

SPECIAL REPORT

Access to Patient-Level Data from GlaxoSmithKline Clinical Trials

Perry Nisen, M.D., Ph.D., and Frank Rockhold, Ph.D.

Efforts are under way to increase the transparency of clinical trial data. Among the efforts are those being undertaken at GlaxoSmithKline, where we have recently made a commitment to provide access to deidentified patient-level data. We are taking this step because it is the right thing to do, both scientifically and for society, and it is in line with our company's commitment to transparency in clinical trial reporting. As of May 2013, investigators are able to request access to deidentified patient-level data from a subset of GlaxoSmithKline–sponsored clinical trials. We expect that research teams requesting data will have sufficient statistical and data-management expertise to evaluate these data sets, and we call for standards in analysis to be defined by the scientific community. This article describes the approach we are adopting and answers some of the most common questions we have received since our announcement of this policy.

WHICH GLAXOSMITHKLINE STUDIES WILL BE AVAILABLE?

The GlaxoSmithKline Clinical Study Requests website (<https://clinicalstudydata.gsk.com>) lists studies available as of May 7, 2013, when the website was launched. These studies are also listed in Section 1 in the Supplementary Appendix, available with the full text of this article at NEJM.org. The list includes clinical studies of medicines that have been approved for specific indications (anywhere in the world) and medicines that have been terminated from development. Manuscripts describing the results of a study — whether they are perceived as positive or negative — are submitted for publication within 18 to 24 months after the last study visit by the last participant enrolled. A study is added to the list once the primary manuscript describing the results has been accepted for publication. Studies that have not been accepted for publication, as well as those that are not submitted or are no longer being submitted for publication, are also available.

Initially, we are including globally conducted clinical studies (phases 1 through 4) started since January 1, 2007; we chose these studies because they used more consistent data standards, making it easier for us to deidentify and provide access to the data. There are approximately 200 studies (Section 1 in the Supplementary Appendix) currently listed on the website. All studies (including those conducted solely in a specific country or region, such as Europe or Asia) initiated after January 1, 2013, will routinely appear on the website when they meet the above criteria for inclusion. Over the next 2 years, we will add our global studies going back to December 2000. We anticipate that, by the end of 2013, there will be approximately 400 studies on the website. The list of studies will be updated at least twice per year; studies will be added within 6 months after meeting the criteria for inclusion. The website will not include observational (or noninterventional) studies, since the majority of these studies were conducted with deidentified patient-level data from third-party databases. Also excluded are clinical trials of medicines for which we do not have the legal authority to share data — for example, because the medicine has been licensed to another company.

Investigators who are preparing research proposals will be able to use a form on the website to ask about the availability of data from GlaxoSmithKline studies that are not yet listed. These studies can be identified by searching the GlaxoSmithKline Clinical Study Register (www.gsk-clinicalstudyregister.com), ClinicalTrials.gov, or literature databases (such as MEDLINE). The online GlaxoSmithKline Clinical Study Register, which contains more than 5000 studies, will be updated regularly with respect to data availability. In rare instances, particularly for older studies, we may have to decline requests that require extensive resources to retrieve documents from archives and deidentify data from paper records.

As detailed below, an investigative team can apply for access to data sets from a given clini-

cal trial or family of clinical trials to address a specific scientific question. Once investigators have received approval for a research proposal from the independent review panel appointed by GlaxoSmithKline and have signed a data-sharing agreement, data will be made available to them.

WHAT INFORMATION WILL BE MADE AVAILABLE?

GlaxoSmithKline will provide the raw data set (the data collected for each patient in the clinical study) and the analysis-ready data set (the data set analyzed by GlaxoSmithKline and provided to regulatory authorities). These electronic data sets do not contain original radiographs, electrocardiograms, or the like; the information derived from these sources is included. If an investigator needs these original materials for their research, we will try to provide them if we have them and if patient privacy and confidentiality can be protected. Clinical study documents will also be provided, including full protocols with amendments, an annotated case-report form (a blank form annotated to describe the data fields and links to information within the database), the reporting and analysis plan for the study (a document defining the methods of analysis, procedures for data handling, and data displays that GlaxoSmithKline used for the study), a document containing data set specifications (e.g., variable labels, variable descriptions, code lists, and formats), and the clinical-study report (appendixes containing patient-level data are not included, because these data are provided in the data sets). Information that is not available or that is available in a language other than English will be noted on the website for each study. In addition, telephone support will be available to help investigators understand and navigate the data.

WHAT INFORMATION WILL INVESTIGATORS BE REQUIRED TO SUBMIT?

It is important that the analyses proposed by investigators petitioning to access a data set have scientific credibility. We believe that there are public health risks if the proposed analyses are not scientifically robust and give rise to erroneous concerns about safety or false hopes of a potential benefit for patients. Therefore, in accordance with the expectations of usual good scientific practice, investigators will be required to submit a brief research proposal with the use of an online form (Section 2 in the Supplementary Appendix), describing their analysis and publication plans, their

management of potential conflicts of interest, and the qualifications and experience of the research team (which should include a statistician).

HOW WILL RESEARCH PROPOSALS BE REVIEWED?

After they have been processed to ensure that the submitted information is complete, proposals will be reviewed by an independent review panel. The panel will initially comprise external experts appointed by GlaxoSmithKline.

The independent review panel will accept or reject proposals on the basis of their scientific rationale and relevance to medical science or patient care. The panel will also consider the qualifications of the investigators and the management of potential conflicts of interest. To make the decision-making process fully transparent, we provide a list of the members of the independent review panel and their charter on the website (Section 3 in the Supplementary Appendix).

If a research proposal is accepted with conditions or is rejected, the reasons will be communicated to the investigator. Assuming that no further information is needed to supplement the request, we anticipate being able to provide the review panel's decision within 30 days after submission of the proposal, although the timing will depend on the volume of requests received. Deidentification of the data after receipt of a signed data-sharing agreement may take up to 5 days per requested study. A report providing an overview of the number of research proposals submitted and the outcome of the panel's reviews, including an overview of the main reasons for rejection of proposals (if any proposals are rejected), will be made publicly available on our website at least annually.

We view our approach as a step in what we hope will be a widespread movement to provide access to research data from clinical trials as well as to other sources of clinical and health-related data. We acknowledge that, as other parties become involved, our process may change to remain in concordance with evolving international standards.

WHAT ARE THE CONDITIONS ON WHICH ACCESS WILL BE PROVIDED?

Investigators will be responsible for obtaining any other approval that may be required for their research (e.g., from ethics committees, institutional review boards, relevant research institu-

tions, or funding bodies). They will also be required to sign a data-sharing agreement that commits them to use the data only for the research purpose described in the accepted research proposal. The name and affiliation of the lead researcher, the title of the research, the requested studies, a summary of the research for a lay audience, the funding source, and any potential conflicts of interest that were provided in the research proposal will be posted on the website. To promote further transparency, and in line with standard scientific practice, investigators will be required to make a summary of their analysis plan publicly available, to post summary results, and to seek publication of the results after study completion. The content of this publication will not be influenced by GlaxoSmithKline or the review panel; investigators will be required to provide copies of manuscripts after they have been submitted to a journal. If their research uncovers any safety concerns, investigators will be required to inform GlaxoSmithKline and regulatory authorities of those concerns as soon as they are identified.

Investigators will retain the rights to any new intellectual property derived from their research, but GlaxoSmithKline will require a no-cost, non-exclusive license to use any such invention. Any additional rights would be negotiated in good faith. Investigators will be able to view the data-sharing agreement template on the website before submitting a research proposal.

HOW WILL PATIENT PRIVACY AND CONFIDENTIALITY BE PROTECTED?

Access to patient-level data must be provided in a way that minimizes the risk of any breach of patient privacy and confidentiality and is in accordance with privacy laws and regulatory guidance.^{1,2} To meet these requirements, we will remove personally identifiable information from the data set, recode identifiers, remove free-text verbatim terms (including information from the medical history that might compromise anonymity), replace date of birth with age, and revise all dates relative to a random anchor. Furthermore, to meet the anonymization requirements of certain European data-protection authorities, the link (key code) between the data set that is provided and the original data set will be destroyed.

Nonetheless, it may be possible to combine deidentified data with other information to iden-

tify individuals. To minimize any such risk, our approach will be to provide access to anonymous patient-level data on a password-protected website that has controls in place to prevent data from being downloaded or transferred. The analytic software packages SAS (SAS Institute) and R (R Foundation for Statistical Computing) will be provided so that investigators can combine data to conduct analyses and download the results. Investigators will have access to the data sets through this system for 12 months, with that period extended when justified. As a third level of protection for privacy, access to data will be provided only when investigators agree that they will not attempt to identify individuals; this requirement is included in the data-sharing agreement.

It could be argued that after the data are made anonymous (and are no longer considered to be personal information), any restriction on future use based on the original informed consent from the patient no longer applies. However, we believe that any use of data must always align with the original permission that research participants provided during the informed-consent process. An alternative would be to obtain new consent from research participants, but we do not consider that to be a practical alternative. In the past, our informed-consent processes and documents have typically sought permission to conduct research related to the medicine or condition that was the subject of the study. Accordingly, we will require that researchers use data from our clinical trials in a way that is related to the medicine or condition that was the subject of the original study; an assessment of compliance with this requirement will be part of the proposal-review process. For studies we conduct from 2013 onward, patients will be asked to give broader consent to accommodate future requests to analyze their data outside the original intent of the particular study.

CAN DATA FROM DIFFERENT SPONSORS BE COMBINED?

This closed-access system helps protect privacy and ensures that the data are used solely for approved research purposes. The obvious downside to providing access within a restricted environment is that investigators will not be able to combine GlaxoSmithKline data with those collected by other sponsors. We believe the next step must be to transition to a system whereby a

body unconnected with the generation of data acts as the custodian of access to the database (or databases), with the expectation that data will be available from multiple companies, public-sector organizations, and funders. This would allow for the types of cross-sponsor meta-analyses that the research community is seeking.

There are advocates for placing these data in the public domain. We are concerned that such an approach has the potential to compromise patient privacy and lead to erroneous scientific conclusions that may have a negative effect on patient care. There are many subtleties and considerations to be taken into account in reaching a common data-sharing standard. Our approach will clearly need to be modified to accommodate the expectations of other sponsors, academia, and regulators. In particular, it will be important to decide whether a centralized or distributed data-access system should be established.

The GlaxoSmithKline Clinical Study Requests website is a first step from which we can all

learn. GlaxoSmithKline is investing substantial resources in building this capability, including provision for continual expansion of the number of studies available on the website and acquisition of necessary software licenses. Risks regarding data privacy and irresponsible use cannot be totally eliminated, and it will be a challenge to accommodate diverse expectations across the scientific and medical community. However, the opportunity to benefit the health of individuals and the public must outweigh these concerns.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

From Research and Development, GlaxoSmithKline, King of Prussia, PA.

1. Data Protection Working Party. Opinion 4/2007 on the concept of personal data (http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf).
2. Other requirements relating to uses and disclosures of protected health information, 45 C.F.R. § 164.514 (<http://www.gpo.gov/fdsys/pkg/CFR-2012-title45-vol1/xml/CFR-2012-title45-vol1-sec164-514.xml>).

DOI: 10.1056/NEJMsrl302541

Copyright © 2013 Massachusetts Medical Society.