HHS Public Access

Author manuscript

Clin Transl Sci. Author manuscript; available in PMC 2016 October 01.

Published in final edited form as:

Clin Transl Sci. 2015 October ; 8(5): 451–459. doi:10.1111/cts.12296.

Developing Common Metrics for the Clinical and Translational Science Awards (CTSAs): Lessons Learned

Doris M Rubio, PhD[1], **Arthur E Blank, PhD**[2], **Ann Dozier, PhD**[3], **Lisle Hites, PhD**[4], **Victoria A Gilliam, MPH**[1], **Joe Hunt, MPH**[5], **Julie Rainwater, PhD**[6], and **William M Trochim, PhD**[7]

[1]Data Center, Center for Research on Health Care, Division of General Internal Medicine, Department of Medicine, School of Medicine, University of Pittsburgh, Pittsburgh, PA

[2]Department of Family and Social Medicine, Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY

[3]Department of Public Health Sciences, University of Rochester, Rochester, NY

[4]Department of Health Care Organization and Policy, University of Alabama at Birmingham, Birmingham, AL

[5]Indiana Clinical and Translational Sciences Institute, Indiana University School of Medicine, Indianapolis, IN

[6]Clinical and Translational Science Center, University of California Davis, Sacramento, CA

[7]Department of Policy Analysis and Management, Cornell University, Ithaca, NY

Abstract

The National Institutes of Health (NIH) Roadmap for Medical Research initiative, funded by the NIH Common Fund and offered through the Clinical and Translational Science Award (CTSA) program, developed more than 60 unique models for achieving the NIH goal of accelerating discoveries toward better public health. The variety of these models enabled participating academic centers to experiment with different approaches to fit their research environment.

A central challenge related to the diversity of approaches is the ability to determine the success and contribution of each model. This paper describes the effort by the Evaluation Key Function Committee to develop and test a methodology for identifying a set of common metrics to assess the efficiency of clinical research processes and for pilot testing these processes for collecting and analyzing metrics. The project involved more than one-fourth of all CTSAs and resulted in useful information regarding the challenges in developing common metrics, the complexity and costs of acquiring data for the metrics, and limitations on the utility of the metrics in assessing clinical research performance. The results of this process led to the identification of lessons learned and recommendations for development and use of common metrics to evaluate the CTSA effort.

Correspondence: Doris M Rubio, PhD, 200 Meyran Ave, Suite 200, University of Pittsburgh, PA 15213, rubiodm@upmc.edu, Phone: 412-692-2023, Fax: 412-246-6954.

INTRODUCTION

The 2010 Institute Of Medicine (IOM) Report "National Cancer Clinical Trials System for the 21st Century: Reinvigorating the National Cancer Institute Cooperative Group Program"[1] states "… the current structure and processes of the entire clinical trials system need to be redesigned to improve value by reducing redundancy and improving the effectiveness and efficiency of trials." Many trials take too long to open, and part of the problem is the number of processes involved. For Phase III trials, Dilts, et al.[2] report almost 300 unique processes possibly needed just to activate a trial, so it should not be surprising that it takes approximately 600 days from the origin of a trial until initiation. Even when a trial is completed, only half of those trials publish within 30 months; the overall publication rate for trials is a dismal 68%.[3]

At the 2012 Clinical and Translational Science Award (CTSA) Evaluation Key Function Committee face-to-face meeting, leadership for the National Center for Advancing Translational Sciences highlighted major concerns about clinical trials, including high costs; failure to start, recruit, and publish; ethics of incomplete studies; and studies that are never published. In the IOM report on the CTSA programs,[4] the need for common metrics was emphasized. The report states that a program-wide evaluation should include metrics reflecting the extent to which CTSAs facilitate clinical studies and reduce delays in clinical trials. The report recognizes the difficulty in developing and implementing common metrics across the CTSA consortium and emphasizes that program accountability cannot be achieved without high-level common metrics. There is an inability to assess the ultimate goal of the program, which is to improve public health; therefore, there is a need to discover indirect ways the CTSAs contribute to research. For example, metrics providing more real-time assessments of progress in advancing clinical and translational research could ultimately improve public health by changing clinical practice.[4]

Developing common metrics that assess the efficiency of clinical research is essential for achieving several CTSA goals. First, common metrics will enable the CTSA institutions to establish benchmarks. The use of common metrics across CTSA institutions in a de-identified manner will allow sites to gauge their status within the consortium. In addition, these benchmarks could be valuable reference points for other efforts attempting to assess clinical research efficiency, both within and outside the National Institutes of Health (NIH). Second, common metrics would enable the CTSAs to undertake systematic process improvement efforts and function as a type of "virtual national laboratory"[4] for clinical and translational science. Finally, common metrics would provide a basis for aggregating results across the entire CTSA initiative, providing greater transparency and accountability to Congress and the public.

The Common Metrics Workgroup was formed as a subgroup of the CTSA Evaluation Key Function Committee, which was comprised of evaluators from all 62 CTSAs. The purposes of the workgroup were to (1) generate potential metrics, (2) define and operationalize the most promising initial candidate metrics, and (3) assess the feasibility of collecting data for the metrics. The purpose of this paper is to describe the pilot study that the Common Metrics Workgroup conducted, the lessons learned, and the potential future directions.

Generation of Metrics

In the spring of 2012, CTSA evaluators and principal investigators (PIs) collaborated on identifying potential metrics for clinical research processes and outcomes. Several hundred metrics were generated as a result. During the October 2012 face-to-face meeting of CTSA evaluators, 15 metrics were identified as being especially promising based on ratings of importance and feasibility. Each of the 15 promising metrics actually constitutes a broad category of metrics which could be operationalized in a number of ways. These metrics were organized into 6 broad categories (Table 1). This initial pilot focused on the top three metrics falling into the category of "Clinical Research Processes"—time from institutional review board (IRB) submission to approval, studies meeting accrual goals, and time from notice of grant award (NOGA) to first accrual.

Definition and Operationalization of Metrics

During a six-month period, the Common Metrics Workgroup examined the top three metrics and recognized a need for a more precise definition and operationalization in order for data to be collected consistently and accurately across institutions. For example, we took the original metric "time from IRB submission to approval" and further defined it as "the number of calendar days from the institution's official IRB proposal receipt date to the official date of IRB approval." In addition, it was necessary to specify protocol inclusion and exclusion rules (e.g., only clinical research, only protocols undergoing full IRB review) and to collect data essential for conducting a reasonable interpretation of the context for the metrics, including descriptive information about the institutional and protocol (e.g., number of IRBs and full-time equivalent support staff, pre-submission assistance in preparing the IRB, number of resubmissions). Similar specificity was applied to the other two metrics. For the metric "studies meeting accrual goals," it was necessary to define both the targeted number of subjects (the goal) and the observed accrual number to determine the difference. In addition, it was vital to record key descriptive variables that could influence or distort accrual estimates (e.g., type of study, age or gender exclusions, target population). Finally, for the metric "time from NOGA to first accrual", it was essential to precisely define key dates and collect descriptive data for the protocol (e.g., type of study, target study population) to enable interpretation of results. For all metrics, a time frame was established to retrospectively collect the data. Consequently, each of the three metrics in this study actually consisted of a separate reporting protocol containing several operationalized measures with a set of accompanying descriptive variables.

PILOT PROJECT

Methods

Study Procedures—Some of the measures in our pilot study required changes in the way data are collected at the participating institutions; therefore, the support of the CTSA PIs was imperative to ensure data were collected correctly.

Thirteen CTSA institutions with members on the Evaluation Leadership Committee and the Common Metrics Workgroup agreed to participate in the initial round of pilot testing. Additionally, evaluators from four other institutions volunteered (N = 17). While there was

an open invitation to institutions to participate in the pilot testing, the number of institutions was purposefully limited to minimize the burden across all CTSAs. By initially pilot testing the metrics on a small number of institutions, we were able to generate useful information about the feasibility of collecting these data and refining the metrics before piloting them on a larger number of institutions.

Each institution was asked to submit a prioritized list of the metric(s) they wished to pilot, as well as explanations for why they might not want to pilot a specific metric. They were also asked to report data on a minimum of 10 protocols that met the inclusion criteria for the metric being piloted

Participating institutions and survey respondents were given access to a ROCKET (Research Organization, Collaboration, and Knowledge Exchange Toolkit)[5] workspace where information about the pilot was centrally located. Information included (1) background on the project, (2) project contact information, (3) a link to the survey to enter new data on a protocol, and (4) the metric definition.

The pilot phase lasted for six weeks. Conference calls were conducted at three weeks to solicit feedback on barriers or challenges needing to be addressed. At the end of the pilot phase, we conducted another conference call with the pilot institutions to solicit their feedback. In addition, each institution was asked to complete a brief survey about the feasibility of collecting data on the metric at their institution.

The CTSA Consortium Coordinating Center version of REDCap (Research Electronic Data Capture)[6] was used to collect data on the metrics, as well as for the survey, to assess the feasibility. All data were kept confidential and no institutional identification was or will be released.

Description of Participation—Seventeen academic health centers with CTSAs participated in the pilot project. Nine institutions volunteered to submit data for the IRB metric, with 67 protocols submitted ranging from 1 to 11 per institution. Four institutions participated in the “study meeting accrual goals” metric, with 39 protocols submitted ranging from 8 to 11 protocols per institution. Seven institutions participated in the “time from NOGA to first accrual” metric, with 52 protocols submitted ranging from 2 to 10 per institution. Three institutions volunteered to collect data on more than one metric. Two institutions submitted data on the IRB metric and “time from NOGA to first accrual” metric. One institution submitted data on the “studies meeting accrual goals” metric and the “time from NOGA to first accrual” metric.

Results

The results of the pilot study are shown in Tables 2–4. While these results may be of substantive interest, it should be noted that the purpose of the pilot study was to test the feasibility of the metrics. Because the sample of institutions is very limited, and the results are not necessarily representative of all CTSA organizations or institutions, the results should be viewed as illustrative and interpreted with caution.

IRB Duration Metrics—Eight institutions participated in the IRB duration metrics pilot, representing 67 studies. This pilot was limited to protocols with full-board reviews and excluded expedited reviews. The key IRB metric was IRB duration, which was successfully completed for all protocols; the median duration was 59 days with a range of 16–328 days. These results are consistent with those of the two previous CTSA-wide IRB studies.

Number of protocol resubmissions gives us insight into how complex a protocol might have been, how prepared investigators were, and how thorough they were in preparations. Results indicate that 65.7% of the protocols were resubmitted one or more times and 18.2% required four or more resubmissions. Ninety-one percent of protocols were answered (59.7% = yes, 31.3% = no) when asked whether the study was multisite. When asked whether the study was regulated by the Food and Drug Administration (FDA), 89.8% of the protocols were answered (60% = yes, 30.8% = no, 9.2% = don't know). All protocols were answered (76.1% = no, 23.9% = yes) when asked whether the study involved an investigational new drug (IND), which is valuable for identifying drug trials that are critically important in CTSA contexts.

Perhaps equally informative as the survey results is which items were *not* practical for system-wide measurement. For example, when asked which institutional resources were available to researchers for assistance with preparing IRB protocols prior to submission, the vast majority of institutions (87.5%) indicated all three types of services listed—centralized, department, and IRB staff—were available. Therefore, either this question does not have enough variability or, more likely, respondents made assumptions about such resources without having any actual evidence in the protocol record. Alternatively, when asked whether the protocol required an investigational device exemption (IDE), every protocol contained an answer but only 1.5% needed an IDE, which reduced the potential value of this metric. While any number of such explanatory variables may be correlated with the key duration metric, and therefore be potentially valuable, it is difficult to know which are relevant in any given situation and whether this differs by institution or type of protocol. Many of the descriptive variables are not regularly collected across institutions, and to require them to be collected would likely pose a considerable burden because data would need to be collected either directly from investigators or by simply reading the protocol, which is even less reliable.

Study Meeting Accrual Goals Metrics—Four institutions participated in the "studies meeting accrual goal" metrics pilot, representing 39 studies with a range of 8–11 protocols per institution. Two institutions reported a total of six multisite studies not eligible for the pilot. For this analysis, they were eliminated, leaving 33 studies as the denominator. Missing data occurred across all four participating institutions.

Access to the requested study-specific information varied by institution. Across all of the studies (aggregated across all sites), 90%–100% provided the number of participants recruited, whether the recruitment target was met, how many participants were needed to analyze the primary research question, and whether there was access to the number of participants needed based on the power analysis. Nearly three-fourths could identify the time between IRB approval and first participant accrual; those who could not were unable to

provide either the IRB approval date or, more commonly, the date of first accrual. Approximately half of the studies were unable to ascertain information on the projected recruitment time, the difference between planned and actual recruitment time, and first accrual date.

Studies selected varied in the number of participants needed, ranging from 6 to 900. Most included both sexes, and one-third had age restrictions. Approximately two-thirds were adult-only studies, and nearly 80% required full-board approval; less than half were FDA regulated, and 13% involved INDs.

Despite only 54% of studies meeting their recruitment target, only 18% reported having a problem recruiting; four studies had established minority recruitment targets. Among the 45% providing information about length of recruitment (2–96 months), seven lasted less than projected, while nine lasted longer than projected. Among the 72% providing information, accrual start date ranged from 41 to 646 days after IRB approval.

NOGA to First Accrual Metrics—Eight CTSA sites volunteered to pilot the metric "time from NOGA to first accrual." Seven institutions reported data on 41 studies; one institution identified no studies meeting criteria. Of the 41 studies reported, 13 studies did not officially meet inclusion criteria (either NOGA was before July 1, 2012, or the study had not yet accrued at least one participant). These studies were excluded, leaving 28 studies from five institutions. The NOGA was available for all 28 studies; however, the date of first accrual was available for only 25.

For the 25 studies that met inclusion criteria, and for which the date of first accrual could be determined, the median number of days from NOGA to first accrual was 203 with a range of 25–380 days. Studies with expedited or "other" types of IRB review had longer intervals than full IRB review. For nearly all studies, the contextual variables were reported. The most common issue noted by the seven institutions that attempted the metric was the amount of time necessary to find the first accrual date for the study. Only one institution found all the necessary data in a single database and, in most cases, the PI or research coordinator had to be contacted to collect the first accrual data element. In addition, doubt on the merit of the metric was indicated in the feedback survey because some studies had a deliberate "planned nonaccrual period" built into the study timeline that the metric did not take into account.

Feasibility

At the mid-point of the pilot study, the Common Metrics Workgroup met via a conference call with 26 representatives from 16 of the 17 institutions to obtain feedback. When the study was complete, we also held another conference call with 23 representatives from 16 institutions. Similar discussions emerged from the two conference calls. Participants were surprised by the amount of time required for data collection. In several instances, the representatives thought they could access the data from an institutional database only to find that they had to access several databases. Even then, they were not always successful at obtaining the necessary data. For most institutions, getting data from the IRB was significantly easier than retrieving data on participant accrual.

The survey was conducted after the institutions finished data collection to obtain additional information about the feasibility of collecting data on the metrics. The results of the survey confirm information obtained during the conference calls. Some data were easy to collect, such as the data regarding IRB receipt or type of IRB submission. Other data, such as accrual of first participant or meeting accrual goals, were much more difficult to collect. In several instances, the institution had to contact the PI of the study to get these data. This was met with challenges, such as low response rate, time to follow-up, and inability to verify data. Many noted that the context variables, such as whether the study involved rare diseases, were harder to collect because many of these data are not tracked.

IRB Duration Metric—The majority of pilot sites were able to report data on the basic metric of "IRB duration" when computing IRB receipt date to final IRB approval date. Other variables that impacted IRB review time and were deemed feasible to report included: type of IRB review (e.g., full committee vs. other), number of resubmissions, and type of study (e.g, multisite, FDA regulated, IND).

Study Meeting Accrual Goals Metric—It was time intensive to find studies that met the inclusion criteria. Challenges identifying studies to meet the inclusion criteria were classified as: (1) not available in a central location (i.e., the IRB) and (2) could not be found in electronic data format, requiring manual collection. Three of the four CTSAs collected data on 10 or more studies for the pilot. Two of the four CTSAs included studies with multisite trials that were excluded from the pilot.

For some sites, it was difficult to obtain data from IRBs, which is where much of the necessary data reside. Challenges included the site's IRB (1) did not electronically collect the data elements, therefore the information had to be manually extracted; (2) did not collect data elements at all or did not use the same definition as delineated by the pilot; and (3) staff was unresponsive, making it difficult to identify time to review and collect data elements.

For some of the data elements, different definitions and data points were used and/or the definitions lacked clarity. For example, the pilot defined "study closure" as "studies that are closed to recruitment (studies may still be collecting data or conducting analyses)." For one site, the trigger that identifies a study as closed is the submission of a study closure or final report to the IRB. An example of unclear definitions includes the establishment of "explicit minority accrual targets." Further explanation of "explicit" and "minority" was requested.

NOGA to First Accrual Metric—It is unlikely that all CTSAs would have enough studies to meet the criteria unless the time frame was lengthened. Eight sites attempted the metric but only five had studies meeting the criteria. CTSAs that had qualifying studies were able to complete the data elements necessary to compute the metric. However, identifying the set of qualifying studies and grants was difficult and involved multiple steps. Most CTSAs that worked on the metric did not think it would be helpful for understanding research efficiency.

Gathering data on the metric often involved manually abstracting data from several databases. In addition, there was a lack of consistency about how data were recorded across the multiple databases being used. This metric almost always required contacting the PI or

research coordinator to obtain the date of first accrual; it is likely too labor intensive for most sites to undertake.

RECOMMENDATIONS

Overall, we found that collecting common metrics has significant value but can be very time consuming, resource intensive, and challenging. We need to be sensitive to the number of metrics developed since a large number can become unfeasible to collect. The approach recommended here is to keep national common metric reporting requirements minimal and to use feedback from these metrics to encourage CTSAs to more closely evaluate what factors might be driving the results at their institutions.

As a rule, two features are of key importance in developing common metrics: burden and value. Ideally, to keep common metrics to a minimum, institutions should focus on metrics that are very low burden to both the investigators and the CTSA but high value to the institution and the CTSA. However, such metrics are not always readily available or identifiable.

First, we recommend keeping cross-institutional common metrics to a minimum, prioritizing those that are low burden and high value, which will inevitably result in the highest compliance. Second, make certain the metrics are clearly defined and standardized. Without identical (or near identical) measures, it is difficult to compare across institutions. To arrive at a useful set of measures, it is important to work iteratively, utilizing a formative evaluation methodology[7]; pilot the selected metrics, revise based on the formative results, pilot again. As a part of this formative evaluation process, the CTSAs should seek feedback from those collecting data on the metrics and incorporate changes on a regular basis to adjust for system fluctuations. By recognizing that collecting common metrics can quickly become labor and resource intensive, it becomes possible to plan for this expense and, at the same time, be realistic about what level of effort is reasonable and appropriate. Some specific recommendations for each metric are noted below.

IRB Duration Metric

To understand the contributing process factors of IRB completion time, which we define as the time from IRB receipt date to final IRB approval date, we recommend that, at a minimum, each institution collect three IRB related variables. The first critical metric is type of review, which our pilot data indicates can be collected with a bivariate measure consisting of full committee review vs. other type of review. Second, it is important that CTSAs track number of resubmissions as this is a key factor in overall duration of the IRB process, regardless of the administrative speed of an individual institution’s IRB. Third, type of study can often predict longer or shorter IRB times. For example, multisite studies can be more complex, as can FDA-regulated studies and IND studies. We recommend that, for now, only these three standardized metrics be collected for each institution, and that this initial formative data be used to determine whether additional local detailed IRB data are needed for process analysis.

The pilot has demonstrated that these recommended metrics are practical and feasible. In addition, they would enable basic benchmarking which would help CTSAs identify variability that might warrant subsequent localized, detailed IRB process analyses.

Studies Meeting Accrual Goals Metric

As with the other standardized metrics, it is essential that a clearly worded standardized definition and operationalization be developed for studies meeting accrual goals so that comparisons can be made across CTSA institutions. After these common metrics are developed, significant time will be required to develop the systems necessary to accurately report data on accrual. For CTSAs with an existing clinical management system, collecting these data will be less of a challenge; however, comparative analyses will not be possible until the entire network decides on and adopts a common set of metrics.

Results of this pilot study indicate that access to data for the metrics and actual performance on metrics is widely variable, even with only four institutions reporting. Establishing common metrics would be difficult, but the process raises awareness about different institutional practices and how others are performing on particular metrics for internal comparison. These formative data are helpful to support informed discussions of best practices regarding both selection of metrics and institutional performance.

NOGA to First Accrual Metric

Results of this pilot study suggest that measuring NOGA to first accrual may only be feasible for institutions with clinical trials management systems or similar databases that include the identified data elements; contacting PIs and research staff is too high a burden for the CTSA given the vast number of studies involved. Accordingly, it is recommended that all CTSA institutions move toward instituting a clinical trials management system and, until then, those without such a system will be limited to capturing a subset of the clinical data.

CONCLUSIONS AND NEXT STEPS

This pilot study constituted an initial concerted effort to explore the feasibility of collecting common metrics that can be used across the CTSA initiative to monitor key processes and outcomes. It built on prior years of discussion of the CTSA Consortium's Evaluation Key Function Committee, which wrote a paper calling for "standardized metrics and cross-cutting analyses that enable aggregation" as part of a "balanced set of evaluation activities and methods."[5] That paper, echoed in the subsequent IOM report,[4] recognized the important role of metrics while acknowledging they are not, by themselves, sufficient to address the evaluation needs of such a complex initiative. Good evaluation metrics typically raise as many questions as they answer. It is seldom possible to tell, from metrics alone, what moves them over time or why they fluctuate between sites; for that, we need contextual data, hypotheses, deliberate interventions, and more controlled evaluations.

But metrics have a critically important signaling value. Even if they don't tell us what is driving them, they tell us where we are and when we are changing direction. Without a set of simple common metrics, the system lacks basic feedback. The most recent CTSA request

for application, responding to the IOM report, recognizes the critical importance of developing such a set of common metrics and calls for efforts like this pilot study.

The general conclusion of this pilot contains both good and bad news. In terms of challenges, this study makes it clear that developing common metrics that could be collected across multiple institutions is a difficult endeavor; simple concepts can be surprisingly difficult to define. For instance, using “IRB receipt date” as the point to start the clock on measuring IRB review processes becomes questionable when some institutions provide significant pre-submission proposal assistance, while others focus on post-submission support. Comparing these two using the same “IRB receipt date” will not provide accurate results and is likely to advantage one institution over the other. Additionally, there are countless factors both within and across institutions (e.g., study type, research subject, target population) that are likely to yield unique portfolios, which will differentially affect how metrics perform.

In addition to the numerous definitional issues, there are contextual challenges in developing and collecting common metrics. This pilot study demonstrated that different institutional processes and legacy data systems make it difficult to assume what is simple for one institution to collect will be equally simple for another. One CTSA may have developed a legacy clinical management information system linking subject accrual data to IRB information. Other institutions may have evolved separate systems requiring considerable labor to connect such data. Without re-engineering both the existing research management processes and information systems across all participating institutions, it is likely the implementation of common metrics will remain a challenge and generate burdens that vary by metric and location.

This pilot also demonstrates the considerable good news regarding common metrics. The fact that this pilot was completed shows that it is possible, even if only on a small scale, to develop and collect several key common metrics across multiple CTSAs. This pilot also provides a template for how the CTSAs might operate as a virtual national laboratory for the development of other metrics going forward. It shows that two of the three metrics identified as the most important can, in fact, be collected consistently. This type of testing made it possible to identify a much simpler and less burdensome set of processes for the next round of collection of these metrics. The third metric, from NOGA to first accrual, was especially difficult to gather because many institutions are still getting information systems in place which standardize subject accrual data across multiple independent clinical studies. Consequently, in many cases, collecting these data required contacting individual investigators; however, that’s not necessarily bad news. Should NIH and the CTSAs decide this is a critically important data point—and its high rating in the metric generation process suggests that it is—a formal requirement to collect this data would send a powerful message to the CTSA institutions (and many others) that data collection processes in this area need to be re-engineered. The CTSA initiative can make major impacts on the clinical research management enterprise across institutions by taking a bold stance on these metrics.

As evaluators in the context of a clinical and translational research enterprise, we are steeped in the tradition of hypothesis testing and experimentation. We are inherently critical of data

and quick to raise questions about how it can be interpreted. The instinctive reaction to this pilot study is to identify all of the ways the metrics fall short. When a study or institution has a longer IRB review time or slower subject accrual, we can immediately generate a long list of legitimate reasons why such a discrepancy might make sense. This is a fundamental challenge to the development of common metrics. This pilot study reminds us that, even if a metric is not perfect, or even if it can be reinterpreted in many different ways depending on contextual factors, there is still an important signaling value in collecting the same metric consistently over time. Such common metrics can provide an empirical anchor, a starting point for raising questions about the factors that might be driving its movement. In this sense, metrics and monitoring are a gateway to the type of deeper evaluation that needs to follow. Metrics provide the data that raise the important questions (e.g., why this study took so long, why this institution has lower accrual rates). Evaluations involve hypothesizing potential performance improvements and assessing whether they affect the metrics and, perhaps, answering the important questions.

The next steps in the evolution of common metrics can build on the foundation provided here. This pilot offers a useful template for how subgroups of CTSAs can, at relatively low cost and researcher burden, identify and collect potential common metrics. One obvious next phase would be to replicate this template with the other 12 metrics initially identified as high in potential, most likely in several independent pilots using subgroups of CTSAs to spread the development costs across the system. Another next step would be wider testing of the metrics investigated here. At least two of these metrics have been simplified and refined based on this pilot and should be ready for wider production. The third metric, date of first subject accrual, probably needs some creative thinking, revision, and another smaller pilot. Additionally, since the landscape of clinical and translational science is continually evolving, we should conduct another round of gathering input from the broader system to check on the longer list of potential metrics developed earlier and identify any newer and emerging priorities that should be considered for metric development. The great news of this pilot study is that it demonstrates that all of this is possible as the CTSAs move toward becoming a virtual national laboratory for reengineering the clinical and translational research enterprise.

Acknowledgments

Support for the work described in this publication was provided by UL1TR000005 (University of Pittsburgh); UL1TR001073 (Albert Einstein College of Medicine); UL1TR000042 (University of Rochester); UL1TR000165 (University of Alabama), UL1TR001108 (Indiana University); UL1TR000002 (UC Davis), and by UL1TR000457 (Weill Cornell Medical College), awards from the National Center for Advancing Translational Sciences (NCATS) at the National Institutes of Health (NIH).

Special thanks to those CTSA institutions who graciously agreed to pilot the metrics and to Meryl Sufian, PhD of the National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD, for her work in preparation of this manuscript.

REFERENCES

1. IOM. National Cancer Clinical Trials System for the 21st Century: Reinvigorating the NCI Cooperative Group Program. Washington, DC: The National Academies Press; 2010.

2. Dilts DM, Sandler AB, Cheng SK, et al. Steps and time to process clinical trials at the Cancer Therapy Evaluation Program. J Clin Oncol. 2009; 27(11):1761–1766. [PubMed: 19255315]

3. Ross JS, Tse T, Zarin DA, Xu H, Zhou L, Krumholz HM. Publication of NIH funded trials registered in ClinicalTrials.gov: cross sectional analysis. BMJ. 2012; 344:d7292. [PubMed: 22214755]

4. IOM. The CTSA program at NIH: Opportunities for advancing clinical and translational research. Washington, DC: 2013.

5. Trochim WM, Rubio DM, Thomas VG. Evaluation Key Function Committee of the CC. Evaluation guidelines for the Clinical and Translational Science Awards (CTSAs). Clin Transl Sci. 2013; 6(4): 303–309. [PubMed: 23919366]

6. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009; 42(2):377–381. [PubMed: 18929686]

7. Quinn Patton, M. Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use. New York, NY: The Guilford Press; 2011.

Table 1

Fifteen Priority Metric Categories

Clinical Research Processes	Time from institutional review board submission to approval
	Studies meeting accrual goals
	Time from notice of grant award to first accrual
Careers	Career development
	Career trajectory
Services	Volume of investigators who used services
	Volume of types of services used
	Satisfaction and needs assessment
Economic Return	Leveraging and return on investment of pilots and KL2 scholars
Collaboration	Researcher collaboration
	Institutional collaboration
Products	Number of technology transfer products
	Time to publication
	Influence of research publication
	Time from publication to research synthesis

Table 2

Institutional Review Board (IRB) Completion Time Data

Item/Field	Respondents, No. (%)	Comments/Findings
Institutional Data (N = 8)		
Was the institution fully accredited by the Association for the Accreditation of Human Research Protection Programs (AAHRPP) during the duration of this survey?	8 (100)	5 (62.5%) yes 3 (37.5%) no
Please indicate which institutional resources below are generally available to researchers for assistance with preparing IRB protocols prior to submission to the IRB.	8 (100)	7 (87.5%) centralized 7 (87.5%) department 7 (87.5%) IRB staff
Approximate number of initial protocols submitted during calendar year 2012 that required at least one review by the fully convened IRB at your institution for which the data are reported.	8 (100)	Responses unusable. Field was a text one and responses ranged from "data cannot be obtained by pilot deadline," to "156 initially submitted protocols/research projects were reviewed by 1 of the 4 full-committee IRB in 2012," to simply "65%." Should restrict field to a number.
Of the new protocols submitted to the IRB, what percentage was submitted electronically in 2012?	6 (75)	1 (12.5%) noted 50% 4 (62.5%) noted 100%
Of the new protocols submitted to the IRB, what percentage was submitted using paper in 2012?	6 (75)	4 (62.5%) noted 0 1 (12.5%) noted 100%
How many full-time equivalent support staff are engaged in the processing of protocol submissions for IRB review? Such persons may be responsible for reviewing and approving research and/or making determinations regarding if projects are exempt human subjects research or do not constitute research or research involving human subjects?	7 (87.5)	Ranged from 5 to 14
Number of IRBs at institution for which data are being reported in this study, excluding external IRBs	7 (87.5)	Ranged from 1 to 5
Estimate % of your new protocols that were reviewed by an external IRB in 2012.	8 (100)	Responses unusable. Field was a text one and responses ranged from "we do not know the exact %," to "unknown, maybe 25%," to ">5%." Should restrict field to a number.
Protocol Data (N = 67)		
Type of IRB review required	66 (98.5)	All were full-board review
IRB receipt date	67 (100)	
Final IRB approval date	67 (100)	
IRB duration	67 (100)	Calculated from above two fields mean = 75.6 days median = 59 days range, 16–328 days
Was the protocol reviewed by an external IRB (e.g., commercial, central IRB)?	67 (100)	1 (1.5%) yes 66 (98.5%) no
Was any pre-submission assistance provided for this protocol before it was submitted to the IRB?	47 (70.1)	47 (100%) no (for respondents)
Were any revisions required by the fully convened IRB to secure final approval?	66 (98.5)	51 (77.6%) yes 14 (20.9%) no
How many times was the protocol resubmitted to the IRB prior to receiving IRB approval?	44 (65.7)	All but one case did not require a revision, so there is only 1 (or 1.5%) true missing value
Research phase	67 (100)	33 (49.3%) Phase I – IV 16 (23.9%) other 18 (26.9%) not specified
Is this a protocol on a rare disease?	67 (100)	4 (6%) yes 37 (55.2%) no 26 (38.8%) don't know

Item/Field	Respondents, No. (%)	Comments/Findings
Study population (check all that apply)	67 (100)	54 (80.6%) adult only 11 (16.4%) children 4 (6%) pregnant 0 (0%) prisoners 2 (3%) judgment impaired 2 (3%) social/ethnic
Are there any sex exclusions?	61 (91)	46 (74.6%) no exclusions 1 (1.5%) females excluded 9 (14.9%) males excluded
Are there any age exclusions?	67 (100)	41 (61.2%) yes 10 (14.9%) no 16 (23.9%) don’t know
Is this a multisite study?	67 (100)	40 (59.7%) yes 21 (31.3%) no 6 (9%) don’t know
Is this an FDA-regulated study?	65 (97)	39 (60%) yes 20 (30.8%) no 6 (9.2%) don’t know
Investigational new drug	67 (100)	51 (76.1%) no 16 (23.9%) yes
Investigational device exemption	67 (100)	66 (98.5%) no 1 (1.5%) yes

Table 3

Study Meeting Accrual Goals Data

Item/Field (N = 39)	Respondents, No. (%)	Comments/Findings
How many participants were recruited for the study?	39 (100)	mean = 87 range, 0–585
How many participants were needed to analyze the primary research question based on the original grant proposal?	39 (100)	mean = 174 range, 6–900
Do you have access to the number of participants needed to analyze the primary research question according to the power analysis?	35 (90)	15 yes 4 did not respond
How many participants were needed to analyze the primary research question based on the power analysis?	10 (26)	10 provided a number low = 15 participants high = 1720 participants
Were there explicit minority accrual targets established?	38 (97)	Only 4 reported having minority target
How many months were projected for participant recruitment?	23 (59)	low = 0 months high = 96 months
How many months did actual study recruitment take?	38 (97)	low = 2 months high = 96 months
Difference between planned and actual recruitment	18 (45)	7 studies took shorter than projected 4 were same as projected 9 were longer than projected (5–50 months)
Number of days from institutional review board (IRB) approval to first participant accrual	72 to IRB approval date 46 to first accrual date	11 (28%) missing IRB approval date 21 (54%) missing first accrual date low = 41 days high = 646 days
Did study meet recruitment targets?	39 (100)	21 (54%) met recruitment target
How long was study open for recruitment?	14 (36)	<1 year = 2 >2 years = 7
Problems with study recruitment	39 (100)	7 (18%) yes
Research phase	39 (100)	18 (46%) not specified 13 (33%) Phase II 4 (10%) Phase III
Required IRB review type	39 (100)	31 (79%) full board 8 (21%) expedited
Is this a protocol on a rare disease?	39 (100)	5 (13%) yes
Study population (check all that apply)	39 (100) (1 did not mark a choice)	24 (62%) adult only 9 (23%) children only 5 (13%) adult and children
Any sex exclusions?	39 (100)	33 (85%) no exclusions 5 (13%) males excluded
Any age exclusions?	39 (100)	27 (69%) yes 10 (26%) don’t know
Is this a multisite study?	39 (100)	24 (62%) yes 9 (23%) no 6 (15%) don’t know
Is this an FDA- regulated study?	39 (100)	15 (39%) yes 19 (49%) no 5 (13%) don’t know
Investigational new drug (IND) or investigational device exemption	39 (100)	13 (33%) IND

Table 4

Notice of Grant Award to First Accrual Data

Item/Field (N = 28)	Respondents, No. (%)	Comments/Findings, No. (%)
Date of notice of grant award (NOGA) - date on the official notice of award to the institution	28 (100)	Earliest NOGA - July 1, 2012 Latest NOGA - June 14, 2013
Are you able to report the date of first accrual?	25 (89)	3 (12) no Reasons included non-responsive principal investigators (PIs), and PIs not involved in recruitment.
Days from NOGA to first accrual (auto-calculated field)	25 (89)	median = 203 days min = 25 max = 380
Research phase	28 (100)	Mode - “other” specifications included “PI did not indicate,” “longitudinal observational study,” “behavioral,” and “pilot.” 2 (7) Phase I 0 (0) Phase I/II 4 (14) Phase II 0 (0) Phase II/III 2 (7) Phase III 0 (0) Phase III/IV 0 (0) Phase IV 0 (0) early (for devices) 11 (39) other 9 (32) not specified
Type of institutional review board review required	26 (93)	16 (61) full board 8 (32) expedited review 2 (8) no response
Is this a protocol on a rare disease?	28 (100)	0 (0) yes 27 (96) no 1 (4) don’t know
Study population (check all that apply)	28 (100)	21 (75) adult 6 (21) children 1 (4) pregnant women 0 (0) prisoners/incarcerated individuals 3 (11) individuals with impaired decision making 0 (0) special social/ethnic groups
Are there any sex exclusions?	26 (93)	21 (82) yes 1 (4) no 2 (7) don’t know 1 (4) no response
Are there any age exclusions?	27 (96)	24 (90) yes 1 (4) no 1 (4) don‘t know 1 (4) no response
Is this a multisite study?	28 (100)	11 (39) yes 16 (57) no 1 (4) don’t know 0 (0) no response
Is this an FDA- regulated study?	28 (100)	3 (10) yes 24 (86) no 1 (4) don’t know 0 (0) no response
Investigational new drug (IND) or investigational device exemption (IED)	28 (100)	4 (14) yes IND 0 (0) yes IED