Measuring the evolution and output of cross-disciplinary collaborations within the NCI Physical Sciences–Oncology Centers Network

Jodi E. Basner1, Katrina I. Theisz2,3, Unni S. Jensen1, C. David Jones1, Ilya Ponomarev1, Pawel Sulima1, Karen Jo3, Mariam Eljanne3, Michael G. Espey3, Jonathan Franca-Koh3, Sean E. Hanlon3, Nastaran Z. Kuhn3, Larry A. Nagahara3, Joshua D. Schnell1 and Nicole M. Moore3,*

1Discovery Logic, a Thomson Reuters business, 2Kelly Government Solutions, Rockville, MD 20852 and 3Office of Physical Sciences – Oncology, Center for Strategic Scientific Initiatives, Office of the Director, National Cancer Institute, Bethesda, MD 20892, USA
*Corresponding author. Email: nicole.moore@nih.gov.

Development of effective quantitative indicators and methodologies to assess the outcomes of cross-disciplinary collaborative initiatives has the potential to improve scientific program management and scientific output. This article highlights an example of a prospective evaluation that has been developed to monitor and improve progress of the National Cancer Institute Physical Sciences—Oncology Centers (PS-OC) program. Study data, including collaboration information, was captured through progress reports and compiled using the web-based analytic database: Interdisciplinary Team Reporting, Analysis, and Query Resource. Analysis of collaborations was further supported by data from the Thomson Reuters Web of Science database, MEDLINE database, and a web-based survey. Integration of novel and standard data sources was augmented by the development of automated methods to mine investigator pre-award publications, assign investigator disciplines, and distinguish cross-disciplinary publication content. The results highlight increases in cross-disciplinary authorship collaborations from pre- to post-award years among the primary investigators and confirm that a majority of cross-disciplinary collaborations have resulted in publications with cross-disciplinary content that rank in the top third of their field. With these evaluation data, PS-OC Program officials have provided ongoing feedback to participating investigators to improve center productivity and thereby facilitate a more successful initiative. Future analysis will continue to expand these methods and metrics to adapt to new advances in research evaluation and changes in the program.

1. Introduction

The shift in science toward collaborative research that crosses traditional disciplines has led to a need for more comprehensive evaluations for the impact of cross-disciplinary collaborations on research productivity and outcomes (Wuchty et al. 2007; Klein 2008). Government research agencies are increasingly investing resources to establish and maintain collaborative cross-disciplinary initiatives that target high-risk problems and bring new perspectives to traditional research (Hall et al. 2012). Supporting and facilitating these collaborative cross-disciplinary teams can be expensive and labor intensive.
Hence, it is important to identify and understand the conditions that strengthen or hinder effective cross-disciplinary interactions to increase investment value.

Due to the complexity of collaborative team networks, social network analysis metrics and network science approaches have emerged as tools for understanding and improving collaborative team science programs (Nagarajan et al. 2011). Network tools typically applied to biological pathways, air traffic control patterns, and the Internet have been found to be relevant for understanding patterns and mechanisms behind scientific collaborations. Social network analysis provides insights into the topological structure, central nodes, statistical properties, and generating mechanisms of scientific networks and can identify areas for strategic integration to increase connectivity. For example, a large study investigating authorship collaboration networks found that biomedical collaboration networks are dominated by many people with few collaborators, whereas other fields, such as physics, are dominated by few with much collaboration (Newman 2001).

Despite the insights gained from network analysis tools, the application of network analysis approaches still has several limitations. First, the definition of collaboration can differ considerably between network participants and therefore analysis is typically limited to just authorship collaborations (Newman 2001; Wuchty et al. 2007). Evaluating only authorship collaborations within a network does not reflect the richness of actual network dynamics over time. Second, there is often a lack of information and methods available for assigning the discipline or career stage of collaborators within the network, making it difficult to examine specific network characteristics, such as the evolution of cross-disciplinary collaborations or the role of young investigators within the network. Collection and analysis of more comprehensive datasets that include all collaborations, disciplines, and seniority can enable several of these questions to be explored. The National Cancer Institute (NCI) Office of Physical Sciences – Oncology (OPSO) has assembled a comprehensive dataset to facilitate robust analysis through a new cross-disciplinary program.

The NCI Physical Sciences – Oncology Centers (PS-OC) Program started in the fall of 2009 and comprises a virtual network of over 260 investigators in 12 centers that is dedicated to building cross-disciplinary teams and infrastructure to better understand and control cancer through research at the intersection of the physical sciences and oncology. In the planning phase of the program, the OPSO built infrastructure for prospective program evaluation allowing program officials to assess the PS-OC Program performance on an ongoing basis and to make adjustments that increase productivity. Based on an initial needs assessment at the start of the PS-OC Program, OPSO and Discovery Logic, a Thomson Reuters business, implemented a data collection plan. Every 6 months, each PS-OC submits a progress report, which includes data on collaborations, in-progress publications, leveraged funds, patents, trainees, courses, meetings, and outreach activities in free form or table format as a progress report. The information from the semi-annual progress reports is entered into an SQL database for data standardization, cleansing, and analysis. Additionally, a one-time survey was distributed to investigators to complement the progress report data collection at the midpoint of the program.

To date, data have been collected for 3 years on the PS-OC Program collaborations and research outputs. This data collection has been supplemented and/or validated with additional information from databases, such as Web of Science, to complete the analysis. The disciplines of all trainees and investigators within the PS-OC Program were determined based on progress reports, survey data, and development of an automated classification algorithm. Therefore, it was feasible to conduct detailed analysis of the evolution of the cross-disciplinary collaborations and outputs within centers and across the PS-OC Network. This network analysis has provided insights into the structure of the program, generating mechanisms of productive collaborations, such as pilot projects, and identified areas for both investigators and program officials to provide strategic directions to increase connectivity in the PS-OCs. This type of evaluation may act as a guide for future collection of progress report information and analysis of large cross-disciplinary programs.

2. Methods

2.1 Data collection and preparation

2.1.1 Progress reports. Semi-annual progress reports provided the majority of data for the PS-OC program grant years (September 2009 to current year). The progress reports are subdivided into five sections: overall center progress, research projects, cores, education and training unit, and outreach and dissemination unit. Within each section, the investigators provide information in a series of subsections that are either free form or in table structure, including collaborations (planning, forming, or executing), in-progress publications, peer-reviewed publications, leveraged funds, patents, trainees, courses, meetings, and outreach activities.

2.1.2 Publication identification methods. A list of publications authored by PS-OC investigators during the pre-award years 2006–08 was generated by name matching algorithms augmented by given author metadata including email addresses to mine ScienceWire®, a merged database of Web of Science and MEDLINE. Publications were identified with a high degree of accuracy that was determined with a precision/recall analysis (92% and
85%, respectively). Post-award publications for the years 2009–12 were extracted from data reported in the progress reports. Note that the PS-OC Program publications may have cited additional funding support and may not necessarily be attributed solely to the PS-OC Program funding.

2.1.3 Software utilized for data collection and analysis. Data were loaded into a suite of Microsoft® SQL Server™ databases, and T-SQL queries were used to conduct analyses. Microsoft Full Text Indexing features were used for extracting words from publication titles and abstracts for the custom analytics.

2.1.4 Interdisciplinary team reporting, analysis, and query resource. The interdisciplinary team reporting, analysis, and query resource (iTRAQR) system was developed to collect and organize the network information reported by participants during the program life cycle. The foundation for the iTRAQR system is a robust data model that conforms to the explicit and implicit structure of the PS-OC semi-annual progress report. To handle cross-center and time series analysis, the data model supports deduplication of reported data and relationships. The iTRAQR’s web-based data entry user interface was developed to facilitate entry of progress report data into a database conforming to the specifications of the data model. When entering center publications into iTRAQR, it is possible to search and link to databases such as MEDLINE and Thomson Reuters Web of Science. iTRAQR was developed as a Microsoft ASP.NET MVC web application, leveraging the entity framework object/relational mapping technology and the Microsoft SQL Server database system. The system complements existing systems available, such as the National Institutes of Health (NIH) Reporter (http://projectreporter.nih.gov/reporter.cfm), used to track NIH grant publications. But, iTRAQR is unique in providing the ability to analyze, search, and interact with information imbedded into the progress reports submitted by investigators.

2.1.5 Survey and self-reported discipline. At the midpoint of the 5 year PS-OC Program, program officials conducted eight anonymous, web-based surveys via Key Survey (WorldAPP, Braintree, MA, USA) consisting of 7–32 questions directed to eight specific respondent populations: principal and senior scientific investigators (PI/SIs), project leads and investigators, trainees, administrators, education unit leaders, outreach unit leaders, advocates, and external scientists (OMB clearance number: 0925-0642). Out of 919 potential respondents, 262 (28.5%) completed the survey. For the purposes of this study, only the results from the investigator group (principal, senior scientific, and project investigators, n = 64) and trainees (n = 75) were included. A primary focus of the survey was cross-disciplinary collaboration, with an emphasis on collaborative group formation (e.g. the number of scientists involved in collaborations and their roles), outputs of collaborations, and barriers to successful collaborations. As part of the survey, respondents were asked to self-identify their field of study or training. The types of disciplines provided were the ones most commonly found within the PS-OC network, and respondents were allowed to ‘select all that apply’. Based on these responses, investigators were characterized as physical scientists or cancer researchers. Of the investigator group, two respondents did not self-identify their disciplines, and were excluded from the results. For those who chose disciplines in both physical sciences and cancer research (20 investigators), the classification was based on the number of physical sciences or cancer research disciplines selected and the investigator was assigned to the category (physical scientist or cancer researcher) with the higher number of disciplines selected.

2.2 Social network analysis

2.2.1 Automated discipline assignment method. Due to the large number of program participants and the lack of a self-reported discipline for all network individuals, program investigators and trainees were categorized into disciplines using a novel automated classification algorithm. A three-step method was developed to automatically assign each PS-OC investigator to a major research field (physical sciences or cancer research) based on aggregated information collected from reported publications. First, standard Web of Science Journal subject categories (JSCs) were manually divided into six broad categories relating to the PS-OC research program (physical sciences, life sciences, medicine, oncology, multidisciplinary, and other). Second, each publication was assigned to a broad category based on the JSC of the publication and the JSCs of the references in the publication. These publication assignments were used to calculate a weighted discipline for each investigator. Using an automated ranking procedure, each investigator was assigned as a physical scientist or cancer researcher based on their weighted discipline value. This method was validated on a sample of investigators using self-reported discipline information and was found to be the most comprehensive for categorizing disciplines permitting inclusion of the largest number of investigators within the PS-OC Program.

2.2.2 Measuring intradisciplinary and cross-disciplinary collaborations. A pre- and post-award analysis of collaborations was conducted using the pre-award publications identified above and post-award publications reported in the progress reports. Utilizing the self-reported disciplines for key investigators and their
publications, the number of each collaboration type, intradisciplinary or cross-disciplinary, was measured and graphed. In the context of this evaluation, intradisciplinary indicates a collaboration either among physical scientists only or cancer researchers only to be distinguished from a cross-disciplinary collaboration among physical scientists and cancer researchers. It is important to keep in mind that the broader disciplines of physical sciences and cancer research are each composed of more closely related disciplines. In another context a collaboration among the disciplines within physical sciences or cancer research may be interpreted as multidisciplinary. Network graphs were generated using NodeXL (Smith et al. 2010).

2.2.3 iTRAQR custom analytics. iTRAQR’s network graph functionality was developed using the Cytoscape Web network visualization library (Smoot et al. 2011), as well as the NodeXL Class Libraries for calculation of graph metrics. Network analysis information captured from the progress reports includes self-reported collaborations, within project collaborations, and collaborations represented through publication co-authorship. iTRAQR supports the capability of visualizing PS-OC collaboration networks in multiple ways, including by progress report period, flexible date range, single or combined collaboration type, by individual PS-OCs or the full network. Interactive capabilities include clicking on nodes to obtain investigator details, clicking on edges for a breakdown of collaborations by type, and manipulating and exporting graph images. Network metrics, such as betweenness centrality, degree, and density, can be exported along with graph details for further analysis.

2.3 Bibliometrics and custom metrics

2.3.1 Field normalized bibliometric percentile. The Bibliometric Percentile is a publication level metric that measures field-specific citation patterns. For this analysis, the 1-year citation counts of all articles published within the same journal, with the same document type and within 6 months of a PS-OC publication were used as a set of field reference publications. The Bibliometric Percentile represents the percentage of reference publications that a PS-OC publication is greater than or equal to based on the citation count.

2.3.2 Field convergence. A set of physical sciences and cancer research terms were curated by analyzing term occurrence in the titles and abstracts of a set of 100,000 publications for each field. Field publications were identified by first mapping Web of Science JSCs to each field and then selecting a random sample of 100,000 publications between the years 2000 and 2008. To be able to identify strong physical sciences and cancer research terms, term occurrences in the physical sciences and cancer research publications were reduced by their occurrence in a set of 100,000 publications randomly selected from biology JSCs. A final manual review resulted in a set of 1,643 physical sciences terms and 571 cancer research terms. An average publication was classified with seven terms that fell into one of these sets. If a PS-OC publication’s title or abstract included \( \geq 1 \) of the physical sciences terms and \( \geq 1 \) of the cancer research terms, then the publication was classified as cross-disciplinary. Increasing these cutoffs to two or three terms did not change the overall trends observed.

3. Results

3.1 Categorizing investigators by discipline

The PS-OC Program has supported over 600 trainees and 350 investigator-level scientists over a 3-year period. For program evaluation, investigators and trainees were categorized into two disciplinary groups, physical scientists or cancer researchers, using several methods of data collection and automated classification algorithms. Information on trainees is self-reported by each PS-OC every 6 months through the progress report. The PS-OC Network trainees comprise \( \sim 75\% \) physical scientists and 25% cancer researchers (Fig. 1).

The information collected from the survey represents a small cohort of investigators and trainees from the PS-OC Program that responded to the survey. Investigator respondents (n = 64) consisted of 56% cancer researchers and 44% physical scientists. The trainee respondents (n = 75) comprised a majority of physical scientists, 55% (Fig. 1). These percentages differ slightly from the progress report data because they only represent a subset of the full list of trainees. Additional PS-OC Program researchers were categorized into discipline using the automated

![Figure 1](image-url). Categorization of PS-OC Program investigators and trainees into two groups, physical scientists (black) and cancer researchers (gray), using three different methods: (a) progress report data on trainees, (b) an automated classification algorithm inferring discipline based on publication history, and (c) surveys of both investigators and trainees.
classification algorithm in order to categorize a larger group of investigators. Based on this algorithm, the program comprises 51% cancer researchers and 49% physical scientists (Fig. 1). The results of the algorithm were validated using manually assigned or self-reported disciplines of 262 key investigators.

3.2 Characterization of cross-disciplinary collaborations

3.2.1 Evolution of cross-disciplinary collaborations. To analyze changes in cross-disciplinary collaborations before and during the PS-OC award, investigator publications from 3 years prior (2006–08) to the program initiation identified in ScienceWire were compared with publications during grant years collected in iTRAQR. Utilizing the self-reported disciplines for 262 PS-OC investigators and the author list for each of the 3,093 pre-award and 779 post-award publications, the number of pairwise intradisciplinary and cross-disciplinary collaborations during each time period were determined (Fig. 2). The publication analysis showed that there was a 16% increase in the proportion of cross-disciplinary collaborations from pre- to post-award years. Increased cross-disciplinary collaborations by both physical scientists and cancer researchers contributed to the overall increase (Fig. 2C). Additionally, within the first 3 years of the program, 40% of the 262 investigators have produced a PS-OC Program publication with a cross-disciplinary authorship collaboration (Fig. 2B).

Network graphs generated in iTRAQR were used to measure changes in collaborations since the start of the PS-OC award. Disciplines for the 262 key PS-OC investigators, 47 participants in the broader PS-OC network (assigned using the automated classification algorithm), and the self-reported disciplines of 677 trainees involved in the PS-OC Program over 3 years were used to color code the graphs. Figure 3A displays collaborations for one center that were reported in each of the three grant years. Strong growth is indicated in this center’s network from 2010 to 2012 with three or four centralized collaboration activities involving interactions between physical scientists and cancer researchers. Within this one center from 2010 to 2012, there are 133 new people, 645 more collaborations, and 91 more cross-disciplinary collaborations.

Investigator level collaboration metrics that can be exported from iTRAQR include the degree and, betweenness centrality in addition to other characteristics of the researcher such as their discipline and role (Freeman 1979). The results indicated that the core collaboration structure of this PS-OC appeared to stabilize after 2 years. In 2011 and 2012, the same top 10 key investigators were identified when researchers were ranked by their degree metric. An increased presence of outside network

Figure 2. (A) Authorship collaboration network graphs. Nodes represent physical scientists (gray) or cancer researchers (black). Edges represent an intradisciplinary collaboration within physical scientists (gray) or cancer researchers (gray) and cross-disciplinary collaborations between physical scientists and cancer researchers (black). The edge width indicates the number of collaborations measured for a pair of scientists. (B) PS-OC investigators involved in cross-disciplinary authorship collaborations. Percentages of the 262 PS-OC (134 physical scientists, 128 cancer researchers) investigators participating exclusively in intradisciplinary collaborations or participating in cross-disciplinary collaborations pre- and post-award. (C) Pairwise collaborations. Percentages of intradisciplinary collaborations [physical scientists (PSs) only or cancer researchers (CRs) only] or cross-disciplinary collaborations out of the total number of pairwise collaborations during the pre- and post-award years.
investigators playing key roles in this center’s collaboration dynamics is indicated by a transition from zero to three outside network investigators among the top 20 key investigators from 2011 to 2012 when researchers are ranked by their betweenness centrality metric (data not shown). The graph density (0.04) remained the same in 2011 and 2012 with a concurrent 72% increase in the number of collaborations (584 in 2011 and 1,008 in 2012). These trends were found to be consistent across seven PS-OCs.

Similar network diagrams were compiled for the PS-OC Network of 12 centers. To understand the influence of physical scientists and cancer researchers within the network, an analysis of the betweenness centrality values of the top 100 physical scientists and 100 cancer researchers nodes was performed after 6 months (2010) of the program and at 3 years (2012) (Fig. 3B). Betweenness centrality (or betweenness) is a measure of a node’s centrality in a network diagram. It describes both the load and importance of a node within the network. A higher load indicates that the node is more essential for forming the network. If all nodes were equally connected, the betweenness for every node would be one. For this analysis, betweenness was normalized to the highest value in the network to account for its growth. At 6 months, the network diagram captures the initial collaboration structure of the PS-OC Network as designed in the research applications. The betweenness centrality values show that the network interactions were initially structured around 10 key physical scientists and 10 key cancer researchers. The normalized betweenness centrality of other investigators (numbers 11–100) were below 0.1 and several investigators had betweenness values of zero indicating no network connectivity. After 3 years, the influence of nodes within the network has evolved and differs by discipline. More physical scientists appear to be key nodes within the network (25 investigators with normalized betweenness >0.25) and the overall betweenness for physical scientists has increased as the network becomes more integrated. In contrast, the betweenness for cancer researchers has remained the same with a modest growth. These data may suggest that more physical scientists have evolved to be vital nodes of the PS-OC Network.

3.2.2 Number of investigators in cross-disciplinary collaborations. PS-OC participants were asked to identify a successful cross-disciplinary collaboration that they were involved in as part of the PS-OC Program and to indicate how many participants were involved in the collaboration (Fig. 4A). The data illustrated a bimodal distribution with a small peak of two-person collaborations and a larger peak of four- to seven-person collaborations. Physical scientists and cancer researchers reported slightly different size collaborations. Cross-disciplinary collaborations from the perspective of the cancer researchers were found to be slightly larger (5–7 investigators) than the collaborations reported by the physical scientists. These differences in the number of investigators per collaboration by discipline may be explained by inherent differences in how each discipline defines a collaborator.

The survey results were compared with the progress report data on collaborations formed and executed and subsequent authorship collaborations. For all types of reported collaborations between 2010 and 2012, the majority was found to involve two people. Over 70% of intradisciplinary collaborations consisted of just two people; whereas, cross-disciplinary collaborations had slightly increased numbers of investigators. Approximately 40% of cross-disciplinary collaborations comprised two people and 60% comprised three or more people.
Therefore, cross-disciplinary collaborations appear to include more investigators per collaboration and the collaborations considered ‘successful’ by the investigators include approximately 4–7 people.

### 3.2.3 Roles of investigators in cross-disciplinary collaborations

Three distinct clusters of roles were identified through the survey responses (Table 1). The first cluster of roles involved leadership and project direction and were mainly taken on by PS-OC principal investigators and to a lesser extent by project investigators. The second cluster of roles involved providing critical technologies and reagents for a collaboration, and these were shared more or less equally among all three groups. The third cluster of roles involved data analysis and project participation, which was primarily fulfilled by trainees and to a lesser extent by project investigators.

### 3.3 Analyzing collaboration outputs

The survey of PS-OC Program investigators helped to define the outcomes of the PS-OC cross-disciplinary collaborations and to identify which components of the collaborations were essential for success. Participants were asked to identify all outcomes from a ‘successful’ cross-disciplinary collaboration. Nearly 80% of respondents indicated that the collaboration was ongoing and had generated new knowledge, suggesting that the majority of collaborations has been productive. Additionally, 40–50% of respondents indicated that collaborations have led to publications, presentations, and new projects or directions, providing some concrete measures of success. Participants were asked to identify critical aspects of the collaboration that were required to facilitate the outcomes described in Table 2. More than 70% of respondents indicated that cross-disciplinary collaborations and support from the PS-OC Program were essential to achieving the outcomes and nearly 60% of respondents indicated that all team members played an essential role. Only 10% or fewer of the respondents indicated that each of these collaboration attributes were not essential for success. These results suggest that while a fraction of the collaborations have measurable outputs, such as publications, grants, and patents, most have generated knowledge that may lead to a measurable output over time.

The measurable outputs of PS-OC Program collaborations have been monitored through progress report data analysis using iTRAQR. The number of publications per collaboration was analyzed by project type (Fig. 5). The PS-OC Program offers a number of different project types and mechanisms to initiate collaborations and each may lead to different outputs. The majority of collaborations stem from stable, long-term research projects and cores that are structural components of each PS-OC. Analysis showed that projects and cores produced about one publication per pairwise collaboration in 3 years, which is similar to findings reported by survey respondents, suggesting that ~70% of these collaborations have resulted in a publication. The measurable outputs of the projects vary for the high-risk pilot and trans-network projects that are funded annually with restricted set-aside funds within the network. Trans-network projects produced over three publications per collaboration and pilot projects produced <0.5 publications per collaboration. Trans-network projects bring together expertise from several PS-OCs, often consist of large groups (4–10 investigators), and receive more funding than pilot projects. In addition, pilot projects usually support young investigators within the PS-OCs. It is possible that the larger teams, more experienced investigators, and funding within the
trans-network projects are leading to higher publications per collaboration. Additional time is required to determine if there is a measurable difference in the overall impact of the publications produced by the various types of projects.

Future studies will examine related other outputs of these project types, in addition to publication number and impact, such as the impact of young investigator involvement in pilot projects on career development. Even though young investigators may not produce as many high impact publications, one may posit that the pilot projects themselves may be beneficial to their professional future, as not every program allows for such development of young investigators. The next phase of the evaluation will look at the long-term impacts of these pilot projects on career transitions and funding of young investigators. Data will also be analyzed to account for publications per investigator to access relative impact on investigators. These analyses yield insights to program officials about the impact of each mechanism on the program.

### 3.4 Impact of cross-disciplinary research collaborations

One key metric of the PS-OC Program is whether the reported cross-disciplinary collaborations actually result in cross-disciplinary research findings as reported in the publications. About 56% of the cross-disciplinary authorship collaborations have resulted in publications with cross-disciplinary content (Fig. 6). However, intradisciplinary collaborations are also producing publications with cross-disciplinary content. About 48% of the intradisciplinary collaborations among physical scientists and 24% among cancer researchers have produced publications with cross-disciplinary content.

As a measure of the impact of these research findings, the average Journal Impact Factor of the cross-disciplinary publications produced by cross-disciplinary and intradisciplinary cancer research collaborators is.
comparable, 11.1 and 10.8, respectively. However, the impact factor of the publications produced by intradisciplinary physical scientists is considerably lower (6.3). In contrast, the average bibliometric percentiles, a field-normalized bibliometric that compares a publication to a set of reference publications in the same field and time period, of the cross-disciplinary content publications are 69%, 67%, and 67%, respectively, for cross-disciplinary collaborations, intradisciplinary physical scientists, and intradisciplinary cancer researchers indicating that all three groups are performing similarly relative to their fields. Taken together, these data suggest that cross-disciplinary research findings are published in higher impact journals when both physical scientists and cancer researchers are authors compared to physical scientists alone. Cancer and biomedical journals on average have higher impact factors than physical sciences-based journals, which may contribute to the measured increase in impact factor. These data demonstrate progress toward attaining a program goal to increase the dissemination of physical sciences in oncology research into the cancer research community and related journals.

3.5 Challenges to forming and tracking productive collaborations

Survey respondents were asked to choose the most significant challenge in establishing effective cross-disciplinary collaborations and to assign a severity score (1–5, with 1 being the least severe to 5 being the most severe) to each of 17 presented challenges (Fig. 7A). The most commonly identified and highest severity concern was a lack of funds to support the collaboration. Other common challenges were related to team interaction and function, including differences in goals, lack of defined roles, and difficulty in communicating across disciplines.

Other challenges faced by the investigators were evident from the semi-annual progress report data. For example, reporting lacked continuity of collaborations in the first 2 years of the PS-OC Program. In the first two reporting periods, only ∼35% of reported collaborations were found to be similar and continued over a 1-year time period (Fig. 7B). The lack of continuous collaborations may indicate challenges in establishing collaborations early in the program, or could in part be attributed to inconsistent reporting by investigators from progress report to progress report. Nevertheless, the number of continuously reported collaborations has increased with time as the PS-OC Program has matured.

4. Discussion

The results collected through a prospective evaluation of the PS-OC Program highlight the evolution and outputs of cross-disciplinary collaborations over the course of 3 years. The insights gained from these data provide information to program officials for strategically designing new mechanisms within the program and informing investigators of center progress toward program goals. In addition, these results compliment and support existing studies investigating the science of team science. To date, several other research programs and scientific networks have been analyzed with techniques used in social networking and bibliometrics. However, several questions remain in the field, including: (1) How are collaborations defined within a developing research network? (2) What is the impact of team size on productivity? (3) Does the cross-disciplinary collaboration affect the type of science generated? and (4) What are the challenges to cross-disciplinary research? Evaluation of the PS-OC Program may help to answer some of these questions by providing new comprehensive datasets, analysis tools, and metrics.
4.1 Defining cross-disciplinary collaborations in a network

Comparative studies of team science are challenging due to the difficulty of effectively measuring and monitoring cross-disciplinary collaborations. A cross-disciplinary collaboration is defined as researchers from different disciplines working together (including hybrids of various disciplines) to generate a shared framework, language, and model(s) (Rosenfield 1992; Stokols et al. 2008a,b; Haines et al. 2011). Most studies measuring cross-disciplinary collaborations only consider co-authored papers that include multiple institutions, which is not a broad definition of collaboration. (Porter et al. 2012). Few studies have defined collaborations more broadly to analyze network interactions through surveys and interviews as has been done for evaluation of the PS-OC Program. As an example, the NCI Cross-disciplinary Research on Energetics and Cancer (TREC) Centers and Cross-disciplinary Tobacco Use Research Centers (TTURCs) characterized scientific collaborations as scientists working together under clearly stated goals (Stokols et al. 2008a).

Among researchers, the definition of collaboration is subjective as anything ranging from several scientific discussions between two scientists which inform their latest work to a long-term project between many investigators in geographically dispersed labs. For the PS-OC Program, investigators were asked to define these collaborations in phases as either forming, planning, or executing and to provide a short summary of the collaboration. However, as evidenced by analysis of progress report data in iTRAQr, there were often discrepancies in reported collaborations because they were not reciprocated by the other said collaborators at different centers. The PS-OC Program evaluation defined a collaboration as two or more investigators working together on a project of any size reported in a progress report or as an authorship collaboration between two investigators. The investigators could be of any discipline and the project did not need to result in a measurable output (e.g., papers and patents). This approach enables investigation of new and developing collaborations in addition to the authorship collaborations. The data points to stabilization of center collaborations after 2 years and highlights the growth of the number of key physical scientists connecting the network over time.

4.2 Measuring evolution of cross-disciplinary collaborations

There are limited numbers of network studies at the investigator level that measure evolving cross-disciplinary collaborations. One study used a network analytic approach, with discipline defined as the field in which the researcher earned their highest academic degree, to compare the interdisciplinary and cross-disciplinary collaboration networks among the top 68 researchers in a Tobacco Harm-Reduction Network (Provan et al. 2008). Although few cross-disciplinary collaborations were reported, they were much more likely to produce publications containing perspectives beyond what the individuals could have developed on their own. The PS-OC Program evaluation has developed the tools and methodology and identified the data necessary to conduct a longitudinal analysis of cross-disciplinary collaborations. Using these techniques, program officials observed the dynamics and stability of developing collaborations and the evolution into robust networks with measurable outcomes. Ongoing analysis is aimed at determining which projects or reported collaborations have resulted in authorship collaborations.
4.3 Measuring impact of team size and collaboration types on productivity

Several reports have investigated the effect of intradisciplinary and cross-disciplinary team size on research productivity and impact. For example, the number of investigators per publication in science and engineering fields has increased from two to four investigators over the past 40 years (Wuchty et al. 2007). These teams of scientists have produced publications that are more frequently cited than individual researchers. However, the study did not examine the impact of cross-disciplinary versus intradisciplinary collaborations. Another report of a large center scale showed that small to medium groups of up to 50 investigators per center were more productive than centers that had more than 50 investigators, and that interacting with several (>10) investigators was central to a successful collaboration (Rhoten 2013). Most findings of similar evaluations suggest that large cross-disciplinary teams experience an initial lag in production of outputs, such as papers, but tend to produce increased outputs over smaller teams (one to two investigators). Evaluators who assessed the NCI TTURC program used similar means to identify the level of collaboration, productivity, and the impact of the papers produced by the program investigators, such as research networks and the number and impact factor of the publications over time. They found that after an initial lag of 2 years, the amount of publications increased exponentially throughout the project years (Hall et al. 2012). Although the average impact factor of journals with TTURC publications were initially lower than that of comparison groups, there was no significant difference after the first 2 years of the program. On-going analysis will clarify for the PS-OC Program the effective impact trends of team size and heterogeneity on productivity and on the broader research community. Early findings suggest that the larger and more heterogeneous teams from the trans-network projects are producing more publications per collaboration. This finding counters recent results reporting less productivity from heterogeneous compared with homogeneous large teams (Cummings et al. 2013). If these differences persist, it will be important to identify the components of the PS-OC Program that strengthen large heterogeneous team size productivity and impact.

For the PS-OC Program, progress reports indicated that the average number of investigators per team was relatively small. But, when the survey respondents were asked to identify the number of investigators on a ‘successful’ cross-disciplinary collaboration, the team size was found to be larger (5–7 investigators). This suggests that the larger teams were viewed more positively by the investigators and is similar to previously reported findings (Rhoten 2013). In addition, it was determined that the cross-disciplinary teams were on average larger than intradisciplinary teams. As reported in this evaluation, greater than half of the cross-disciplinary co-authorship collaborations resulted in cross-disciplinary publications with cross-disciplinary content (Fig. 6). Challenges of publishing cross-disciplinary research include the following: a lack of sufficient cross-disciplinary journals as a venue for publishing, particular needs around the peer review of cross-disciplinary research, and the longer time required for cross-disciplinary collaborations to mature into publishable results (Kueffer et al. 2007). The impact of the cross-disciplinary publications produced during this initial reporting phase was measured using bibliometric percentiles and indicate that the publications rank in the top third of their respective field. Interestingly, trans-network collaborations produced 3-fold more publications than research projects and core projects alone, raising the question of whether different laboratory perspectives combined with different discipline perspectives further promote productivity. Multivariate statistical analysis will be conducted in the future to better understand the relationships among the PS-OC Program variables.

4.4 Understanding the challenges to cross-disciplinary-based initiatives

The PS-OC Program investigators identified four challenges to cross-disciplinary research: (1) lack of funds; (2) members prioritized their personal goals before the overall team goal; (3) responsibilities, roles, and expectations were not clear; and (4) difficulties in communication across scientific disciplines. The highest rated challenge (lack of funds) is common for team-based research and is not considered unique to cross-disciplinary science. Additionally, prioritization of one’s own goals (challenge 2) may seem more pronounced in cross-disciplinary research because the goals of different groups may be somewhat divergent. The last two challenges, however, are communication difficulties commonly found in cross-disciplinary science (Stokols et al. 2005; Masse et al. 2008; Hall et al. 2012).

There are repeated instances of difficulties in communication across scientific disciplines in cross-disciplinary science outside of the PS-OC Program. For example, evaluators of the TTURC program found that investigators who were geographically closer (within the same room/lab, on the same floor, or in the same building) took less time to start their centers than those whose infrastructure did not allow for much face time (Stokols et al. 2005). The TTURC investigators at the geographically dispersed centers reported that the distance made it more difficult to learn the languages of the other scientific disciplines involved, and that working out administrative responsibilities took longer to resolve as well. Additionally, the time it took for the TTURC investigators to communicate effectively greatly increased with the number of disciplines involved, which added further delays as compared to centers with fewer disciplines. For
the PS-OC Program, a lag was observed in forming continuous reported collaborations, and the survey respondents described communication issues quite similar to those described in the TTURC evaluation. In both programs, this may be attributed to communication barriers at the onset of the collaborations.

4.5 Future analysis of network collaborations
To provide insight into the unique features of the program that impact cross-disciplinary research, future analysis will examine differences among groups comparable with the PS-OC Program. The analysis will include both a comparison with NIH/NCI Research Project Grant (e.g. R01 or P01) awardees who are conducting research on similar topics as well as large center grants (e.g. P50 or U54) with comparable program objectives to generate breakthrough research by facilitating cross-disciplinary collaborations.

5. Conclusion
Comprehensive data sources, analysis methods, and iTRAQ-like information systems are enabling prospective grants program monitoring and evaluation. By tracking progress in near real-time and developing new metrics in line with program goals, program officials and investigators are provided with ongoing feedback to formulate a focused trajectory aligned with program goals and contribute to successful initiatives. The PS-OC evaluation is advancing social network analysis of scientific collaboration networks beyond co-authorships with the incorporation of self-reported collaboration data and providing insights into the dynamics of evolving collaborations from inception. Incorporation of researcher discipline information into the network analysis provides novel measures focused specifically on understanding and improving cross-disciplinary collaborations. Findings from these metrics have provided new insights into generating mechanisms of collaborations, such as an increased emphasis on trans-network projects, and has led to new strategic directions to increase collaborations and productivity by investigators. To date, several PS-OC projects have already changed personnel or structure based on feedback from the network analysis. These findings support previous studies and give edification for improving team science-based initiatives through new reporting and analysis tools.

Acknowledgements
The authors would like to thank James Corrigan and Larry Solomon at the NCI Office of Science Planning and Assessment and Lisa Krueger and Michelle Vos at the NCI Knowledge Management and Special Projects Branch for their discussion and advice in designing and executing the analytics; Brian Zuckerman and colleagues at the Science and Technology Policy Institute for assistance in designing the evaluation plan, and other NCI program officials for their feedback. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

Funding
This work was supported in part by contract funding from the National Cancer Institute at the National Institutes of Health.

References
Cummings, J. N. et al. (2013) ‘Group Heterogeneity Increases the Risks of Large Group Size: A Longitudinal Study of Productivity in Research Groups’, Psychological Science, 24/6: 880–90.
Freeman, L. C. (1979) ‘Centrality in Social Networks Conceptual Clarification’, Social Networks, 1/3: 215–39.
Haines, V. A., Godley, J. and Hawe, P. (2011) ‘Understanding Interdisciplinary Collaborations as Social Networks’, American Journal of Community Psychology, 47/1–2: 1–11.
Hall, K. L. et al. (2012) ‘Assessing the Value of Team Science A Study Comparing Center- and Investigator-Initiated Grants’, American Journal of Preventive Medicine, 42/2: 157–63.
Klein, J. T. (2008) ‘Evaluation of Interdisciplinary and Transdisciplinary Research: A Literature Review’, American Journal of Preventive Medicine, 35(Suppl. 2): S116–23.
Kueffer, C. et al. (2007) ‘Towards a Publication Culture in Transdisciplinary Research’, Gaia-Ecological Perspectives for Science and Society, 16/1: 22–6.
Nagarajan, R., Lowery, C. L. and Hogan, W. R. (2011) ‘Temporal Evolution of Biomedical Research Grant Collaborations across Multiple Scales—a CTSA Baseline Study’, AMIA Annual Symposium Proceedings, 2011, 987–93.
Masse, L. C. et al. (2008) ‘Measuring Collaboration and Transdisciplinary Integration in Team Science’, American Journal of Preventive Medicine, 35/2: S151–60.
Newman, M. E. J. (2001) ‘The Structure of Scientific Collaboration Networks’, Proceedings of the National Academy of Sciences of the United States of America, 98/2: 404–9.
Porter, A. L., Garner, J. and Crowd, T. (2012) ‘Research Coordination Networks: Evidence of the Relationship between Funded Interdisciplinary Networking and Scholarly Impact’, Bioscience, 62/3: 282–8.
Provan, K. G., Clark, P. I. and Huerta, T. (2008) ‘Transdisciplinarity Among Tobacco Harm-Reduction Researchers - A Network Analytic Approach’, American Journal of Preventive Medicine, 35/2: S173–81.
Rhoten, D. (2013) ‘A Multi-Method Analysis of the Social and Technical Conditions for Interdisciplinary Collaboration’, <http://www.hybridvigor.net/interdis/pubs/hv_pub_interdis-2003.09.29.pdf> accessed May 3.
Rosenfield, P. L. (1992) ‘The Potential of Transdisciplinary Research for Sustaining and Extending Linkages Between the Health and Social Sciences’, Social Science and Medicine, 35/11: 1343–57.
Smith, M. et al. (2010) ‘NodeXL: A Free and Open Network Overview, Discovery and Exploration Add-in for Excel 2007/
2010’, http://nodexl.codeplex.com/ from the Social Media Research Foundation, http://www.smrfoundation.org.
Smoot, M. E. et al. (2011) ‘Cytoscape 2.8: New Features for Data Integration and Network Visualization’, Bioinformatics, 27/3: 431–2.
Stokols, D. et al. (2005) ‘In Vivo Studies of Transdisciplinary Scientific Collaboration - Lessons Learned and Implications for Active Living Research’, American Journal of Preventive Medicine, 28/2: 202–13.
——. (2008a) ‘The Science of Team Science - Overview of the Field and Introduction to the Supplement’, American Journal of Preventive Medicine, 35/2: S77–89.
——. (2008b) ‘The Ecology of Team Science - Understanding Contextual Influences on Transdisciplinary Collaboration’, American Journal of Preventive Medicine, 35/2: S96–115.
Wuchty, S., Jones, B. F. and Uzzi, B. (2007) ‘The Increasing Dominance of Teams in Production of Knowledge’, Science, 316/5827: 1036–9.