-to-day implementation of the Common Metrics. These three roles were identified as providing a comprehensive view of each hub's experience as they represented leadership, management, and on-the-ground implementation perspectives. We refer to these participant roles as "Principal Investigator", "Administrator", and "Implementer." We piloted each version of the guide during mock interviews with personnel from Tufts CTSI. After each interview, the qualitative team members debriefed and revised the interview guide as needed to clarify content and improve the flow of the interview.

## **Organizational Resources Survey**

This cross-sectional survey focused on the period following the training and coaching period to approximate activities and personnel that typically will be involved in the annual updates required by the Operational Guidelines. Questions pertained to the most recently required Common Metrics update, which was due in August, 2017. Using an open-text format, the survey asked about the following topics:

- Hub activities to implement Common Metrics and performance management (up to five)
- Personnel involved in carrying out the activities (up to five) and approximate hours used
- Optional open-ended comments on time and effort involved, planned changes in approach, and recommendations for NCATS

The survey was pretested and revised to reduce participant burden.

## Administrative Data

We also collected data on hub size and funding cohort from NCATS and confirmed it through publically available sources when possible. Hub size was defined as total funding from U, T, K, and/or R grants for 2015-2016. Hub funding cohort was calculated based on the year it was first funded.

## **Data Collection and Hub Participation**

## Surveys

All hubs received each survey, making a sampling plan unnecessary. Surveys were selfadministered online using REDCap software. For the cohort assignment, baseline, and follow-up surveys, we sent an invitation email to one Principal Investigator per hub, who was instructed to assign one person to complete the survey with input from others at the hub. For the organizational resources survey, we sent the invitation email to the Executive Director or Administrator at each hub, who was instructed to either complete it entirely or ask others who had the information necessary to respond to all or part of the survey.

To maximize response rate, we sent reminder emails from Evaluation Team members and Tufts CTSI's Principal Investigator. The promptness of hub responses differed across surveys. Depending on the survey, we sent up to two, four, or five reminders to encourage participation.

Hub participation in surveys ranged from 82% - 100% (Figure 11). All eligible hubs responded to the cohort assignment and baseline surveys. The two follow-up surveys received 95% response (57 out of 60) and 98% response (59 out of 60), respectively. Although slightly fewer hubs responded to the organizational resources survey (82%, 49 out of 60), the response rate was sufficient to lend confidence to the results.



#### Figure 11. Data collection instruments and response rates

\*At the time of the Cohort Assignment survey, there were 59 eligible hubs.

#### **Semi-structured Interviews**

#### Selecting Hubs

We conducted semi-structured qualitative interviews with a nested sub-sample of hubs. The sampling frame was comprised of hubs eligible for the surveys. The sampling plan sought a balanced portrayal across the range of hubs' experience with performance improvement activities and, secondarily, across key hub characteristics. To achieve balance across experience with performance improvement activities, we created a matrix of hub scores on the study's primary outcome at two time points. The first score ("prior experience") was defined as the extent of completion of metric computation and performance improvement activities for any local metric in the five months prior to the Tufts Common Metrics Implementation program (i.e., the metric reported on at baseline). The second score ("initial progress on a Common Metric") was defined as the extent of completion of metric computation and performance improvement activities for one of the Common Metrics at the end of the training and coaching period (i.e., the metric reported on in the first follow-up survey). At each time point, we trichotomized hub scores into three levels: minimal (0.0 - 4.5), moderate (5.0 - 8.5), significant (9.0 - 10.0). Three categories of scores at two time points created a sampling matrix of hubs in nine cells (Table 8).
Prior Experience	Initial Progress on a Common Metric					
Number of prior performance improvement activities <sup>**</sup> completed for any local metric (Baseline survey)	Number of performance improvement activities <sup>**</sup> completed for one Common Metric by the end of the coaching period (Follow-up survey 1)					
(Duschine survey)	Minimal	Moderate	Significant			
Minimal	5	8	6			
Moderate	2	11	6			
Significant	1	6	12			

### Table 8. Semi-structured interviews: sampling frame (N=57 hubs\*)

\*Three hubs did not respond to one or more of the surveys and were excluded from the sampling frame. \*\* Activities range from 0-10. Minimal = 0.0-4.5, moderate=5.0-8.5, significant=9.0-10.0.

With a target sample size of half of the eligible hubs participating (i.e., 30), the goal was to select three or four hubs for each of the nine cells. For the two cells with fewer than three hubs ("significant-minimal" and "moderate-minimal"), we selected all hubs in those cells for inclusion in the sample. The remaining cells had more than four members, and hubs were selected through simple random sampling. To stay within our planned sample of 30 hubs, we targeted four hubs in each of these cells except for the "significant-significant" category, for which we targeted three hubs.

The rationale for targeting fewer hubs in the "significant-significant" category was that these hubs reported a high level of activities at both time points, which would have provided less insight into challenges and facilitators to implementation compared to hubs that experienced change in the level of completion or inability to improve. Review of preliminary themes during data collection with the three hubs initially selected for this cell indicated that thematic saturation had been reached and the likelihood that a fourth hub would yield significant new information was relatively low.

The resulting sample was reviewed 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). Balance was assessed both overall and within cells to the extent possible.

Three hubs from the original sample for qualitative interviews did not participate and were replaced using the same sampling strategy. Reasons for not participating were an inability to provide participants for three interviews (Principal Investigator, Administrator, and Implementer) or lack of response to recruitment attempts. One of the hubs that did not respond to recruitment attempts was the only hub in the "significant-minimal" category. To ensure sufficient insight into facilitators and strategies for overcoming challenges, we replaced this hub by sampling an additional hub from the "minimal-significant" cell. The final sample included 30 hubs (Table 9).

<b>Prior Experience</b> Number of prior performance improvement activities <sup>*</sup> completed for any metric (Baseline survey)	Initial Progress on a Common Metric Number of performance improvement activities <sup>*</sup> completed for one Common Metric by the end of the coaching period (Follow-up survey 1)					
	Minimal	Significant				
Minimal	4	4	5			
Moderate	2	4	4			
Significant	0	4	3			

# Table 9. Semi-structured interviews: final sample of hubs (N=30)

\* Activities range from 0-10. Minimal = 0.0-4.5, moderate=5.0-8.5, significant=9.0-10.0.

# **Recruiting Participants**

Recruitment for qualitative interviews began by sending an initial email invitation to a Principal Investigator at each hub in order to obtain agreement for the hub to participate. When needed, follow-up emails to the Principal Investigator and hub Administrator were sent to attempt to obtain agreement to participate. Up to three telephone calls were made to reach the hub Principal Investigator or Administrator.

Upon agreeing to participate, the Principal Investigator or Administrator were asked to nominate individuals in the other two roles to be invited for interviews. In cases in which more than one individual at a hub filled one of the roles of interest, we relied on hub leadership to choose who (within each role) would be most knowledgeable of the hub's experience with implementing the Common Metrics. Email invitations were sent to those nominated, and follow-up emails and telephone calls were made as needed. If interviews for all three roles could not be scheduled, another hub was selected.

All 30 participating hubs provided participants for the three role-specific interviews, totaling 90 interviews. For eight hubs, the Implementer interview had more than one participant so as to share a wider range of experiences (for example, when an individual participant had experience with only one of the three Common Metrics).

# Conducting Interviews (N=90)

One interviewer conducted each study interview by telephone using a semi-structured format. Interviews with the three participant roles (Principal Investigator, Administrator, and Implementer) lasted approximately 20, 40 and 60 minutes, respectively. Upon scheduling a time for an interview, each participant was emailed an information sheet describing the study. Prior to beginning the interview, participants provided their verbal consent to participate in the interview and have it audio recorded. Interviewers also took notes during interviews and conducted reflective journaling after each interview. These reflections and ongoing experiences using the interview protocol were discussed during weekly meetings with the wider qualitative team to identify emerging themes and assess saturation.

# Interviewer Training and Quality Assurance

Interviewer training entailed mock interviews and debriefing to ensure consistency. The more experienced interviewer conducted the cognitive testing interviews, followed by debriefing with the study team. This interviewer then conducted the first several study interviews, and these were listened to by the second study interviewer, followed by a debriefing between the two interviewers. The second interviewer then conducted additional mock interviews with Tufts CTSI personnel prior to conducting study interviews, followed by similar debriefing sessions.

Additionally, during weekly meetings, qualitative team members discussed participants' experiences with interview questions. Following procedures for qualitative interviewing, the team identified additional language to further facilitate future interviews (e.g., clarifying the difference between different types of calls that were part of the CMI training process). Finally, the interviewers periodically reviewed a subset of later interviews conducted by each interviewer to ensure consistent application of the interview protocols.

### **Analytic Approach**

Quantitative and qualitative data were analyzed independently, and results were merged to develop a full description of hub experiences implementing the first three Common Metrics.<sup>3</sup> Overall, 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). Qualitative data also provided insights to help explain associations seen in statistical analyses.

#### **Statistical Analysis**

Hub characteristics were described overall and by Implementation Group using means and standard deviations for continuous variables and proportions for categorical variables. We tested for differences in these characteristics by Implementation Group using t-tests and chi-squared tests, where appropriate. Similar numeric summaries were used to describe the frequencies of completion of performance improvement activities. As each hub was assessed on each of three metrics, we tested for differences in mean completion by 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 (Table 10). For the multivariable linear regression model, a stepwise variable selection procedure using AIC was performed, starting with a full model including all covariates and proceeding with both backward and forward selection.

#### Table 10. Hub characteristics used in linear regression modeling, by domain

Characteristic
Hub attributes
Size at start of Common Metrics Implementation* Initial funding cohort
Experience with metric-based performance improvement
Maturity of performance management system** Extent of automated data collection Extent of data stored in centralized database
Participation in the Tufts Implementation Program
Attendance of one or more person(s) from the hub at each training session (n=7) Attendance of one or more person(s) from the hub at each coaching session (n=6) Self-selected metric the hub focused on during coaching sessions Primary coach

\* CTSA size is defined as total funding from U, T, K, and/or R grants for 2015-2016.

\*\* Created through factor analysis; see Table 11.

# Factor Analysis

To summarize a hub's prior experience with metric-based performance management, we conducted a factor analysis to create an experience factor score. The factor analysis used 10 items from the cohort assignment survey (Table 11). Each response category was assigned a numerical value such that a higher value indicated a higher level of experience. For questions with multiple parts, the "yes" responses were summed to create a single score for that item.

#### Table 11. Survey items used in factor analysis

Survey Item	Values
In 2015, to what extent did your CTSA collect data for use as performance metrics?	1-4
In 2015, to what degree were these performance metrics used to quantify progress on specific CTSA goals, objectives and/or performance targets?	1-4
In 2015, which CTSA components were monitored through one or more performance metrics: research services/resources, education/training programs, central administration/management, and special initiatives/projects?	0-4
In 2015, did your CTSA have procedures in place for disseminating and reviewing results on performance metrics with: CTSA faculty and staff, internal CTSA oversight committees and boards, any other key stakeholders?	0-3
In 2015, considering all of the key decisions that were made regarding CTSA opera- tions, how much did you rely on performance metrics to make these decisions?	1-4
In 2015, considering all of the key strategic decisions that were made regarding CTSA strategy, how much did you rely on performance metrics to make those decisions?	1-4
In 2015, to what degree was the data collection process automated?	1-3
In 2015, how much of your CTSA's performance data were stored in a centralized CTSA database?	1-5
Does your CTSA currently use at least one improvement method to address the results on performance metrics?	0-1
How would you rate your CTSA's capability to use performance metrics to monitor and/or manage your CTSA in 2015?	1-5

All 10 dimensions were used in an exploratory factor analysis with a two-factor model based on the proportion of variance explained. A single factor score was created to represent the "maturity of a performance management system," 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 Interviews

Interview recordings were transcribed verbatim by a professional transcription company. Transcripts were uploaded into the NVivo qualitative data analysis software<sup>5</sup> to facilitate coding and analysis.

The codebook was developed using a two-stage consensus-based process. First, the

qualitative team developed an initial codebook. The main topics of the interview protocol were used as a set of pre-identified nodes for the coding scheme. Then, analysts reviewed two transcripts, interview notes, and reflections to identify emergent concepts. These preidentified nodes and emergent concepts were merged into a single initial codebook. This codebook was reviewed by the team for clarity and consistency.

Second, analysts applied the initial codebook to two rounds of transcripts to ensure definitions were clear and were being applied consistently. During the first round, five analysts applied the initial codebook to one transcript and met to review coding and resolve discrepancies. The codebook was revised as needed. During the second round, three analysts who would be involved in the coding process applied the codebook to three additional transcripts, one from each participant role. Again, the team met to review the coding and resolve discrepancies, and the codebook was revised as needed.

Once consensus was reached, one team member coded the interviews using the codebook. Another team member periodically reviewed coded transcripts for fidelity to the codebook. The full qualitative team discussed all potential new themes or revisions as a group before any changes were made to the codebook.

Over the course of coding transcripts, themes were grouped into four domains: metricdesign and content, stakeholder engagement, hub engagement, and perceived value of implementing Common Metrics. Each domain was assigned to one analyst to perform a summary analysis of the range of themes in that domain. The analyst read all quotes for the codes within the domain, sorted them into specific facilitators and challenges, and then summarized the range of themes, including illustrative quotations. Each analyst also identified any intersections among themes that emerged, which were discussed by the full team and incorporated into the presentation of results. Sub-analyses were conducted to understand whether two topics of particular interest to the evaluation (hubs' engagement with the Common Metrics Implementation and the hubs' perceived value of the initiative) differed by participant role (Principal Investigator, Administrator, Implementer).

#### **Open-ended Survey Responses**

Both the second follow-up survey and the organizational resources survey included open-ended survey responses. Coding and analysis of these responses followed similar consensus-based procedures as for the longer semi-structured interviews.

For responses from the second follow-up survey, a team of two analysts independently reviewed the text to develop initial codes. These analysts met to discuss their coding schemas and develop an initial codebook. Then, each analyst applied the codebook to a subset of responses to ensure definitions were clear and codes were being applied consistently. The analysts met to discuss and resolve discrepancies, and the codebook was modified as needed. After nine meetings, the analysts were applying the codebook consistently. At that point, one analyst coded the remaining responses and noted any questions to discuss and resolve with the other analyst. Given the relatively straightforward nature of the responses, codes were summarized using frequencies and illustrative quotations.

For responses to the organizational resources survey, two analysts each reviewed a sample of the completed questionnaires (n=15) and developed a preliminary set of codes. Codes generally were labels of broad thematic topics, with sub codes nested underneath the broader codes. The analysts reviewed and discussed their codes, and based on the discussion some codes were revised to improve for clarity and to reduce overlap. One analyst then coded all cases according to the finalized codebook. 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 (Table 12). At the beginning of the Common Metrics Implementation Program, these hubs ranged in size of their annual budgets from \$4.56 million to \$24.26 million and their initial funding years ranged from 2006 to 2015. 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. Hubs had slightly more variation in the number of stakeholder groups for which procedures were in place to review metric results (standard deviation of 1.10 on a 0-3 scale).

Since Implementation Groups were assigned based on a hub's preference, differences in the composition of the groups on key characteristics were possible. Although Implementation Groups did not differ significantly in size (i.e., funding level at the start of the Common Metrics Implementation Program) or on 10 indicators of experience with metric-based performance improvement, they differed 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.

Survey results indicated wide participation in the Tufts Implementation Program. On average, at least one person from each hub attended most of the training and coaching sessions (Table 13), although Implementation Group 2 attended fewer of both types of sessions. Across Implementation Groups, more hubs focused on the IRB Review Duration or Pilot Funding Publications metrics than the Careers metric during coaching. The number of hubs assigned to each coach differed across groups. By the end of the coaching period, hubs achieved an average score of 23.8 on a 52-point assessment scale, with Implementation Group 1 achieving the highest average score of 28.1.

# Table 12. Hub attributes and related experience, overall and by Implementation Group (N=59 hubs\*)

Characteristic	All Hubs	Implementation Group			
		1	2	3	p-Value
		(n=20)	(n=17)	(n=22***)	
Attribute (n, %)					
Size at start of CMI program, (2015-2016), by tertile <sup></sup>					0.23
< \$4.56 million	20 (34.5)	8 (40.0)	3 (17.6)	9 (42.9)	
\$4.56-8.04 million	19 (32.8)	6 (30.0)	9 (52.9)	4 (19.0)	
≥\$8.05 million	19 (32.8)	6 (30.0)	5 (29.4)	8 (38.1)	
Initial funding cohort, by tertile <sup>…</sup>					0.02
2010-2015	18 (31.0)	5 (25.0)	3 (17.6)	10 (47.6)	
2008-2009	19 (32.8)	9 (45.0)	9 (52.9)	1(4.8)	
2007 or earlier	21 (36.2)	6 (30.0)	5 (29.4)	10 (47.6)	
Experience with metric-based performance	improvement (m	iean, SD)			
Variables comprising factor called "Maturity o	f performance m	anagement system	"		
Extent data were collected for performance metrics (range: 1-4)	3.4 (0.75)	3.2 (0.83)	3.6 (0.70)	3.4 (0.68)	0.24
Degree metrics were used to quantify progress (range: 1-4)	3.3 (0.71)	3.3 (0.75)	3.3 (0.67)	3.4 (0.74)	0.90
Number of CTSA components monitored with metrics (range 0-4)	3.4 (0.97)	3.4 (1.23)	3.6 (0.70)	3.4 (0.92)	0.67
Number of stakeholder groups for which procedures were in place to review results (range: 0-3)	1.4 (1.10)	1.6 (1.15)	1.2 (0.92)	1.4 (1.21)	0.56
Extent of reliance on metrics for operations decisions (range: 1-4)	3.1 (0.80)	3.1 (0.78)	3.1 (0.94)	3.1 (0.73)	0.92
Extent of reliance on metrics for strategy decisions (range: 1-4)	3.1 (0.75)	3.0 (0.75)	2.9 (0.85)	3.2 (0.68)	0.56
Self-rated capability of hub to use metrics to monitor/manage (range: 1-5)	3.2 (1.09)	3.5 (1.07)	3.3 (1.02)	2.9 (1.14)	0.25
Other variables in factor analysis					
Degree data collection was automated (range: 1-3)	1.9 (0.34)	1.8 (0.37)	2.0 (0.35)	1.9 (0.30)	0.39
Extent of data storage in centralized CTSA database (range: 1-5)	3.2 (0.98)	3.0 (1.00)	3.5 (1.01)	3.1 (0.94)	0.33
Current use of performance method (range: 0-1)	0.7 (0.47)	0.8 (0.42)	0.6 (0.50)	0.6 (0.50)	0.43

SD=Standard Deviation; CMI=Common Metrics Implementation

\* One hub did not respond to the follow-up survey and is excluded from analyses.

\*\* CTSA size is defined as total funding from U, T, K, and/or R grants for 2015-2016.

\*\*\* Data for CTSA size and initial funding year are missing for one hub.

Characteristic	All Hubs	Implementation Group				
		1 (n=20)	2 (n=17)	3 (n=22)	p-Value	
Participation in Tufts Implementation	on Program					
Self-selected coaching metric (n, %)					0.69	
Careers in translational science	11 (18.6)	4 (20.0)	4 (23.5)	3 (13.6)		
IRB duration	23 (39.0)	9 (45.0)	7 (41.2)	7 (31.8)		
Pilot publications	25 (42.4)	7 (35.0)	6 (35.3)	12 (54.5)		
Program attendance** (mean, SD)						
Training sessions (n=7)	6.3 (1.1)	6.9 (0.37)	6.1 (0.75)	6.0 (1.5)	0.02	
Coaching sessions (n=6***)	5.6 (0.75)	5.7 (0.67)	5.2 (1.0)	5.9 (0.35)	0.01	
Primary coach (n, %)					0.01	
Coach A	27 (45.8)	15 (75.0)	6 (35.3)	6 (27.3)		
Coach B	32 (54.2)	5 (25.0)	11 (64.7)	16 (72.7)		
Assessment score at end of coaching period, 0-52 possible <sup>s</sup> (mean, SD)	23.8 (5.7)	28.1 (5.1)	20.2 (4.0)	22.6 (4.8)	<0.001	

# Table 13. Hub participation in Tufts Implementation Program, overall and byImplementation Group (N=59 hubs`)

SD=Standard Deviation

\* One hub did not respond to the follow-up survey and is excluded from analyses.

\*\* Attendance at a training or coaching session is defined as at least one person from the hub attended.

\*\*\* Based on 6 coaching sessions. Implementation Groups 1 and 2 were offered 7 sessions; Implementation Group 3 was offered 6 sessions.

§ Score refers to the coach's assessment of hub progress on achieving meaningful application of the RBA framework at the end of the coaching period.

# Hub Progress in Implementing Common Metrics and Performance Improvement Activities

By January 2018, all hubs reported that they had begun the work of implementing the Common Metrics and performance improvement for each of the first three metrics. However, less than one-third of hubs (17 of 59) had completed all activities at least once for each metric (score of 30; Figure 12). 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).

# Figure 12. Completion of Common Metrics and performance management activities per hub: three metrics combined

0-30 points possible



Comparing completion scores by metric, a larger percentage of hubs completed more activities for the Careers and Pilots metrics compared to the IRB metric (Figure 13).





To understand the variation in hub completion scores, we computed average scores overall and by cluster of activities. Overall, hubs achieved an average score of 23.7 out of 30 (Table 14). On average, hubs completed almost all activities related to creating metric results (average score of 5.9 out of 6) and the vast majority of activities related to understanding current performance (average score of 5.5 out of 6). The variation began to appear for activities related to developing performance improvement plans, for which hubs achieved an average score of 6.4 out of 9. These activities were completed less often for the IRB metric compared to the Careers and Pilots metrics (average score of 1.9, 2.3, and 2.3, respectively, out of 3).

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 (average score of 4.1 out of 6). Fully documenting a metric result and the four elements of the improvement plan was lowest, with an average score of 1.8 out of 3.

Mean (SD),	Overall Sum		By Metric						
Range	Possible	Actual	Possible		Actual				
				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		

# Table 14. Completion of Common Metrics and performance management activities (N=59 hubs\*)

SD=Standard Deviation

<sup>\*</sup>One hub did not respond.

"Composition of clusters: 1) Creating metric result entails data collection and computing metric according to Operational Guideline; 2) Understanding current performance 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, and identifying partners, potential actions, and planned actions. Of note, the Implementation Team's interim assessment of hub progress at the end of the small-group coaching period was associated with progress by the end of the longer evaluation period in January 2018 (Table 15). The interim assessment was made for the one Common Metric that hubs chose to focus on during coaching sessions. For every 10 point increase in the interim assessment score (52 points possible), the average number of completed activities by the end of the study period was higher by 17% for all metrics combined (5.1 of 30 points) and between 16% and 19% per metric (1.6 and 1.9 of 10 points, respectively).

# Table 15. Association of interim assessment and completion of Common Metrics and performance improvement activities by January 2018

	Change in Hub Score						
	Overall Sum	By metric					
	(0-30)	Careers (0-10)	IRB (0-10)	Pilots (0-10)			
10-point increase in interim assessment score (out of 52 points possible) $^{\rm \$}$	5.1***	1.6***	1.9***	1.6***			

\*\*\*<0.01

<sup>s</sup>Tufts Implementation Program assessment of progress toward meaningful application of the RBA-based framework on one metric at the end of the 18-week intensive period of training and coaching.

# A Closer Look at Lack of Completion

To further understand activities that hubs did not complete, we looked within clusters to assess rates of completion of individual activities among those eligible to be completed. This approach accounted for logical skip patterns in order to focus on those activities that could have been completed but were not. For example, only hubs that developed a performance improvement plan were 'eligible' to implement a plan.

Overall, completion of eligible activities varied within clusters and, to some extent, across metrics (Table 16 and Appendix I). Similarly, hub reasons for lack of completion also differed somewhat across specific activities (Table 17 and Appendix J).

Although the vast majority of hubs reported collecting data and computing the metric result for all three metrics, these activities were not complete in four cases, three of which related to their IRB data systems. One of the four cases pertained to data collection. One hub had not begun collecting data for the IRB metric due to awaiting execution of the relevant data system. The remaining three cases pertained to computing the metric result according to the Operational Guideline; two were a result of lack of alignment between the hub's data system and definitions in the Operational Guideline.

Regarding understanding current performance, the activity of specifying underlying reasons for the metric result was incomplete for only 6% of cases, most commonly for the IRB metric. Reasons included that improvement in the metric result was not needed, the

#### Table 16. Completion for individual activities, overall and by metric

Cluster a	nd Activity	Completed Activities Among Eligible Activities					
Indentatio	ns indicate reduction in number of eligible	All Metrics		By Metric			
activities aue to survey skip patterns		(n=177*) %	Careers (n=59*) %	IRB (n=59*) %	Pilots (n=59*) %		
Creating N	letric Result						
Began coll	ecting data	99.4	100.0	98.3	100.0		
	If began collecting data, then computed result according to Operational Guideline	98.3	100.0	96.6	98.3		
Understan	ding Current Performance						
Compared	result to another data source	83.6	78.0	91.5	81.4		
Created fo	recast of future performance	63.8	67.8	61.0	62.7		
Specified u	inderlying reasons	93.8	94.9	89.8	96.6		
	If specified underlying reasons, then included hub <u>or</u> other stakeholders	100.0	100.0	100.0	100.0		
	<ul> <li>Included hub leaders/faculty/staff</li> </ul>	98.8	100.0	96.2	100.0		
<ul> <li>Included external stakeholders</li> </ul>		92.2	92.9	94.3	89.5		
Developin	g Performance Improvement Plan						
Developed	performance improvement plan	76.8	79.7	69.5	81.4		
	If developed plan, then included hub <u>or</u> other stakeholders	100.0	100.0	100.0	100.0		
	<ul> <li>Included hub leaders/faculty/staff</li> </ul>	99.3	100.0	97.6	100.0		
	- Included external stakeholders	93.4	93.6	97.6	89.6		
	If developed plan, then specified actions for improvement	95.6	95.7	92.7	97.9		
	If specified actions, then prioritized them	90.0	93.3	89.5	87.2		
	If prioritized actions, considered likely feasibility or effectiveness	100.0**	100.0	100.0	100.0**		
Implemen	ting Performance Improvement Plan						
	If developed plan, then asked partners for help in carrying it out	86.8	87.2	90.2	83.3		
	If developed plan, then began implementing it	89.7	89.4	87.8	91.7		
Document	ing Metric Result and Plan Fully						
	If began collecting data, specified underlying reasons, and developed performance improvement plan, then documented five aspects of the process <sup></sup>	80.5	80.4	82.1	79.2		

<sup>\*</sup> For the three metrics combined, each hub could complete an activity three times (59 hubs x 3 metrics = 177 potential activities). Among potential activities, eligible activities reflect skip patterns in the survey. For example, if "began collecting data" was not completed, then the hub was not asked about computing the metric result. See Appendix I for number of eligible activities per activity for three metrics combined and individual metrics.

"Data for "when prioritizing actions, considered feasibility or effectiveness" are missing for one metric at one hub.

"Documenting fully entails entering the following in the CTSA Consortium Scorecard: metric result, underlying reasons, potential partners, potential improvement actions, planned improvement actions. hub did not have sufficient time or resources, or stakeholders were difficult to engage. In contrast, it was less common for hubs to compare metric results to another data source (16% of cases incomplete) or create a forecast (36% of cases incomplete). Comparing the metric result to another data source was incomplete more often for the Careers and Pilots metrics than for the IRB metric (22%, 18%, and 9% cases incomplete, respectively). The main reason for not comparing the result was lack of comparison data. Creating a forecast was not completed for one-third of cases. Hubs cited lack of confidence in the accuracy of a forecast, views that forecasting as described in the training was not necessary, and opinions that the metric as defined was not relevant or useful for their hubs.

Developing a performance improvement plan was incomplete in 23% of cases. Lack of an improvement plan was more common for the IRB metric compared to the Careers and Pilots metrics (incomplete: 31%, 20%, and 19%, respectively). The most commonly cited reasons were that improvement on the metric result was not needed and insufficient time and resources.

When performance improvement plans were developed, some type of stakeholder – from the hub or external to the hub – was generally included. Engaging stakeholders from outside the hub was slightly less common than engaging those within the hub (incomplete cases: 7% for stakeholders outside the hub and 1% for internal stakeholders) and was less common for the Pilots and Careers metrics compared to the IRB metric (incomplete cases: 10%, 6%, and 2%, respectively).

Performance improvement plans generally included actions that could be taken to improve the metric result; this activity was not completed for only 4% of cases. When actions were not specified, the main reason was lack of time or resources. When actions were specified, they were prioritized in all but 10% of cases, and prioritization considerations included feasibility or effectiveness of potential actions. The main reason for lack of prioritization was an assessment that it was not necessary (e.g., the plan was sufficiently simple to not need prioritization of activities).

When performance improvement plans were developed, hubs reported that they had begun to implement them in all but 10% of cases. The main reasons for not having begun to implement a plan were lack of time or resources and difficulty engaging stakeholders. Among the improvement plans that were developed, hubs asked partners for help in implementation for 87% of them. Whether hubs asked for help to implement a plan appeared to differ somewhat by metric. This activity was not completed for 17% of improvement plans for Pilot Funding, 13% of plans for Careers, and 10% of plans for IRB Review Duration. The most common reasons for not asking for help were that it was not necessary (i.e., all partners were internal to the CTSA) and lack of time on the part of partners.

According to self-report survey information, documentation in Scorecard of five main elements (the metric result, underlying reasons, partners with a role to play, potential actions, and planned strategies) was not complete in part or whole in about 20% of cases. A main reason was that documentation of all five elements could not be achieved if any of the related activities were incomplete. In addition, some hubs indicated completing an activity but not documenting it.

# Table 17. Most common reasons for lack of completion, by activity\* (three metrics combined)

Activity and Most Common Reason(s)	Illustrative Quotation
Began collecting data	
Accurate data not available	"We are awaiting final implementation of our clinical trials management system before reporting on this metric."
Computed result according to Operation	nal Guideline
Metric definition not relevant / useful	"Some of the specific types of research categories identified in the Operational Guidelines are not tracked by the IRB'ssystem, so the Operational Guidelines could not be followed to the letter."
Activity in process	"We plan to compute the metric, but this is the third priority of getting our Common Metrics work initiated"
Compared result	
Accurate data not available	"We are interested in benchmarking this metric; however, we have yet to identify or access good comparison data."
Created forecast	
Lack of confidence in forecast	"Forecasting for this metric is difficult. We cannot predict the number of papers that will be published because it varies by the research being done."
Not necessary / not pursued	"Our focus has been on quality improvement processes and we have not yet used the data for forecasting."
Metric not relevant / not useful	"The NCATS Common Metric is just one time point in the whole protocol approval timeline, and we use a more holistic approach to forecasting from the date of protocol submission (which precedes IRB submission) to full protocol approval."
Specified underlying reasons	
Improvement not needed	(closed-ended response categories - quotations not available)
Lack of time / resources	
Difficulty engaging stakeholders	
Developed improvement plan	
Improvement not needed	"Because of our structure, we cannot improve our measure any better than a median 14 days"
Lack of time / resources	"Hub is currently in a no cost extension period."
Specified actions	
Lack of time / resources	"We did not have sufficient time to operationalize the plan in order to derive desired outcomes."
Prioritized actions	
Not necessary / not pursued	"The simplicity of our performance improvement plan didn't require us to assign priorities to different actions."
Asked partners for help in carrying out	improvement plan
Partners are internal	"Our current improvement plan includes key stakeholders who are part of the hub. The hub has not needed to reach out to others"
Lack of time/resources	"Lack of time to cooperate by key individuals."
Began to implement improvement plan	
Lack of time/resources	"IRB staffing deficit - they are necessary for process."
Difficulty engaging stakeholders	(closed-ended response category – quotations not available)

\* For each activity, reasons cited for least 20% of incomplete cases are displayed in this table; see Appendix J for full list of reasons, frequencies, and range of illustrative quotations.

Combining all activities, the most common reason for not completing an activity was the lack of time and/or resources (Figure 14 and Appendix K). Other reasons given more than 10% of the time were: lack of accurate data, improvement not needed, or the hub determined that the particular activity was not necessary for that metric.



#### Figure 14. Reasons for lack of completion: all metrics combined (N=241 reasons)

Although average completion of activities did not differ significantly by metric (Table 14), the most common reasons for not completing an activity appeared to vary by metric (Table 18). For the Careers metric, the three most common reasons for not completing activities were lack of accurate data (21.2%), no need to improve on the metric result (18.2%), and a decision that an activity was unnecessary (16.7%). For IRB Review Duration and Pilot Publications, the most common reason was lack of time and/or resources (24.7% and 23.3%, respectively). For the IRB metric, this was followed by no need to improve (14.6%) and difficulty engaging stakeholders (12.4%), whereas for Pilot Publications it was followed by lack of accurate data (18.6%) and a decision that an activity was unnecessary (14.0%).

Self-reported Reason	Percentage Among All Reasons Reported*						
	Overall		By Metric				
	N=241^ %	Career n=66* %	IRB n=89* %	Pilots n=86* %			
Insufficient time/resources	21.6	15.2	24.7	23.3			
Accurate data not available/hub newly funded	15.8	21.2	9.0	18.6			
Improvement on metric result not needed	13.3	18.2	14.6	8.1			
Activity not necessary/not pursued	12.4	16.7	7.9	14.0			
Metric not relevant/not useful	8.7	7.6	6.7	11.6			
Difficulty engaging stakeholders	7.5	0.0	12.4	8.1			
Lack of confidence in forecast	6.2	4.5	9.0	4.7			
Activity in process	5.0	12.1	1.1	3.5			
Lack of authority/input from process owner	4.1	0.0	11.2	0.0			
No need for external partners	3.7	3.0	1.1	7.0			
Used other data to forecast future performance	0.8	0.0	1.1	1.2			
Priority of actions was predetermined	0.4	0.0	1.1	0.0			
Don't know	0.4	1.5	0.0	0.0			

# Table 18. Reasons for lack of completion of activities, overall and by metric

 $^{\star}$  More than one reason could be given for each incomplete activity.

# **Understanding Hub Progress and Experiences**

### Effect of Hub Characteristics Measured Quantitatively

The conceptual model (Figure 9) identified key factors that varied across hubs and were expected to influence the implementation of Common Metrics and performance improvement activities. Three organizational factors (hub size/funding level, funding cohort, and prior performance improvement experience) and differences in hub participation in the Tufts Implementation Program could be assessed quantitatively. A fourth organizational factor (diverse organizational aims, structures, and cultures) was explored through qualitative interviews and open-ended responses.

Hub characteristics measured quantitatively explained between 16% and 21% of the variation in completing improvement activities across hubs and metrics (Table 19). Moreover, some of the statistical results were somewhat counterintuitive without further explanation. This section reviews the quantitative results, and the next section integrates quantitative and qualitative results to expand the interpretation (see Deeper Inspection of Local Context, Challenges, and Facilitators).

### **Basic Attributes**

# Hub Size

As noted above, the most common reason hubs cited for not completing performance improvement activities was lack of time and resources. Hubs were funded at very different levels, and this appeared to have some effect, particularly for the Pilots metric. Compared to the smallest hubs, mid- 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. Additionally, the trend toward larger hubs completing more activities was statistically significant for the Pilot Publications metric but not for the other two metrics.

# **Funding Cobort**

Although we expected that hubs funded earlier would have more established processes and stakeholder relationships to conduct the work of Common Metrics, hubs in the middle-funded cohort completed an average of 15% more activities than the earliest funded hubs (a difference of 4.75 out of 30 points). 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.

#### Prior Experience with Metric-Based Performance Improvement and Data Systems

Similarly, although we anticipated that prior experience with metric-based performance improvement would facilitate completion of performance improvement activities, the factor representing level of maturity of a hub's performance management system appeared to have a small negative effect on completion activities. Although this effect disappeared after accounting for other characteristics in multivariable models, similarly unexpected effects related to existing data collection and storage appeared more robust for the IRB and Pilots metrics.

# Table 19. Results of testing for effects of hub characteristics on completion of performance improvement activities (N=59 hubs $^{\epsilon}$ )

	UNIVARIABLE MODELS		MULTIVARIABLE MODELS					
Characteristic	(	Change in I	Hub Score			Change in	Hub Score	9
	Overall		By Metric		Overall		By Metric	
	Sum (0-30)	Careers (0-10)	IRB (0-10)	Pilots (0-10)	Sum (0-30)	Careers (0-10)	IRB (0-10)	Pilots (0-10)
Model N					55	55	55	55
Model Adjusted R <sup>2</sup>					0.17	0.16	0.20	0.21
Basic attributes								
Size <sup>ℓ</sup> 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)								
Experience with metric-based	performan	ice impro	vement					
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 Impleme	entation Pro	ogram						
Attendance <sup>¥</sup>								
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

 ${}^{\varepsilon}$  One hub did not respond.

<sup>*E*</sup> CTSA size is defined as total funding from U, T, K, and/or R grants for 2015-2016.

<sup>2</sup> Attendance 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.

First, for the IRB metric, it was unexpected that the presence of more automated data collection was correlated with fewer completed performance improvement activities. For every step up in the extent of automated data collection (none to partial, partial to full), the mean number of activities completed for the IRB metric decreased by 28% (2.76 out of 10 points). This effect remained significant when accounting for other hub characteristics in the multivariate models.

Second, for the Pilots metric, automated data collection had the expected effect of completion of more activities, but the extent of centralized data storage had the opposite effect. For every step up in the extent of automated data collection, the mean number of activities completed increased slightly, and, when accounting for other hub characteristics in multivariate models, the size of the increase grew to 17% (1.73 out of 10 points). Also when accounting for other characteristics in multivariable models, a small overall negative effect of centralized data storage remained significant for the Pilots metric (a decrease of 0.47 out of 10 points).

# Participation in Tufts Implementation Program

### Attendance

Hub attendance at training and coaching sessions appeared related to completion of activities. 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.

Attending more training sessions was particularly important for the Pilots metric, an effect that persisted when accounting for other characteristics in the multivariable models. Attending more coaching sessions appeared important for the IRB and Pilots metrics (1.10 and 0.72 more activities out of 10, respectively), but more coaching remained significant only for the IRB metric when accounting for other characteristics.

# Self-selected Coaching Metric

For the Careers and IRB Review Duration metrics, receiving coaching while working on that metric was associated with completing more performance improvement activities for it. For the Careers metric, hubs that did not focus on this metric during coaching completed fewer activities for it by the end of the evaluation (1.89 and 1.26 fewer activities out of 10 for hubs focusing on IRB and Pilots, respectively). This effect of fewer activities completed on the Careers metric when not focusing on it during coaching was statistically significant for hubs that focused on the IRB metric during coaching, even when accounting for other characteristics in the multivariable models.

A similar but less strong trend appeared for completing performance improvement activities for the IRB metric. 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. Similarly, hubs that focused on the Pilots metric during coaching completed fewer activities for the IRB metric. However, the positive effect of focusing on the IRB metric during coaching on completing performance activities for that metric was not statistically significant and reduced in size when accounting for other characteristics in the multivariable models.

Overall, these results suggest that receiving coaching on metrics may facilitate the completion of performance improvement activities on those metrics. Of note, because characteristics of the Common Metrics Implementation necessitated a descriptive study design, we were not able to test whether the content of the training and coaching program had a causal effect. Hub feedback on the Tufts Implementation Program provided additional insight (see Role of Tufts Implementation Program).

# Deeper Inspection of Local Context, Challenges, and Facilitators

Qualitative interviews and open-ended questions that elicited information about hubs' local contexts, challenges, and facilitators clarified results for hub characteristics measured quantitatively and helped extend our understanding. This section reflects the main themes hubs identified as impacting their progress on implementing Common Metrics. (See Appendices L, M, and N for combined lists of challenges and facilitators.)

# **Section contents:**

# Themes that clarify hub characteristics measured quantitatively

- Resources: funding, data systems, and personnel and expertise
- Hub engagement and its association with funding cohort
- Prior experience with metric-based performance improvement

# Themes that further extend understanding

- Availability of accurate data
- Starting with "low-hanging fruit"
- Metric clarity and usefulness at a local level
- Engaging stakeholders
- Hub authority and control

#### Resources

Quantitative analyses showed that a smaller funding award did not fully account for hubs' reports that lack of resources was the most common reason for not completing performance improvement activities. Funding (including investment from hubs' home institutions and periods of interrupted funding), alignment of existing data systems with the needs of the Common Metrics, and 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.

# Funding

Availability of resources to implement the Common Metrics varied as a result of different levels of institutional investment (Table 20). Implementing the Common Metrics typically required new resources, particularly for information technology to track processes and collect data. If an institution was not able or willing to make these investments, hubs had to manage or vie for scarce resources, hindering or delaying progress on collecting and reporting measures. Hubs that reported being able to leverage existing institutional resources reported that they "lean on" (*Principal Investigator*) a variety of institutional functions to support hub activities, such as regulatory and evaluation functions.

For a small number of hubs, interrupted funding was a factor. A period of no-cost extension, for example, substantially limited what a hub could accomplish. The impact was not just on the hub's present ability to meaningfully engage in implementing process improvement plans, but it also could impact the longer-term metric results for the hub as programs were limited or even paused.

### Table 20. Challenges and facilitators for implementation: funding

#### Main Themes with Illustrative Quotations\*

#### Challenges

Lack of institutional investment<sup>†</sup>

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're 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. – Principal Investigator\*\*

#### Interrupted funding

...[G] iven our no-cost extension status, ...we don't 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. - Implementer

#### Facilitators

Availability of institutional resources<sup>†</sup>

Our evaluation group actually also has institutional funding.... All centers and other programs now—or many, I should say—require an evaluation piece. So from a school-level we created ... a core resource. – Principal Investigator

... 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. – Principal Investigator

<sup>&</sup>lt;sup>+</sup> Unless stated otherwise, themes manifested in more than one way; a quotation represents one manifestation.

<sup>&</sup>quot;Participant is affiliated with a medical center that functions as a CTSA without current CTSA funding.

<sup>&</sup>lt;sup>†</sup>Indicates that the challenge has a corresponding facilitator.

#### Table 21. Challenges and facilitators for implementation: data system resources

#### Main Themes with Illustrative Quotations\*

#### Challenges

Lack of data system or an existing system that was not aligned with the Common Metrics definitions created more effort for effective tracking<sup> $\dagger$ </sup>

...we never implemented a system to collect [publications] so we have to go back one by one and contact everyone who received a pilot and find out if they published. – Administrator

...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. – Principal Investigator

#### Facilitators

Alignment of Common Metrics with and ability to use existing data collection tools<sup>†</sup>

[Existing data collection tools helped] to start to track that data. -Administrator

With our K scholars we've always had constant communication with them, but that was already established. We have a social media network that's set up specifically for them; that's been set a couple years now. So, I think that helped us in being able to stay in contact with them easier. -Administrator

\* Unless stated otherwise, themes manifested in more than one way; a quotation represents one manifestation.
 <sup>†</sup> Indicates that the challenge has a corresponding facilitator.

#### Data Systems and Resources

When data collection systems did not exist or were not aligned with Common Metrics data requirements, more effort was required to implement Common Metrics and performance improvement. Hubs lacking existing data collection systems faced a higher level of manual work to collect metric data. For hubs with existing data systems that were not aligned with metric needs, making changes to the data collection systems required time and resources. On the other hand, hubs with data systems that aligned with metric needs identified that as an important facilitator for completing the work of the Common Metrics (Table 21).

#### Personnel and Expertise

Hubs reported wide variation in CTSA personnel and effort allocated for implementing the Common Metrics and performance improvement framework. This variation was apparent across hubs and metrics.

Although each hub reported only up to five CTSA-funded positions with important roles in their most recent update, collectively, hubs identified positions across 12 distinct types of personnel, or position groups (Table 22), and 33 individual positions (Appendix O). Of the 12 position groups, only one quarter (n=3) were reported by more than half of hubs: evaluators (65%), content experts (53%), and CTSA leadership (51%).

Position Group (coded)	Hubs		Approximated Hours	
	n	%	Mean	SD
Evaluator	32	65.3	95	93
Content expert	26	53.1	45	59
CTSA leadership	25	51.0	40	59
Manager/coordinator	21	42.9	166	273
Data collector/analyst/support	20	40.8	53	67
Common Metrics champion	4	8.2	89	141
Informatics	4	8.2	152	113
Clinical research personnel	2	4.1	48	46
Performance improvement expert	2	4.1	38	31
Institutional leaders	1	2	16	
Biostatistics	1	2	6	
Other interested individual	1	2	33	

Table 22. CTSA-funded personnel with important roles in most recent update (up to five) and approximate hours spent (N=49 hubs\*)

\* Eleven hubs did not respond about this topic.

Although only 65% of hubs mentioned an evaluator as one of the five positions reported, qualitative interviews suggested that most hubs located the Common Metrics implementation tasks within the hub's evaluation function. Specific personnel included an evaluation lead and/or evaluation-related manager(s) or coordinator(s).

Approximations of the hours spent on the last Common Metrics update (August 2017) reveal substantial variation of effort within position groups across hubs, demonstrated by large standard deviations around the mean number of hours per position group. Of note, about 40% of hubs indicated that their reports underestimated the effort expended for the Common Metrics, and 14% indicated likely overestimations (Table 23). Additionally, most hubs did not have detailed effort logs to refer to when providing information on effort. Still, even imprecise estimates indicate wide variation.

	Hu	bs
	<b>n</b> **	%
Reasons reports of effort are underestimations	19	38.8
Survey does not provide full picture of effort invested to date because it limited reporting to the most recent annual update	9	18.4
Effort to familiarize/educate others about process and allay fears is not included	5	10.2
Hub had insufficient staff to conduct all activities	3	6.1
Effort by IRB leaders and staff is not included	2	4.1
Effort to support continuity, including training new personnel, is not included	2	4.1
Persons involved spanned institutional programs/departments, which are not included	1	2.0
Positions supporting leaders and others responsible for conducting process are not included	1	2.0
Effort related to strategic and action planning is not included	1	2.0
Reason reports of effort are overestimations	7	14.3
Difficulty parsing overall effort to report only on the most recent update	7	14.3

#### Table 23. Hub reasons that approximated hours are under- or over-estimated (N=49 hubs\*)

\* Eleven hubs did not respond about this topic.

\*\*Within each category, hubs could mention more than one sub-category.

As evidenced by the reported effort, personnel involved in implementing Common Metrics frequently split their time among multiple projects and roles. This was true for both larger and smaller hubs. For example, an Administrator may split time between work for the hub and another large grant. Even evaluation leads who were fully funded within the hub may have several roles within the hub, with the Common Metrics as "just part of their portfolio" (*Principal Investigator*). In the words of an Administrator, "Everyone is partially dedicated."

*Variation by Metric* – Hubs also indicated that personnel and effort varied by metric (Figure 15). About 60% of hubs reported that personnel varied "some" or "a lot" by metric, and 74% said that personnel hours varied to the same degree.



Figure 15. Variation in personnel and hours: "To what degree did personnel/hours vary by metric?" (N=49 hubs)

Additionally, 82% of hubs reported that one of the three metrics "consumed a great deal more time and resources" than the others, with the highest number of hubs identifying the Pilots metric followed by the IRB and Careers metrics (Figure 16).





*Personnel-related Challenges and Facilitators* – Hubs identified several challenges and facilitators related to personnel (Table 24). For many hubs, other priorities against which the Common Metrics must compete for time, attention, and effort presented a "big challenge" (*Principal Investigator*).

#### Table 24. Challenges and facilitators for implementation: personnel

#### Main Themes with Illustrative Quotations\*

#### Challenges

#### **Competing priorities**

From the standpoint where you have to divert effort to comply with new mandates, that does impede progress on the Common Metrics because some of the same people...are now tasked to do these other things. - Principal Investigator

Any time the IRB is contemplating changes, their attention and personnel are deflected from their day-to-day work... [T]here is so much changing with the Common Rule and everything, I think their personnel were distracted, especially their leadership. – Principal Investigator<sup>\*\*</sup>

#### Lack of adequate staffing<sup>†</sup>

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 doesn't have the capability to do this herself independently. This is where it all kind of breaks down. Nobody really thought about what impact it was going to have on the time allocation for the leadership that was responsible for evaluation when this concept of Turning the Curve was unleashed. - Principal Investigator

...[A] lot of lip service is given to the importance of evaluation, but resources aren't [provided for] folks who actually support it. ...So having two tenure-track faculty members and a PhD staff member spending time on [the Common Metrics] when there are other people who could collect it, like the [masters level] individual if she had full time working at this, where we could oversee—that would have helped. – Implementer\*\*

#### Lack of evaluation and other specific expertise<sup>†</sup>

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. – Administrator

[With staff turnover,] ...that historical knowledge is lost every time. And so for us, getting everybody up to speed every time a new person comes onboard has been... a big issue. – Administrator

#### **Facilitators**

#### Adequate evaluation and other specific expertise<sup>†</sup>

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. - Principal Investigator

We work closely with our IRB, and we have a member of our team who worked in the IRB for about eight years. She had really deep knowledge of that system. She was very important, and also she has technical knowledge. - Implementer

#### Leveraging extended teams<sup>†</sup>

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'm 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 I created a parallel group of advisers, people who were interested in the Common Metrics. - Implementer

\* 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.
- <sup>†</sup> Indicates that the challenge has a corresponding facilitator.

Implementing the Common Metrics required effort, and "anything that requires a lot of effort means it's going to be pulled away from other things" (*Principal Investigator*). These competing priorities were often described as occurring within the hub but also played a role when attempting to engage stakeholders external to the hub.

Availability of both adequate personnel and related expertise was important for successful implementation. A few hubs reported lacking adequate resources to assign evaluation personnel or support staff to the Common Metrics effort. As a result, senior personnel were conducting activities that could have been delegated to more junior staff had they been available. Some hubs also reported lacking personnel with the particular expertise required for implementing the Common Metrics and performance improvement framework. Insufficient expertise in evaluation was especially problematic.

In contrast, other hubs described how the availability of evaluation and other expertise such as for IRB and regulatory issues, informatics, as well as institutional knowledge and general administrative support greatly facilitated implementation of Common Metrics and performance improvement.

Because Common Metrics teams typically were rather small, some hubs leveraged other individuals and groups within their hub and home institution. They reported that these extended teams facilitated completion of data collection and performance improvement activities. In a number of cases, the work with these other stakeholders was organized into extended teams organized around the different metrics to facilitate regularized collaboration and sustained commitment.

#### Hub engagement and its association with funding cohort

During the course of qualitative interviews, participants indicated different levels of engagement with Common Metrics and performance improvement activities:

- active engagement: folding metrics and the RBA-based performance improvement framework into a hub's standard work processes,
- **compliance:** approaching the Common Metrics as only fulfilling an external requirement,
- mix of active engagement and compliance-based engagement.

For some hubs, the three participants in qualitative interviews held the same perspective on the hub's level of engagement; in other cases, participants' viewpoints differed (Figure 17).

# Figure 17. Hubs' self-described engagement (N=30 hubs)

(Coded from qualitative data)



### Active Engagement

At one-third of hubs (10 of 30), all participants characterized the hub integrating the Common Metrics and RBA-based performance improvement into their standard work processes.

[I]t really has made sense and become kind of part of our culture and fabric at our hub. – Principal Investigator

...And then also diving deeper into understanding...what some of the root cause issues are... So we've refined the IRB metrics not just to say, "Are things getting better or worse," but [to] start to break out the components of who's responsible for the efficiencies or inefficiencies. – Principal Investigator

Hubs already performing at high levels on the topic areas of the Common Metrics were able to expand performance improvement planning and identify future improvement opportunities.

I would say that it's helpful to the extent that we're able to expand, especially the targets of what it is we're trying to improve. So,...we've essentially said, "All right-there's a Common Metric, but then there are all these other ancillary metrics that matter to us with regard to program performance." And so that's what we've done...with our Turn the Curve planning, is to kind of talk through how do we do better the things that matter to us, to the extent that the original Turn the Curve plan was all tied back to a single Common Metric or a family of metrics that we had some concerns with. – Implementer

Other hubs that initially were not well-positioned for the Common Metrics saw opportunities to improve their data systems and then developed plans to improve them. Well, we did a couple of things. One is we looked at what seemed to be an evolving set of criteria around the Common Metrics, and we had mapped them against the information that we currently collect on a regular basis. And then, secondarily mapped that to what we are currently doing with the data that we are getting... We also looked at data that we collect to round out our evaluation, because we collect a lot more data around the Common Metrics than is involved in the Common Metrics. – Administrator

Participants at 53% of hubs (16 of 30) indicated a mixture of engagement levels. In some cases, a leader (either the Principal Investigator or an Administrator) reported active engagement whereas the implementer reported a compliance approach. In other cases, individual participants reported both active engagement and a compliance approach (e.g., active engagement with one or two of the metrics but not all of the metrics).

#### Compliance

Some participants described their approach to implementing the Common Metrics as complying with an external requirement. At four of 30 hubs, all three participants in qualitative interviews agreed their hub was just "checking a box."

...the Common Metrics [Implementation] is a box we check, and we're doing all of our work anyhow. So, the box that we checked on that was tied to the timetable for the deadline.... –Implementer

For some, taking a compliance approach was not accompanied by negative opinions of the Common Metrics implementation, but others taking a solely compliance-based approach expressed assessments that implementing the Common Metrics produced little value.

I didn't have any negative thoughts about [Common Metrics], but other than that, it was something that we would just have to record and see whether we're moving, again, in the right direction. – Principal Investigator

I guess I just felt it was kind of an administrative task, a little bit of an administrative burden. I don't feel like it's making a huge impact. That's okay. It doesn't seem like anybody is really paying a lot of attention to them. – Administrator

#### **Challenges for Maintaining Engagement**

Overall, participants identified three main challenges to maintaining a higher level of engagement with Common Metrics (Table 25).

#### Table 25. Challenges for maintaining a higher level of engagement

Theme and Illustrative Quotation(s)*
Challenges
Annual reporting cycle induces bursts of effort
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's not good as a management tool. [In order] to use them in terms of Turning the Curve plans, you really should be generating Turn the Curve plans on a much more frequent basis. – Principal Investigator
Interrupted funding
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 outlinedWe are only working with the IRB as a collaborative partner and trying to push our agendaget them to implement some of our suggested action plans. – Implementer
Reduced motivation due to lack of alignment with existing processes or unclear definitions
[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. And we did it because it's cooperating with the Common Metrics of the national consortium. But if I was to sit here and take the amount of time that was spent on that versus the amount of time that we could spend on something else, I'm not sure I would necessarily say it was time well spent. – Administrator

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

#### Association of Engagement with Completion of Activities and Funding Cohort

Not surprisingly, hubs in which participants all reported active engagement or were mixed in their approach completed more performance improvement activities than hubs in which all participants reported a compliance-based approach (Table 26 and Appendix P). Although these differences did not reach statistical significance in this sample, the small number of hubs in which all participants reported a compliance approach may have hindered detection of a statistical difference. Still, it is notable that, for the Careers and Pilots metrics, the hubs in which all participants reported active engagement had similar scores to those in which the leader reported active engagement and the implementer reported a compliance approach. In contrast, for the IRB metric, hubs in which a leader reported active engagement while the implementer reported a compliance approach tended to complete the most activities, although this difference did not reach statistical significance. 

 Table 26. Results of testing for effects of hub engagement on completion of performance improvement activities (N=30 hubs)

Engagement Category	N	Hub Score (Mean, SE)			
Coded from qualitative interviews		Overall Sum	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					

Ref=reference group SE=standard error \*p≤0.10

Additionally, hub engagement was associated with funding cohort. 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 (Figure 18).



Figure 18. Association of hub engagement and funding cohort\* (N=29 hubs)

Hub engagement reflected hubs' levels of willingness or ability to adjust processes to accommodate the requirements of the Common Metrics, which differed across funding cohorts.

• Latest funded cohort (2010-2015): Hubs funded in the latest cohort were less likely to have firmly established processes, which could make the introduction of a performance improvement system useful. Yet, these hubs sometimes had difficulties with resources or contextual issues (e.g., developing relationships with stakeholders). For example, a hub that was actively engaged struggled with resource issues for collecting data on one of the Common Metrics:

The IRB thing I think is easy and the education thing we did fine. But the Pilot metric, we wish we would have had more funding in there for someone to run that information down. And so now we are trying to create a position...to be a resource to our investigators in the library but also to collect this data. But we're trying to find the money and rebudgeting to pay for that. – Principal Investigator

• Earliest funded cohort (2006-2007): Hubs funded in the earliest cohort more likely had established processes. If these processes were aligned with the Common Metrics, then the work could be completed based on existing workflows; if their processes were not aligned, then adaptation of existing processes could present difficulties. For example, participants at two hubs with a compliance-oriented approach spoke of implementing the Common Metrics according to existing processes:

It's part of a component of a larger number of metrics that we collect to monitor our dashboards. This is data we already collected, and we wouldn't collect it unless we thought it was useful to running our hub. So, in that regard, I mean whether it's Common Metrics or not Common Metrics, we use the data. – Principal Investigator

[T]he good news is that we had just recently finished what we call the alumni survey. We only do this every couple of years. We go back to every single person in the KL2 program, from the beginning, and ask them what are they currently doing? ... If we hadn't had all of that information recently collected, it may not have been as easy to find out... – Implementer

• Middle cohort (2008-2009): 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). Additionally, their existing processes and systems appeared not quite as firmly established as those funded earlier, making it easier to adapt to Common Metrics. For example, actively engaged hubs in the middle funding cohort identified existing partnerships and the willingness to make changes or accelerate current processes.

But overall, the process has really been enlightening for us to be able to have a transparency across the institution, and we've been able to utilize the partnership with the IRB but also the information coming out to make some pretty significant changes. – Principal Investigator

It has encouraged us, I would say, also to do more of what we do, which is we dashboard a lot of things... So, it certainly accelerated that in some of these areas, in addition to the deeper engagement ... – Principal Investigator

#### Prior Experience with Metric-Based Performance Improvement: Importance of Alignment

Similar to the effect of funding cohort, the potential benefit of prior experience was related to whether hubs' existing approaches were aligned with the framework used for the Common Metrics.

As described in the Resources section, the mere presence of a data system did not necessarily facilitate completion of performance improvement activities related to Common Metrics. When existing data collection systems were not aligned with metric requirements, changes were needed, which required time and resources.

The importance of alignment extended to existing performance improvement frameworks. Some hubs had established approaches to performance improvement that were locally seen as being beneficial. When aligned, existing frameworks were more likely to facilitate implementation of Common Metrics and related performance improvement activities. Conversely, when hubs perceived lack of alignment between their existing local approach and the framework used for the Common Metrics, there appeared to be resistance to adopting a new approach and/or modifying the pre-existing approach:

It hasn't had a lot of impact because, again, ...in a more rigorous fashion and I think in a way that can demonstrate whether we have improvement or if we're worsening, we've been collecting all that data that we can collect and examining it on an ongoing basis. – Principal Investigator

...A Turn the Curve plan and the various factors to it, that part is useful for individuals who don't already have some sort of established reporting plan or plan for an evaluation project because it helps them think about each part, like the stakeholders and the drivers behind this and all these things. ...No one wants to use that [Scorecard] software here. – Implementer\*

#### Availability of Accurate Data

Obtaining accurate data that followed the Common Metric Operational Guideline was challenging for some hubs (Table 27). As noted, lack of alignment with existing data systems required more effort, and also could lead to a lack of accurate data. Additionally, tracking investigators, which was required for the Careers and Pilots metrics, was more difficult when those investigators had left the home institution. Strong data collection systems and strategies increased success, but the challenge remained. Additionally, lack of line authority over data needed for computing metric results created difficulties for obtaining needed data. This issue was most common with the IRB Review Duration metric.

<sup>\*</sup> Participant is affiliated with a medical center that functions as a CTSA without current CTSA funding.

In the near-term, an existing data system that was aligned with Common Metric requirements facilitated obtaining accurate data. For hubs without an existing system, the ability to use existing tools, such as REDCap, allowed hubs "to start to track that data" (*Administrator*) and capture information in the form of "progress reports" (*Administrator*).

In the long-term, the capacity to develop a data system that would be aligned with Common Metrics requirements could create efficiencies for data collection and performance improvement. The immediate challenges of not having a data system offered the long-term advantage of allowing that a new system's design would reflect the needs of the Common Metrics. Obtaining a commitment to create an aligned system was facilitated by the external requirement to implement Common Metrics and the belief that, once the system was created, maintenance would be easier than cobbling together pieces of related data from a pre-existing system to meet Common Metrics requirements.

#### Table 27. Challenges and facilitators for implementation: availability of accurate data

Main Themes with Illustrative Quotations*
Challenges
Lack of alignment of Common Metrics with existing data systems $^{\dagger}$
The [first] IRB is separate from the [second] IRB. And while we got most of the data from the [second] IRB,the availability of data is different in the two IRBs and so we had to do a little bit more of estimating based on the [first] IRB's data. So, you know, had those databases been linked together under one system, that would have made it easier for us to do this comprehensive approach. – Principal Investigator
Difficulties with effective tracking due to the nature of the topic
Careers, you know, once your KL2 Scholars are out in the world, especially if they've left your institution, it can be challenging to track them down – Administrator
Lack of line authority for needed data
One issue with the CTSAs, particularly in a decentralized organization like ours, is we're responsible for outcomes but don't have authority over them. It's an exercise I'm trying to lead from the middle. – Principal Investigator
Facilitators
Alignment of Common Metrics with and ability to use existing data collection tools $^{\scriptscriptstyle \dagger}$
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 – Principal Investigator
Capacity to build an aligned data system for long-term use will create future efficiencies
So, one of the things that was helpful withworking on the Common Metricswas figuring out what kind of an electronic system we could put in place to track how much time it takes for a protocol from initiation of an idea to development of the protocol to submission to the IRB to approval, and how do you track that electronically. – Administrator
* Unless stated otherwise, themes manifest in more than one way; a quotation represents one manifestation.

<sup>&</sup>lt;sup>†</sup> Indicates that the facilitator has a corresponding challenge.

#### Starting with "Low-hanging Fruit"

Participants from all three roles–Principal Investigators, Administrators, and Implementers– mentioned that the first three Common Metrics were a relatively easy place to begin. Introducing a common framework for performance improvement using "lowhanging fruit...that could be measured" (*Implementer*) created a challenge for some hubs but facilitated the roll-out of Common Metrics for others (Table 28).

On one hand, some hubs found it difficult to pursue improvement when they were already meeting their targets for a metric. Participants reported confusion about the need to develop improvement plans in areas of high performance and questioned whether this was the most efficient use of limited resources. On the other hand, other hubs found the starting point to be useful because they were learning and developing trust in the Common Metrics process (data collection and RBA-based improvement plans). These hubs generally reported confidence and eagerness for addressing more difficult outcomes and adding other metrics.

#### Table 28. Challenge and facilitator for implementation: starting with "low-hanging fruit"

#### Main Themes with Illustrative Quotations\*

#### Challenge

Spending limited resources on areas that did not need improvement was not helpful<sup>†</sup>

If it was an issue, it would be addressed. But, doing a Turn the Curve [plan] to say, "Hey, I recommend you try to get a little bit tighter or get a day better," I don't think that would be a good use of time because we have other groups that maybe aren't in the Common Metrics that need more help in other areas. – Administrator

#### Facilitator

Addressing "low-hanging fruit" allows for smooth start-up in preparation for more challenging metrics  $^{\scriptscriptstyle \dagger}$ 

So I thought they [the first three Common Metrics] were very good choices because ... [they] were relatively low-hanging fruit, meaning they were the ones that generally every CTSA was collecting some sort of data around ... [T]hey were good choices and... it helped prepare our hubs, certainly, for the more difficult ones that I'm sure are to come; the ones that are more complicated to gather. – Administrator

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

<sup>&</sup>lt;sup>†</sup> Indicates that the facilitator has a corresponding challenge.
## Metric Clarity and Usefulness at a Local Level

Clarity of definitions and usefulness for local improvement were important for accurate data collection and computation, for the completion of performance improvement activities, and for maintaining a higher level of hub engagement. Usefulness was facilitated or impeded depending on alignment with local goals and institutional priorities (Table 29).

## Table 29. Challenges and facilitators for implementation: metric clarity and usefulness

Main Themes with Illustrative Quotations*
Challenges
Changing metric definitions and clarifications created duplicative work $^{\dagger}$
We redid the Pilot stuff five, six, seven times, to get it to exactly what Tufts wanted it to look like, and it just was incredibly frustrating. It took me out of my normal job and this is added work that was not anticipated or budgeted for. – Implementer
Lack of usefulness for local improvement
So, we're really looking at the questions that we want to ask rather than barely reporting on the Common Metrics. And we don't have anything particularly against the Common Metrics. It's just we have an opportunity here with other people that are tracking folks to bring it all together to allow us to ask what we consider to be better questions [than the Common Metrics]. – Administrator
Time lag for assessing current performance
[T]he percentage of graduates who are doing clinical translational science, that's a curve that is not going to turn quickly at all. I mean, it's probably, it's years from action to seeing the result. – Principal Investigator
Lack of alignment with institutional priorities $^{\dagger}$
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're 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. –Principal Investigator
Facilitators
Clear metric definition allows focus on improvement work $^{\scriptscriptstyle \dagger}$
We got [the performance improvement plan] togetheronly because the metric was easy to understand. I think there wasn't really conflict in definitions. – Administrator
Alignment with institutional priorities $^{\dagger}$
The institution is very interested in this. So, I think that this is something the institution is highly invested in doing well on. – Principal Investigator
So, I think that since these are important for our institution, it was just easy for us to provide the data and to show how we're improving. Our VP of Research and our research officers believed that the IRB is important to them, our Pilot program does the tracking, that's important to them and Career Development and the K wasa priority to the renewal. – Principal Investigator**
* 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.

 $<sup>^{\</sup>dagger}$  Indicates that the facilitator has a corresponding challenge.

# Metric Clarity

Participants differed in their perspectives on the clarity of metric definitions. Some reported that the metric definitions were clear, well-defined, and measured success, and this facilitated their use. Others reported that metric definitions were not always clear or that, while a definition may be clear in and of itself, it becomes less so given that local processes vary across hubs. These participants indicated that successive efforts to follow the Operational Guidelines within their local context created a challenge for using the metric efficiently.

# Usefulness for Local Improvement

Even when the definition was clear and the metric was accurately computed, the metric may not reflect success in the way that these hubs would define it or be perceived as helpful for identifying opportunities to improve.

During this ten-year period ...we've made great success in studies that have changed the practice of medicine, we've had great success in bringing in a diverse population of both research patients and junior faculty who are being trained to be the next generation of leaders. We are very concerned about metrics that begin to paint the story that's anything other than that. – Principal Investigator

For the Pilot metric and the [Careers] metric, by making it a cumulative [metric] year after year...the denominator just keeps getting bigger and bigger and bigger and so, as the years go on, you have very little opportunity to actually move the curve. And the way that we've been looking at the data is by cohort and then over time by each cohort... – Administrator

Additionally, the Careers and Pilots metrics necessitate a waiting period for assessing performance. For scholars and trainees, it takes time to complete their programs and get situated in the next career position. For Pilot awardees, achieving publication or subsequent funding also requires a waiting period, albeit shorter than for the Careers metric. These lag times make it difficult to identify current opportunities to improve.

*IRB Metric* For the IRB metric, contextual characteristics at the local level made the metric less useful for individual hubs. These characteristics included the number of local IRBs at the hub, the "case mix" of local protocols (e.g., innovative versus follow-up projects, relative complexity of protocols), and representativeness of the metric for the overall review and study start-up workflow.

...the way that Common Metrics is setup, it really only measures 5% of our [IRB's] total annual submissions. – Administrator

If the idea of the metric really is [that] we want to know what all of the IRBs at your institutions are doing in terms of their turnaround times, then I think you are putting together so many different types of IRBs and types of studies, that it loses real meaning. – Principal Investigator *Careers Metric* Some hubs perceived the Careers metric to define success too narrowly by focusing on traditional career paths to becoming an independently funded investigator. These hubs considered other successful outcomes to include scholars and trainees who participate in team science or work in health care systems or industry. When these other forms of success were perceived to not be captured by the Careers metric, hubs viewed this metric as possibly underestimating the impact of the hubs.

But the other thing that bothered me about it, and still does, if you look at the definition of what the KL2 scholar program is supposed to do, and you put that next to team science not everybody is supposed to be an R01 leader. Some people are destined, and should be destined, to participate in team science. ...And maybe our best example of such a circumstance is a pathologist who developed a series of techniques which are vital to the performance of many studies. So, he happens to be a co-investigator on innumerable grants because he brings a talent that is very hard to come by and is necessary for the success of these grants. ...And I don't think, therefore, that what we have as a so-called measure of success is truly measuring success, because we haven't defined, in reality, what success truly is. – Principal Investigator

[T]he spectrum of careers that people go into, which would still be considered scholarly, is fairly broad....I think making sure we understand what "success" means for each of these folks is a bit of a challenge. – Principal Investigator

*Pilots Metric* Participants at some hubs indicated that the Pilots metric was limited and failed to drive more ambitious long-term goals of local pilot funding programs. Suggestions included accounting for the type of study in relation to the likelihood of near-term publication or subsequent funding, representing the total number of publications for each award, and recognizing the value of a variety of outcomes.

But if you look at the immediate success, as the Common Metrics calls you to do, it's going to be very different if you're looking at T3 to T4 pilot awards versus T0 to T2. – Principal Investigator

...First of all, they wanted, any publications, not how many. And to us, whether or not there were publications was important, but also how many publications came out of various pilots is also important. So, we felt like the quantity was being taken off the table, and that didn't make sense to us. – Administrator

I think it's too limited. It may be that some pilot studies are going to provide key preliminary data that would facilitate winning a great big grant, a high-impact clinical trial. It could be a pilot study could lead to a patent that would be important for solving a translational research problem—which is one of the goals of pilot studies—by developing a new method or a new device or something like that. There could be other goals for pilot studies other than "getting a publication." – Principal Investigator

# Alignment and Integration with Institutional Priorities

The extent to which the metrics were aligned with priorities of the home institutions and research communities affected their local usefulness. Such alignment was important for institutional investment (see Resources section above), and higher or lower institutional priority served as a facilitator or challenge for implementation. Some hubs reported that, although their institutions generally were supportive of the hub's work, there was less institutional priority or visibility around the Common Metrics. This relative lack of institutional priority created a challenge for implementing the Common Metrics.

# **Engaging Stakeholders**

Engaging stakeholders is a fundamental aspect of implementing Common Metrics using RBA-based performance improvement. As noted above, some of the broad challenges hubs faced at their institutions—such as Common Metrics not being an institutional priority or competing with other priorities—could have a consequence of impeding stakeholder engagement in the work of the Common Metrics. This section focuses more specifically on challenges, facilitators, and strategies for involving stakeholders in the work of the Common Metrics.

Challenges for engaging stakeholders included lack of an existing line of consistent communication with other units, difficulty securing initial buy-in, or sustaining cooperation over time (Table 30). Difficulty with initial buy-in could result 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), an institutional culture of cooperation, 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 could facilitate engaging stakeholders.

## Table 30. Challenges and facilitators for implementation: engaging stakeholders

#### Main Themes with Illustrative Quotations\*

#### Challenges

Lack of a direct line of communication within institution

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. – Principal Investigator\*\*

Securing needed buy-in or cooperation from key stakeholders

I think there's probably been more resistance or pushback or concern about quantifying that [IRB metric] just because people at the IRB—and rightfully so—believe that they're not the only ones responsible for the turnaround time. – Principal Investigator

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's a constant struggle... – Principal Investigator

#### **Facilitators**

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. – Principal Investigator

Integration of Common Metrics with institutional priorities<sup>†</sup>

This has been embraced...as a barometer at the institution. ...I think having metrics that have the measure at the institution is a good thing. ...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. That's been our goal. – Principal Investigator

Hub location and 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. – Principal Investigator

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. – Principal Investigator

<sup>1</sup>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.

<sup>†</sup>Indicates that the facilitator has a corresponding challenge.

Hubs also identified proactive strategies for enhancing their abilities to successfully engage stakeholders in the Common Metrics (Table 31). A key strategy for engaging stakeholders was first to understand the various sets of stakeholders for the Common Metrics and different methods for engaging them. Stakeholders and methods were observed to vary by metric, as well as over time. Part of obtaining the buy-in and cooperation of other units and offices relied on persuading and showing how those partners also benefit from helping implement the Common Metrics and Turn the Curve Plans. Other hubs described how engagement is a process of creating avenues and forums for discussion, dialogue, and

feedback with stakeholders. Examples include taking a "deeper dive" with stakeholders around the metrics and the review of the work process, as well as making sure to listen to stakeholders at the "ground level," not only leaders. Hubs also noted that engagement of stakeholders may occasionally require a degree of persistence that occasionally may be construed as "nagging" or "bugging." Participants variously described positioning their CTSA as a "hub" in the literal sense of serving as a "bridge" or "liaison" to engage stakeholders across their institution, even at times incorporating key stakeholders from other parts of the institution into roles within the hub to ensure engagement.

### Table 31. Strategies for engaging stakeholders

#### Main Themes with Illustrative Quotations\*

#### Identifying the diversity of stakeholders

For each metric, you have to think through that process, which would be who is going to be in charge of the data? Who are the stakeholders that should be most interested in this and who is sort of that organization or program official that kind of cares about the metric and will push the team to turn the curve. So it does take some thought about what is the team for each metric and being intentional about getting people together. - Principal Investigator

#### Demonstrating benefit

So, that helps us in terms of engaging our partners, because they see, "Whoa, we benefit. Our trainees are getting into your mentor career development program and benefitting from that..." - Principal Investigator

#### Developing dialogue and feedback with stakeholders

As far as engaging stakeholders, that's one of the bonuses of having metrics and it has definitely had us take a deeper dive with our leadership of the IRB... it definitely was through the engagement of the stakeholders that some of that creativity was allowed to come out. -Principal Investigator

A general point would be to always ensure that there's dialogue and that you're listening to the people who are at the ground level, because they have important perspectives. – Principal Investigator

#### Persistence

Because, again, we don't control the IRB, we end up sort of nagging the IRB or other groups where we're trying to influence what they do. – Principal Investigator

#### Building bridges outwards and inwards

Well, most of the stakeholders, like I said, the head of our IRB is involved in our CTSA. ...The [senior leader] at the School of Medicine is one of the co-PIs of our CTSA. The [senior leader] of one of our hospital affiliates runs our [Program name]. I mean most of the people we need to grease the wheels are at the table with us as part of the organization, so we don't have too many issues. – Principal Investigator

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

## **Hub Authority and Control**

Participants described diverse ways in which hubs were situated relative to their home institutions. Many "administratively live" (Principal Investigator) within their universities' schools of medicine—as one of the school's divisions or attached to the school's administration, such as the Dean's Office. A number of other hubs, though, were not "ensconced in the School of Medicine" (Principal Investigator). The other arrangements positioned the hub leaders "a little bit more directly in terms of line of authority" (Principal Investigator), directly in contact with a wide array of stakeholders, or between several intuitional hierarchies. Across these arrangements, participants identified key challenges and a strategic facilitator related to hub authority and control over factors related to implementing the work of the Common Metrics (Table 32).

A main challenge was lack of authority over key organizational components and processes within their institutions that determine hub outcomes. Participants especially emphasized this challenge in relation to their IRB and IRB metrics. Additionally, the complexity of processes related to the Common Metrics made it difficult to understand and measure all the interdependent aspects involved. As a result, it was more difficult to pinpoint drivers of performance and identify improvement strategies. The complexity also exacerbated

Main Themes with Illustrative Quotations*
Challenges
Lack of line authority over key drivers
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? – Principal Investigator
Metric topics, particularly IRB Duration, are part of complex processes
it's complicated. I think these are good metrics to assess. It's just difficult sometimes to understand what all of the causal factors are and whether we can directly influence those factors. – Principal Investigator
if an investigator submits an IRB protocol, if the IRB responds quickly and gets them back and asks for minor changes, if the investigator sits on those minor changes for six months, I can't stop that. So that's going to look really bad on those metrics, and we have nothing to do with thatSo, I think we all take our licks. I mean, there's always going to be protocols that will be delayed which will affect this number. – Principal Investigator
Facilitator
Occupying institutional and integrated leadership roles affords some influence
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. – Principal Investigator
So administratively we are a separate center even though I'm 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'm going to also hold a title of Associate Dean for Research, as did my predecessor, and that's by design. – Principal Investigator

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

the challenge of some factors being outside of hubs' direct control, such as investigators' response times to IRB stipulations and the need to coordinate with multiple IRBs.

Although the problem of lack of direct authority could not be fully mitigated, some hubs identified a strategy that provided some help. That is, 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.

# Local Teams and Organization of Common Metrics Work

# **Core Teams**

Participants, particularly those serving in managerial roles related to implementing Common Metrics, often reported forming a core team to work on the Common Metrics and related performance improvement activities. These teams were intended to provide an organized approach to implementation activities, including 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.

I would say, establish the team early...so that you can make sure that you have more than one person listening to the calls, reading the material, attending the training, so that you can have an organized approach to [the Common Metrics]. But, yeah, I think the team is the most significant thing. – Administrator

Core teams were typically relatively small, often leveraging other personnel within the hub or institution to contribute to metrics data collection. In smaller hubs, core teams may be particularly lean and exhibit less differentiation in roles related to Common Metrics Implementation. Although not typically part of the core team, directors of hub programs related to Common Metrics' specific topic areas also played critical roles in the overall Common Metrics team. In addition to supporting Common Metrics work, their ownership of the data and/or familiarity with the processes in their topic areas were considered valuable for implementing improvement strategies.

[T]he program directors are the ones who are in the trenches; they understand their program very well. The PI understands the program but doesn't see the day-to-day difficulties that are going on. And so, it's really good to have the main person involved in each of the different metrics... to be the one sort of leading the charge because they really understand what the barriers and what the facilitators are. – Administrator

I was just saying that really changes the dynamic when potentially people who would be responsible for carrying out a strategy to improve IRB performance kind of own the data and then generating the statistics themselves, it was really easy to get buy-in. – Implementer Participants identified three facilitators related to 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 3) a collaborative team climate with effective communication (Table 33).

## Table 33. Facilitators for implementation: core team

Main Themes with Illustrative Quotations*
Facilitators
Team member(s) take ownership of implementation
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. – Implementer
It's really good to have the main person involved in each of the different metrics, either be it Careers, IRB, or Pilots or Informatics, to be the one leading the charge because they really understand what the barriers and what the facilitators are. – Administrator
Local champion on the team
Our project manager is a real believer and a true champion for this process. Both he and our overall evaluation lead have been out there beating the drum for the overall process. I'd give a lot of credit to them for embracing this because it really has helped us change our culture – Principal Investigator
Effective team climate
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. – Administrator
Each week in rotation different members of our senior staff report to the whole group, some of which relate[s] to the Common Metrics. So we keep up-to-date by having very, very free flowing communication and reporting back. – Principal Investigator

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

## Hub Principal Investigator Role

Although hub Principal Investigators typically split their time among multiple responsibilities and leadership roles, participants noted a variety of facilitative roles that Principal Investigators played to promote the implementation of the Common Metrics (Table 34). For example, Principal Investigators provided strategic and operational guidance to the Common Metrics team at regular meetings. Principal Investigators offered valuable "higher level" views on the purpose and objectives of the Common Metrics Initiative, helped trouble-shoot implementation issues, and assisted in interpreting data, reviewing improvement plans, and making recommendations. In many hubs, the Principal Investigator was described as a "champion" of the Common Metrics process who kept the Common Metrics "on the agenda," ensured the Common Metrics work progressed "as best it can," and acted as "ambassador for CMI [the Common Metrics Initiative] to constituents throughout the institution." On occasion, members of Common Metrics teams described their Principal Investigator as an important source of encouragement and enthusiasm for implementation. Even when not specifically referred to as a "champion," the hub Principal Investigator often played a prominent role in engaging stakeholders for the Common Metrics process. The visibility and clout of the hub Principal Investigator was observed as helping to enhance their effectiveness with engaging stakeholders.

In overseeing the implementation of Common Metrics in their hubs, hub Principal Investigator's involvement ranged from a hands-on, in-the-weeds approach to more handsoff management approach. The level of involvement in the details of implementation varied depending on the Principal Investigator's own expertise and areas of interest, or the stage of implementation of the Common Metrics at the hub. A hands-off approach was considered by Common Metrics staff in a number of such hubs as benign or even appropriate.

# **Organization of Common Metrics Work**

To better understand how hubs conducted the work of implementing Common Metrics and performance improvement, each hub was asked to identify up to five main activities undertaken for the most recent update (August 2017). As expected, many hubs characterized their activities in terms of data tasks (collection, analysis, and reporting) and

# Table 34. Facilitators for implementation: principal investigator involvement

Main Themes with Illustrative Quotations*
Facilitators
Providing strategic guidance
We would report the statistics to him, or the Common Metrics, and turn the proofs into him before we finalized them and of course, before we presented them to the program officer. [He] would review the Turn the Curve plans and make recommendations, and I'd say he's very involved. He 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?" – Administrator
Serving as champion
I would say our PI, I think he has the role of champion on our Common Metrics team and he has definitely I think been that. So he welcomes, I think, 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. – Administrator
Our PI has been a source of great encouragement and support He has helped to make connections where we've needed themIt's not as exciting as some of the other things that are happening in biomedical research on our campus, but he provides a lot of encouragement and enthusiasm for these activities, which I think goes a long way. – Implementer
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. – Administrator
Providing hands-on oversight during start-up
[The Principal Investigator] 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. But that's not terribly uncommon for new initiatives that we have to roll out. I think as the Common Metrics continue, he would need to be less involved on a day-to-day basis. – Administrator

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

# Table 35. Hub characterizations of main activities involved in most recent update (N=49 hubs<sup>•</sup>)

Main Asticity has find a second		Hubs	
Main Activity by Category	n	%	
High-level championing and/or oversight			
Championing/visioning	10	20.4	
Overseeing execution	1	2.0	
Project management			
Planning/preparation	19	38.8	
Ongoing monitoring	10	20.4	
Creating timelines	6	12.2	
Identifying personnel	4	8.2	
Data collection, analysis, and/or reporting	I		
Data collection and metric computation	29	59.2	
Reporting/entering metric result	19	38.8	
Needs assessment and procedures	17	34.7	
Disseminating results	3	6.1	
Data analysis	1	2.0	
Implementing performance improvement framework	1		
"Turn the Curve" Includes reporting	39	79.6	
Results Based Accountability (RBA) Includes training and reporting	13	26.5	
Collaborating and integrating feedback			
Collaboration and communication	31 12	63.3 24 5	
	12	2710	
Training and Tufts evaluation study			
Training/orientation <sup>***</sup>	13	26.5	
Responding to surveys	1	2.0	

\*Eleven hubs did not respond about this topic.

"Participants could provide up to five main activities.

<sup>•••</sup>Denotes training or orientation in advance of the implementation phase; for most hubs, this refers to the start-up period, not the update due in August 2017.

implementing a performance improvement framework (Table 35).

A high percentage of hubs also characterized one of their main activities as collaboration and communication (63.3%). Between one-fifth and one-quarter of hubs identified the activity of integrating feedback and lessons learned (24.5%) and championing/visioning related to the Common Metrics (20.4%). Project management also was a common group of activities, including planning/preparation (38.8%), ongoing monitoring (20.4%), and to a lesser extent, creating timelines (12.2%) and identifying personnel (8.2%).

# **Planned Changes to Approach**

Almost half of responding hubs (46.9%) reported that they planned to change their approach to completing the annual Common Metrics updates (Table 36). The type of changes varied, with the most common being a plan to change processes for collecting data (16.3%), to begin the required activities earlier (14.3%), and to change personnel roles and responsibilities (12.2%). Of note, slightly more than half of responding hubs planned no change after just the second reporting period of the Common Metrics Initiative.

	Hubs	
	n**	%
Any change planned	23	46.9
Evolve processes for collecting and requesting data	8	16.3
Begin process or portions of process earlier	7	14.3
Allocate roles and responsibilities differently	6	12.2
Evolve current measurement methods to metric requirements	4	8.2
Alter approach to performance improvement plans	2	4.1
Augment with hub-specific metrics	2	4.1
Type of change depends on leadership decisions (e.g., review of efficiency, alignment with other activities)	2	4.1
No changes planned	26	53.1

Table 36. Hubs' anticipated	d changes to their a	approaches (N=49 hubs*)
-----------------------------	----------------------	-------------------------

\*Eleven hubs did not respond about this topic.

"Participants could indicate more than one type of change.

# **Role of Tufts Implementation Program**

Although the need to provide every CTSA the best possible training and support did not allow for a randomized design to test of the effect of the Tufts Implementation Program on hub progress, results suggest that the program of training and coaching played a positive role. As noted above, attendance at the training and coaching sessions appeared to have a positive association with completion of performance improvement activities, but it is unclear whether the *content* of the training and coaching program had a causal effect. Participant assessments provide additional evidence that the program itself played an overall positive role.

Eighty percent reported satisfaction with the program (5% were extremely satisfied, 48% moderately satisfied, and 28% slightly satisfied), and many fewer (20%) reported some level of dissatisfaction (Figure 19).



Figure 19. Satisfaction with Tufts Implementation Program (N=59 hubs)

When asked about the program's effectiveness, 27% found it extremely or very effective and another 44% found it moderately effective. The 25% who reported that the program was slightly or not at all effective indicate that the needs and/or preferences of some hubs were not met (Figure 20).



Figure 20. Effectiveness of Tufts Implementation Program (N=59 hubs)

Tufts CTSI Common Metrics Report - Part II: Evaluation Study

To better understand the unmet needs, hubs were asked about the amount of knowledge and proficiency gained from the program relative to what was needed. Overall, the vast majority of hubs gained the knowledge and proficiency they needed, or more, to carry out the work of the Common Metrics (an average of 93% and 86%, respectively). Although hubs gained the requisite knowledge and proficiency to proceed, results revealed that hubs needed different levels of training and support (Figure 21). On average, 61% gained the amount of knowledge needed about collecting metric data and RBA, but 39% reported a level of need different from what was gained. Specifically, 32% gained more or much more than needed and 7% gained less or much less than was needed. The trend was similar for proficiency gained, with an average of 57% gaining the amount of proficiency needed to collect metric data and implement RBA, and 42% reporting a different level of need (29% gaining more or much more proficiency than needed, 13% gaining less or much less than needed). These results did not differ by Implementation Group (Appendix Q).



Figure 21. Knowledge and proficiency gained from Tufts Implementation Program (N=59 hubs)

Qualitative interview results shed light on the reasons behind hub opinions. Overall, hubs indicated that the Tufts program was helpful for Common Metrics implementation. Hubs that used the supports offered by Tufts identified them as a facilitator. On the other hand, hubs that did not use the Tufts support acknowledged that, had they done so, it could have helped their implementation efforts (Table 37).

## Table 37. Challenge and facilitator for implementation: Tufts Implementation Program

Theme and Illustrative Quotation(s)*
Challenge
Lack of use by some hubs <sup>†</sup> [T]hat [reaching out to Tufts for individual support] might have been useful. And maybe that's there, but I just didn't reach out. – Implementer
Facilitator
Useful training and support $^{\dagger}$
I find that the trainings did accomplish for us what they needed to, in terms of helping us to use the language and facilitate the discussions with the various hub stakeholders. I thought that the time and the follow-up activities were pretty efficient and helpful. So overall, the training experience I think was well-conducted. –Implementer
they [Tufts] were always just really responsive and helpful and flexible – Implementer

<sup>+</sup>Unless stated otherwise, themes manifest in more than one way; a quotation represents one manifestation. <sup>+</sup>Indicates that the facilitator has a corresponding challenge.

However, there was concern that the training took too much time overall and relative to what was needed.

I think it was too onerous. It was overkill quite a bit, and I also feel like a lot of it could have been done on one's own and not having to spend time with the webinar. ... So that, and then the fact that I had so many staff that were involved in that. It just was not a good use of our time. So, I feel like that just wasn't the most satisfying experience. I think it actually in some respects overcomplicated what was being done. – Implementer

Smaller hubs may have felt this pressure more acutely than large hubs. One small hub described the reason as follows:

So, there's a disproportionate burden put on small hubs for, not only training, but the Common Metrics [Implementation] overall, because we just don't have the person power to do it. So, it doesn't matter if you have five full-time people working on program evaluation or you have less than one, you all have to do the same thing for the Common Metrics. – Implementer

Participants recognized the value of offering a variety of types of supports and indicated that coaching and discussing with other hubs was most helpful.

...I'm talking about the initial trainings; [they weren't as helpful]. The coaching calls, the follow-up calls, those were more useful. Those were definitely worth attending. – Implementer\*

<sup>\*</sup> Participant is affiliated with a medical center that functions as a CTSA without current CTSA funding.

One thing I did like on the training calls was...that they grouped us with other CTSAs so that we could hear some of their strategies. ... It was the slides and all the presentations, and it was a little repetitive and just a huge time commitment. – Administrator

Some hubs pointed out that, depending on a hub's pre-existing knowledge, a more tailored approach may be appropriate.

...I think that both trainings are necessary for some people, particularly because you might have people that are inexperienced with different kinds of programs... But I understand that everyone is different in how they take in that information, so I can see where it would be valuable to some people, but for my own purposes, it just kind of dragged on a little bit. – Implementer

## Feedback on Results Based Accountability Framework and Scorecard Software

After receiving training on and using the Results Based Accountability (RBA) framework and the related Scorecard software for more than a year, participants identified both benefits and limitations of each (Table 38).

## **Results Based Accountability Framework**

Participants often discussed their understanding of the performance improvement framework in concert with their sense of its usefulness. Generally, participants found the RBA framework to be straightforward. That said, some found it more useful than others.

Many participants focused specifically on the "Turn the Curve" plans in regard to the usefulness of RBA. They described these plans as a blueprint for performance improvement ideas and actions. While some type of evaluation was already occurring at all hubs, for many hubs the RBA framework formalized a process for performance improvement. That said, some hubs indicated that the framework could have offered more guidance to help them explore ideas for performance improvement. Additionally, hubs with a pre-existing approach that worked locally were less likely to engage with the RBA framework, as they often perceived their existing approach led them to dig deeper than this framework would require.

## **Operationalization Via the Scorecard Software**

The framework and the Turn the Curve plans specifically were documented using the "Scorecard" software. Hubs either designed new processes or modified pre-existing ones so that they could generate the information needed for the Scorecard. For some hubs, the software facilitated their experience with implementing the Common Metrics because it was simple to use. Participants found it useful for documenting their plans and progress as well as for reporting. On the other hand, they felt that there were various ways in which the software could be improved (Table 39).

Aside from the software challenges and desires for improvement, it is also important to note that there were additional challenges due to lack of alignment between the metric definitions and the template for reporting and visualizing results in the Scorecard software.

My challenge was also figuring out the target value. Because sometimes the way I thought the way the graph was going to show based on the numbers of putting in, I had to really come back and go line by line, thinking "What exactly am I putting in?" – Implementer

# Table 38. Benefits and limitations of Results Based Accountability framework and Scorecard software

#### Main Themes with Illustrative Quotations\*

#### **Benefits**

#### RBA framework provides a blueprint $^{\dagger}$

I think it's a good thing, [the Turn the Curve plan], that's one of the things I feel Administrators are always charged with process improvement – how do we do the same thing more efficiently and more effectively? I think the Turn the Curve plans helps us to develop that blueprint to accomplish those things. So, I see a valid benefit from the Turn the Curve plans, it's a resource and it's a working document for us to benefit from. – Administrator

Scorecard offers easy, common platform<sup>†</sup>

Because it is actually a fairly easy-to-use, not terribly sophisticated interface that doesn't try to do data management. So we can handle all the data management locally but that [the Scorecard software] actually facilitates the process itself, plans it, then produces fairly easy-to-use graphics. And the fact that we can use it for a variety of different metrics and have a common platform for displaying them is helpful ...so, I became more positive as the process progressed. – Implementer\*\*

#### Limitations

#### Lack of depth of RBA<sup>†</sup>

[RBA] doesn't really give you any next steps or it doesn't really push you to explore the holes in the framework necessarily. So, I don't know... I feel like it's a framework, it's very clear as to what steps you should take and what steps you need to do, but I don't feel like it really guides me to what is missing or what other... [steps] I need to take. – Implementer

Technical limits of Scorecard software<sup>†</sup>

I think the Scorecard is just clunky. It's very difficult to navigate through. It's difficult to download stuff and get clean data. -Implementer

...[finding] some additional ways that we can look at the data and display it, that would be more useful. - Administrator

\* 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.

 $^{\scriptscriptstyle \dagger}$  Indicates that the facilitator/benefit has a corresponding challenge/limitation.

# Table 39. Desired improvements for the Scorecard software

Desired Improvement and Illustrative Quotation(s)*
Expand software capabilities to include data collection and storage
[l]t doesn't allow one to automatically download. And so, it's created a lot of work for people to do manual data entry when contemporary software products really should allow automatic download of data. So that's been a substantial frustration to our evaluation team – Principal Investigator
[I]f Scorecard was also a way to be able to collect the data-because you still have to collect the data and then enter it into Scorecard. So, it's a little bit redundant in that regard [C]an't it also have a component of being able to be a database to collect all of this data and store that data versus us having to go ahead and create that or create something like a REDCap database? – Implementer
<b>Enhance user experience</b> Examples: speed, intuitiveness, number of clicks required to navigate
They said the software was easy to use, but it tended to be slow and could be cumbersome with too many clicks to get to the target page. – Principal Investigator
I would get rid of the Scorecard and create some sort of just general dash, [something] more user friendly for people who are actually collecting the data They handed it over to me, and they just let me run with it and had minimal input – Implementer
[I]t was somewhat challenging to figure out how to input the numbers to show how the curve should look [W]e knew how it should look, especially if you were doing it in Excel format, but getting it into the actual program could sometimes be challenging – Implementer
There's a piece of operation on the Scorecard that completely confuses me. So, I have my Story Behind the Curve, and I have my performance measures, and then I have all these other boxesand I'm afraid to delete them. They look empty, but you seem to be able to open a limitless number of duplicate things, which can be very confusing. It seems like for a given project there should be one Story Behind the Curve. – Implementer
<b>Improve visualization capability</b> Examples: create more display options, display multiple metrics simultaneously
We'll probably go more towards a one-page dashboard where we could put multiple metrics onone sheet of paper, so it could be more easily reviewed by a leadership teamin one sitting, versus kind of clicking through multiple metrics – Implementer
[T]he software used to develop the graphic representations of the metric data was not as good as it could have been [For example,] instead of picking a point, you know 2016 or whatever, it would have been more useful to develop these cohort curves, so the trainees that completed in 2012, this is their experience, 2013—so you can look at trends in the dataI think there could have been easier or better, more informative ways to display the information. – Principal Investigator
The Results Scorecard is a fine way for us to submit data obviously, but we can do better things with local visualization tools—for example, like Tableau—than we can do with Results Scorecard. So other than communicating with you guys, we haven't found Results Scorecard to be the benefit of this whole endeavor. – Principal Investigator
<sup>•</sup> Unless stated otherwise, themes manifest in more than one way; a quotation represents one manifestation.

Furthermore, there was the sense on the part of some hubs that their pre-existing systems were just as good, if not better, and that some were not proprietary, so using the Scorecard felt like an unnecessary drain for a limited resource. First, the number of licenses provided to each hub is limited (unless specifically requested in the hub's budget). Second, hubs were not certain that those licenses that were provided in the past would continue to be provided into the future.

...[R]ight now, we're not paying for it, but if it's going to be a requirement going forward for CTSAs, are we going to have a licensing fee to use the software? ... We haven't put [that] into our budget... – Principal Investigator

They were concerned that if they expended a lot of effort to engage with the Scorecard and then the software was no longer supported, that their effort would be lost.

# **Perceived Value of Common Metrics Implementation**

## **Overall Self-assessment**

Self-assessments at two time points-the beginning and end of the study period-revealed that hubs considered their abilities to manage performance to have improved (Figure 22 and Appendix R). This improvement was statistically significant for the ability to assess

### Figure 22. Self-assessment over time (N=58 hubs\*)



\*One hub did not respond to the second follow-up survey. A second hub was dropped from this analysis due to missing data. \*\*p=0.02

p=0.02 \*\*\*p=0.01

 $^{E}$ 0-100 scale; higher score reflects a better self-assessment.

See Appendix R for differences in the wording of some items between the baseline and final follow-up surveys.

current performance and identify actions with the potential to improve performance. The improvement was also statistically significant for hubs' abilities to assess likely future performance, which is notable given that the specific activity of forecasting metric results was incomplete in more than one-third of cases (Table 16). Assessment of the ability to advance clinical and translational science remained unchanged.

# Areas of Value Added

Participants in qualitative interviews shared their perspectives on the perceived value of the Common Metrics Initiative. They articulated ways in which it added value, both locally and for the CTSA Program, and voiced concerns about obtaining value from the initiative (see Areas of Concern About Value section).

Participants from all hubs in the qualitative interview sample reported at least one—and typically multiple—types of perceived value added from the Common Metrics Initiative. These included providing a formal structured process, enabling strategic conversations, facilitating improvements and tracking progress, and providing an external impetus for improvements and justification for continued funding for CTSAs.

# **Providing a Formal Structured Process**

Principal Investigators, Administrators, and Implementers remarked on the value of a formal structured process to "think through" performance improvement.

[T]he process of having to formally think through what is leading to the current level of performance and what you could improve I think is a useful thing. So, we've been trying to approach the Turn the Curve plan with sort of an open mind and thinking that this is probably a good thing for us to go through formally. – Principal Investigator

It's given us more of a structured approach to defining metric-specific plans moving forward. Like I said, we always collected data and collected metrics and we use those to base our strategic planning on. But we never did the in-depth analysis of the metric itself and looking beyond the obvious factors. Looking into the secondary and tertiary analysis I think has been beneficial...because some of those things we found that we were actually able to address and to change. – Administrator

Other participants emphasized the value of the structured approach in providing sets of clearly defined metrics on which hubs could focus.

I think the process of sort of having clearly-defined metrics, whether they're the ultimate metrics or not, sort of in looking at the context behind those, I do think it adds value and just helps us maintain a focus on some of the key programs and how we can best help facilitate them. And also recognize changes when they're happening a little more readily. – Administrator

Some participants noted that the structured metrics and process had helped to enhance accountability for improvement and integrate evaluation activity "more deeply" into all the hub's activities.

[I]t opens the door for communication, and it made communication more frequent because now we have to make sure that the Turn the Curve plan is up-to-date and that we're following the plan that we've implemented. – Implementer

Yes, over time I have perceived the metrics as being increasingly useful to my push to integrate evaluation activity more deeply into all of [our organization's] programs and services. – Implementer

## **Enabling Strategic Conversations**

Participants at a number of hubs noted that the conversations stimulated by implementing Common Metrics enabled strategically-oriented discussions about current performance and improvement opportunities. As one participant described, implementing Common Metrics helped the hub to "take a step back" to reflect over all the hub's programs, even those performing well.

I mean it's a learning curve for us with the Results-Based Accountability, but we find it very effective actually, and it's actually a good way to lead discussions among the program directors on where they want to take their metrics... It's been really good to facilitate structured conversation and discussion about how we can improve everything that we do here. –Administrator

Others commented on how the Common Metrics Implementation encouraged "broader strategic conversations" and the "broad picture," including the national scope of the Common Metrics Implementation.

It's much more about the conversation around the data. There are questions about what else kind of is going into the program in terms of process and outcomes. And then a provider conversation of how can we do better. And to me that's a success; the data allow us to have these broader strategic conversations around pilots, around education, around IRB that we wouldn't have been able to do if it was just kind of a regular routine kind of process. –Implementer

It's broad picture versus "Do I just stay in my own little hole, put my head in the sand, and I do my area of subspecialty expertise" in how I'm contributing to this overall grant. But it's making sure that all leadership is involved and is aware of "Here's the existing Common Metrics that we're working nationwide, part of the broader team, here's what we're doing." It's good for them to see that overview. – Administrator

### **Facilitating Improvements in Immediate Outcomes and Processes**

Participants from a variety of hubs pointed to improved processes and services realized through the implementation of Common Metrics. Although improvement in the metric result may not be evident, a Principal Investigator described observing notable improvement in services and immediate outcomes.

I realized that these things are just absolutely essential to an institution, to build a clinical research program, and there is zero, zero question that I've seen things improve. The scholars are better. They're better prepared. They're sending out their first grants that are a lot better. They've got infrastructure support for statistics that used to be, you know, ad hoc, nonexistent or worse. – Principal Investigator

Other hub participants noted impact on improving how performance data are collected and identifying specific gaps, such as with compliance issues.

...[O]ne thing the Common Metrics has done, it's made us think aboutoutside of the Common Metrics-how we collect our local metrics and what they should be. And so, I think it's been a good thing, because that's the only way that we can ever really change our environment. – Administrator

So, going through this and the pilot really shined a light on what some of our inefficiencies were, where some of our gaps were, and where we had some real exposure to potential compliance issues as well. So that really has actually, if anything, given us ammunition to actually say "We need this, we have to do this, we're at risk if we don't do this, and besides our clinical research portfolio is going to suffer if we don't implement some of these improvements, structural and system improvements." So that's been helpful. – Administrator

## Providing External Impetus for Improvement and Justification for Funds Invested

Participants at a few hubs noted that the external mandate and requirements of the Common Metrics, although chaffing for some, may serve as a useful "impetus" or rationale to justify and motivate hub stakeholders to expend effort on the topic areas addressed by the Common Metrics.

[E]ven though I'd hoped that we would do that anyway, ...the fact that we have to report these outcomes is an impetus to do it. I mean, the external pressure sometimes, taking advantage of this external requirement to do it is useful, so just the mere fact of stipulating to do it. – Principal Investigator

It's also put deadlines on reporting and forced the program areas to be actively involved. Not that we needed people to be forced, but it gives more dedicated focus, so much going on and having a deadline that's required for these things, it's helpful. – Administrator

Participants at a number of other hubs emphasized the role of Common Metrics in justifying the value of CTSA activities to a variety of key stakeholders and funders within and outside their institutions.

The way I see [the Common Metrics'] biggest value is, number one, it gives us—and that's you and me and the whole Consortium—a way to show Congress, a way to show the public, a way to show the rest of NIH, "Look

*at what the* NCATS *program is doing or the* CTSA *program is doing. This is good stuff.* We're *doing good stuff.*" – Principal Investigator

Also, running the finances of an institution, especially in higher education, I am a big [return-on-investment] person. I am a big "We have to show your bang for your buck" [person], especially someplace like the CTSAs [that] are so heavily subsidized by our institution. – Administrator

## Areas of Concern About Value

While participants across all hubs found some value in implementing the metrics, concerns with receiving value from or demonstrating value of the Common Metrics remained. Participants at more than two-thirds of hubs taking part in qualitative interviews expressed concerns with the usefulness of the chosen metrics and performance improvement framework and/or the overall value relative to the effort expended.

### **Overall Usefulness**

## Usefulness at the local level

As described above, participants expressed concerns about the usefulness of the chosen Common Metrics for local improvement (see Metric Clarity and Usefulness section), and these concerns were cited as a reason for not completing performance improvement activities. Participants indicated that the Common Metrics did not necessarily reflect local definitions of success, incorporated lag time that made assessing current performance difficult, and did not align with local context or priorities.

Certainly, the areas that were initially proposed and the new areas generally going forward make a great deal of sense. They are completely reasonable. Having said that, I have no idea exactly yet what their value added to us is because we obviously maintain many, many, many more and different kinds of metrics than these for internal purposes. So, it's a little unclear to me as PI kind of where this is going to go across the hub. – Principal Investigator

These concerns raised a tension between the need of the CTSA program for standardized metrics versus the need of hubs to make performance measurement meaningful for local improvement efforts within their own institutional context.

In order for it to be more meaningful, the metric would have to come out of the hub itself. Like, this is where you get into that balance of the CTSA program as a whole versus each individual CTSA in that we do function according to the needs of our institutions. And in order for a metric to be more useful for the institution itself, it's going to need to be tailored to the institution. So, there's limited value in a metric that's going to be used across all CTSAs in terms of improving a specific CTSA... – Administrator

Acknowledging this tension did not necessarily mean that the metrics were perceived as useful at the Consortium level.

# Usefulness at Consortium Level

Concerns about the value of the metrics at the Consortium level included the limited scope of the metrics and variability in measurement across hubs.

*Limited Scope* Similar to concerns raised at the local level, an overall concern at the Consortium level was that the Common Metrics were currently too narrow in scope to give a useful portrait of performance.

...[Y]ou can't pick everything, so...if your point is to accelerate clinical trials, then create a dashboard for yourself that can benchmark CTSAs in that function. Just having IRB review is not going to do that. If your goal is to try to bring the next generation forward, then create what that dashboard would be. But right now, these little spot-weld things, I mean I don't know how anyone centrally can look at that across six institutions and be able to say anything. So, it probably needs to be fleshed out in my opinion. – Principal Investigator

*Variation in Measurement* Perceived lack of clarity of metric definitions or lack of alignment with local systems allowed for potential variability in how hubs computed metric results. Participants recognized the potential for lack of standardization and "interpreting [a metric] and kind of doing it in the ways that will be most favorable for their organization" (*Administrator*).

Devil's in the details... - Principal Investigator

There are ways to distort statistics. - Principal Investigator

*'[I]f you've seen one CTSA, you've seen one CTSA.' And so, everyone was collecting things a little bit differently. – Principal Investigator* 

...[H]ere on paper, we have a Common Metric that says everybody's doing this, when in fact I know that that's not happening. But rather than discuss it and figure out a way to deal with that, we just sort of passed it forward and people reported something. As long as people reported something, there was less interest in sort of making sure that everyone was reporting exactly the same thing... – Administrator

IRB Metric – Variability in the time period being measured existed for two reasons. First, for some hubs, the time points for the definition were not clear.

There was some confusion with the terminology that was used, so we ended up having to determine what we were going to use for date of enrollment, date of contracts... So, we felt that maybe you couldn't really compare it across CTSAs because it sounded like others were having some of the same issues with the terminology a little bit. – Administrator\*

Second, hub workflows differed regarding when in the local review process the clock started ticking for the IRB metric. For some hubs, all other review processes were performed before the protocol entered the IRB.

<sup>\*</sup> Participant is affiliated with a medical center that functions as a CTSA without current CTSA funding.

An example of the first case is when the IRB review process only starts after a pre-review period is completed and the protocol is identified as ready for IRB review.

[I]n our system [the IRB metric] is meaningless because we spend most of our time with new protocol submissions on working with the research team to prepare the proposal for submission to the IRB. So, in reality, if you looked at our [entire] turnaround time for IRB new protocols, it would be about, oh, 35 or 40 days. If you look at what data is being collected in the Common Metrics, it's eight days...Which is more meaningful? – Principal Investigator

For other hubs, the other reviews were folded into the IRB process such that the IRB duration not only included but also depended upon the efficiency of those other processes.

Because contracting often takes so long-our IRB is very efficient, contracting is not as efficient...-now we're looking at not getting IRB approval until a contract has been signed. So, there are sometimes perverse incentives that are put in because of the way the metrics hit. And there's really no way around it. I mean, that's the problem with pretty much anything that you measure. How you measure it will incentivize things, hopefully that you want, but sometimes incentivize things that you don't want. – Principal Investigator

Pilots Metric – For the Pilots metric, participants suggested that hubs varied in the types of pilots included in data collection. Some hubs only house CTSA-funded pilots, while other hubs house or are connected with a larger number of pilot programs. Hubs in the latter category may report the Pilots metrics specifically for their CTSA-funded pilot program or for all of the pilot programs that they manage. If some hubs are doing the former and others are doing the latter, there is concern that the metric is not reporting on the same program.

And so, I actually think you got a hodgepodge going in there of what's being reported. And the only reason I know that is I ... explicitly asked a bunch of people who were telling me they run [many] pilot programs, and I'm saying, "There's no way you have enough money to run [that many] different kinds of pilots or whatever." And they said, 'Oh, well, we do ... we're really running this, but we're running it for that group, and our pilot manager does it, so we report that." – Principal Investigator

Although similar concerns about variability in measurement for the Careers metric were not expressed in the qualitative interviews, the general issue of lack of systematic definitions may apply.

# **Demonstrating Sufficient Value Relative to Effort Expended**

Some participants expressed concern with the ability to demonstrate enough value of the Common Metrics relative to the resources invested, and the consequences of not doing so (e.g., maintaining buy-in and momentum).

I would just say it's not a small amount of time and effort that folks are putting towards this, and I think we are concerned as a hub of how to maintain momentum given there's a bit of missing information with regards to return on investment. – Administrator

Several hubs described less than desired value of the Common Metrics at the current time but a willingness to wait to see if more value would be realized, either for the content areas of the current metrics or of future ones.

I can clearly see how it's of importance to the CTSA Consortium because you want to be able-and NCATS wants to be able-to provide objective and tangible evidence of how much we're succeeding at our goals. Eventually, it will be useful for us as individual hubs to see changes over time and determine whether we're succeeding in, as you would say, Turning the Curve or just improving what we're doing. I think it's too early in the process yet to really see that, at this point. – Principal Investigator

I don't know. My first reaction is I don't feel like it's added value, but I think if you just step back and look at it, knowing we have these specific numbers which is something we didn't pull out before and look at, I think that could be useful especially once we've got, say, three years of it to look at to understand where we're going... – Implementer

We see value in the Common Metrics, but specifically the way that it is constructed right now for these few metrics that are in the system, not the most valuable to us, but we are certainly optimistic on how all of the other ones are going to unfold and how we respond to those. – Administrator

Additionally, some hubs questioned the value of the RBA-based performance improvement framework. In their experience, the performance improvement plans required a great deal of work, but produced disappointing outcomes that did not appear worth the effort.

Well, I can tell you my only comment about this was collecting the metrics wasn't particularly difficult, but for our staff this concept of changing the curve actually ended up being a lot work and a lot of effort. And I'm not sure that I know in my own mind that I feel like it was worth the effort put into it. – Principal Investigator

Taken together, the areas of value added held alongside the concerns about usefulness and sufficient value suggested that the Common Metrics Implementation was perceived to be a valuable endeavor with its own opportunities for improvement.

# **Additional Concerns and Hub Recommendations to NCATS**

In addition to the concern about sufficient usefulness and value of the Common Metrics, participants also voiced concerns about balancing local and national needs, issues with data collection and comparability of results, and using a proprietary software program (Table 40). These concerns represented feedback to inform adjustments to enhance the impact of Common Metrics and hubs' experiences implementing them.

#### Table 40. Hub concerns about the Common Metrics Initiative

#### Theme and Illustrative Quotation(s)\*

#### Local vs. national needs

Metrics meet NCATS needs but do not help hubs learn/improve

So we're really looking at the questions that we want to ask rather than barely reporting on the Common Metrics. And we don't have anything particularly against the Common Metrics. It's just we have an opportunity here with other people that are tracking folks to bring it all together to allow us to ask what we consider to be better questions [than the Common Metrics]. – Administrator

Focusing on improving areas of high performance is unhelpful

I remember even asking this, "If you have 100% across these different metrics, what is the Turn the Curve plan supposed to be about?" ...And what the Tufts person told me was, "The Turn the Curve plan should be about what your CTSA intends to do to continue that high level of success." ...The people in my CTSA...kind of found those data not really helpful. – Administrator

#### **Data collection and results**

Comparisons across hubs may not be valid

As long as people reported something, there was less interest in sort of making sure that everyone was reporting exactly the same thing; at least that was my interpretation. – Administrator

... there was confusion about how it was worded in the Common Metrics and what our IRB then decided to use. So we felt that maybe you couldn't really compare it across CTSAs because it sounded like others were having some of the same issues with the terminology a little bit. -Administrator\*\*

Data collection is difficult when hubs do not control data at their institution

The data was really hard for us to get, and that took a lot of time to sift through and get the information. -Principal Investigator

Need for acknowledgement that affecting change will take time

...[S]ome of those people are still training [or they] are just starting with their first faculty position ... because clinical training is so prolonged. ... there's not a lot to report. - Principal Investigator

#### **Scorecard software**

The Scorecard software is a proprietary program that requires individually purchased licenses

... [E]veryone really needs to be very mindful of the fact that the Scorecard system is not free and we only get a certain set of licenses. And if you really want to implement it throughout the university and want everybody to have an account, then these accounts need to be paid for and there need to be resources within each individual hub's budget to do that. ... I'm not sure why it's focusing so much on a proprietary system at the beginning instead of using just a model that anybody could use. -Implementer

\* 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.

## **Hub Recommendations to NCATS**

Their concerns notwithstanding, hubs also provided recommendations to NCATS for improving the Common Metrics Implementation moving forward. Their recommendations focused on providing useful benchmarking, disseminating best practices, promoting peer-to-peer learning, and sharing evaluation results throughout the CTSA Consortium (Table 41).

Benchmarks were of substantial interest among a number of hubs in order to help them understand their performance relative to other hubs, calibrate their progress, and identify areas still in need of improvement. At the same time, some participants warned of hazards to avoid in order to ensure benchmarking would be useful at the local and national levels. Specifically, these participants advocated some type of process to ensure metric results across hubs are comparable and emphasized use of the metrics for local improvement rather than ranking CTSAs.

Of equally high interest, hubs recommended that NCATS, and/or another entity, cull and disseminate successful strategies and best practices for implementing Common Metrics and achieving improvement.

Related to this, hubs recommended the creation of more opportunities to learn from each other, particularly hubs that are similar to each other. Suggested mechanisms included i) connecting similar types of hubs who likely face similar contexts and challenges, or ii) pairing experienced hubs with those more recently established.

Hubs also recommended expanding the use of data and research to guide Common Metrics work. Suggestions included using additional hub data to test for predictors of Common Metrics outcomes, collecting data about what metrics hubs use in order to identify commonalities across hubs, and sharing assessments of the Common Metrics Implementation with hubs.

### Table 41. Hub recommendations for Common Metrics Implementation

#### Theme and Illustrative Quotation(s)\*

#### Provide useful benchmarking to inform improvement

Benchmarks would help calibrate performance and identify areas for improvement

I never know what "performing well" is because there's not necessarily a benchmark.

- Principal Investigator

I suppose if I saw data that made us look like we were not doing as well as our sister hubs, I would then be very interested in what they're saying they're doing; and maybe we'd adjust our strategies accordingly. – Administrator

Benchmarks are useful if they are accurate and facilitate local improvement

[I]t is so critical...to develop Common Metrics and define very strict parameters on...what that data should look like that's put into that system, so really we can make valid comparisons across the Consortium. - Administrator

The main recommendation I would have is that these [metrics] should be used primarily to help each site improve...and used longitudinally. There should be less emphasis on comparing one site to another... - Principal Investigator

#### **Disseminate drivers and strategies for improvement**

I think [it'd be] helpful [to] see what's really working well at another place. And then if it looks interesting and it's something we could implement here, then we...have a little more...data or a plan of what works someplace else to show the leadership at [our hub]... – Administrator

#### **Promote peer-to-peer learning**

...[L]et's say that they reported stats for...about 10 or 15 CTSAs that were comparable to ours...and you knew who the institutions were within that group, you could reach out to all of them and ask..., "How are you doing it?" ...It just seems like it would be a conversation starter. – Administrator

#### Use research and assessments to inform decision making

Expand data-driven decision making within the Common Metrics Initiative

The whole reason for having metrics...is to be making data-driven decisions... [W]ith all of the granular information [at hubs]....we ought to have the ability to...see what things might be predictive of "better outcomes" or "shorter IRB times." – Principal Investigator

...[I]t would be nice if there was a survey to find out what [metrics] people actually are collecting, that we could find "common" metrics, common ground in the data... – Implementer

#### Share assessments throughout the CTSA Consortium

I think that people are eager to hear about the outcomes of the Common Metrics Initiative and to see the aggregated data, ... examples of Turn the Curve plans, and...internal evaluations of whether or not this is worth it and how. [D]emonstrate to the PIs that there is some serious self-reflection that's going on, and that includes warts and all. – Implementer

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

# DISCUSSION

# **Hub Progress**

The Common Metrics Implementation was a CTSA Program Consortium-wide effort, and every hub was tasked to participate fully. Implementation began in June 2016, and every hub was expected to complete all activities for each metric by the end of the data collection period, January 2018. In practice, by the end of data collection, all hubs had begun the work of implementing the Common Metrics, but only about one-third of hubs had completed each activity at least once for all three metrics.

The vast majority of hubs reported that they computed metric results following the Operational Guidelines and undertook activities to understand current performance. Differences in completion among hubs began to appear in the development of their performance improvement plans. On average, three-quarters of hubs reported developing improvement plans, which were completed less often for the IRB Review Duration metric than the Careers and Pilots metrics.

# **Factors Affecting Implementation**

A variety of local factors, challenges, and facilitators affected implementation of Common Metrics and performance improvement activities across hubs. The most common reason hubs cited for not completing an activity was limitations of available resources, including personnel. The size of the hub's funding award did not fully account for this challenge. Investment from home institutions, periods of interrupted funding, alignment of existing systems with the needs of the Common Metrics, 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.

Hubs varied in their engagement with Common Metrics work. Types of engagement included actively folding Common Metrics and the RBA-based performance improvement framework into standard work processes, complying with an external requirement, or some mixture of these approaches within the hub and/or its staff. A hub's type of engagement was associated with the degree to which it completed the performance improvement activities. Not surprisingly, hubs in which all participants reported only a compliance-based approach to the Common Metrics completed fewer activities related to Common Metrics and performance improvement than hubs in which one or more participants reported active engagement.

Across hubs, alignment (or lack thereof) of the Common Metrics and performance improvement framework with a hub's local conditions and needs affected implementation. 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 and/ or performance improvement frameworks could facilitate implementation. But, this circumstance was not a given. 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 the existing data system was not aligned with the metric definition, modifying the existing system to follow the metric's Operational Guidelines absorbed a great deal of time and resources. For a few hubs, accurate metric data continued to be unavailable, sometimes due to lack of alignment of data systems.

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.

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 or processes related to Common Metrics experienced challenges in implementing performance improvement. 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. Even so, some hubs expressed concern about a seeming lack of recognition across the CTSA Consortium that affecting change would take time and require substantial effort, particularly when data and processes are not under their control.

Overall, there was substantial heterogeneity across hubs in their data systems, existing processes and personnel, organizational structures, and local priorities of home institutions. Implementing Common Metrics across varied local contexts created disparate experiences across hubs.

## Summary Recommendation 2:

## Maximize usefulness to hubs by selecting metrics that align with local needs.

- **2a** Select metrics that better align with local CTSA and home institution needs and priorities. For example:
  - i. Consider clustering similar CTSAs to address selected metric topics rather than creating Consortium-wide requirements.
  - **ii.** Acknowledge and communicate to hubs that local priorities can influence performance targets.

# **Summary Recommendation 7:**

Sustain engagement by facilitating solutions to barriers due to resources and authority, accounting for hub heterogeneity, and ensuring effective communication.

- **7a** Facilitate solutions to limited resources and personnel and use multiple strategies to account for heterogeneity across hubs. For example:
  - **i.** Consider aligning Common Metrics reporting with other required reporting (e.g., annual reporting).
  - **ii.** Consider an explicit process to weigh the value of a metric with the effort to obtain data.
  - **iii.** Consider a designated budget allocation to support Common Metrics work.
  - iv. Use a software platform that does not limit the number of users due to fees.
- 7b Account for heterogeneity of hub data, processes, and local priorities. For example:
  - **i.** Consider clustering similar CTSAs to address selected metric topics rather than creating Consortium-wide requirements.
  - ii. Offer expanded flexibility in choice of performance improvement framework.
- **7d** Promote peer-to-peer learning about successful strategies for affecting change in the home institution.

## **Strategies to Facilitate Performance Improvement**

Providing training, coaching, and opportunities for hubs to share experiences and best practices was useful for hubs. For each metric, Tufts Implementation Team training and/ or coaching were related to completing more performance activities. Increased attendance was a factor for the IRB and Pilot metrics, and receiving coaching while focusing on the Careers metric was useful. Although there was evidence of facilitation by these services, not completely clear is whether this relates 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. Hubs reported that opportunities for learning from other hubs were particularly valuable. Most hubs reported that they gained the knowledge and proficiency needed to carry out the work of the Common Metrics from the training and support. Future efforts might benefit from having mechanisms to adjust for different levels of need.

Hubs identified effective local teams as an essential element for successful implementation and suggested "best practices" in this regard. Most hubs convened a small core team to organize the work of the Common Metrics. Participants identified three facilitators that maximized the effectiveness of the team: 1) identify one team member who takes ownership of the project, 2) include a local champion on the team, and 3) attend to team climate and interactions. Additionally, the involvement of hub Principal Investigators was found to be useful, particularly to provide strategic guidance and oversight, champion the project, and facilitate stakeholder engagement. The RBA framework and Scorecard software received mixed reviews. Hubs without an existing, effective improvement framework appreciated having a blueprint for activities and a common software platform. However, some perceived RBA to be less useful for guiding in-depth analyses and discussions. Although hubs recognized the value of a common software platform, they reported technical limitations of Scorecard in data collection and storage, user experience, and visualization capabilities. Some hubs also questioned the long-term value of using a proprietary software platform.

# **Summary Recommendation 4:**

Equip hubs to fully implement each metric and performance management by providing peer-to-peer learning and training, coaching, and assistance for varying levels of experience.

**4a** Provide training and coaching that meets the needs of adult learners with different learning styles and various levels of prior experience in performance management.

## **Summary Recommendation 5:**

Support implementation by promoting metric-specific teams, allowing for capacity-building periods, providing accurate benchmarks, and updating performance drivers and best practices.

- **5b** Promote hub-identified facilitators for building effective teams, including identifying one team member who takes ownership of the project and a local champion on the team, and attending to team climate and interactions.
- **5c** Encourage involvement of the CTSA Principal Investigator to provide strategic guidance and oversight, to champion the project, and to facilitate stakeholder engagement.
- **5f** Provide useful, accurate benchmarking data to help hubs better target areas for improvement.
- 5i Promote peer-to-peer learning and disseminate best and promising practices.

# **Summary Recommendation 6:**

# Maximize usefulness of the reporting platform by enhancing functionality, visualization options, and user experience.

- **6a** Expand software features and functionality to support data collection, storage, and quality checks.
- **6b** Enhance the software's user experience (e.g., speed, intuitiveness, number of clicks required to navigate), and improve visualization capability (e.g., create more display options, display multiple metrics simultaneously).

# **Perceived Usefulness and Added Value**

Overall, hubs experienced some value in the Common Metrics Implementation, but they also expressed continuing concerns about the Common Metrics Initiative. Hubs did report an increase in their abilities to conduct specific performance improvement activities (e.g., assess current performance, assess future performance, and identify actions to improve). They reported value of the Common Metrics Implementation in providing a structured performance improvement process if one did not exist previously. This enabled strategic conversations, facilitated improvements in processes and immediate outcomes, and provided an external requirement that helped hubs justify recommendations for changes or additional funding. At the same time, hubs' self-assessed abilities to advance clinical and translational science did not change between the start and end of the study period. Among hubs taking part in the qualitative interviews, participants at more than two-thirds expressed concerns about the usefulness of the first three metrics and their overall value relative to the effort expended.

The usefulness of the first three Common Metrics varied across hubs for several reasons. As noted, alignment with local institutional priorities was a key factor. Additional challenges included perceived lack of clarity or stability of metric definitions, lack of ability to inform improvement at the local level, and the intrinsic time lag of some metrics for informing assessments of current program performance. Although allowing hubs to start with metrics that they perceived as "low-hanging fruit" helped some to prepare for implementing more challenging metrics, hubs already performing well on these metrics were faced with spending limited resources on areas that did not need improvement from their perspectives.

Specific to the initial three metrics, the Careers and Pilots metrics were considered by some hubs to be too narrow in scope to fully capture the goals of their local programs. For the IRB metric in particular, usefulness of the metric at the local level varied depending on the number of IRBs, the types of protocols reviewed, and the work process of reviews for ethics, feasibility, and budgets and contracts.

At the national level, some expressed concern that variation in data collection and computation at the local level would make combining or comparing metric results across hubs invalid. Knowledge of lack of alignment of existing data systems and perceived lack of clarity of metric definitions gave rise to concerns about the comparability of results across hubs. Although many hubs expressed a desire for benchmarks to help target areas for improvement, concern about comparability of metric results among hubs raised questions about the validity of aggregating hubs' results to identify performance benchmarks.

Questions about the usefulness of the first three metrics gave rise to concerns about a tension between local and national needs. Although the first three metrics may have met the needs of NCATS and NIH, many hubs found it difficult to use them for local improvement, particularly if the hub was performing well.

Given their continuing questions and concerns, some hubs indicated modifications to the process of the Common Metrics Implementation. There was a desire for more communication about the long-term plans for the Common Metrics Initiative and the expected timeframe to realize change. Such information could affect perceptions of the overall value of the effort. Additionally, a few hubs proposed a broader use of existing data at hubs. For example, understanding the metrics already being collected by hubs may reveal commonalities on which to build. Similarly, existing data at hubs beyond what is needed to compute the Common Metric results might be used to identify key drivers of performance. Lastly, hubs expressed interest in using the findings of this evaluation study to inform future directions of the Common Metrics Initiative.

# Summary Recommendation 1:

Develop metrics using robust pilot testing, and engage stakeholders in ongoing review

**1d** Periodically engage hubs in a review of each metric for completeness, clarity, usefulness, and required effort.

## Summary Recommendation 3:

Maximize usefulness to the National CTSA Consortium by ensuring validity of aggregation and comparison reporting.

**3b** If aggregation or comparison of hubs' metric results is pursued, ensure results are comparable across hubs.

## Summary Recommendation 7:

Sustain engagement by facilitating solutions to barriers due to resources and authority, accounting for hub heterogeneity, and ensuring effective communication.

- **7c** Maintain realistic expectations about the amount of improvement that can be achieved and the pace of change, particularly when the CTSA leader does not have line authority over the target processes.
- 7f Inform hubs of future directions for the Common Metrics Initiative.

# **Summary Recommendation 8:**

# Expand use of data to inform future directions of the Common Metrics Initiative and the CTSA Program.

- **8a** Use hub data beyond what is needed to implement the Common Metrics (e.g., other clinical and operational data) to inform the selection of metrics and to identify potential drivers of outcomes.
- **8b** Use discussion of the Common Metrics evaluation results to catalyze a broader conversation about other high impact research projects to drive data-driven decisions related to the structure of CTSAs and the CTSA Program.

# **Limitations and Future Opportunities**

This mixed method post-test evaluation study provided a multi-faceted understanding of hubs' progress and the related contextual factors, challenges, and facilitators. Still, three notable limitations provide opportunities for additional future study.

First, limitations related to study design reflected trade-offs in balancing competing needs. The goal of achieving the most robust implementation for every hub in the same time frame did not allow for a controlled comparison group design. As a result, this descriptive study was not able to test for causal predictors of variation in completion of performance improvement activities. Our descriptive findings of the importance of organizational variation, such as different data systems, processes, local institutional priorities, and Principal Investigator positions in the home institution's organizational structure, suggest opportunities to further our understanding of organizational "best practices" in relation to Common Metrics and performance improvement. For example, a future study could identify a typology of CTSA organizational structures in order to test their predictive value for implementing data-driven performance management, for Common Metrics and more generally.

Second, although we used both quantitative and qualitative data to provide rich descriptions, an apparent discrepancy between the study results and the experiences of the Tufts Implementation Team remained unexplained. Specifically, as part of the evaluation study, the vast majority of hubs reported that they computed metric results according to the Operational Guideline for all three metrics. However, when the Implementation Team reviewed metric results entered into the Scorecard software, they observed many instances in which the result clearly was not consistent with the Operational Guideline. Some of the participants in qualitative interviews also expressed the opinion of variation in how hubs were collecting data and computing results, and these opinions often were based on their knowledge of disparate systems and approaches.

Because the comparability of metric results across hubs is important for their usefulness at the national level, we explored potential explanations for the discrepancy. The timeframe
of the Implementation Team's review of metric results differed from evaluation study data. The Implementation Team completed its review of metric results in July, 2017. This was before the August, 2017 annual update of Common Metric results (at which time hubs could have changed prior results) and six months prior to the last round of data collection for the evaluation study in January, 2018. As a result, the discrepancy in information between the Implementation and Evaluation components could have been due to a time lag. However, a comparison of dates that hubs provided for computing metric results according to the Operational Guidelines suggested that some hubs believed they were adhering to the Guidelines at about the time that the Implementation Team was finding discrepancies. To address this question of technical alignment with the Operational Guidelines, hub self-report of adherence would require confirmation by an audit of underlying data. Neither the Tufts Implementation Team nor the Evaluation Team audited metric results as of January, 2018 because CLIC had assumed responsibility in the fall, 2017 for leading implementation of, and liaising with hubs about, Common Metrics. A future audit of metric results would inform the continuing question of whether hubs are actually computing metric results as intended.

Finally, the timeframe of the evaluation study was shorter than what would be needed to observe a change in performance on metric results. The outcome we measured– completion of activities related to Common Metrics and performance improvement– provided important insights into organizational factors affecting progress and hub engagement. Future studies could build on these insights to examine predictors of improvement on metric results.

### INTEGRATED SUMMARY AND CONCLUSIONS

The Tufts Common Metrics Implementation Program provided structure and support for implementing Common Metrics across the CTSA Consortium between June, 2016 and December, 2017. The Tufts Common Metrics Evaluation Study developed evidence and insights to assess and reflect on the Common Metrics Initiative. Together, they provided quantitative and qualitative data and insights about the metrics, hub progress on conducting performance improvement activities, challenges and facilitators they faced, and perspectives on the overall value of the Initiative to date. Findings and conclusions speak most directly to future directions of the Common Metrics Initiative, but they can also inform other CTSA Program initiatives and similar networks that plan to embark on implementing shared metrics and performance improvement frameworks. The findings also provide a window into the strengths and challenges of CTSAs that relate to their overall function and impact.

The following sections summarize the main findings and summary recommendations to inform future efforts (Table 42).

### SELECTING AND DEVELOPING METRICS

The first three Common Metrics (Careers in Clinical and Translational Research, Pilot Funding Publications, and Institutional Review Board [IRB] Review Duration) and a fourth metric yet to be implemented (Clinical Trial Median Accrual Ratio) were developed by separate workgroups, which specified the requirements in Operational Guidelines. Pilot tests were conducted for each, focusing on feasibility of data collection for the first three metrics and expanding to feasibility and usefulness for the fourth metric. Both types of pilot testing identified gaps in metric definitions, and the more robust approach taken for the fourth metric also uncovered confusion about computation and challenges to usefulness for strategic management that were important to address before widespread implementation.

Usefulness of the first three Common Metrics was intended for two levels: local CTSA hubs and the national CTSA Consortium. At the local level, the usefulness of the first three metrics varied across hubs for multiple reasons, including: extent of alignment with local institutional priorities, ability of metric definitions to directly inform local improvement, time lags embedded in the metrics for assessing performance, and perceived lack of clarity or stability of metric definitions. Starting with metrics that many hubs

perceived as "low-hanging fruit" for them helped some to prepare for implementing more challenging metrics. However, others reported spending limited resources on areas that did not need improvement. Although hubs experienced some value in implementing the first three metrics, many continued to have concerns about whether the metrics provided enough benefit to justify the required effort.

At the national level of the CTSA Consortium, there were continuing concerns about between-hub variation in how metric data were collected and computed. Lack of comparability would undermine any benefit from comparing or aggregating hub results for benchmarking or other purposes.

### **IMPROVING PERFORMANCE**

By the end of the evaluation study, the vast majority of hubs self-reported that they had computed the metric result and engaged in activities to understand current performance, but there was more variation in developing and implementing performance improvement plans. The experience of implementing the first three metrics identified three types of tools for assisting hubs in carrying out performance improvement work.

First, to equip hubs with the required knowledge and proficiency, Tufts Implementation Team provided training, small-group and individualized coaching, and technical assistance. Participants appreciated the opportunity that coaching sessions provided for peer-to-peer learning, and some hubs valued a structured approach to meeting project milestones. The vast majority of hubs reported that they gained the knowledge and proficiency they needed, or more, to carry out the work of the Common Metrics.

Second, the Tufts Implementation Team and hubs alike identified a range of useful supports. Important supports included metric-specific teams (with subject matter experts, local champions, and strategic partners), involvement of the hub's Principal Investigator, time to build relevant systems and processes prior to reporting, availability of benchmarks or other targets to help hubs identify areas needing improvement, and dissemination of performance drivers and best practices to inform choices about strategies to pursue.

Third, hubs and the Implementation Team recognized the importance of a common software platform to support data collection, reporting, and visualization. To implement the first three Common Metrics, hubs used the Scorecard software. Although the value of a shared reporting platform was understood, hubs did not use all features related to performance improvement strategies. Instead, they requested enhancements to its functionality related to data collection and management, visualization capabilities, and user experience.

### ADDRESSING BARRIERS AND SUSTAINING ENGAGEMENT

Not surprisingly, hubs reporting active engagement in implementing Common Metrics completed more performance improvement activities than those reporting a compliancebased approach. Many factors could affect a hub's level of engagement. By the end of the evaluation period, the most common reason hubs cited for not completing an activity was lack of resources or personnel to devote to Common Metrics. Other common challenges included lack of alignment of the Common Metrics with existing data systems or processes, lack of alignment of Common Metrics with institutional priorities, and a hub leader's lack of line authority over data or processes related to the Common Metrics. The heterogeneity of hub data systems, processes, organizational structures, and institutional priorities meant that experiences and needs differed across hubs.

Regardless of the challenges facing a particular hub or group of hubs, effective communication was key to implementation efforts. Communication strategies were multifaceted during the implementation period, including online postings and in-person events. In terms of content, some hubs indicated a desire for more information about future plans for the Common Metrics Initiative, particularly when weighing its overall value.

### **EXPANDING DATA-DRIVEN DECISION MAKING**

The Tufts Implementation Program and Evaluation Study revealed many structural and cultural aspects of CTSA organizations that affected hubs' abilities to engage with and complete Common Metrics and performance improvement activities. These results suggest two opportunities to expand data-driven decision making in the CTSA Program. First, recognizing opportunities to improve the Common Metrics Initiative, several hubs indicated a desire to learn from other available data and research results. Second and more broadly, insights about structural and cultural factors affecting implementation of Common Metrics likely apply to other current and future initiatives. The importance of these factors suggests a ripe opportunity to address additional questions about how CTSAs operate and strive to move Clinical and Translational Research forward.

Expanding the "science of the conduct of science" approach to governing the CTSA Consortium would address additional questions more deeply. For example, are there "best practices" for CTSA organizational structures and ways of linking to home institutions that expedite performance improvement or other types of initiatives? How did implementation of Common Metrics affect CTSAs and their relationships with home institutions? Are there metrics or processes that are commonly used across hubs that could inform selection of future Common Metrics or other priority initiatives?

### **METRIC-SPECIFIC OPPORTUNITIES TO IMPROVE**

Having implemented and/or piloted four Common Metrics, an important opportunity now exists to review the metrics with hubs to ensure they are useful relative to required effort and address any needed modifications to the Operational Guidelines.

### **First Three Common Metrics**

Implementing the first three metrics and hub input during the evaluation study identified needed clarifications to the Operational Guidelines and questions about the usefulness of these metrics.

#### Metric: IRB Review Duration

Hubs continued to have questions about how to apply the IRB metric's Operational Guideline. Indeed, differences in how hubs applied definitions raised concerns about comparability of metric results across hubs. Developing a strategic management plan in conjunction with the IRB often required overcoming organizational boundaries, which was more difficult when the hub's Principal Investigator did not have line authority over the relevant processes. Usefulness of the IRB metric at the local level varied depending on the number of IRBs, the types of protocols reviewed, and the work process of reviews for ethics, feasibility, and budgets and contracts.

#### **Metric: Pilot Funding Publications**

Hubs identified two main challenges with interpreting and using this metric. First, some hubs reported that the cumulative nature of the metric made interpreting results at the hub level difficult, particularly for hubs with large numbers of pilot awards. Second, some hubs considered the Pilot metric too narrowly focused on publications, which did not fully capture the goals of their local programs.

#### Metric: Careers in Clinical and Translational Science

Hubs varied greatly in the extent to which they had been tracking graduates and their career statuses over time, and if they had, with what method and frequency. Lack of clarity and agreement with elements of the Operational Guideline created concerns about comparability of data across hubs. Additionally, a number of hubs reported that the cumulative nature of these metrics made interpretation of metric results difficult at the hub level. Moreover, some hubs considered the Careers metric too narrow in scope to capture the goals of their local programs.

### **Pilot Test: Clinical Trial Accrual Metric**

Although pilot hubs recognized the importance of improving accrual to clinical trials, all eight hubs that piloted this metric faced numerous challenges collecting the metric data and developing strategic management plans. Use of a clinical trial management system (CTMS) did not mitigate the challenges. Only one of eight pilot hubs was able to assess the piloted accrual ratio for all eligible trials, and all hubs had difficulty creating a central list of trials at their institution to use for a sampling frame. Although most hubs had limited their sampling frames to a group of trials for which they believed metric data were present, they were still not able to determine inclusion/exclusion criteria for all of the trials in their samples. The exclusion criteria of trials with fewer than 10 targeted participants removed many otherwise potentially-eligible clinical trials from the sampling frame. This was particularly the case at smaller primary institutions with smaller pools of clinical trial participants from which to draw, or at those institutions with large numbers of multi-site trials for which they were recruiting a subset of participants. Hubs' concerns about the quality of their data limited the usefulness of the metric for strategic management from their perspectives.

#### Table 42. Summary recommendations

#### \*For a full list of recommendations, see Table 1

#### **Summary Recommendation**

- 1. Develop metrics using robust pilot testing, and engage stakeholders in ongoing review.
- 2. Maximize usefulness to hubs by selecting metrics that align with local needs.
- 3. Maximize usefulness to the National CTSA Consortium by ensuring validity of aggregation and comparison reporting.
- 4. Equip hubs to fully implement each metric and performance management by providing peer-to-peer learning and training, coaching, and assistance for varying levels of experience.
- 5. Support implementation by promoting metric-specific teams, allowing for capacitybuilding periods, providing accurate benchmarks, and updating performance drivers and best practices.
- 6. Maximize usefulness of the reporting platform by enhancing functionality, visualization options, and user experience.
- 7. Sustain engagement by facilitating solutions to barriers due to resources and authority, accounting for hub heterogeneity, and ensuring effective communication.
- 8. Expand use of data to inform future directions of the Common Metrics Initiative and the CTSA Program.
- 9. Make improvements to the first three Common Metrics by clarifying Operational Guidelines and assessing usefulness with hubs.
- 10. Use the results and recommendations detailed in the full Accrual Metric Pilot report to determine the future direction of metric implementation.

### REFERENCES

<sup>1</sup>IOM (Institute of Medicine, now National Academy of Medicine). The CTSA Program at NIH: Opportunities for Advancing Clinical and Translational Research. Washington, DC: The National Academies Press. June, 2013. Accessed from: <u>http://www.nationalacademies.org/hmd/Reports/2013/The-CTSA-Program-at-NIH-Opportunities-for-Advancing-Clinical-and-Translational-Research.aspx</u>.

<sup>2</sup>Friedman M. Trying Hard Is Not Good Enough: How to Produce Measurable Improvements for Customers and Communities. FPSI Publishing, 2005. ISBN: 1-4392-3786-7.

<sup>3</sup>Fetters MD, Curry LA, Creswell JW. Achieving Integration in Mixed Methods Designs—Principles and Practices. Health Services Research, 2013, 48:6 (Part II):2134-56.

<sup>4</sup>Tomoaia-Cotisel, Andrada, Debra L. Scammon, Norman J. Waitzman, Peter F. Cronholm, Jacqueline R. Halladay, David L. Driscoll, Leif I. Solberg et al. "Context matters: the experience of 14 research teams in systematically reporting contextual factors important for practice change." The Annals of Family Medicine 11, no. Suppl 1 (2013): S115-S123.

<sup>5</sup>QSR. 2012. NVivo qualitative data analysis software, Version 10. Melbourne, Australia: QSR International Pty Ltd.

# **Appendices**

## Appendix A. Hub experiences with specific elements of the Tufts Implementation Program

N=59 Hubs		Total	Implei	Implementation Group		
			1 (n=20)	2 (n=17)	3 (n=22)	
Time allocated for:		n(%)		n(%)		
Training on RBA methodology and Common Metrics operational guidelines	Much less than needed Less than needed About what was needed	1 (2) 2 (3) 32 (54)	0 (0) 2 (10) 10 (50)	0 (0) 0 (0) 9 (53)	1 (5) 0 (0) 13 (59)	
	More than needed Much more than needed Do not know	9 (15) 15 (25) 0 (0)	2 (10) 6 (30) 0 (0)	3 (18) 5 (29) 0 (0)	4 (18) 4 (18) 0 (0)	0.534
Coaching sessions with small groups of hubs	Much less than needed Less than needed About what was needed More than needed Much more than needed	1 (2) 4 (7) 36 (61) 10 (17) 8 (14)	0 (0) 3 (15) 12 (60) 2 (10) 3 (15)	0 (0) 0 (0) 11 (65) 4 (24) 2 (12)	1 (5) 1 (5) 13 (59) 4 (18) 3 (14)	0.632
	Do not know	0 (0)	0 (0)	0 (0)	0 (0)	
Support involving your hub only	Much less than needed Less than needed About what was needed More than needed Much more than needed	0 (0) 9 (15) 39 (66) 3 (5) 3 (5)	0 (0) 5 (25) 12 (60) 1 (5) 2 (10)	0 (0) 2 (12) 11 (65) 2 (12) 1 (6)	0 (0) 2 (9) 16 (73) 0 (0) 0 (0)	0.360
Learning session webinars open to all	Much less than needed	5 (8) 3 (5) 5 (8)	1 (5) 3 (15)	0 (0)	2 (9) 2 (9)	
hubs	About what was needed More than needed Much more than needed	35 (59) 6 (10) 8 (14)	2 (10) 2 (10) 2 (10)	2 (12) 4 (24)	2 (9) 13 (59) 2 (9) 2 (9)	0.662
	Do not know	2 (3)	1 (5)	0 (0)	1 (5)	

See Appendix Q for hub overall experiences with the Tufts Implementation Program

### Appendix A, continued. Hub experiences with specific elements of the Tufts Implementation Program

N=59 Hubs		Total	Implei	Implementation Group		
			1 (n=20)	2 (n=17)	3 (n=22)	
Teaching and assistance methods		n(%)		n(%)		
The amount of didactic	Much less than needed	0 (0)	0 (0)	0 (0)	0 (0)	
training involved	Less than needed	4 (7)	3 (15)	1(6)	0 (0)	
	About what was needed	28 (47)	9 (45)	7 (41)	12 (55)	
	More than needed	12 (20)	2 (10)	5 (29)	5 (23)	
	Much more than needed	15 (25)	6 (30)	4 (24)	5 (23)	0.428
	Do not know	0 (0)	0 (0)	0 (0)	0 (0)	
The amount of	Much less than needed	1 (2)	1 (5)	0 (0)	0 (0)	
discussion involved	Less than needed	13 (22)	4 (20)	3 (18)	6 (27)	
	About what was needed	28 (47)	11 (55)	9 (53)	8 (36)	
	More than needed	5 (8)	0 (0)	1(6)	4 (18)	
	Much more than needed	12 (20)	4 (20)	4 (24)	4 (18)	0.459
	Do not know	0 (0)	0 (0)	0 (0)	0 (0)	
The amount of	Much less than needed	0 (0)	0 (0)	0 (0)	0 (0)	
homework assigned	Less than needed	2 (3)	1 (5)	1(6)	0 (0)	
	About what was needed	27 (46)	9 (45)	9 (53)	9 (41)	
	More than needed	14 (24)	5 (25)	3 (18)	6 (27)	
	Much more than needed	14 (24)	5 (25)	4 (24)	5 (23)	0.933
	Do not know	2 (3)	0 (0)	0 (0)	2 (9)	
The amount of personalized help provided to your hub	Much less than needed	1(2)	0 (0)	0 (0)	1 (5)	
	Less than needed	5 (8)	2 (10)	2 (12)	1 (5)	
	About what was needed	44 (75)	14 (70)	14 (82)	16 (73)	
	More than needed	3 (5)	2 (10)	1(6)	0 (0)	
	Much more than needed	2 (3)	2 (10)	0 (0)	0 (0)	0.420
	Do not know	4 (7)	0 (0)	0 (0)	4 (18)	

See Appendix Q for hub overall experiences with the Tufts Implementation Program

## Appendix B. Tufts Implementation Team's interim assessment tool for meaningful application of the Results Based Accountability framework

Criterion	1-Not Meeting*	2-Approaching	3-Meeting	4-Exceeding
1. Team formation	on, training, and sta	keholder involvement		
A core team of at least 2 individuals is formed at hub and members participate in training. Core team trains other hub staff.	No participation in training	A key role did not participate in training. Little or no training of hub staff.	All key roles participated in training. Training conducted for 1-2 hub staff based on hub needs.	Broad participation in training across core team and hub staff
Relevant stakeholders/ metric topic experts are identified and engaged in the Turn the Curve (TTC) process	No stakeholder, topic expert involvement beyond core team	Ad hoc involvement of 1-2 stakeholders/ metric topic experts	Regular involvement of stakeholders/ metric topic experts	Broad involvement of >2 stakeholders/ metric topic experts
2. "How are we d	doing?"			
Common Metric (CM) data are collected and entered to the online Scorecard system	No CM data have been entered	Estimated values for current CM data have been entered	Actual values for current CM data have been entered	Historical values for CM data have been entered
The graph includes a baseline forecast (i.e., that assumes no new actions)	No baseline forecast	A baseline forecast is present but incorrect	A baseline forecast is present	
3. "What is the	story behind the cu	ırve?"		
Identifies the factors/ underlying causes that influence the value of the performance measures (includes factors that are positive and negative, internal and external, current and anticipated)	No or minimal number (e.g., 1-2) of factors identified. No exploration to identify root causes.	Factors are identified but not in a range of categories. Partial exploration to identify root causes.	Factors are identified in a range of categories. Exploration to identify root causes.	Multiple factors are described in multiple categories. Complete exploration to identify root causes.
Factors are prioritized according to strength of influence ("leverage")	Factors are not yet identified	Factors are identified but priorities are incomplete or unclear	Clearly prioritizes factors	Rigorous process for prioritizing factors
4. "Who are part	ners who might hav	e a role to play in turni	ng the curve?"	
Identifies partners who might have a role to play in turning the curve (e.g., people who could affect change)	No partners or roles are identified	Relatively few partners and roles identified	Appropriate number and roles of partners identified	Extensive involvement of partners

## Appendix B, continued. Tufts Implementation Team's interim assessment tool for meaningful application of the Results Based Accountability framework

Criterion	1-Not Meeting*	2-Approaching	3-Meeting	4-Exceeding
5. "What works to	turn the curve?"			
Identifies potential actions (including low-cost and/or no cost actions) related to high priority factors identified in the story behind the curve (SBtC)	No actions identified or a few actions that are not related to priorities in SBtC	Actions present but unclear relationship to priorities in SBtC and/or do not include low-cost or no cost items	Actions and priorities in SBtC are clearly related. Actions include low- cost or no cost items.	Innovative actions identified
Identifies where additional information will be needed to sufficiently understand what works to turn the curve	No description of adequacy of available information	Additional information needed but no plan to gather	All needed information available or plan to gather is described	Identifies and implements plan for additional data
6. "What do we pro	opose to do to turn	the curve?"		
Actions are selected related to high priority factors identified in the story behind the curve	No actions are present or they are not related to high- priority factors	Actions are present but unclear how or whether the actions are related to high-priority factors	Actions are present that are clearly related to high- priority factors	Actions for lesser factors as well as high priority factors are identified
Action plans provide specifications necessary to be implemented (e.g., deliverables, person(s) responsible, necessary resources, and deadlines)	No action plans or action plans are vague	Action plans are insufficiently detailed for implementation	Action plans are clear with amount of detail necessary for implementation	Action plans are clear, strong project management structure in place
7. Selecting perfor curve of a Comm	mance measures (P Ion Metric	Ms) for managing a co	mponent of a strate	egy to turn the
PMs are clearly defined	No PMs defined	Incomplete measure definition	Complete and clear measure definition	Data collection for PMs has begun
PMs have TTC plans	No turn-the-curve plan associated with PMs	Partial TTC plan developed	Full TTC plan developed	Plan for ongoing management developed

\*Score criteria as 0 if unable to assess using Turn the Curve plan, i.e., RBA framework step is blank.



Tal	ble of Contents
Sec	tion 1   Introduction & Background
	What is a Change Package?
	How Was This Change Package Developed?
	Strategic Management Method4
Sec	tion 2   IRB Duration Driver Diagram
Sec	tion 3   Strategies & Examples7
	Help investigators create high-quality IRB applications8
	Eliminate waste and redundancy
	Improve processes based on feedback10
Ref	erences

Tufts CTSI Toffs Clinical and Translational Science Institute Common Metrics Implementation

### Section 1 | Introduction & Background

The <u>Clinical and Translational Sciences Awards (CTSA) Consortium</u>, led by the <u>National</u> <u>Center for Advancing Translational Sciences (NCATS)</u>, is charged with accelerating and improving clinical and translational research. So far, the potential of the CTSA Program is only partially realized. In order to maximize the Consortium's impact, NCATS has implemented the <u>Common Metrics Initiative</u>, which employs a set of common metrics for use in collaborative management based on the principles of the <u>Results-Based</u> <u>Accountability (RBA)</u> framework. The Common Metrics Initiative is using a set of common metrics for three initial topics, IRB Duration, Pilot Funding Publications and Grants, and Careers in Clinical and Translational Research, to help to focus activities as a network and at the individual CTSA Program hubs on making significant improvements in research translation and workforce development. <u>This change package outlines</u> <u>potential strategies for hubs to use as they begin or advance strategic management efforts for IRB Duration</u>.

### What is a Change Package?

A change package is a concise and practical document that includes ideas and inspiration for teams seeking to apply methods to increasing the effectiveness and efficiency of their processes and outcomes. Change packages focus on a specific metric or process, and generally include background material; a summary of evidence or best practices; and specific strategies, tools and examples that can be applied to the work.



## How Was This Change Package Developed?

This initial change package and the Key Driver Diagram within was developed by the Tufts CTSI Common Metrics Implementation Team, funded by NCATS grant UL1TR001064. It was informed by research findings when available, as well as strategies implemented or planned by hubs participating in the Common Metrics Initiative. It will be revised as additional learning surfaces.



### Strategic Management Method

The work of hub teams participating in the Common Metric Initiative is guided by the Results-Based Accountability framework. Developed by Mark Friedman and described in his book *Trying Hard is Not Good Enough*, RBA is used by organizations to improve the performance of their programs or services. RBA starts with ends and works backward, towards means. RBA provides a step-by-step process to get from ends to means. This process is called "Turn the Curve" thinking.

Hubs in the Common Metric Initiative use Scorecard software to enter and graph their common metric results and facilitate their Turn the Curve planning process.











Tufts CTSI Turks Circled and Common Metrics Implementation

### DRIVER: APPROPRIATE AND SUFFICIENT IRB STAFF & **REVIEW COMMITTEES WITH OPTIMIZED WORKLOADS** Rationale: **EXAMPLE STRATEGIES** Proper staffing is crucial to the effective operation of an IRB Tip Sheets to help organizations write Human Regulatory authorities (e.g., Research Protection Program (HRPP) policies and FDA, OHRP) require procedures\* infrastructure support for the IRB, and the development and The HRPP Operations Manual can serve as a dissemination of policies and reference for staff, reviewers and others procedures Revise the review committee meeting schedule and duration (and a presentation by the Wake Forest CTSI about this initiative) \*these reference accreditation standards but may also be helpful models when accreditation is not being sought **DRIVER: IDENTIFY AND ELIMINATE** WASTE AND REDUNDANCY Rationale: **EXAMPLE STRATEGIES** It is important to look beyond symptoms to uncover the true causes of delays An example from the Indiana CTSI of using process mapping to identify potential waste in IRB Minimizing non-value added activities and reducing processes variation can eliminate A tutorial on Cause and Effect diagrams – a rework and bottlenecks and method for conducting a root cause analysis improve satisfaction

Tufts CTSI Turbs Clinical and Turbs Common Metrics Implementation

### DRIVER: IMPROVE PROCESSES BASED ON RESEARCHER AND SYSTEM FEEDBACK



#### Rationale:

Transparent disclosure of IRB duration metrics can help build will for improvement efforts and manage researcher expectations for outcomes

### **EXAMPLE STRATEGIES**

A <u>dashboard</u> of metrics to enable tracking and monitoring of IRB performance

Reporting of Number of Days to IRB Approval

#### Rationale:

Researcher feedback can provide a valuable source of information about ways in which performance can be improved and the researcher community be better served

### EXAMPLE STRATEGIES

Positive and negative <u>feedback</u> elicited from the researcher community

10

Tufts CTSI Turks Circleal and Turks Common Metrics Implementation

<b>R</b> eferences	
Andrews JE, Moore JB, Means P, Weinberg RB. <u>Overhauling an academic medical cent</u> <u>HRPP with simple process changes</u> . J of Clin Res Best Pract. 2014;10(9).	<u>er</u>
Blustein J, Regenstein M, Siegel B, Billings J. <u>Notes from the field: jumpstarting the IF</u> approval process in multicenter studies. Health Serv Res. 2007;42(4):1773-82.	<u>(B</u>
Brassil D, Kost RG, Dowd KA, Hurley AM, Rainer TL, Coller BS. <u>The Rockefeller Universe</u> Navigation Program: a structured multidisciplinary protocol development and education program to advance translational research. Clin Transl Sci. 2014; 7(1):12-19	<u>sity</u> mal
Friedman M. Trying Hard Is Not Good Enough: How to Produce Measurable Improvem for Customers and Communities. FPSI Publishing, 2005. ISBN: 1-4392-3786-7	ents
	11
This CTSI tofa Cinicia and	11



Ta	ble of Contents
Se	ction 1   Introduction & Background
	What is a Change Package?
Se	ction 2   Pilot Funding Publications Driver Diagram5
Se	ction 3   Strategies & Examples7
	Improve pilot team effectiveness and timeliness
Re	<b>ferences</b> 14
	Tufts CTSI Tracticities and Common Metrics Implementation

### Section 1 | Introduction & Background

The <u>Clinical and Translational Sciences Awards (CTSA) Consortium</u>, led by the <u>National</u> <u>Center for Advancing Translational Sciences (NCATS)</u>, is charged with accelerating and improving clinical and translational research. So far, the potential of the CTSA Program is only partially realized. In order to maximize the Consortium's impact, NCATS has implemented the <u>Common Metrics Initiative</u>, which employs a set of common metrics for use in collaborative management based on the principles of the <u>Results-Based</u> <u>Accountability (RBA)</u> framework. The Common Metrics Initiative is using a set of common metrics for three initial topics, IRB Duration, Pilot Funding Publications and Grants, and Careers in Clinical and Translational Research, to help to focus activities as a network and at the individual CTSA hubs on making significant improvements in research translation and workforce development. This change package outlines potential strategies for hubs to use as they begin or advance strategic management efforts for Pilot Funding Publications.

### What is a Change Package?

A change package is a concise and practical document that includes ideas and inspiration for teams seeking to apply methods to increasing the effectiveness and efficiency of their processes and outcomes. Change packages focus on a specific metric or process, and generally include background material; a summary of evidence or best practices; and specific strategies, tools and examples that can be applied to the work.



## How Was This Change Package Developed?

This initial change package was informed by research findings when available, as well as strategies implemented or planned by CTSA hubs participating in the Common Metrics Initiative. It will be revised as additional learning surfaces.



### Strategic Management Method

The work of hub teams participating in the Common Metric Initiative is guided by the Results-Based Accountability framework. Developed by Mark Friedman and described in his book *Trying Hard is Not Good Enough*, RBA is used by organizations to improve the performance of their programs or services. RBA starts with ends and works backward, towards means. RBA provides a step-by-step process to get from ends to means. This process is called "Turn the Curve" thinking.

Hubs in the Common Metric Initiative use Scorecard software to enter and graph their common metric results and facilitate their Turn the Curve planning process.









### Strategies

Starting on the next page are specific examples for a number of the strategies in the driver diagram that may yield to improvement at the level of a CTSA hub. These strategies are organized around the drivers outlined above.

A Strategy answers the question "What are we going to do?" [to Turn the Curve].

--Phil Lee, Clear Impact





### DRIVER: EFFECTIVE PILOT-FUNDED TEAMS COMPLETE PROJECTS IN A TIMELY MANNER



#### Rationale:

Even in high-performing, well-established teams, room for improvement likely exists in areas such as meeting productivity, team communications, and resolving team conflicts.

### EXAMPLE STRATEGIES

UC Irvine's research acceleration and facilitation team (<u>RAFT</u>) proactively assists researchers with timely project execution

Navigators, individuals or teams who are expert in all phases of translational research, can monitor award progress and connect researchers with resources and services. Example Navigator programs at hubs include:

- Indiana CTSI
- <u>Tufts CTSI</u>
- University of Rochester CTSI



### DRIVER: PILOT-FUNDED INVESTIGATORS RECEIVE HIGH-QUALITY MENTORSHIP

Tufts CTSI Turbs Circleal and Translational Science Institute Common Metrics Implementation

#### Rationale:

Mentoring: Beginning and midlevel members are assisted by and collaborate with established scholars.

In clinical and translational science, mentorship is particularly important for helping trainees develop, integrate, and apply skill sets related to basic science research and to clinical research.

### EXAMPLE STRATEGIES

University of Illinois at Chicago CCTS provides a number of <u>mentoring resources</u>, including best practices and tools

Entering Mentoring training materials, developed by the Institute for Clinical and Translational Research at University of Wisconsin-Madison, for use with clinical and translational science award mentors.

The UCSF CTSI <u>Mentor Training Program</u> includes online and in person components designed to train mid-career and early senior faculty to be effective clinical and translational research mentors



"Strong mentorship has been linked to enhanced mentee productivity, self-efficacy, and career satisfaction."

--Pfund, et al

### DRIVER: AWARDEES ACCESS CTSI-SPONSORED RESOURCES AND SERVICES

#### Rationale:

Faculty with minimal scholarship production experience need training and personal support to become proficient.

Writing skills, particularly of junior investigators, limit production of publications (and probably acceptance rates).

Many hub resources and services exist to help investigators plan, conduct and publish their research.

### EXAMPLE STRATEGIES

Please submit strategies for this driver that you are willing to share with the Consortium to the Tufts Implementation team at: <u>TuftsMCCMIHelpdesk@tuftsmedicalcenter.org</u>



### DRIVER: PILOT AWARDS ARE MADE FOR PROJECTS WITH ATTRIBUTES ASSOCIATED WITH HIGHER RATES OF PUBLICATION

#### Rationale:

Research findings resulting from CTSA-supported research make important contributions to the field through publication.

Variables associated with the project or the award itself may make it more or less likely that the funded award results in publication.

### EXAMPLE STRATEGIES

UCLA CTSI has assessed <u>factors associated with</u> <u>return</u> (including publication) on funded pilot awards. UCLA utilized the <u>Longitudinal Scientific</u> <u>Achievement Survey</u> (LSAS) in this study. Originally modelled after the Rockefeller University's Graduate Tracking Survey System, it has undergone significant revision. Currently version 3.0, additional revisions are planned.







### funding in email footer messages University of New Mexico Health Sciences Center provides example citations and answers to Frequently Asked Questions about citing and





To help ensure resources are available for future research, please cite the ICTS grant.



### DRIVER: PUBLICATIONS WITH PILOT FUNDING CITATION ARE IDENTIFIED AND TRACKED

### EXAMPLE STRATEGIES

#### Rationale:

For any research project that received any benefit from the hub, including pilot award funding, it is an NIH requirement to acknowledge the CTSA grant number of the institution providing support in any related publications. This acknowledgement is tracked and reported to the NIH and is a key metric of success.



The web-based electronic <u>Rockefeller University</u> <u>Graduate Tracking Survey System</u> pre-populates a graduate's information, including publications, from public data sources, facilitating survey completion and aggregation of data.

Developed by Weill Cornell CTSC, <u>PROMPTR</u>, a component of WebCAMP, offers the ability to survey pilot awardees about publications, has search and import tools for PubMed publications, and can provide summary & drill-down capabilities to understand the Story Behind the Curve.

The University of Massachusetts CCTS inquires about submitted publications in each <u>Pilot Award</u> <u>Progress Report</u>.

An example root cause analysis using the <u>Five Whys</u> technique to examine why pilot awardees are not reporting their publications

The <u>Pilot Metric Worksheet</u> provides an example of how to calculate the Pilot Funding Publications Common Metric.

Encouraging awardees to register for and use an <u>ORCID persistent digital identifier</u> may improve ease of tracking and associating publications with Pilot funding.

Tufts CTSI Twis Clinical and Translational Science Institute Common Metrics Implementation
### **Appendix D: Common Metrics Pilot Publications change package**





# **Table of Contents** Strategic Management Method ......4 Section 2 | Careers in Clinical and Translational Research Driver Diagram......6 High quality mentorship......9 Career development resources.....10 Facilitation of networking.....11 Scholar research skills.....12 Tracking graduates and career status.....14 References ......15

Tufts CTSI Twis Clinical and Translational Science Institute Common Metrics Implementation

# Section 1 | Introduction & Background

The <u>Clinical and Translational Sciences Awards (CTSA) Consortium</u>, led by the <u>National</u> <u>Center for Advancing Translational Sciences (NCATS)</u>, is charged with accelerating and improving clinical and translational research. So far, the potential of the CTSA Program is only partially realized. In order to maximize the Consortium's impact, NCATS has implemented the <u>Common Metrics Initiative</u>, which employs a set of common metrics for use in collaborative management based on the principles of the <u>Results-Based</u> <u>Accountability (RBA)</u> framework. The Common Metrics Initiative is using a set of common metrics to help to focus activities as a network and at the individual CTSA hubs on making significant improvements in research translation and workforce development. <u>This change package outlines potential strategies for hubs to use as they begin or advance strategic management efforts for Careers in Clinical and Translational <u>Research</u>.</u>

## What is a Change Package?

A change package is a concise and practical document that includes ideas and inspiration for teams seeking to apply methods to increasing the effectiveness and efficiency of their processes and outcomes. Change packages focus on a specific metric or process, and generally include background material; a summary of evidence or best practices; and specific strategies, tools and examples that can be applied to the work.



# How Was This Change Package Developed?

This change package was informed by research findings when available, as well as strategies implemented or planned by CTSA hubs participating in the Common Metrics Initiative. It will be revised as additional learning surfaces.



# Strategic Management Method

The work of hub teams participating in the Common Metric Initiative is guided by the Results-Based Accountability framework. Developed by Mark Friedman and described in his book *Trying Hard is Not Good Enough*, RBA is used by organizations to improve the performance of their programs or services. RBA starts with ends and works backward, towards means. RBA provides a step-by-step process to get from ends to means. This process is called "Turn the Curve" thinking.

Hubs in the Common Metric Initiative use Scorecard software to enter and graph their common metric results and facilitate their Turn the Curve planning process.

An example Turn the Curve plan for the Careers in Clinical and Translational Research Common Metric is provided <u>here</u>.





	Drivers	Charles in
	Drivers	Strategies
		<ul> <li>Provide mentorship for guidance, support, resources, connections, and funding</li> <li>Help develop mentor networks rather than hierarchical</li> </ul>
Common Metric	[]	<ul> <li>dyads*</li> <li>Add a category to faculty profiles that designate which facul have a focus on and expertise with mentoring underrepresented persons (URPs) *</li> </ul>
Improve: •the number and	1. Scholars receive high-quality mentorship	Offer mentorship training to mid-career and early senior faculty     Incorporate implicit bias training into mentor training*
percent of institutional scholars and trainees who completed the KL2 and TL1 programs,		<ul> <li>Ensure that all mentors learn strategies to advise on caree life balance*</li> <li>Identify and select for generic and CTR-specific mentor competencies</li> </ul>
respectively, who are currently engaged in clinical and translational research (CTP)		<ul> <li>Provide monetary support for mentors; acknowledge mentoring toward promotion</li> <li>Develop systems for mentor accountability</li> </ul>
•the number and percent of institutional scholars and trainees		Evaluate and give feedback to mentors
who are women or under-represented persons who completed the KL2 and TL1 programs, respectively, who are currently engaged in CTR.		<ul> <li>Conduct a pre-KL2 application session to provide information about institutional career development, answer questions</li> <li>Provide career development seminars and activities that emphasize necessary pace and steps in progression to independence</li> </ul>
	2. Resources specifically for career development are provided	<ul> <li>Provide scholars with opportunities for external experiences (externships, etc.)</li> <li>Partner with Research Administration and Development Of staff to create systems to identify and communicate funding opportunities</li> <li>Survey or interview program graduates on strategies they found helpful</li> </ul>
*While interventions do not need to be gender- or race- specific, these drivers and strategies may contribute to meeting		<ul> <li>Require scholars to develop/implement a personal professional development plan</li> <li>Provide training in: Negotiation skills, Maintaining work-life balance, Maintaining career trajectory after an absence from academia, and Resilience to respond to career criticism or rejection*</li> </ul>
*While interventions do not need to be gender- or race- specific, these drivers and strategies may contribute to meeting	career development are provided	<ul> <li>staff to create systems to identify and communicate fund opportunities</li> <li>Survey or interview program graduates on strategies th found helpful</li> <li>Require scholars to develop/implement a personal professional development plan</li> <li>Provide training in: Negotiation skills, Maintaining work balance, Maintaining career trajectory after an absence f academia, and Resilience to respond to career criticism rejection*</li> </ul>

	Driver D	jagram - Part 9
	Driver D	lagi alli - 1 al t 2
	Drivers	Strategies
Common Metric Aim Improve: •the number and percent of institutional scholars and trainees who completed the KL2 and TL1 programs, respectively, who are currently engaged in clinical and translational research (CTR). •the number and percent of institutional scholars and trainees who are women or under-represented persons who completed the KL2 and TL1 programs, respectively, who are currently engaged in CTR.	<ul> <li>3. Networking is facilitated between current scholars, alumni, and other successful CTR researchers in their discipline</li> <li>4. Scholars develop research skills</li> <li>5. Scholars receive exposure to and training in team science</li> </ul>	<ul> <li>Host networking seminar for incoming and current scholars and alumni to share suggestions about being a KL2 scholar</li> <li>Provide a Networking Forum for trainees and alumni</li> <li>Hold "Lunch and Learn" or other sessions for K cohort support</li> <li>Provide encouragement or support for scholars to attend national events for networking</li> <li>Assist scholars to identify networking opportunities</li> <li>Develop special interest groups for people from underrepresented backgrounds to promote peer mentorship*</li> <li>Provide opportunities for Information Networking (e.g., helping identify funding opportunities) to increase recruitment and reduce attrition of URPs*</li> <li>Ensure women and URP seminar speakers to represent having a research career*</li> <li>Provide a library of funded grants to help illustrate strong grant writing</li> <li>Provide "pink sheets" or reviewer comments from non-funded grants</li> <li>Conduct grant pacing workshops</li> <li>Host mock study sections to simulate NIH review sessions</li> <li>Partner with the Business School in activities around "pitching" one's research</li> <li>Conduct scholar exit interviews to assess satisfaction with training and skills</li> <li>Utilize a multidisciplinary team model for training and scholar development</li> <li>Provide mentor and mentee training on team science</li> </ul>
*While interventions do not need to be gender- or race- specific, these drivers and strategies may contribute to meeting the Common Metric Aim for women and	6. Recruitment, marketing, and applicant review strategies target more diverse applicants*	<ul> <li>Appoint a review committee diversity advocate to ensure applicant reviews are equitable and free from bias*</li> <li>Ensure diversity (i.e., gender, race, ethnicity) of selection committee members*</li> <li>Partner with school (e.g., Medicine, Engineering) and/or institution-wide Office of Diversity and Inclusion*</li> <li>Recruit at historically black colleges and universities (HBCU). &amp; Hispanic (HSI) and Asian American Native American Pacific Islander-serving institutions (AANAPISI) *</li> </ul>

# Section 3 | Strategies & Examples

# Strategies

Starting on the next page are specific examples for a number of the strategies in the driver diagram that may yield to improvement at the level of a CTSA hub. These strategies are organized around the drivers outlined above.

A Strategy answers the question "What are we going to do?" [to Turn the Curve].

--Phil Lee, Clear Impact





# DRIVER: SCHOLARS RECEIVE HIGH-QUALITY MENTORSHIP



Rationale:

The availability of outstanding mentors has been proposed as an essential means of ensuring a pipeline for training researchers and for recruiting and retaining clinician-scientists. (Feldman, et al)

However, it is very improbable to find a single person who can fulfill all the diverse mentoring needs of another individual.



# EXAMPLE STRATEGIES

The University of Utah CCTS Matrix Mentoring Model moves away from "dyadic mentoring", allowing an individual to create a network where different mentors contribute unique skills and empowers individuals, especially women and URPs, with tools to thrive in academic medicine. A presentation about the model is <u>here</u>; a published article is <u>here</u>.

Designed to create national networking opportunities for underrepresented minorities, the <u>National Research Mentoring Network</u> links skilled mentors from various disciplines with diverse mentees.

University of Illinois at Chicago CCTS provides a number of <u>mentoring resources</u>, including best practices and tools.

<u>Entering Mentoring</u> training materials, developed by the Institute for Clinical and Translational Research at University of Wisconsin-Madison, for use with CTR mentors.

The UCSF CTSI <u>Mentor Training Program</u> is designed to train faculty to be effective clinical and translational research mentors.

A set of competencies for mentors of clinical and translational scholars is outlined <u>here</u>.

The <u>Mentoring Competency Assessment</u> (MCA) skills inventory has been shown to be reliable and valid. Mentoring evaluation forms developed by the University of Wisconsin Institute for CTR are available <u>here</u>.

Tufts CTSI Ture Cloked and Transformed Science Institute Common Metrics Implementation

# DRIVER: RESOURCES SPECIFICALLY FOR CAREER DEVELOPMENT ARE PROVIDED

# **EXAMPLE STRATEGIES**

#### Rationale:

Scholars and trainees reported that their overall interest in a career involving CTR increased as a result of their participation in the CTSA education and training programs.

– <u>Evaluation report</u> of CTSA program

Penn State CTSI requires a personal professional development plan as part of the <u>KL2 application</u>, and in the annual <u>KL2 Progress Report</u>.

The <u>education and career development program</u> at the Oregon Clinical and Translational Research Institute was designed around major inflection points in the research career trajectory.

A number of CTSA institutions are participating in the NIH-funded <u>BEST</u> (Broadening Experiences in Scientific Training) program; an effort to explore ways of improving biomedical career development. Information provided by Emory University about BEST is available <u>here</u>.





# DRIVER: NETWORKING IS FACILITATED BETWEEN CURRENT SCHOLARS, ALUMNI, AND OTHER SUCCESSFUL CTR RESEARCHERS IN THEIR DISCIPLINE

#### Rationale:

Networking can help the scholar or graduate make connections and form collaborations; develop new insights, perspectives, and approaches; and increase visibility and build their CV.



# **EXAMPLE STRATEGIES**

<u>The Network for Women in Science</u> program provides support, guidance, and opportunity for female scientists at Scripps Research Institute.

The University of Pittsburgh developed the <u>Sunrise</u> <u>Series</u>, an early morning venue for women faculty, fellows and students to network across schools and departments.

NCATS has started a <u>LinkedIn page</u> to feature KL2 scholars who have led innovative projects in the areas of workforce development, lifespan, methods and processes, collaboration and engagement, and informatics.

New York University CTSI hosted a <u>Science Café</u> to encourage researchers at the medical center to collaborate with researchers from engineering.

#### Networking

Interacting with others to exchange information and develop professional or social contacts. --Oxford English Dictionary

Tufts CTSI Tufts Clinical and Translational Science Institute Common Metrics Implementation

# DRIVER: SCHOLARS DEVELOP RESEARCH SKILLS

### Rationale:

The ability to develop and maintain a CTR career requires skills to conduct, interpret, evaluate and apply research.

1	Hone Your	6.0
	Research	1.1
~	Skills	.18

# **EXAMPLE STRATEGIES**

The Vanderbilt Institute for Clinical and Translational Research provides a <u>grant pacing</u> <u>workshop</u> to help scholars learn to coordinate all elements of grant preparation.

The Penn State CTSI leverages offerings of the business school to provide scholars a chance to learn to <u>"pitch" their research</u>.



# DRIVER: SCHOLARS RECEIVE EXPOSURE TO AND TRAINING IN TEAM SCIENCE

Tufts CTSI Turbs Circleal and Translational Science Institute Common Metrics Implementation

#### Rationale:

CTR generally requires an interdisciplinary approach. Scholars benefit from team membership, collaborations across interdisciplinary boundaries, and learning to address complex or challenging problems that require integration across multiple disciplines.

# **EXAMPLE STRATEGIES**

The UTMB Institute for Translational Sciences utilizes a <u>Multidisciplinary Translational Team</u> (<u>MTT</u>) model for training and development of translational research investigators. An article about the MTT is <u>here</u>.

Exposure to team science, a key strategy in clinical and translational research, is uneven.

--2012 CTSA National Evaluation <mark>Final Report</mark>



# DRIVER: GRADUATES AND THEIR CAREER STATUS ARE TRACKED OVER TIME

#### Rationale:

Tracking graduates and their current involvement in research will not change the underlying rate at which they are engaged in CTR, but is necessary in order to collect the data for the Careers metric and understand the Story Behind the Curve.



# EXAMPLE STRATEGIES

The web-based electronic <u>Rockefeller University</u> <u>Graduate Tracking Survey System</u> can be used to track the careers and accomplishments of graduates in a comprehensive and standardized manner.

Developed by Weill Cornell CTSC, <u>PROMPTR</u>, a component of WebCAMP, supports alumni surveys in which KL2, TL1 and other training program graduates can be followed over time with questions about current involvement in research.

The <u>Careers Metric Worksheet</u> provides an example of how to calculate the Careers in Clinical and Translational Research Common Metric.

Tufts CTSI Toth Cloked and Transistical Science Institute Common Metrics Implementation



### **Careers in Clinical and Translational Research**

Abedin Z, Biskup E, Silet K, et al. <u>Deriving competencies for mentors of clinical and</u> <u>translational scholars</u>. Clin Transl Sci. 2012;5(3):273-80.

Abedin Z, Rebello TJ, Richards BF, Pincus HA. <u>Mentor training within academic health</u> <u>centers with Clinical and Translational Science Awards</u>. Clin Transl Sci. 2013;6(5):376-80.

Anderson L, Silet K, Fleming M. <u>Evaluating and giving feedback to mentors: new evidence-based approaches</u>. Clin Transl Sci. 2012;5(1):71-7.

Burnham EL, Schiro S, Fleming M. <u>Mentoring K scholars: Strategies to support research</u> <u>mentors</u>. Clin Transl Sci. 2011; 4(3): 199–203.

Byington CL, Keenan H, Phillips JD, et al. <u>A Matrix Mentoring Model that effectively supports</u> clinical and translational scientists and increases inclusion in biomedical research: lessons from the University of Utah. Acad Med. 2016;91(4):497-502.

Dilmore TC, Rubio DM, Cohen E, et al. <u>Psychometric properties of the mentor role</u> <u>instrument when used in an academic medicine setting</u>. Clin Transl Sci. 20103(3):104-8.

Feldman MD, Huang L, Guglielmo BJ, et al. <u>Training the next generation of research</u> mentors: the University of California, San Francisco, Clinical & <u>Translational Science</u> <u>Institute Mentor Development Program</u>. Clin Transl Sci. 2009;2(3):216-221

Fleming M, House S, Hanson VS, et al. <u>The Mentoring Competency Assessment: validation</u> of a new instrument to evaluate skills of research mentors. Acad Med. 2013;88(7):1002-8.

Friedman M. Trying Hard Is Not Good Enough: How to Produce Measurable Improvements for Customers and Communities. FPSI Publishing, 2005. ISBN: 1-4392-3786-7

Meagher E, Taylor L, Probsfield J, Fleming M. <u>Evaluating research mentors working in the</u> <u>area of clinical translational science: a review of the literature</u>. Clin Transl Sci. 2011;4(5):353-8.

Pfund C, House SC, Asquith P, et al. <u>Training mentors of clinical and translational research</u> scholars: a randomized controlled trial. Acad Med. 2014; 89(5):774–782.

Ripley E, Markowitz M, Nichols-Casebolt A, Williams L, Macrina F. <u>Training NIH K Award</u> recipients: the role of the mentor. Clin Transl Sci. 2012;5(5):386-393.

Romanick M, Ng K, Lee G, Herbert M, Coller BS. <u>The Rockefeller University Graduate</u> <u>Tracking Survey System</u>. Clin Transl Sci. 2015;8(4):326-329.





# **Appendix F. Collaborative Learning Session agendas**

## Learning Session 1: August 30, 2016

- Opportunities to Advance Translational Science via the Common Metrics Initiative Redonna Chandler, NCATS
- *Indiana CTSI IRB Process Improvement Project* Brenda Hudson and Joe Hunt, Indiana CTSI

# Learning Session 2: Tuesday, September 27, 2016

- CTSI IRB Duration Quality Improvement Process, Wake Forest CTSI
- *Cause and Effect: Enhancing your Story Behind the Curve* Denise Daudelin, Tufts Implementation Team

## Learning Session 3: Tuesday, November 8, 2016

- *Break It Down to Build It Better* (grant pacing workshops for KL2 and TL1 scholars) Katherine Hartmann Vanderbilt CTSI
- *Strengths and Barriers of Sharing Scorecards across Hubs* Pat Barlow and Lena Swander, University of Iowa
- Share Seamlessly/Steal Shamelessly Shout-out (UC-Irvine's e-mail footer) facilitated by Laura Peterson, Tufts Implementation Team

## Learning Session 4: Tuesday, December 13th, 2016

- Funding Innovation in Clinical and Translational Science: What Predicts Return in a Large Multi-institutional CTSA Pamela Davidson, UCLA CTSI
- IRB Driver Diagram and Change Package Laura Peterson
- *Share Seamlessly/Steal Shamelessly Shout-out* (Penn State Elevator Pitch seminar) Denise Daudelin

# Learning Session 5: Tuesday, January 10, 2017

- Update on Plans for NCATS Program Director Calls Redonna Chandler, NCATS
- The Matrix Model of Mentoring (M<sup>3</sup>) to Improve Clinical and Translational Career Development Carrie Byington and Erin Rothwell, Utah
- Pilot Publications Driver Diagram and Change Package Laura Peterson

# Learning Session 6: Tuesday, February 14, 2017

- Real-time IRB CTSI of Southeast Wisconsin
- CTSI Connect (CTSA Science Café) Deborah Keeling, NYU CTSI
- *New Resources* Laura Peterson:
  - Careers Driver Diagram / Change Package
  - Careers Metric Worksheet
- Revised Operational Guidelines with numerator/denominator statements

# Appendix G. Tufts Implementation Team's decision rules for data quality checks of hub metric results

### **Metric: Pilot Funding Publications and Subsequent Funding**

# of pilot projects expending funds:

- Data must be present for 2015

If more data than 2015 are present:

- No data earlier than 2012 should be included (as this denominator for this metric is cumulative from 2012-2015)

- The number for the first year should be >0

- The number for each subsequent year should increase or stay the same (not decrease) compared to the previous year (metric is cumulative)

- If the number does not increase for any time interval >1 year, determine from the Turn the Curve (TTC) plan or hub if additional pilot funding was put on hold

% Pilots with at least one publication:

- Data must be present for 2015

If more data than 2015 are present:

- No data earlier than 2012 should be included

- The first year the % is reported should be "0" or very low, as pilots first expending funds in 2012 are unlikely to have published in 2012. If a hub had many pilots that started in 2011 and also expended funds in 2012, the rate of publications could possibly be higher but should be confirmed with the TTC plan or the hub directly.

- The rate of publications should not decrease to "0".

% Pilots with at least one subsequent research award (optional metric score):

- Data may be present for 2015 (or 2012-2105)

If more data than 2015 are present:

- No data earlier than 2012 should be included

- The first year the % is reported should be "0" or very low, as pilots first expending funds in 2012 are unlikely to have achieved additional funding in 2012. If a hub had many pilots that started in 2011 and also expended funds in 2012, the rate of subsequent funding could possibly be higher but should be confirmed with the TTC plan or the hub directly.

- The rate of subsequent funding should not decrease to "0".

As an additional rule, the absolute number of research publications (the number of pilots expending funds times the % with at least one publication) cannot go down from a previous year, as the metric is cumulative.

### **Metric: IRB Review Duration**

This Common Metric requires one metric score, the median number of days of IRB review duration (i.e., "IRB Turnaround Time").

Rules:

- Data must be present for 2015
- Data may be present for 2012-2014
- Range: greater than 0, less than 100 (values will be revised after 2015 data is submitted)

# Appendix G, continued. Tufts Implementation Team's decision rules for data quality checks of hub metric results

### **Metric: Careers in Clinical and Translational Science**

This Common Metric can yield up to 12 scores, depending on whether a hub has a TL1 program in addition to their KL2. For each:

# and % of graduates currently engaged in CTR:

- Data must be present for 2015
- No data earlier than 2012 should be included (metric denominator is cumulative from 2012-2015)

# and % of URPs:

- Data must be present for 2015
- No data earlier than 2012 should be included
- In any given year, # of URPs currently engaged cannot exceed total # of grads currently engaged

# and % of women:

- Data must be present for 2015
- No data earlier than 2012 should be included

- In any given year, the # of women currently engaged should not exceed the total # of graduates currently engaged

As an additional rule for the Careers metric:

-The total # of URPs engaged in CTR and total # of women engaged in CTR should likely not be the same. While it is possible, women should not be counted as URPs unless they are also an URP for a reason that is \*not\* their gender.

# **Appendix H. Transition timeline**

# Timeline for transition of Common Metrics activities from Tufts Implementation Team to the Center for Leading Innovation and Collaboration (CLIC)

	2017					2018	
	Jul Aug Sep Oct Nov Dec					Jan	
Communication Plan							
CLIC announced as new Coordinating Center	7/15						
Tufts provides message on technical assistance transition							
Manual of Operations							
Tufts provides Manual of Operations			9/1				
Scorecard System							
CLIC obtains additional licenses and administrator rights							
CLIC staff complete advanced Scorecard training							
CLIC provides hub user Scorecard technical support							
Accrual Metric Pilot Test							
Tufts conducts Accrual metric pilot test, prepares report							
CLIC shadows Tufts Implementation Team							
Data Cleaning and Reporting for Metrics 1-3							
Tufts provides CLIC with data cleaning materials, data cleaning spreadsheet, reports of hub status of data entry, report of ranges for metric values							
Data for 2016 due from hubs to Scorecard		8/31					
CLIC conducts data cleaning for 2016 data							
Technical Assistance via Website and Help Desk							
Tufts provides TA to hubs							
Tufts provides copies of materials (e.g., worksheets, Operational Guidelines, change packages, training materials, FAQs, Learning Collaborative materials) to CLIC				10/2			
"Go-live" cut-over from Tufts to CLIC as primary Common Metrics website						12/1	
CLIC provides TA to hubs							

Denotes activities to be performed by CLIC with Tufts available for questions

# Appendix I: Self-reported lack of completion, by activity

### All metrics combined

Activity	Eligible Activities <sup>*</sup>	Completions	Non-Completions	
	N	n	n	%
Creating metric result				
Began collecting data	177	176	1	0.6
Computed result according to Operational Guideline	176	173	3	1.7
Understanding current performance				
Compared result	177	148	29	16.4
Created forecast	177	113	64	36.2
Specified underlying reasons	177	166	11	6.2
Included hub -or- others in specifying underlying reasons	166	166	0	0.0
<ul> <li>Included hub leaders/faculty/staff in specifying reasons</li> </ul>	166	164	2	1.2
<ul> <li>Included external stakeholders in specifying reasons</li> </ul>	166	153	13	7.8
Developing performance improvement plan				
Developed performance improvement plan	177	136	41	23.2
Included hub -or- others in developing improvement plan	136	136	0	0.0
<ul> <li>Included hub leaders/faculty/staff in developing plan</li> </ul>	136	135	1	0.7
<ul> <li>Included external stakeholders in developing plan</li> </ul>	136	127	9	6.6
Specified actions	136	130	6	4.4
Prioritized actions	130	117	13	10.0
When prioritizing actions, considered feasibility or effectiveness <sup>…</sup>	117	116	0**	0.0
Implementing performance improvement pla	an			
Asked partners for help in carrying out improvement plan	136	118	18	13.2
Began to implement improvement plan	136	122	14	10.3
Documenting fully				
Documented five aspects of process	133	107	26	19.5

<sup>\*</sup> For the three metrics combined, each hub could complete an activity three times (59 hubs x 3 metrics = 177 potential activities). Eligible activities reflect skip patterns in the survey. For example, one hub did not collect data on one metric and therefore was not "eligible" to compute the result for that metric.

"Data for "when prioritizing actions, considered feasibility or effectiveness" are missing for one metric at one hub.

# Appendix I, continued: Self-reported lack of completion, by activity

### **Careers metric**

Activity	Eligible Activities*	Completions	Non-Completions		
	Ν	n	n	%	
Creating metric result					
Began collecting data	59	59	0	0.0	
Computed result according to Operational Guideline	59	59	0	0.0	
Understanding current performance					
Compared result	59	46	13	22.0	
Created forecast	59	40	19	32.2	
Specified underlying reasons	59	56	3	5.1	
Included hub -or- others in specifying underlying reasons	56	56	0	0.0	
<ul> <li>Included hub leaders/faculty/staff in specifying reasons</li> </ul>	56	56	0	0.0	
<ul> <li>Included external stakeholders in specifying reasons</li> </ul>	56	52	4	7.1	
Developing performance improvement plan					
Developed performance improvement plan	59	47	12	20.3	
Included hub -or- others in developing improvement plan	47	47	0	0.0	
<ul> <li>Included hub leaders/faculty/staff in developing plan</li> </ul>	47	47	0	0.0	
<ul> <li>Included external stakeholders in developing plan</li> </ul>	47	44	3	6.4	
Specified actions	47	45	2	4.3	
Prioritized actions	45	42	3	6.7	
When prioritizing actions, considered feasibility or effectiveness	42	42	0	0.0	
Implementing performance improvement plan	ı				
Asked partners for help in carrying out improvement plan	47	41	6	12.8	
Began to implement improvement plan	47	42	5	10.6	
Documenting fully					
Documented five aspects of process	46	37	9	19.6	

\*Eligible activities reflect skip patterns in the survey. For example, if a hub did not collect data on a metric, it was not "eligible" to compute the result for that metric.

# Appendix I, continued: Self-reported lack of completion, by activity

# **IRB Review Duration metric**

Activity	Eligible Activities*	Completions	Non-Con	ompletions	
	N	n	n	%	
Creating metric result					
Began collecting data	59	58	1	1.7	
Computed result according to Operational Guideline	58	56	2	3.4	
Understanding current performance					
Compared result	59	54	5	8.5	
Created forecast	59	36	23	39.0	
Specified underlying reasons	59	53	6	10.2	
Included hub -or- others in specifying underlying reasons	53	53	0	0.0	
<ul> <li>Included hub leaders/faculty/staff in specifying reasons</li> </ul>	53	51	2	3.8	
<ul> <li>Included external stakeholders in specifying reasons</li> </ul>	53	50	3	5.7	
Developing performance improvement plan	I				
Developed performance improvement plan	59	41	18	30.5	
Included hub -or- others in developing improvement plan	41	41	0	0.0	
<ul> <li>Included hub leaders/faculty/staff in developing plan</li> </ul>	41	40	1	2.4	
<ul> <li>Included external stakeholders in developing plan</li> </ul>	41	40	1	2.4	
Specified actions	41	38	3	7.3	
Prioritized actions	38	34	4	10.5	
When prioritizing actions, considered feasibility or effectiveness	34	34	0	0.0	
Implementing performance improvement p	lan				
Asked partners for help in carrying out improvement plan	41	37	4	9.8	
Began to implement improvement plan	41	36	5	12.2	
Documenting fully					
Documented five aspects of process	39	32	7	17.9	

\*Eligible activities reflect skip patterns in the survey. For example, one hub did not collect data on one metric and therefore was not "eligible" to compute the result for that metric.

# Appendix I, continued: Self-reported lack of completion, by activity Pilot Funding Publications metric

Activity	Eligible Activities*	Completions	Non-Completions		
	N	n	n	%	
Creating metric result					
Began collecting data	59	59	0	0.0	
Computed result according to Operational Guideline	59	58	1	1.7	
Understanding current performance					
Compared result	59	48	11	18.6	
Created forecast	59	37	22	37.3	
Specified underlying reasons	59	57	2	3.4	
Included hub -or- others in specifying underlying reasons	57	57	0	0.0	
<ul> <li>Included hub leaders/faculty/staff in specify- ing reasons</li> </ul>	57	57	0	0.0	
<ul> <li>Included external stakeholders in specifying reasons</li> </ul>	57	51	6	10.5	
Developing performance improvement plan					
Developed performance improvement plan	59	48	11	18.6	
Included hub -or- others in developing improvement plan	48	48	0	0.0	
<ul> <li>Included hub leaders/faculty/staff in devel- oping plan</li> </ul>	48	48	0	0.0	
<ul> <li>Included external stakeholders in developing plan</li> </ul>	48	43	5	10.4	
Specified actions	48	47	1	2.1	
Prioritized actions	47	41	6	12.8	
When prioritizing actions, considered feasibility or effectiveness <sup>…</sup>	41	40	0**	0.0	
Implementing performance improvement plan					
Asked partners for help in carrying out improvement plan	48	40	8	16.7	
Began to implement improvement plan	48	44	4	8.3	
Documenting fully					
Documented five aspects of process	48	38	10	20.8	

\*Eligible activities reflect skip patterns in the survey. For example, if a hub did not collect data on a metric, it was not "eligible" to compute the result for that metric.

\*\* Data for considered feasibility or effectiveness when prioritizing actions are missing for this metric at one hub.

Activity*	Eligible	Non-	Completions	Reasons (s)	Count**	Illustrative Quotation(s)
	N	n	%		n	
Collected data	177	1	0.6	Accurate data not available	1	"We are awaiting final implementation of our clinical trials management system before reporting on this metric."
Computed result according to Operational Guideline	176	3	1.7	Metric definition not relevant/useful	2	"Some of the specific types of research categories identified in the Operational Guidelines are not tracked by the IRB's electronic research application system so the Operational Guidelines could not be followed to the letter."
			-	In process	1	"We plan to compute the metric, but this is the third priority of getting our Common Metrics work initiated"
Compared result	177	29	16.4	Accurate data not available	22	"We are interested in benchmarking this metric; however, we have yet to identify or access good comparison data."
						"We are a young CTSA, have no prior data to compare to. Have not seen any data from other hubs."
				Not necessary/ not pursued	4	"We have not yet shared these results outside of our institution, other than informal discussions with other hubs and formal presentation to our EAC."
				Lack of authority	1	"To our knowledge, these data have not been shared outside of our hub. Data were provided to the Office of Research Integrityto allow for comparison with prior time periods, but it is uncertain whether this has occurred."
			-	Lack of time/ resources	1	"The IRB office has been staffed at ~65% throughout including a vacancy at the IRB Director position."
			-	Metric not relevant/not useful	1	"The Common Metric was measured differently than our standard education tracking, making comparisonsfrom prior time periods difficult."
Created forecast	177	64	36.2	Lack of confidence in forecast	15	"There are a number of important developments at the national level (e.g., the common rule, single IRB) that will have a yet-to-be-determined impact on IRB review times, so any forecast of this metric will be surrounded by considerable uncertainty."
						"Forecasting for this metric is difficult. We cannot predict the number of papers that will be published because it varies by the research being done."
						"It is hard to forecast if former trainees will enter CTS careers or remain in those careers. The changing job markets and demands on researchers make it hard to speculate regarding job placement."
				Not necessary/ not pursued	14	"Our focus has been on quality improvement processes and we have not yet used the data for forecasting."
						"Our leadership has not embraced the forecasting methodology."

Activity*	Eligible	Non	-Completions	s Reasons (s)	Count**	Illustrative Quotation(s)
	N	n	%		n	
Created forecast, continued	Created forecast, continued		Metric not relevant/not useful	13	"The NCATS Common Metric is just one time point in the whole protocol approval timeline, and we use a more holistic approach to forecasting from the date of protocol submission (which precedes IRB submission) to full protocol approval."	
						"To have the data collected and presented in a cumulative fashion is not useful for us to forecast future results. We use the data in an annual format in order to make comparisons across years and thus forecast future results."
				Accurate data not available	9	"We do not have enough data points to make accurate predictions at this time."
				"So far, we have only had one round of CTSA funded pilots that have completed their awardsThese pilots still need some time to yield publications."		
				Improvement on metric result not needed	5	"We have consistently met our target goals for subsequent pubs and funding by our pilot awardees. The metrichas not resulted in any changes to future forecasts, i.e. we seek to maintain our current success rate."
			Lack of time/ resources	5	"We are currently on a no-cost extension, waiting on NOA and have not started a new round of pilots since collecting the metric."	
		Lack of authority	3	"CTSI has no authority over the Human Subjects Research Office. CTSI has to work collaboratively and defer to their priorities and timeline."		
				In process	2	"While we have collected this data throughout our CTSA funding, we have not fully implemented the Common Metrics into a comprehensive evaluation frameworkThis is our primary activity now that we have been refunded."
				Used other data sources	2	"We do have results from our non-CTSA funded pilots that we are using to set realistic targets for our CTSA-funded pilots."
Specified underlying reasons	177	11	6.2	Improvement on metric result not needed	6	
				Lack of time/ resources	6	(closed-ended response categories- quotations not available)
				Difficulty engaging stakeholders	5	- -
				Accurate data not available	1	"Computation is not completed."
				In process	1	"We are in start-up mode, so re-establishing our programs has taken considerable time."
				Lack of authority	1	"Our HSD is independent of our hub and we do consult."

Activity*	Eligible	Non-O	Completions	Reasons (s)	Count**	Illustrative Quotation(s)						
	N	n	%		n							
Developed improvement plan	177	41	23.2	Improvement on metric result not needed	20	"Because of our structure, we cannot improve our measure any better than a median 14 days"						
				Lack of time/	15	"Hub is currently in a no cost extension period."						
					resources		"Efforts have been focused on filling vacant positions, with the top priority of hiring an IRB DirectorOnce the office is somewhat stabilized in terms of staffing, we will be able to refocus efforts toward the review duration metric"					
				Difficulty engaging stakeholders	7	(closed-ended response category- quotations not available)						
					Metric not relevant/no useful	Metric not relevant/not useful	5	"This metric has not added value to our continuous improvement efforts. We have a variety of 'flavors' of pilot programs, and we use a wide variety of metrics and indicatorsdepending on the flavor."				
								"The operational definition only includes NCATS-funded scholars; this underestimates the success of the program due to not including institutionally-funded scholars."				
									"Metric, as defined, does not reflecthub performance. In cumulative reporting, the denominator increases each year while the relative amount of time to attain subsequent fundingdecreases, which inevitably results in a misleading downward curve."			
				-	Accurate data not available	4	"We have yet to report on this metric, hence we have not developed a formal improvement plan as defined by the CMI [Common Metrics Implementation]. We do have improvement plans in place that were developed using other means."					
							"Pending implementation of a new electronic IRB system for additional metrics."					
				In process	3	"It will result in a plan if appropriate; we are simply not at that point in development given our start-up at this time."						
			-	Not necessary/	3	"Focus is on parallel processing."						
				not pursued	not pursued	not pursued	not pursued	not pursued	not pursued	not pursued		"We are consistently working on improving publication tracking, but it is not specifically related to this metric."
						"This metric is closely monitored by institutional leadership, as the institution is evaluating the results of performance improvement measures taken within the IRB."						
				Lack of authority	1	"Lack of control over our IRB."						
Specified actions	136	6	4.4	Lack of time/ resources	3	"We did not have sufficient time to operationalize the plan in order to derive desired outcomes."						
				Misunderstood question	2	"I'm not sure what you mean by performance improvement plan - do you mean the turn the curve plans?"						
						"I actually don't know. We have actions in the TTC [Turn the Curve] plan, but I don't know if this is the same."						
				In process	1	"Substantive discussion specifically around these issues have only recently been developing."						

Activity*	Eligible	Non-Co	mpletions	Reasons (s)	Count	Illustrative Quotation(s)
	N	n	%		n	
Specified actions, continued				Not necessary/ not pursued	1	"Recommendations were entered global."
Prioritized actions	130	13	10.0	Not necessary/ not pursued	7	"The simplicity of our performance improvement plan didn't require us to assign priorities to different actions."
						"We are not intending to use Scorecard as our program management tool. We have other systems currently in place"
				In process	2	"Still in process of prioritizing for next grant year"
						"The implementation of the performance improvement plan hasbeen on hold until a private vendor with which our hub is partnering was identified in January 2018."
			Lack of authority	Lack of authority	2	"I actually don't know this because this is the IRB We don't manage the IRB."
				Lack of time/ resources	2	"in a no cost extension periodWhile we outlined action plans to improve this metric, we did not prioritize them as we await decision on our reapplication."
						"Time constraints have kept us from getting to this point of the process for this metric."
				Accurate data not available	1	"Not really a plan to improve performance as related to Pilot funding outcomes, versus getting more information from pilot awardees to confirm outcomes."
				Predetermined priority	1	"Institutional direction and priorities determined that conversion to e-IRB was the priority activity."
Asked partners for help in carrying out improvement plan	136	18	13.2	Partners are internal	9	"Our current improvement plan includes key stakeholders who are part of the hub. The hub has not needed to reach out to others at this point."
				Lack of time/	4	"Lack of time to cooperate by key individuals."
				resources		"Resources were fully devoted to e-IRB implementation."
				Not necessary/ not pursued	2	"Did not need their assistance"
				Improvement on metric result not needed	1	"This metric does not require significant improvement at our hub. So we are not devoting extraordinary effort into improving it."
				In process	1	"In process."
				Lack of authority	1	"We don't manage the IRB"
				Don't know	1	"I don't know."

Activity*	Eligible	Non-	Completions	Reasons (s)	Count**	Illustrative Quotation(s)
	N	n	%		n	
Started to implement improvement plan	136	14	10.3	Lack of time/ resources	16	"IRB staffing deficit - they are necessary for process."
				Difficulty engaging stakeholders	6	(closed-ended response category- quotations not available)
				Not necessary/ not pursued	2	"No new plan enacted, no changes as a result of the metric findings."
				In process	1	"Substantive discussion are just now getting organized."
				Lack of authority	1	"We don't manage the IRB."

 $^{\star}\mbox{Only}$  activities not completed by at least one hub for one metric are included.

\*\*More than one reason could be given for each activity not completed.

Overall (N=241 Reasons*)			By Activity				
Reason	n	%	Activity	Total	Careers	IRB	Pilot
Lack of time/resources	52	21.6	Began to implement	n 16	2	7	7
				15	Л	Λ	7
			Specified underlying reasons	6		ד 2	, 2
			Created forecast	5	, 2	2	- 1
			Asked partners for help in carrying out improvement plan	4	-	2	2
			Specified actions	3	-	3	-
			Prioritized actions	2	1	-	1
			Compared result	1	-	1	-
Accurate data not available	38	15.8	Compared result	22	10	2	10
			Created forecast	9	3	3	3
			Developed improvement plan	4	1	2	1
			Prioritized actions	1	-	-	1
			Began collecting data	1	-	1	-
			Specified underlying reasons	1	-	-	1
Improvement on metric	32	13.3	Developed improvement plan	20	7	8	5
result not needed			Specified underlying reasons	6	2	3	1
			Created forecast	5	2	2	1
			Asked partners for help in carrying out improvement plan	1	1	-	-
Activity not necessary /	30	12.4	Created forecast	14	5	2	7
not pursued			Prioritized actions	4	1	-	3
			Compared result	4	1	2	1
			Developed improvement plan	3	-	2	1
			Asked partners for help in carrying out improvement plan	2	2	-	-
			Started to implement improvement plan	2	2	-	-
			Specified actions	1	-	1	-
Metric not relevant/not	21	8.7	Created forecast	13	3	3	7
useful			Developed improvement plan	3	1	1	1
			Computed result according to Operational Guideline	2	-	2	-
			Specified underlying reasons	2	-	-	2
			Compared result	1	1	-	-

# Appendix K: Reasons for lack of completion, by reason

# Appendix K, continued: Reasons for lack of completion, by reason

Overall (N=241 Reasons*)			By Activity				
Reason	n	%	Activity	Total n	Careers n	IRB n	Pilot n
Difficulty engaging	18	7.5	Developed improvement plan	7	-	5	2
stakeholders			Started to implement improvement plan	6	-	2	4
			Specified underlying reasons	5	-	4	1
Lack of confidence in forecast	15	6.2	Created forecast	15	3	8	4
In process	12	5.0	Developed improvement plan	3	2	1	-
			Created forecast	2	2	-	-
			Prioritized actions	2	1	-	1
			Asked partners for help in carrying out improvement plan	1	-	-	1
			Started to implement improvement plan	1	1	-	-
			Specified actions	1	1	-	-
			Specified underlying reasons	1	1	-	-
			Computed result according to Operational Guideline	1	-	-	1
Lack of authority	10	4.1	Created forecast	3	-	3	-
			Prioritized actions	2	-	2	-
			Developed improvement plan	1	-	1	-
			Asked partners for help in carrying out improvement plan	1	-	1	-
			Started to implement improvement plan	1	-	1	-
			Specified underlying reasons	1	-	1	-
			Compared result	1	-	1	-
Partners are internal	rtners are internal 9 3.7 a		Asked partners for help in carrying out improvement plan	9	2	1	6
Used another data source to forecast	2	0.8	Created forecast	2	-	1	1
Priority of actions was 1 predetermined		0.4	Prioritized actions	1	-	1	-
Don't know	1	0.4	Asked partners for help in carrying out improvement plan	1	1	-	-

\*More than one reason could be given for each activity not completed.

# Appendix L. Combined challenges for implementing Common Metrics and performance improvement activities

### Challenge with Illustrative Quotation(s)\*

#### Funding

Lack of institutional investment  $^{\dagger}$ 

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're 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. – Principal Investigator\*\*

### Interrupted funding

...[G] iven our no-cost extension status, ...we don't 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. - Implementer

#### Data system resources

Lack of data system or an existing system that was not aligned with the Common Metrics definitions created more effort for effective tracking<sup>†</sup>

...we never implemented a system to collect [publications] so we have to go back one by one and contact everyone who received a pilot and find out if they published. – Administrator

...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. – Principal Investigator

### Personnel

### **Competing priorities**

From the standpoint where you have to divert effort to comply with new mandates, that does impede progress on the Common Metrics because some of the same people...are now tasked to do these other things. – Principal Investigator

Any time the IRB is contemplating changes, their attention and personnel are deflected from their day-to-day work... [T]here is so much changing with the Common Rule and everything, I think their personnel were distracted, especially their leadership. – Principal Investigator\*\*

### Lack of adequate staffing $^{\!\dagger}$

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 doesn't have the capability to do this herself independently. This is where it all kind of breaks down. Nobody really thought about what impact it was going to have on the time allocation for the leadership that was responsible for evaluation when this concept of Turning the Curve was unleashed. – Principal Investigator

...[A] lot of lip service is given to the importance of evaluation, but resources aren't [provided for] folks who actually support it. ...So having two tenure-track faculty members and a PhD staff member spending time on [the Common Metrics] when there are other people who could collect it, like the [masters level] individual if she had full time working at this, where we could oversee—that would have helped. - Implementer\*\*

# Appendix L, continued. Combined challenges for implementing Common Metrics and performance improvement activities

### Challenge with Illustrative Quotation(s)\*

Lack of evaluation and other specific expertise  $^{\dagger}$ 

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. -Administrator

[With staff turnover,] ...that historical knowledge is lost every time. And so for us, getting everybody up to speed every time a new person comes onboard has been... a big issue. – Administrator

### Maintaining a higher level of engagement

Annual reporting cycle induces bursts of effort

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's not good as a management tool. [In order] to use them in terms of Turning the Curve plans, you really should be generating Turn the Curve plans on a much more frequent basis. - Principal Investigator

### Interrupted funding

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... ...We are only working with the IRB as a collaborative partner and trying to push our agenda...get them to implement some of our suggested action plans. -Implementer

Reduced motivation due to lack of alignment with existing processes or unclear definitions

...[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. And we did it because it's cooperating with the Common Metrics of the national consortium. But ... if I was to sit here and take the amount of time that was spent on that versus the amount of time that we could spend on something else, I'm not sure I would necessarily say it was time well spent. – Administrator

### Availability of accurate data

Lack of alignment with existing data systems<sup>†</sup>

...The [first] IRB is separate from the [second] IRB. And while we got most of the data from the [second] IRB,...the availability of data is different in the two IRBs and so we had to do a little bit more of estimating based on the [first] IRB's data. So, you know, had those databases been linked together under one system, that would have made it easier for us to do this comprehensive approach. – Principal Investigator

Difficulties with effective tracking due to the nature of the topic

Careers, you know, once your KL2 Scholars are out in the world, especially if they've left your institution, it can be challenging to track them down... – Administrator

# Appendix L, continued. Combined challenges for implementing Common Metrics and performance improvement activities

### Challenge with Illustrative Quotation(s)\*

Lack of line authority for needed data

One issue with the CTSAs, particularly in a decentralized organization like ours, is we're responsible for outcomes but don't have authority over them. It's an exercise I'm trying to lead from the middle. – Principal Investigator

### Starting with "low-hanging fruit"

Spending limited resources on areas that did not need improvement was not  $helpful^{\dagger}$ 

If it was an issue, it would be addressed. But, doing a Turn the Curve [plan] to say, "Hey, I recommend you try to get a little bit tighter or get a day better," I don't think that would be a good use of time because we have other groups that maybe aren't in the Common Metrics that need more help in other areas. – Administrator

### Metric clarity and usefulness

Changing metric definitions and clarifications created duplicative work  $^{\dagger}$ 

We redid the Pilot stuff five, six, seven times, to get it to exactly what Tufts wanted it to look like, and it just was incredibly frustrating. It took me out of my normal job and this is added work that was not anticipated or budgeted for. – Implementer

Lack of usefulness for local improvement

So, we're really looking at the questions that we want to ask rather than barely reporting on the Common Metrics. And we don't have anything particularly against the Common Metrics. It's just we have an opportunity here with other people that are tracking folks to bring it all together to allow us to ask what we consider to be better questions [than the Common Metrics]. – Administrator

Time lag for assessing current performance

[T]he percentage of graduates who are doing clinical translational science, that's a curve that is not going to turn quickly at all. I mean, it's probably, it's years from action to seeing the result. -Principal Investigator

## Lack of alignment with institutional priorities<sup>+</sup>

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're 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. – Principal Investigator

# Appendix L, continued. Combined challenges for implementing Common Metrics and performance improvement activities

### Challenge with Illustrative Quotation(s)\*

#### **Engaging stakeholders**

Lack of a direct line of communication within institution

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. – Principal Investigator<sup>\*\*</sup>

Securing needed buy-in or cooperation from key stakeholders

I think there's probably been more resistance or pushback or concern about quantifying that [IRB metric] just because people at the IRB—and rightfully so—believe that they're not the only ones responsible for the turnaround time. – Principal Investigator

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's a constant struggle... – Principal Investigator

### Hub authority and control

Lack of line authority over key drivers

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? – Principal Investigator

Metric topics, particularly IRB Duration, are part of complex processes

...it's complicated. I think these are good metrics to assess. It's just difficult sometimes to understand what all of the causal factors are and whether we can directly influence those factors. – Principal Investigator

...if an investigator submits an IRB protocol, if the IRB responds quickly and gets them back and asks for minor changes, if the investigator sits on those minor changes for six months, I can't stop that. So that's going to look really bad on those metrics, and we have nothing to do with that. ...So, I think we all take our licks. I mean, there's always going to be protocols that will be delayed which will affect this number. – Principal Investigator

#### **Tufts Implementation Program**

Lack of use by some hubs<sup>†</sup>

[T]hat [reaching out to Tufts for individual support] might have been useful. And maybe that's there, but I just didn't reach out. - Implementer

\* 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.

<sup>&</sup>lt;sup>†</sup>Indicates that the challenge has a corresponding facilitator.
## Facilitator with Illustrative Quotation(s)\*

#### Funding

Availability of institutional resources<sup>†</sup>

Our evaluation group actually also has institutional funding.... All centers and other programs now—or many, I should say—require an evaluation piece. So from a school-level we created ... a core resource. – Principal Investigator

... 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. – Principal Investigator

#### Data system resources

Alignment of Common Metrics with and ability to use existing data collection tools<sup>†</sup>

[Existing data collection tools helped] to start to track that data. - Administrator

With our K scholars we've always had constant communication with them, but that was already established. We have a social media network that's set up specifically for them; that's been set a couple years now. So, I think that helped us in being able to stay in contact with them easier. – Administrator

#### Personnel

Adequate evaluation and other specific expertise<sup>†</sup>

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. – Principal Investigator

We work closely with our IRB, and we have a member of our team who worked in the IRB for about eight years. She had really deep knowledge of that system. She was very important, and also she has technical knowledge. – Implementer

#### Leveraging extended teams<sup>†</sup>

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'm 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 I created a parallel group of advisers, people who were interested in the Common Metrics. – Implementer

# Facilitator with Illustrative Quotation(s)\*

## Availability of accurate data

Alignment of Common Metrics with and ability to use existing data collection tools  $^{\dagger}$ 

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... – Principal Investigator

Capacity to build an aligned data system for long-term use will create future efficiencies

So, one of the things that was helpful with...working on the Common Metrics...was figuring out what kind of an electronic system we could put in place to track how much time it takes for a protocol from initiation of an idea to development of the protocol to submission to the IRB to approval, and how do you track that electronically. – Administrator

# Starting with "low-hanging fruit"

Addressing "low-hanging fruit" allows for smooth start-up in preparation for more challenging metrics  $^{\scriptscriptstyle \dagger}$ 

So I thought they [the first three Common Metrics] were very good choices because ... [they] were relatively low-hanging fruit, meaning they were the ones that generally every CTSA was collecting some sort of data around ... [T]hey were good choices and... it helped prepare our hubs, certainly, for the more difficult ones that I'm sure are to come; the ones that are more complicated to gather. - Administrator

# Metric clarity and usefulness

Clear metric definition allows focus on improvement work<sup>†</sup>

We got [the performance improvement plan] together...only because the metric was easy to understand. I think there wasn't really conflict in definitions. – Administrator

Alignment with institutional priorities<sup>†</sup>

The institution is very interested in this. So, I think that this is something the institution is highly invested in doing well on. – Principal Investigator

So, I think that since these are important for our institution, it was just easy for us to provide the data and to show how we're improving. Our VP of Research and our research officers believed that the IRB is important to them, our Pilot program does the tracking, that's important to them... and Career Development and the Ks was...a priority to the renewal. – Principal Investigator\*\*

## Facilitator with Illustrative Quotation(s)\*

#### **Engaging stakeholders**

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. – Principal Investigator

Integration of Common Metrics with institutional priorities

This has been embraced...as a barometer at the institution. ...I think having metrics that have the measure at the institution is a good thing. ...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. That's been our goal. – Principal Investigator

Hub location and 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. – Principal Investigator

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.. – Principal Investigator

### Hub authority and control

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. – Principal Investigator* 

So administratively... we are a separate center even though I'm 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'm going to also hold a title of Associate Dean for Research, as did my predecessor, and that's by design. – Principal Investigator

# Facilitator with Illustrative Quotation(s)\*

#### Core team

### Team member(s) take ownership of implementation

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. – Implementer

It's really good to have the main person involved in each of the different metrics, either be it careers, IRB or pilots or informatics, to be the one leading the charge because they really understand what the barriers and what the facilitators are. – Administrator

#### Local champion on the team

Our project manager is a real believer and a true champion for this process. Both he and our overall evaluation lead have been out there beating the drum for the overall process. I'd give a lot of credit to them for embracing this because it really has helped us change our culture...- Principal Investigator

### Effective team climate

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. – Administrator

Each week in rotation different members of our senior staff report to the whole group, some of which relate[s] to the Common Metrics. So we keep up-to-date by having very, very free flowing communication and reporting back. – Principal Investigator

# Principal investigator involvement

# Providing strategic guidance

We would report the statistics to him, or the Common Metrics, and turn the proofs into him before we finalized them and of course, before we presented them to the program officer. [He] would review the Turn the Curve plans and make recommendations, and I'd say he's very involved. He 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?". – Administrator

#### Serving as champion

I would say our PI, I think he has the role of champion on our Common Metrics team and he has definitely I think been that. So he welcomes, I think, 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. – Administrator

Our PI has been a source of great encouragement and support... He has helped to make connections where we've needed them...It's not as exciting as some of the other things that are happening in biomedical research on our campus, but he provides a lot of encouragement and enthusiasm for these activities, which I think goes a long way. – Implementer

## Facilitator with Illustrative Quotation(s)\*

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. – Administrator

### Providing hands-on oversight during start-up

[The Principal Investigator] 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-today initiatives and day-to-day work, he was more hands-on with the Metrics than he is with some of the other things. But that's not terribly uncommon for new initiatives that we have to roll out. I think as the Common Metrics continue, he would need to be less involved on a day-to-day basis. – Administrator

#### **Tufts Implementation Program**

Useful training and support<sup>†</sup>

I find that the trainings did accomplish for us what they needed to, in terms of helping us to use the language and facilitate the discussions with the various hub stakeholders. I thought that the time and the follow-up activities were pretty efficient and helpful. So overall, the training experience I think was well-conducted. – Implementer

...they [Tufts] were always just really responsive and helpful and flexible... - Implementer

\* 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.

<sup>†</sup>Indicates that the facilitator has a corresponding challenge.

# Appendix N. Paired facilitators and challenges

Facilitator	Challenge					
Metric definitions, data collection, and computation						
Clear metric definition allows focus on improvement work	Changing metric definitions and clarifications created duplicative work					
Alignment of Common Metrics with and ability to use existing data collection tools	Lack of data system or an existing system that was not aligned with the Common Metrics definitions created more effort for effective tracking					
Addressing "low-hanging fruit" allows for smooth start-up in preparation for more challenging metrics	Spending limited resources on areas that did not need improvement was not helpful					
Implementation support, performance improvement framework, and software						
Useful training and support	Lack of use of training and support by some hubs					
Results Based Accountability framework provides a blueprint <sup>*</sup>	Lack of depth of Results Based Accountability framework <sup>*</sup>					
Scorecard offers easy, common platform	Limitations of the Scorecard software					
Resources and team structure						
Adequate evaluation and other specific expertise	Lack of evaluation and other specific expertise					
Leveraging extended teams	Lack of adequate staffing					
Institutional context						
Availability of institutional resources Alignment with institutional priorities	Lack of institutional investment Lack of alignment with institutional priorities					

'Labeled as a benefit or limitation in the main text.

# Appendix O. CTSA-funded personnel\* and approximate hours spent

N=49 Hubs	Hubs**		Approximated Hours	
Position* (coded)	n	%	Mean	SD
Evaluator	32	65.3	95	93
Evaluation Director/leader	26	53.1	93	105
Evaluator, non-director	13	26.5	49	30
Content expert	26	53.1	45	59
Topic expert/core leader	26	53.1	45	59
CTSA leadership	25	51.0	40	59
Executive Director/Administrator/Associate Director	15	30.6	41	66
Principal Investigator	14	28.6	20	23
Director of Strategy/Planning	2	4.1	47	24
Chief Operating Office	2	4.1	5	1
Co-Principal Investigator	1	2	30	
Associate Director of Operations	1	2	4	
Manager/coordinator	21	42.9	166	273
Program manager/coordinator	11	22.4	164	326
Project leader/manager	9	18.4	159	187
Research administrator	2	4.1	10	
Program liaison	1	2	60	
Component coordinators	1	2	10	
Data collector/analyst/support	20	40.8	53	67
Core administrative/data support	8	16.3	20	11
Data collector	6	12.2	56	45
Data analyst	3	6.1	112	92
Data manager	2	4.1	54	71
Analyst	1	2	120	
Program data manager	1	2	3	
Common Metrics champion	4	8.2	89	141
Project champion	3	6.1	108	166
Core metric champion	1	2	30	

N=49 Hubs	Hubs**		Approximated Hours	
Position* (coded)	n	%	Mean	SD
Informatics	4	8.2	152	113
Director of Information Technology/Informatics	2	4.1	216	34
Lead Software Engineer	1	2		
Research Information Technology Director	1	2	24	
Clinical research personnel	2	4.1	48	46
Clinical research facilitator	1	2	40	
Clinical research officer	1	2	40	
Clinical Research Operations Director	1	2	15	
Performance improvement expert	2	4.1	38	31
Quality control expert	1	2	60	
Results-Based Accountability expert	1	2	16	
Institutional leaders	1	2	16	•
Assistant Vice President	1	2	16	
Biostatistics	1	2	6	
Biostatistics, Epidemiology, and Research Design (BERD) Director	1	2	6	
Other	1	2	33	•
Interested individual	1	2	33	

# Appendix O, continued. CTSA-funded personnel\* and approximate hours spent

\* Up to five CTSA-funded personnel with important roles in the most recent update could be reported.

\*\*An individual hub could report more than one position within each group of positions.

# Appendix P. Effect of hub engagement on completion of performance improvement activities, overall and by metric

The following four figures provide a visual comparison of hub completion of Common Metrics and performance improvement activities (overall sum and per metric) across the four categories of engagement. Higher scores reflect completion of more activities.

For each category of engagement, the "box" represents the middle 50% of scores. The line within the box marks the median score; if a line is not present, then the median is equivalent to one end of the box. The "whiskers" extending above and below the box represent the full range of scores for that category of engagement.





Appendix P, continued. Effect of hub engagement on completion of performance improvement activities, overall and by metric





# Appendix Q. Role of Tufts Implementation Program

Hubs reported on their experience with the Tufts Implementation Program. Survey participants had sufficient knowledge of the program to report accurately. Most participants reported personally participating in the program components (range: 89%-100% depending on component), and 95% could consult with at least one other person at the hub who attended some part of the program.

N=59 Hubs		Imple	ementation	Group	p-Value
		1	2	3	
		n=20	n=17	n=22	
Overall assessment			n (%)		
Satisfaction with the	Extremely satisfied	1 (5)	1(6)	1(5)	
program	Moderately satisfied	12 (60)	6 (35)	10 (45)	
	Slightly satisfied	4 (20)	5 (29)	7 (32)	
	Slightly dissatisfied	0 (0)	2 (12)	2 (9)	
	Moderately dissatisfied	2 (10)	1(6)	2 (9)	
	Extremely dissatisfied	1 (5)	2 (12)	0 (0)	0.745
Effectiveness of the	Extremely effective	1 (5)	1(6)	1 (5)	
program	Very effective	7 (35)	4 (24)	2 (9)	
	Moderately effective	8 (40)	4 (24)	14 (64)	
	Slightly effective	2 (10)	7 (41)	3 (14)	
	Not at all effective	2 (10)	1(6)	0 (0)	0.081
	Do not know	0 (0)	0 (0)	2 (9)	
Knowledge gained about:			n (%)		
Collecting metric	Much less than needed	1 (5)	0 (0)	0 (0)	
data per Operational Guidelines	Less than needed	0 (0)	3 (18)	1(5)	
	About what was needed	13 (65)	11 (65)	13 (62)	
	More than needed	4 (20)	2 (12)	4 (19)	
	Much more than needed	2 (10)	1(6)	3 (14)	0.494
	Do not know	0 (0)	0 (0)	0 (0)	

# Hub experience with Tufts Implementation Program by Implementation Group

N=59 Hubs		Imple	Implementation Group		
		1 n=20	2 n=17	3 n=22	
RBA	Much less than needed	0 (0)	0 (0)	0 (0)	
	Less than needed	1(5)	1(6)	1(5)	
	About what was needed	11 (55)	11 (65)	12 (55)	
	More than needed	4 (20)	4 (24)	5 (23)	
	Much more than needed	4 (20)	1(6)	3 (14)	0.953
	Do not know	0 (0)	0 (0)	1(5)	
Proficiency gained in:			n (%)		
Collecting metric data per Operational Guideline	Much less than needed	1(5)	0 (0)	0 (0)	
	Less than needed	0 (0)	4 (24)	3 (14)	
	About what was needed	14 (70)	9 (53)	10 (45)	
	More than needed	3 (15)	4 (24)	7 (32)	
	Much more than needed	2 (10)	0(0)	2 (9)	0.239
	Do not know	0 (0)	0 (0)	0 (0)	
Implementing RBA	Much less than needed	0 (0)	0 (0)	0 (0)	
	Less than needed	1(5)	3 (18)	3 (14)	
	About what was needed	12 (60)	10 (59)	12 (55)	
	More than needed	2 (10)	3 (18)	2 (9)	
	Much more than needed	5 (25)	1(6)	2 (9)	0.595
	Do not know	0 (0)	0 (0)	3 (14)	

# Appendix Q, continued. Role of Tufts Implementation Program

# Appendix R. Hub self-assessment over time

N=58 Hubs*		Time Point		p-Value
To what extent is your hub able to?		Baseline	Final	
Assess whether current performance	Not at all	1(2)	0 (0)	
is on track to meet its goals, aims, and	A little	2 (3)	1(2)	
objectives	Some	23 (40)	13 (22)	
Baseline item:	A lot	31 (53)	43 (74)	0.109
is on track to meet its goals	Not sure	1(2)	1(2)	
	Mean Score**	82.5	91.2	0.016
Assess whether future performance is	Not at all	2 (3)	1(2)	
likely to be on track to meet its goals,	A little	11 (19)	4 (7)	
anns, and objectives	Some	25 (43)	21 (36)	
Baseline item:	A lot	18 (31)	30 (52)	0.074
likely to be on track to meet its goals	Not sure	2 (3)	2 (3)	
	Mean Score <sup>**</sup>	68.5	81.0	0.011
Engage hub leaders, faculty, and staff in discussions about operational or	Not at all	1(2)	0 (0)	
	A little	2 (3)	4 (7)	
	Some	13 (22)	8 (14)	
Baseline item: same	A lot	42 (72)	46 (79)	0.386
	Not sure	0 (0)	0 (0)	
	Mean Score <sup>**</sup>	88.5	90.8	0.545
Engage stakeholders outside the hub	Not at all	7 (12)	5 (9)	
in discussions about operational or	A little	9 (16)	6 (10)	
	Some	25 (43)	31 (53)	
Baseline item: same	A lot	16 (28)	15 (26)	0.657
	Not sure	1(2)	1(2)	
	Mean Score <sup>**</sup>	62.6	66.1	0.539
Identify actions that have potential to	Not at all	1(2)	0 (0)	
improve performance	A little	6 (10)	2 (3)	
Baseline item:	Some	18 (31)	13 (22)	
Identify actions and activities that have potential to influence performance	A lot	32 (55)	43 (74)	0.144
potential to influence performance	Not sure	1(2)	0 (0)	
	Mean Score**	80.7	90.2	0.020

<b>Appendix R</b>	, continued.	Hub	self-assessment	over	time
-------------------	--------------	-----	-----------------	------	------

N=58 Hubs*		Time	p-Value	
To what extent is your hub able to?		Baseline	Final	
Efficiently address performance issues	Not at all	1(2)	1(2)	
Baseline item	A little	4 (7)	5 (9)	
Be efficient in addressing performance	Some	35 (60)	25 (43)	
issues	A lot	16 (28)	27 (47)	0.207
	Not sure	2 (3)	0 (0)	
	Mean Score**	72.6	78.2	0.193
Effectively address performance issues	Not at all	2 (3)	1(2)	
Baseline item: Be effective in addressing performance	A little	6 (10)	4 (7)	
	Some	28 (48)	24 (41)	
issues	A lot	20 (34)	29 (50)	0.447
	Not sure	2 (3)	0 (0)	
	Mean Score**	72.6	79.9	0.115
Advance clinical and translational science	Not at all	1(2)	0 (0)	
Baseline item: same	A little	1(2)	3 (5)	
baseline item: same	Some	18 (31)	17 (29)	
	A lot	36 (62)	37 (64)	0.565
	Not sure	2 (3)	1(2)	
	Mean Score**	86.3	86.5	0.950

\*One hub did not respond to the second follow-up survey. Another hub was dropped from the analysis due to missing data on the self-assessment questions at baseline, despite completing other parts of the survey.

\*\* 0-100 scale; higher score reflects a better self-assessment.