The Oncologist®

Commentary

The Project Data Sphere Initiative: Accelerating Cancer Research by Sharing Data

Angela K. Green,[a] Katherine E. Reeder-Hayes,[a] Robert W. Corty,[b] Ethan Basch,[a] Mathew I. Milowsky,[a] Stacie B. Dusetzina,[c,d] Antonia V. Bennett,[d] William A. Wood[a]

[a]UNC Lineberger Comprehensive Cancer Center, [b]School of Medicine, Division of Hematology and Oncology, [c]Eshelman School of Pharmacy, Division of Pharmaceutical Outcomes and Policy, and [d]Gillings School of Global Public Health, Department of Health Policy and Management, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

Disclosures of potential conflicts of interest may be found at the end of this article.

Key Words. Project Data Sphere • Data sharing • Prostate cancer • Comparative effectiveness research

Abstract

Background. In this paper, we provide background and context regarding the potential for a new data-sharing platform, the Project Data Sphere (PDS) initiative, funded by financial and in-kind contributions from the CEO Roundtable on Cancer, to transform cancer research and improve patient outcomes. Given the relatively modest decline in cancer death rates over the past several years, a new research paradigm is needed to accelerate therapeutic approaches for oncologic diseases. Phase III clinical trials generate large volumes of potentially usable information, often on hundreds of patients, including patients treated with standard of care therapies (i.e., controls). Both nationally and internationally, a variety of stakeholders have pursued data-sharing efforts to make individual patient-level clinical trial data available to the scientific research community.

Potential Benefits and Risks of Data Sharing. For researchers, shared data have the potential to foster a more collaborative environment, to answer research questions in a shorter time frame than traditional randomized control trials, to reduce duplication of effort, and to improve efficiency. For industry participants, use of trial data to answer additional clinical questions could increase research and development efficiency and guide future projects through validation of surrogate end points, development of prognostic or predictive models, selection of patients for phase II trials, stratification in phase III studies, and identification of patient subgroups for development of novel therapies. Data transparency also helps promote a public image of collaboration and altruism among industry participants. For patient participants, data sharing maximizes their contribution to public health and increases access to information that may be used to develop better treatments. Concerns about data-sharing efforts include protection of patient privacy and confidentiality. To alleviate these concerns, data sets are deidentified to maintain anonymity. To address industry concerns about protection of intellectual property and competitiveness, we illustrate several models for data sharing with varying levels of access to the data and varying relationships between trial sponsors and data access sponsors.

The Project Data Sphere Initiative. PDS is an independent initiative of the CEO Roundtable on Cancer Life Sciences Consortium, built to voluntarily share, integrate, and analyze comparator arms of historical cancer clinical trial data sets to advance future cancer research. The aim is to provide a neutral, broad-access platform for industry and academia to share raw, deidentified data from late-phase oncology clinical trials using comparator-arm data sets. These data are likely to be hypothesis generating or hypothesis confirming but, notably, do not take the place of performing a well-designed trial to address a specific hypothesis. Prospective providers of data to PDS complete and sign a data sharing agreement that includes a description of the data they propose to upload, and then they follow easy instructions on the website for uploading their deidentified data. The SAS Institute has also collaborated with the initiative to provide intrinsic analytic tools accessible within the website itself.

As of October 2014, the PDS website has available data from 14 cancer clinical trials covering 9,000 subjects, with hopes to further expand the database to include more than 25,000 subject accruals within the next year. PDS differentiates itself from other data-sharing initiatives by its degree of openness, requiring submission of only a brief application with background information of the individual requesting access and agreement to terms of use. Data from several different sponsors may be pooled to develop a comprehensive cohort for analysis. In order to protect patient privacy, data providers in the U.S. are responsible for deidentifying data according to standards set forth by the Privacy Rule of the U.S. Health Insurance Portability and Accountability Act of 1996.

Using Data Sharing to Improve Outcomes in Cancer: The "Prostate Cancer Challenge." Control-arm data of several studies among patients with metastatic castration-resistant

Correspondence: Angela K. Green, M.D., 102 Manning Drive, CB #7500, Suite 3100, Chapel Hill, North Carolina 27599, USA. Telephone: 571-277-1452; E-Mail: Angela.Green@unchealth.unc.edu Received October 31, 2014; accepted for publication March 20, 2015. ©AlphaMed Press 1083-7159/2015/$20.00/0 http://dx.doi.org/10.1634/theoncologist.2014-0431

prostate cancer (mCRPC) are currently available through PDS. These data sets have multiple potential uses. The "Prostate Cancer Challenge" will ask the cancer research community to use clinical trial data deposited in the PDS website to address key research questions regarding mCRPC.

General themes that could be explored by the cancer community are described in this article: prognostic models evaluating the influence of pretreatment factors on survival and patient-reported outcomes; comparative effectiveness research evaluating the efficacy of standard of care therapies, as illustrated in our companion article comparing mitoxantrone plus prednisone with prednisone alone; effects of practice variation in dose, frequency, and duration of therapy; level of patient adherence to elements of trial protocols to inform the design of future clinical trials; and age of subjects, regional differences in health care, and other confounding factors that might affect outcomes.

Potential Limitations and Methodological Challenges. The number of data sets available and the lack of experimental-arm data limit the potential scope of research using the current PDS. The number of trials is expected to grow exponentially over the next year and may include multiple cancer settings, such as breast, colorectal, lung, hematologic malignancy, and bone marrow transplantation. Other potential limitations include the retrospective nature of the data analyses performed using PDS and its generalizability, given that clinical trials are often conducted among younger, healthier, and less racially diverse patient populations. Methodological challenges exist when combining individual patient data from multiple clinical trials; however, advancements in statistical methods for secondary database analysis offer many tools for reanalyzing data arising from disparate trials, such as propensity score matching. Despite these concerns, few if any comparable data sets include this level of detail across multiple clinical trials and populations.

Conclusion. Access to large, late-phase, cancer-trial data sets has the potential to transform cancer research by optimizing research efficiency and accelerating progress toward meaningful improvements in cancer care. This type of platform provides opportunities for unique research projects that can examine relatively neglected areas and that can construct models necessitating large amounts of detailed data. The full potential of PDS will be realized only when multiple tumor types and larger numbers of data sets are available through the website. ***The Oncologist*** 2015;20:464–e20

Introduction

In this paper, we wish to provide background and context regarding the potential for a data-sharing platform, Project Data Sphere, to transform cancer research and improve patient outcomes. Project Data Sphere, LLC, is an initiative of the CEO Roundtable on Cancer Life Sciences Consortium, built to voluntarily share, integrate, and analyze comparator arms of historical cancer clinical trial data sets to advance future cancer research. Such data are likely to be hypothesis generating or hypothesis confirming and do not take the place of a well-designed trial to address a specific hypothesis. We also discuss a group of data sets in metastatic castration-resistant prostate cancer (mCRPC) as an initial disease area that illustrates how a platform like this might be used.

The Need: Optimizing Efficiency in Cancer Research

In 2013, a total of 1,660,290 new cancer cases and 580,350 cancer deaths were estimated to occur in the U.S. Cancer death rates have decreased by 1.8% per year in men and by 1.5% per year in women from 2005 to 2009 and declined overall by 20% from their peak in 1991 to 2009 [1]. In contrast, deaths from cardiovascular disease decreased by 32.5% within a comparable period of time [2]. A new research paradigm is needed to accelerate the emergence of therapeutic approaches for oncologic diseases. The gold standard for evidence needed to change clinical management often comes from large randomized controlled trials (RCTs), which are time consuming and expensive. Despite the resources invested in RCT completion, data generated from these efforts are not maximally leveraged by the cancer research community. Complete clinical trial data sets have the potential to facilitate new studies to guide future drug development and RCT planning, to answer key scientific questions, and to inform clinical practice [3].

A Potential Solution: Clinical Trial Data Sharing

What Is Data Sharing?

Nationally and internationally, multiple efforts have been undertaken by stakeholders (including researchers, government regulators, funding organizations, and medical publishers) to make individual patient-level clinical trial data available to the scientific research community, a process commonly referred to as "data sharing." Phase III clinical trials, which constitute more than 90% of the cost of a drug's development, generate large volumes of potentially usable information [4]. Data are collected on a variety of demographics, exposures, pretreatment factors, toxicities, complications, and outcomes, often on hundreds of patients including "controls," or patients treated with standard of care therapies. This underutilized information provides more detailed and granular information than that contained within traditional observational data sources. This level of detail would allow researchers to answer a variety of scientifically relevant questions using data from clinical trials that were designed for separate and specific purposes.

Potential Benefits and Risks of Data Sharing

Clinical trial data sharing offers many potential benefits to academic researchers and industry and, most importantly, to the patients who participate in the studies. Health services and outcomes researchers can conduct comparative effectiveness and cost-effectiveness research on a shorter timeline than traditional randomized intervention studies. Shared data among researchers across institutions and disciplines can foster an environment of collaboration among scientists, engaging participants to share ideas and projects with one another, in contrast to the often "siloed" nature of current biomedical research. Sharing data through an open platform has the potential to reduce duplication of effort and improve efficiency of future trials.

For industry participants, providing access to clinical trial data can enhance research and development through several potential mechanisms including validation of surrogate endpoints [5], development of prognostic or predictive models, selection of patients for phase II trials, development of strategies for stratification, and identifying patient subgroups to target for development of novel therapies or approaches. Data transparency efforts can also help promote a public image of collaboration and altruism among industry participants, ameliorating current negative perceptions [6].

Data sharing will most importantly benefit patients by increasing access to information that can be used to develop new research initiatives and better treatments. In addition to direct potential benefit for their own conditions, patients often enroll in trials with the hope of advancing science so that treatments can be improved for future patients. Expanding the use of patient-level data may maximize the contribution of trial enrollees to public health and scientific advancement.

Concerns about data-sharing efforts have included protection of patient privacy and confidentiality. To alleviate these concerns, data sets are deidentified to maintain anonymity. Previous work has shown that the risks of reidentification are largely limited to data for which deidentification has not been done to existing standards [7]. When deidentification is effective and data use agreements are provided in a setting in which anticipated societal benefit is substantial, data-sharing efforts are consistent with the U.S. Department of Health and Human Services standard of “a reasonable balance between the risk of identification and the usefulness of the information” [8]. Other potential challenges of data sharing relate to industry concerns around protection of intellectual property (IP) and competitive concerns. There are now several different models for data sharing with varying levels of access to the data and different types of relationships between trial sponsors and data access sponsors [9]. These models include a more restrictive approach, in which an intermediary performs analyses and returns results for submitted queries, and a broad-access, open model in which trial sponsors post data sets to be available for download. Although open access models have the greatest theoretical IP risks for industry participants, these models also maximize the research potential of the data sets and have greater administrative efficiency than models that require extensive review for each research initiative. Examples of different models are provided:

- **Black Box, or Database Query Model.** The potential user submits a research query to the data holder, and the data holder runs the query and returns the results. This is the least transparent and most restrictive model but may be suitable for some types of very sensitive health information.
- **Gatekeeper Model.** An applicant submits a research query to the data generator. An independent review board assesses the application for (a) sound science and analytical plan, (b) risks related to privacy and intellectual property, and (c) expertise to carry out the proposed analysis.
- **Open Access Model.** There is no applicant review panel. The criterion is, fundamentally, a responsible-use attestation. The data generator routinely posts appropriately deidentified data from trials once they are publicly reported, along with documentation to facilitate the use of data. IP restrictions may be more relaxed to encourage innovation among researchers. The volume of data and the therapeutic area may be more limited than that of the gatekeeper model. In addition, this model may not be suitable for small trials or for very sensitive data for which privacy risks may be higher.

Examples of Data Sharing

Several regulatory efforts to promote public access to clinical trial data have been made over the past several years. In the U.S., all clinical trials must be registered at ClinicalTrials.gov, according to the U.S. Food and Drug Administration Amendments Act (FDAAA) [10]. Nonetheless, results of trials tend to be under-reported in the peer-reviewed literature [11, 12]. Under the Trial and Experimental Studies Transparency Act of 2012, results of all interventional trials including adverse effects are to be reported as open public knowledge to the online clinical trial registry databank [10]. Beyond registration and reporting requirements, regulators are moving toward promoting access to primary clinical trial data. The European Medicines Agency recently drafted policy for the release of patient-level data submitted to the agency after March 2014 [13]. All trial data, with personal information deidentified, can be made available on request for research purposes within the scope of the original informed consent and with the agreement not to reidentify participants [13].

The NIH, according to its Data Sharing Policy, requires research applicants seeking more than $500,000 in direct support in any given year to submit a data-sharing plan with the application or to indicate why data sharing is not possible [14]. In November 2014, the NIH proposed a draft policy that expects all NIH-funded clinical trials to be submitted to ClinicalTrials.gov, even if they are not subject to the FDAAA.

Medical publishers are also promoting data sharing. As of 2013, in an effort to encourage clinical trial reporting for independent evaluation, *BMJ* committed to publishing only clinical trials that make patient-level data available by reasonable request [15]. In addition, *JAMA*, the journal of the American Medical Association, called on pharmaceutical companies to make patient-level data publicly available to qualified researchers for secondary analyses [16].

In response to such appeals, multiple recent efforts have been made by industry sponsors to share clinical trial data. GlaxoSmithKline published a special report in August 2013 detailing its efforts to allow access to a subset of clinical trials, including both raw data sets and analysis-ready data [17]. GlaxoSmithKline and 10 other industry sponsors have now joined the data-sharing platform ClinicalStudyDataRequest.com, which allows researchers to submit requests for access to patient-level data from studies listed on the sponsor’s website or identified through study registers [18].

The Project Data Sphere Initiative: Open Access Data Sharing in Cancer Research

Development of the Project Data Sphere Platform

PDS is funded by financial and in-kind contributions from members of the CEO Roundtable on Cancer. The aim is to provide a neutral, broad-access platform on which industry and academia can share raw, deidentified data from both successful and failed late-phase oncology clinical trials using comparator-arm data sets. PDS collaborates with multiple cancer trial sponsors, including industry participants, clinical trial cooperative groups, and academic institutions, to identify and facilitate upload of relevant data sets. Prospective providers of data to PDS complete and sign a data sharing agreement that includes a description of the data they propose to upload, and then they follow easy instructions on the website for uploading the deidentified data. The SAS Institute has collaborated with the initiative to provide intrinsic analytic tools accessible within the website itself. A major goal of the initiative is to enable researchers to "link data, skills, technology, and ideas" through high-powered information and sophisticated statistical analysis [19].

As of October 2014, the PDS website had available cancer trial data from 14 trials including 9,000 subjects, with hopes to further expand the database to include more than 25,000 subject accruals within the next year. PDS differentiates itself from other data-sharing initiatives by its degree of openness. Other initiatives have regulated access by requiring research proposals and subsequent evaluation by a review panel to ascertain scientific credibility of the specific proposals. In contrast, PDS requires submission of only a brief application with information about the background of the individual requesting access and an agreement to terms of use. On approval of the application, authorized users have access to all data sets on the website. Users can search within the website for specific trials using key words. Trial summaries are provided with links to SAS-encoded data sets that can be downloaded directly for use through personal statistical software or within the website itself using a unique SAS Analytics program. This overcomes the difficulty of applying for a single data set at a time and allows third-party researchers to pool data from several different sponsors to develop a comprehensive cohort for analysis. In order to protect patient privacy, U.S. data providers are responsible for deidentifying data according to standards set forth by the Privacy Rule of the U.S. Health Insurance Portability and Accountability Act of 1996.

Sample Applications: Using Data Sharing to Evaluate Outcomes in Prostate Cancer

The PDS platform has the potential to accelerate improvements in cancer outcomes by enabling access to individual subject data from cancer clinical trials. PDS plans to incorporate data from a broad array of trials in multiple different malignancies. The control arms of studies among subjects with mCRPC have been made available for research use through PDS (Table 1). There are multiple potential uses of control arms from these data sets to build on the current state of knowledge of the treatment of prostate cancer. The Prostate Cancer Challenge will be an open challenge to the cancer research community to use clinical trial data deposited on the Project Data Sphere data-sharing platform. The following examples are major themes that could be explored by the cancer research community.

Disease Biology: Prognostic Models

Prognostic models can be created using pooled data from trial control arms to evaluate the effect of a wide variety of pretreatment factors on endpoints such as overall survival, progression-free survival, treatment toxicities and adverse events, and patient-reported outcomes. Prognostic models aid in the development of risk groups to better enhance the clinical application of cancer therapies and the selection of patients for participation in clinical trials. Prognostic models have been used most commonly in the initial diagnostic setting. Examples of well-known prognostic models developed from large cancer clinical data sets include the AdjuvantOnline! and Predict models for breast cancer and the international prognostic index for non-Hodgkin's lymphoma, all of which use patient and tumor characteristics to calculate estimated survival among newly diagnosed patients [28–30].

In 2003, Halabi et al. developed a prognostic model to predict survival among mCRPC patients utilizing data from six phase III trials conducted by Cancer and Leukemia Group B (CALGB) between 1992 and 1998 [31]. Lactate dehydrogenase, prostate-specific antigen, alkaline phosphatase, Gleason sum, Eastern Cooperative Oncology Group performance status, hemoglobin, and presence of visceral disease were identified as significant predictors of survival. More recently, in 2013, Halabi et al developed a prognostic model of overall survival for mCRPC patients in the postdocetaxel setting using patient-level data from the TROPIC trial, with control-arm data that have now been made available through the PDS platform. Data from the SPARC trial, which evaluated the efficacy of satraplatin plus prednisone versus placebo plus prednisone, was then used for external validation. Two new prognostic factors were identified through this study, time since docetaxel use and duration of hormone therapy, providing potential insights into tumor disease biology [32].

PDS provides a well-suited environment within which to develop prognostic models for metastatic prostate cancer, with the robustness of these models expected to increase as additional prostate cancer data sets are added to the pool. In addition, because of the granular clinical trial quality data included within these data sets, prognostic models can be constructed using the breadth of data from multiple trials. Questions could be addressed: Which coincident comorbid illnesses or concurrent medications influence the risk for death or disease progression? How might prognostic models at study baseline vary at key follow-up time points for patients on clinical studies, and how does the emergence of clinical toxicities during follow-up further influence risk?

With increasingly sophisticated prognostic models, investigators could develop therapies among enriched populations to achieve clearer efficacy signals within shorter time horizons. Prognostic models could also help to minimize variance in outcomes for well-defined subsets of patients. In turn, these models could then lead to more effective identification of outliers. Further study of these outliers could lead to identification of previously unmeasured confounding variables.

Clinical Management: Comparative Effectiveness Research

After publication of late-phase clinical trials, practice patterns in the community may or may not change for a variety of reasons. Many times, competing standards of care exist for patients. Although little incentive exists for industry to sponsor

Table 1. Clinical trials performed among patients with metastatic castrate-resistant prostate cancer currently available in Project Data Share as of January 2015

Identifier	Official title	Outcome
TROPIC, NCT00417079 [20]	A Randomized, Open Label Multi-Center Study of XRP6258 at 25 mg/m^2 in Combination With Prednisone Every 3 Weeks Compared to Mitoxantrone in Combination With Prednisone For The Treatment of Hormone Refractory Metastatic Prostate Cancer Previously Treated With A Taxotere®-Containing Regimen	Increased survival after docetaxel among patients treated with cabazitaxel over mitoxantrone plus prednisone, although toxicity was significantly more frequent in the cabazitaxel arm
TAX 327 [21]	A multicenter phase III randomized trial comparing TAXOTERE administered either weekly or every three weeks in combination with prednisone versus mitoxantrone in combination with prednisone for metastatic hormone-refractory prostate cancer (TAX 327)	Showed superiority of every-3-week docetaxel over mitoxantrone plus prednisone in prolonging overall survival and improving quality of life.
VENICE, NCT00519285 [22]	A Multicenter, Randomized, Double-Blind Study Comparing the Efficacy and Safety of Aflibercept Versus Placebo Administered Every 3 Weeks in Patients Treated with Docetaxel / Prednisone for Metastatic Androgen-Independent Prostate Cancer	Showed no superiority in adding the VEGF inhibitor aflibercept to first-line docetaxel plus prednisone therapy
SUN 1120, NCT00676650 [23]	A Multicenter, Randomized, Double-Blind, Phase 3 Study Of Sunitinib Plus Prednisone Versus Prednisone In Patients With Progressive Metastatic Castration-Resistant Prostate Cancer After Failure Of A Docetaxel-Based Chemotherapy Regimen	Showed no superiority of treatment with the angiogenesis-targeting agent sunitinib compared with prednisone
NCT00638690 [24]	A Phase 3, Randomized, Double-blind, Placebo-Controlled Study of Abiraterone Acetate (CB7630) Plus Prednisone in Patients with Metastatic Castration-Resistant Prostate Cancer Who Have Failed Docetaxel-Based Chemotherapy	Observed significant increase in median overall survival among patients receiving abiraterone acetate plus prednisone versus placebo plus prednisone
NCT00385827 [25]	A Phase 2, Multicenter, Open-Label Study of CNTO 328 (Anti-IL-6 Monoclonal Antibody) in Combination With Mitoxantrone Versus Mitoxantrone in Subjects With Metastatic Hormone-Refractory Prostate Cancer (HRPC)	Trial was stopped for futility after IDMC evaluation
Mainsail, NCT00988208 [26]	A Phase 3 Study to Evaluate the Efficacy and Safety of Docetaxel and Prednisone With or Without Lenalidomide in Subjects With Castrate-Resistant Prostate Cancer (CRPC)	Trial was stopped for futility after IDMC evaluation
ASCENT-2, NCT00273338 [27]	A Phase 3, Randomized, Open-Label Study Evaluating DN-101 in Combination With Docetaxel in Androgen-Independent Prostate Cancer (AIPC) (ASCENT-2)	Study was terminated, no further details provided

Abbreviations: IDMC, independent data monitoring committee; VEGF, vascular endothelial growth factor.

head-to-head comparison trials of approved standard of care therapies, there has been recent national interest in using large observational data sets for this purpose, a type of investigation known as “comparative effectiveness” research. To date, most comparative effectiveness research has been performed using large observational registry or claims-based data sets (e.g., Medicare, Kaiser), which have limited clinical information on potentially important determinants of outcome such as comorbidities, medical history, and treatment-related toxicity. The availability of comparator-arm data sets in PDS presents a unique opportunity to conduct comparative effectiveness research using rich clinical trial data, allowing for improved quality and depth of data compared with other data sources.

Furthermore, the ability to access and combine multiple data sets through the PDS website increases the power for more robust subgroup analyses. Differences in specific outcomes within control populations could be evaluated among subsets of patients with common comorbidities such as type II diabetes or coronary artery disease, genetic phenotypes, tumor biology, and lifestyle factors. Such analyses could highlight populations that may experience increased drug toxicity with standard of care therapy. Patients with mCRPC patients and type II diabetes at baseline, for example, may be more susceptible to cardiotoxicity with mitoxantrone or neurotoxicity with docetaxel. In addition, more comprehensive analyses may be performed among less commonly represented patient populations in trials, such as women, minorities, and older adults, by pooling studies to include higher numbers of participants from under-represented groups [33]. This may also provide insight into the potential selection bias that may be introduced in clinical trial design and its impact on outcomes.

An example of a comparative effectiveness study using PDS is represented by the comparison of mitoxantrone plus prednisone with prednisone alone in advanced mCRPC [34–37], as published concurrently in this issue of ***The Oncologist*** [38]. As other data sets are added, additional competing analyses of standards of care could be examined. With the current growing interest in efficient resource use and individualized treatment decision making in the U.S. health care environment, the granularity of the data contained within PDS could also facilitate cost effectiveness and resource utilization analyses for comparison of various treatment strategies.

Health Services Research: The Effects of Practice Variation Among and Within Protocols on Process and Outcome

Information on optimal dosing for chemotherapeutic agents may be limited because of inherent constraints within clinical trial design. Only a limited number of experimental arms, for example, can be constructed while maintaining sufficient power to assess for differences in efficacy. The effect of different choices in dose, frequency, and duration of therapy between trials on outcomes including survival, toxicity, and patient compliance could guide dosing in future trials and inform clinical practice with regard to dose adjustment. A secondary analysis of data from breast cancer patients in the CALGB 8541 study, for example, demonstrated that obese women treated at full weight-based dosing did not experience excess toxicity relative to their normal weight counterparts, but obese women whose dose was capped because of weight experienced inferior survival outcomes relative to those treated at full dose [39]. This finding ultimately led to modification of national guidelines clearly warning against dose capping for obese chemotherapy patients. Similarly, comparator-arm data in PDS could be used to further evaluate the influence of body mass index (BMI) and other patient characteristics on the relationship between treatment dose and efficacy to determine whether dose adjustments are warranted. Although a national guideline exists for a general approach to chemotherapy dosing in obese cancer patients, questions regarding the appropriate treatment of prostate cancer patients at different BMI percentiles remain important areas for investigation, with an inadequate level of evidence accumulated to date.

Although clinical trial protocols result in a somewhat homogenized approach to treatment, nonadherence to protocol and early discontinuation of therapy remain important issues within clinical trial populations, particularly those with advanced disease. PDS data could be used to evaluate trial participant adherence to study elements such as scheduled blood tests, procedures, and follow-up physician appointments to understand levels of adherence and factors associated with adherence. These findings could inform the design of future clinical trials. If, for example, adherence to a particular lengthy or unpleasant test is noted to be lower than adherence to other elements of the protocol, or if protocols including the collection of certain samples or endpoints appear to have disproportionate dropout rates, these factors could be considered in designing future trials to minimize missing data.

Geographic variation may play a role in the efficacy or toxicity of standard of care therapy for reasons that are sometimes not immediately apparent. A pooled analysis of clinical trial data from patients receiving the monoclonal antibody cetuximab for metastatic colorectal cancer was the first study to detect a large degree of geographic variation in rates of anaphylactic infusion reactions to the drug, a finding that was ultimately traced to a particular IgE antibody to pollens found in the southeastern U.S. [40, 41]. The large clinical trials available through PDS were performed in multiple centers, thus the effects of center sites, urban versus rural settings, or large geographic regions on the association between standard of care treatments and clinical outcomes can be evaluated, depending on the geographic identifiers contained within the particular data sets. Effects may vary within different settings because of factors such as quality and quantity of access to basic health care services, access to pre- and post-treatment anticancer therapies, variation in clinical practice standards or quality of care, socioeconomics, population genetics, environmental exposures, and differences in lifestyle. Although the data in PDS may not be able to account specifically for all of these factors, significant geographic variation may be hypothesis generating and may lead to future studies designed to further explore these differences.

Methodological Challenges and Potential Limitations

Important methodological challenges are present in working with individual participant data from multiple clinical trials using PDS. The diverse sources and the wide time span of the clinical trial data may be accompanied by different formatting and coding standards for different data sets and should be investigated when planning a research project using these data. Furthermore, caution must be exercised when combining individual patient data from multiple clinical trials. An important issue concerns whether data should be combined in a “one-step” or “two-step” meta-analysis, with at least one recent publication finding similar results with both methods in an analysis of 24 randomized controlled trials evaluating antiplatelet agents for the prevention of pre-eclampsia in pregnancy [42]. Researchers interested in exploring other potential benefits and important challenges contained within individual participant data studies are referred to several useful reviews in the literature [43, 44].

Thanks to advances in pharmacoepidemiology and statistical methods for secondary database analysis, many tools can be used for reanalyzing data arising from disparate trials. Although the data may arise from randomized trials, the benefits of randomization are lost when comparing arms from separate studies. In order to ensure that comparisons are valid, PDS users may want to use propensity score methods to balance the characteristics of patients included in analyses from separate trials. These methods are helpful for two reasons. First, they allow the researcher to determine whether there are characteristics that are dissimilar between the two trials by demonstrating factors that are strong predictors of being in one trial versus another. The process of generating a propensity score and evaluating the strength of the variables used in the propensity score can help the researcher to identify coding differences (which would indicate a need for further data management) or differences in inclusion criteria between the trials (suggesting that the two trials may be inappropriate for pooling). Second, propensity score methods will reduce the need for individual covariate adjustment, which can be problematic with small sample sizes and large numbers of measured covariates. Application of the propensity score can also provide guidance to the researcher regarding the similarity of patients included in the trials that are to be compared. If the researcher chooses to match patients from one trial to the next, a low match rate would indicate that few patients would be eligible for both trials (again, suggesting that they may be inappropriate for comparison). We have provided a sample study in this issue of ***The Oncologist*** [38] that demonstrates the use of propensity score methods for analyzing studies in PDS.

The potential scope of research using PDS is currently limited by the clinical trials available; however, the number of trials in PDS is projected to grow exponentially based on a robust plan to engage with data providers across industry and academia. In addition, these analyses are retrospective in nature in that both data and outcomes have been collected prior to the start of research. Accumulation of new data are limited only by the willingness to participate; however, unlike traditional retrospective studies, for which data are often incomplete, covariate information is comparatively comprehensive on the PDS platform because it has been derived from clinical trials. Generalizability is also a concern, given that trials are often conducted among younger, healthier, and less racially diverse patient populations. Patients in trials, including those in control arms, also receive more clinical monitoring than patients receiving routine care off trial, and that also could influence outcomes. Despite these concerns, few if any comparable data sets include this level of detail across multiple clinical trials and populations. This leaves PDS with substantial and significant advantages over other data resources for performing the types of research studies that have been described in this paper.

Conclusion

Providing access to large, late-phase, cancer-trial data sets has the potential to transform cancer research by optimizing research efficiency and accelerating progress toward meaningful improvements in cancer care. This type of platform provides opportunities for unique research projects that can examine relatively neglected areas, such as outcomes on standard of care therapies, and that can construct models necessitating large amounts of detailed data, such as subgroup analyses and prognostic indices. The Prostate Cancer Challenge and the four currently available prostate cancer data sets on the PDS website help demonstrate a pilot proof-of-concept project, although the full potential of PDS will be realized only when multiple tumor types, larger numbers of data sets, and, ideally, data from experimental arms are available. Through projects like these, Project Data Sphere, LLC, seeks to create an environment of research collaboration among industry sponsors, cancer researchers across institutions, the public, and other stakeholders, unified for the goals of improving cancer outcomes and protecting public health.

Acknowledgments

We thank Dr. Kald Abdallah, chief Project Data Sphere officer, for his input in the early stages of manuscript development. The other four officers overseeing Project Data Sphere, LLC are Christopher A. Viehbacher, president; William C. Weldon, first vice president; Robert A. Ingram, second vice president; and Dr. Martin J. Murphy, Jr., treasurer. The CEO Roundtable on Cancer is a nonprofit corporation composed of chief executives from more than 30 major American companies representing diverse industries, from nonprofit organizations committed to fighting cancer, and from several National Cancer Institute-designated cancer centers. The CEO Roundtable on Cancer was founded by Robert A. Ingram in 2001 at the request of former President George H.W. Bush.

Author Contributions

Conception/Design: Angela K. Green, Katherine E. Reeder-Hayes, Robert W. Corty, Ethan Basch, Matthew I. Milowsky, Stacie B. Dusetzina, Antonia V. Bennett, William A. Wood
Collection and/or assembly of data: Angela K. Green
Data analysis and interpretation: Angela K. Green, Katherine E. Reeder-Hayes, Robert W. Corty, Ethan Basch, Matthew I. Milowsky, Antonia V. Bennett, William A. Wood
Manuscript writing: Angela K. Green, Katherine E. Reeder-Hayes, Ethan Basch, Matthew I. Milowsky, Stacie B. Dusetzina, Antonia V. Bennett, William A. Wood
Final approval of manuscript: Ethan Basch, William A. Wood

Disclosures

The authors indicated no financial relationships.

References

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. CA Cancer J Clin 2013;63:11–30.

2. QuickStats: Age-adjusted death rates for heart disease and cancer — United States, 1999–2009. Available at http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6021a6.htm. Accessed October 31, 2014.

3. Pfister DG. The just price of cancer drugs and the growing cost of cancer care: Oncologists need to be part of the solution. J Clin Oncol 2013;31:3487–3489.

4. Roy A. Stifling new cures: The true cost of lengthy clinical drug trials. Available at http://www.manhattan-institute.org/pdf/fda_05.pdf. Accessed October 31, 2014.

5. Eichler HG, Pétavy F, Pignatti F et al. Access to patient-level trial data—a boon to drug developers. N Engl J Med 2013;369:1577–1579.

6. The price of drugs for chronic myeloid leukemia (CML) is a reflection of the unsustainable prices of cancer drugs: From the perspective of a large group of CML experts. Blood 2013;121:4439–4442.

7. El Eman K, Jonker E, Arbuckle L et al. A systematic review of re-identification attacks on health data. PLoS One 2011;6:e28071.

8. Standards for privacy of individually identifiable health information. Office of the Assistant Secretary for Planning and Evaluation, DHHS. Final rule. Fed Regist 2000;65:82462–82829.

9. Mello MM, Francer JK, Wilenzick M et al. Preparing for responsible sharing of clinical trial data. N Engl J Med 2013;369:1651–1658.

10. Drazen JM. Transparency for clinical trials—the TEST Act. N Engl J Med 2012;367:863–864.

11. Mathieu S, Boutron I, Moher D et al. Comparison of registered and published primary outcomes in randomized controlled trials. JAMA 2009;302:977–984.

12. Ross JS, Mulvey GK, Hines EM et al. Trial publication after registration in ClinicalTrials.gov: A cross-sectional analysis. PLoS Med 2009;6:e1000144.

13. Publication and access to clinical-trial data. Available at http://www.ema.europa.eu/docs/en_GB/document_library/Other/2013/06/WC500144730.pdf. Accessed October 31, 2014.

14. NIH data sharing policies. Available at http://www.nlm.nih.gov/NIHbmic/nih_data_sharing_policies.html. Accessed October 31, 2014.

15. Godlee F. Clinical trial data for all drugs in current use. BMJ 2012;345:e7304.

16. Bauchner H, Fontanarosa PB. Restoring confidence in the pharmaceutical industry. JAMA 2013;309:607–609.

17. Nisen P, Rockhold F. Access to patient-level data from GlaxoSmithKline clinical trials. N Engl J Med 2013;369:475–478.

18. ClinicalStudyDataRequest.com. Available at http://www.clinicalstudydatarequest.com/Default.aspx. Accessed February 8, 2015.

19. Hede K. Project Data Sphere to make cancer clinical trial data publicly available. J Natl Cancer Inst 2013;105:1159–1160.

20. Bahl A, Oudard S, Tombal B et al. Impact of cabazitaxel on 2-year survival and palliation of tumour-related pain in men with metastatic castration-resistant prostate cancer treated in the TROPIC trial. Ann Oncol 2013;24:2402–2408.

21. Berthold DR, Pond GR, Roessner M et al. Treatment of hormone-refractory prostate cancer with docetaxel or mitoxantrone: Relationships between prostate-specific antigen, pain, and quality of life response and survival in the TAX-327 study. Clin Cancer Res 2008;14:2763–2767.

22. Tannock IF, Fizazi K, Ivanov S et al. Aflibercept versus placebo in combination with docetaxel and prednisone for treatment of men with metastatic castration-resistant prostate cancer (VENICE): A phase 3, double-blind randomised trial. Lancet Oncol 2013;14:760–768.

23. Michaelson MD, Oudard S, Ou YC et al. Randomized, placebo-controlled, phase III trial of sunitinib plus prednisone versus prednisone alone in progressive, metastatic, castration-resistant prostate cancer. J Clin Oncol 2014;32:76–82.

24. de Bono JS, Logothetis CJ, Molina A et al. Abiraterone and increased survival in metastatic prostate cancer. N Engl J Med 2011;364:1995–2005.

25. A safety and efficacy study of siltuximab (CNTO 328) in male subjects with metastatic hormone-refractory prostate cancer (HRPC). Available at https://clinicaltrials.gov/ct2/show/NCT00385827.

26. Study to evaluate safety and effectiveness of lenalidomide in combination with docetaxel and prednisone for patients with castrate-resistant prostate cancer (Mainsail). Available at https://clinicaltrials.gov/ct2/show/results/NCT00988208?term=NCT00988208&rank=1.

27. Scher HI, Jia X, Chi K et al. Randomized, open-label phase III trial of docetaxel plus high-dose calcitriol versus docetaxel plus prednisone for patients with castration-resistant prostate cancer. J Clin Oncol 2011;29:2191–2198.

28. A predictive model for aggressive non-Hodgkin's lymphoma. The International Non-Hodgkin's Lymphoma Prognostic Factors Project. N Engl J Med 1993;329:987–994.

29. Wishart GC, Bajdik CD, Azzato EM et al. A population-based validation of the prognostic model PREDICT for early breast cancer. Eur J Surg Oncol 2011;37:411–417.

30. Olivotto IA, Bajdik CD, Ravdin PM et al. Population-based validation of the prognostic model ADJUVANT! for early breast cancer. J Clin Oncol 2005;23:2716–2725.

31. Halabi S, Small EJ, Kantoff PW et al. Prognostic model for predicting survival in men with hormone-refractory metastatic prostate cancer. J Clin Oncol 2003;21:1232–1237.

32. Halabi S, Lin CY, Small EJ et al. Prognostic model predicting metastatic castration-resistant prostate cancer survival in men treated with second-line chemotherapy. J Natl Cancer Inst 2013;105:1729–1737.

33. Murthy VH, Krumholz HM, Gross CP. Participation in cancer clinical trials: Race-, sex-, and age-based disparities. JAMA 2004;291:2720–2726.

34. Kantoff PW, Halabi S, Conaway M et al. Hydrocortisone with or without mitoxantrone in men with hormone-refractory prostate cancer: Results of the Cancer and Leukemia Group B 9182 study. J Clin Oncol 1999;17:2506–2513.

35. Osoba D, Tannock IF, Ernst DS et al. Health-related quality of life in men with metastatic prostate cancer treated with prednisone alone or mitoxantrone and prednisone. J Clin Oncol 1999;17:1654–1663.

36. Tannock IF, Osoba D, Stockler MR et al. Chemotherapy with mitoxantrone plus prednisone or prednisone alone for symptomatic hormone-resistant prostate cancer: A Canadian randomized trial with palliative end points. J Clin Oncol 1996;14:1756–1764.

37. Berry W, Dakhil S, Modiano M et al. Phase III study of mitoxantrone plus low dose prednisone versus low dose prednisone alone in patients with asymptomatic hormone refractory prostate cancer. J Urol 2002;168:2439–2443.

38. Green AK, Corty R, Wood WA et al. Comparative effectiveness of mitoxantrone plus prednisone versus prednisone alone in metastatic castrate-resistant prostate cancer after docetaxel failure. ***The Oncologist*** 2015;20:516–522.

39. Rosner GL, Hargis JB, Hollis DR et al. Relationship between toxicity and obesity in women receiving adjuvant chemotherapy for breast cancer: Results from Cancer and Leukemia Group B study 8541. J Clin Oncol 1996;14:3000–3008.

40. O'Neil BH, Allen R, Spigel DR et al. High incidence of cetuximab-related infusion reactions in Tennessee and North Carolina and the association with atopic history. J Clin Oncol 2007;25:3644–3648.

41. Commins SP, Platts-Mills TA. Allergenicity of carbohydrates and their role in anaphylactic events. Curr Allergy Asthma Rep 2010;10:29–33.

42. Stewart GB, Altman DG, Askie LM et al. Statistical analysis of individual participant data meta-analyses: A comparison of methods and recommendations for practice. PLoS One 2012;7:e46042.

43. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: Rationale, conduct, and reporting. BMJ 2010;340:c221.

44. Simmonds MC, Higgins JPT, Stewart LA et al. Meta-analysis of individual patient data from randomized trials: A review of methods used in practice. Clin Trials 2005;2:209–217.

Editor's Note: See the related article, "Comparative Effectiveness of Mitoxantrone Plus Prednisone Versus Prednisone Alone in Metastatic Castrate-Resistant Prostate Cancer After Docetaxel Failure," on page 516 of this issue.