The Rockefeller University Graduate Tracking Survey System

Michelle Romanick, B.A.¹, Kwan Ng, B.S.², George Lee, B.S.², Matthew Herbert, B.S.², and Barry S. Coller, M.D.³

Abstract

Background: It is essential to track the careers and accomplishments of the graduates of translational research training programs to assess the impact of the programs and to improve them. The major obstacle is the lack of a convenient method to collect the information in a comprehensive and standardized manner.

Methods: We have developed a Web-based electronic Graduate Tracking Survey System (GTSS) that prepopulates the graduate’s information on publications, grants, patents, and clinical trials from public data sources, thus insuring a uniform data format, facilitating survey completion, and facilitating the aggregation of data at individual or multiple sites. GTSS questions are designed to assess whether trainees make important contributions that improve human health, and to track related “surrogate” career development indicators of likely future success.

Results: The GTSS has been in use at Rockefeller University since 2011 and has been adopted by 21 other Clinical and Translational Science Award programs.

Conclusions: The GTSS provides an efficient and convenient mechanism to track the graduates of a wide variety of training programs. It has the potential to aggregate standardized data across institutions, thus providing benchmarks for the assessment of individual training programs and data for program improvement. Clin Trans Sci 2014; Volume 8: 326–329

Keywords: trainee career development, education, graduate tracking, translational research

Introduction

One of the important metrics to both assess and improve training programs is the success of the graduates of the program in achieving their career goals. It is crucial, therefore, for training programs to have a mechanism to track the career achievements of their graduates. In fact, the recent Institute of Medicine report on the Clinical and Translational Science Award (CTSA) program emphasized the importance of going beyond a limited number of conventional measures of success (e.g., number of trainees, conversion from training to independent investigator grants, number of publications, and number and types of degrees completed), to assessing “the professional career trajectory” as part of a comprehensive evaluation program.¹ The value of such reporting is underscored by the important insights about M.D.-Ph.D. degree programs derived from the study of Brass et al., who provided data on the careers of graduates of 24 M.D.-Ph.D. programs.² This survey was only a single snapshot, however, and required integrating data from program-specific questionnaires and Website searches of alumni and public databases. Data were available from only approximately 45% of trainees and some programs could not supply all of the requested data.

There are, however, a number of challenges in achieving this goal, including: (1) defining the core information to track, (2) developing a tool to reach an ever-growing number of graduates who are geographically dispersed, (3) standardizing the format for collecting information such as publications, grants, patents, and participation in clinical trials so that the data from multiple trainees can be aggregated, (4) minimizing the time and inconvenience required to provide the information so as to maximize the response rate from busy graduates, (5) providing flexibility to track additional information for a limited or extended period of time, (6) making the tool sufficiently versatile so that it can track trainees in different training programs at a single institution as well as trainees in programs at multiple institutions, (7) choosing the optimal methods to analyze and display the data. In addition, since aggregating data from different institutions would allow for benchmarking and the potential to identify best practices, it would be desirable for the system to be designed so that it could be easily modified and adopted by multiple institutions. In the case of the CTSA program, this would also facilitate reporting to National Institutes of Health (NIH), Congress, and other interested parties on the impact of the program’s training programs.

This paper describes the Rockefeller University Graduate Tracking Survey System (GTSS), which was designed to meet these goals, along with data from our program’s graduates and information on other institutions that have adopted the system. While we have focused on CTSA training programs, we believe that the system is also suitable for tracking graduates of many other training programs, including M.D.-Ph.D. programs, medical and surgical residencies, medical and surgical fellowships, and a range of career development programs sponsored by government and philanthropy.

Methods

The anatomy and functional aspects of the GTSS

Since the GTSS was designed to address the challenges posed above in tracking graduates, it is convenient to describe its design elements in relationship to each of the challenges.

Defining the core information to track

The philosophical basis for selecting questions to include in the GTSS is that the essential criterion to judge the success of a translational science training program is whether graduating trainees go on to improve human health. As a result, many questions are designed to assess this directly. However, since there is almost always a time lag between when a trainee completes a training program and when she or he improves human health, other questions are designed to assess “surrogate” indicators that may provide valuable interim measures of likely success. The initial set of questions developed at Rockefeller underwent broad review.

¹Hospital Administration, Rockefeller University, New York, New York, USA; ²Department of Information Technology, Rockefeller University, New York, New York, USA; ³Laboratory of Blood and Vascular Biology, Rockefeller University, New York, New York, USA.
Correspondence: Michelle Romanick (mromanick@rockefeller.edu)
DOI: 10.1111/cts.12238
and modification by CTSA training program directors, initially from the New York area, and then throughout the CTSA program.

**Developing mechanisms for reaching an ever-growing number of graduates who are geographically dispersed**

We chose a Web-based electronic infrastructure that accommodates email-based reminders to facilitate communication with trainees throughout the world. The GTSS uses LAMP technology, which is a combination of free, open source software that includes Linux, Apache HTTP Server, MySQL, and PHP. Linux is the operating system installed on the GTSS server. Apache is the Web server software program used to deliver the Web service. All GTSS data are stored using MySQL database software. The entire GTSS application is coded using the PHP programming language. An email list management application is incorporated into the system, including a unique profile for each graduate (username, email address, password, and institution). The graduates are grouped according to the training program in which they participated and the year of their graduation. The email management system is linked to a survey scheduling system that automatically sends the proper survey at the appropriate time.

Since the GTSS collects trainee-specific information, we submitted the GTSS survey to the Rockefeller Institutional Review Board (IRB). Based on the recommendations of the IRB, the first screen that our graduates see each time they enter the GTSS is an informed consent form that requests their approval before proceeding.

**Standardizing the format of trainee information that would benefit from aggregating, including publications, grants, patents, and participation in clinical trials**

The GTSS application is a "smart" system in that it prepopulates fields by querying data from four external databases (PubMed, NIH RePORTER, PatentLens, and ClinicalTrials.gov) and displays the results matching the survey respondent's name as a list with checkboxes on the core survey. As a result, survey respondents only have to confirm the information by checking the appropriate boxes.

For PubMed publications, the latest NCBI data (http://www.ncbi.nlm.nih.gov/books/NBK25501/) are queried in real time. The results, in XML format, are parsed using "PHP Simple HTML DOM Parser."

For patent information, real-time data are obtained from PatentLens (http://www.patentlens.net/) via an RSS feed. Results are returned in XML format and then parsed as above.

For clinical trials information, XML data files originating from the NIH Clinical Trials database (http://clinicaltrials.gov) are downloaded manually every 6 months and then parsed to a local database. GTSS queries are run against the Rockefeller University local database.

Finally, for NIH grants, CSV data files originating from NIH RePORTER (http://projectreporter.nih.gov/) are downloaded manually to a Rockefeller University local database using a MySQL command line every 6 months. GTSS queries are then run against the local database.

**Minimizing the time and inconvenience required to provide the information so as to maximize the response rate from busy graduates**

Two elements were designed to minimize the time and inconvenience of completing the survey. The first is the "smart" design, which saves the responder from having to collect and display information on PubMed publications, patents, registered clinical trials, and NIH grants. The second is splitting the system into an initial survey, which includes demographic data that will not change over time, and a yearly update survey that focuses only on new information since the last survey was completed. One of the limitations of the "smart" approach is that it requires based on the graduate's name, which is of limited value for individuals

![Figure 1. Publications per Clinical Scholar. Each bar indicates an individual Clinical Scholar. Publications from the time the Clinical Scholar entered the Clinical Scholars Program at Rockefeller University are shown.](image-url)
felt are required for assessing graduates’ career development and achievements. Since one of the potential goals of the system is to facilitate aggregating data across institutions to permit benchmarking, the core questions cannot be modified. At the same time, we realized that there may be good reasons to ask supplementary questions to obtain additional information, either for a limited or extended period of time. For example, as part of a research project, an investigator may want to query graduates about a particular aspect of their training or achievements relating to mentoring that is not included in the core questions. This could be accommodated by adding a question for a limited period of time. Alternatively, since we designed the GTSS so that it could be readily adopted by multiple institutions, we recognized that some programs may wish to add questions in an ongoing way that reflect unique aspects of their curricula or the competencies they set as goals. Thus, we included a “build-a-survey” application in the system so that each institution can modify the survey by adding additional questions to the core survey. An online manual for system administrators describes in detail how to use this application.

Choosing the optimal methods to assess survey completion and to analyze and display the data
The email list allows an administrator to identify trainees who have not started the survey, those who have started but not completed the survey, and those who have completed the survey. The data in the MySQL database can be downloaded, analyzed, and displayed using a variety of programs. All of the raw data can be exported and aggregated, including “hidden data” contained in the downloads from the public databases. For example, although not displayed in the GTSS, data from RePORTER include the Congressional district in which a study is conducted, and that information may be valuable for some reporting purposes.

Providing versatility so that the instrument can track trainees in different training programs and at different institutions
The system is designed to be easily modified to allow for tracking trainees or graduates from multiple training programs at a single institution, and for tracking graduates at different institutions. The online manual contains information on how to make the modifications. For example, we use the GTSS to track the graduates of our Master’s degree program, our Certificate in Clinical and Translational Science program for basic scientists, and our medical student research program.

Providing flexibility to track additional information for a limited or extended period of time
The GTSS was designed with a core set of questions that represents the cardinal data that the training program directors with common names. There are ways of enhancing the targeting of queries for such individuals by incorporating additional information, such as the individual’s institutional affiliation, but these need to be customized and may still be less robust. There are a number of initiatives to assign unique personal identifiers to individuals so that the authors of publications can be unequivocally assigned, and if any succeed in achieving wide adoption, they can be incorporated into the GTSS.

Figure 2. Total number of publications by Clinical Scholars by year of publication.

Figure 3. Number of grants awarded to Clinical Scholars by year that the grant began.
The system can provide a detailed report on a graduate’s publications, grants, inventions, clinical trials, patents, and approved drugs, devices, and diagnostics. Additional data reports are under development. As an incentive for graduates to complete the survey, they can download the completed survey in Word format or in a curriculum (CV) format.

Results and Discussion

GTSS data from the Rockefeller Clinical Scholars Program

Since the beginning of our CTSA-supported Master’s degree Clinical Scholars Program was initiated in 2006, 27 individuals have graduated, and 26 of the graduates have completed the survey (95%), including 5 who just recently graduated in June 2014. Figure 1 presents data on the 22 Clinical Scholars who have published at least one paper after joining the program. The wide range of number of publications per Scholar reflects our selecting some individuals with extensive research experience and others with relatively little research experience who are still completing medical training. It also reflects differences in when Scholars graduated and our requirement that Clinical Scholars design and conduct a human subjects protocol under a senior mentor as the central education component of the program since such studies sometimes require a long period of time to gain IRB approval, recruit participants, conduct the laboratory assays, and analyze the data. Figure 2 provides aggregated data on the number of Scholar publications by year of publication. Figure 3 provides data on the number of grants awarded to Clinical Scholars by year of grant start date, and Figure 4 provides data on the number of clinical trials registered in clinicaltrials.gov on which Clinical Scholars are listed as investigators.

Adoption of the GTSS by other institutions

We have offered the GTSS, along with the online manual and personalized training, to other institutions participating in the CTSA program free of charge. At present, 21 other CTSA have adopted the system. We recently queried 17 administrators overseeing the system in different institutions about their experience in implementing and deploying the system. Of the nine administrators who responded, eight indicated that they would recommend the GTSS to other administrators.

Future opportunities

The GTSS has the potential to aggregate data from multiple institutions for reporting to NIH and the public. These aggregated data would represent national benchmarks for comparing the results of individual training programs and identifying best training practices and opportunities to improve training programs. As the current users gain more experience with the system, we plan to discuss with them the benefits and potential drawbacks of data aggregation. The GTSS might also prove valuable as a research tool. For example, to define the specific value added by CTSA training programs, the GTSS could be used to simultaneously track for comparison a control group of trainees who did not enroll in a CTSA training program.

We conclude that the GTSS provides a robust electronic infrastructure for assessing the success of individual CTSA programs in training the next generation of clinical and translational scientists. It also has the potential to provide a global assessment of groups of training programs that can be used to analyze and improve them.

Conflict of Interest

The authors reported no conflicts of interest related to this study.

Acknowledgments

This publication was supported in part by grant UL1 TR000043 from the National Center for Research Resources and the National Center for Advancing Translational Sciences (NCATS), a part of the NIH. We thank the Education and Career Development Key Function Committee and CTSA Principal Investigators who provided valuable comments and suggestions in developing the GTSS. We thank Dr. Daniel Rosenblum of NCATS for his insightful comments and support for this project, and thank Dr. Melissa Begg of Columbia University for her detailed review of GTSS and valuable suggestions in the development of the tracking system.

References

1. IOM (Institute of Medicine). The CTSA Program at NIH: Opportunities for Advancing Clinical and Translational Research. Washington, DC: The National Academies Press; 2013.

2. Brass LF, Akabas MH, Burnley LD, Engman DM, Wiley CA, Andersen OS: Are MD-PhD program meeting their goals? An analysis of career choices made by graduates of 24 MD-PhD programs. Acad Med. 2010; 85(4): 692–701.