Proceedings

Issues and Case Studies in Clinical Trial Data Sharing: Lessons and Solutions

17th May 2013 • 7:30 AM to 5:00 PM

Wasserstein Hall, Millstein East • Cambridge, MA

TABLE OF CONTENTS

MEETING PARTICIPANTS

Albert J. Allen, MD, PhD*
Senior Medical Fellow
Bioethics & Pediatric Capabilities
Eli Lilly and Company

Mark Barnes, JD, LLM*
Partner, Ropes & Gray
Harvard University MRCT

Barbara Bierer, MD*
Sr. Vice President of Research
Brigham & Women's Hospital
Professor of Medicine
Harvard Medical School
Harvard University MRCT

Martha Brumfield, PhD*
Interim President, Chief Executive Officer
Critical Path Institute

Jeffrey M. Drazen, MD*
Editor in Chief, New England Journal of Medicine

Parker B. Francis Professor of Medicine
Harvard Medical School
Harvard School of Public Health
Brigham and Women's Hospital

Claudia I. Emerson, PhD
Senior Scientist, Sandra Rotman Centre
University of Toronto
Assistant Professor, McMaster University

Yaniv Erlich, PhD*
Andrea and Paul Heafy Family Fellow and Principal Investigator
Whitehead Institute for Biomedical Research

Susan Forda, BPharm, MSc, PhD
Vice President Global Regulatory Affairs, International
Eli Lilly and Company

Jeff Francer, JD, MPP*
Assistant General PhRMA

Sabine Haubenreisser, MSc, PhD*
European Medicines Agency Liaison
United States FDA

Kristen M. Henderson, JD
Regulatory Affairs and Privacy Attorney
Quintiles

Robin Jenkins, MBA*
Head, Strategic Planning
North America Medical Affairs
Sanofi Pharmaceuticals

Aaron S. Kesselheim, MD, JD, MPH
Assistant Professor of Medicine
Harvard Medical School

Richard Kingham, JD*
Partner, Covington and Burling, LLP

Agnes Klein, MD, DPH
Director, Centre for Evaluation of Radiopharmaceuticals and Biotherapeutics
Health Canada

Rebecca Li, PhD
Executive Director
MRCT

Mark D. Lim, PhD
Innovations Program Manager
FasterCures

Elizabeth Loder, MD, MPH*
Senior Research Editor, BMJ
Brigham and Women's Hospital
Harvard Medical School

Holly Fernandez Lynch, JD, MBE
Executive Director, Petrie-Flom Center
Harvard Law School

Michelle M. Mello, JD, PhD*
Professor of Law and Public Health
Harvard School of Public Health

Justin McCarthy, JD*
Senior Vice President
Associate General Counsel
Chief Counsel Worldwide Research and Development
Pfizer

Jules T. Mitchel, PhD, MBA*
President, Target Health, Inc.

Sandra A. Morris, PhD, PMP
Vice President, Strategy Realization
Johnson and Johnson

Sally Okun, RN, MMHS
Vice President, Advocacy, Policy & Patient Safety
PatientsLikeMe

Robert O'Neill, PhD
Senior Statistical Advisor to CDER, OTS; FDA

John J. Orloff, MD
Chief Medical Officer
Senior Vice President, Global Development
Novartis Pharmaceuticals

Evgeny Rogoff, MD, PhD, JD
Clinical Trials Control Dept.
Roszdravnadzor of Russia

Benjamin Roin, JD
Hieken Assistant Professor of Patent Law
Harvard Law School

Jessica Scott, MD, JD*
Director, North America Medical Advocacy and Policy
GlaxoSmithKline

Ira Shoulson, MD
Professor and Director of the Program for Regulatory Science and Medicine
Georgetown University

Patricia Teden, MBA*
Principal, Teden Consulting LLC

Toshiyoshi Tominaga, PhD
Professor and Director, Food and Drug Evaluation Center
Osaka City University Hospital

James Ware, MA, PhD
Frederick Mosteller Professor of Biostatistics
Associate Dean for Clinical and Translational Science
Harvard School of Public Health

Mark Wilenzick, JD
Senior Advisor
Harvard University MRCT

Deborah A. Zarin, MD*
Director, ClinicalTrials.gov
National Library of Medicine

* Speaker

AGENDA

Time	Topic	Moderator/Presenter
7:30 – 8:00 AM	Breakfast and Registration	
8:00 - 8:15 AM	Welcome Remarks	Barbara Bierer Mark Barnes
8:15 – 9:10 AM	***Keynote Speakers***	
	Access to Patient-Level Data from Clinical Trials: What Data? For Whom? For What Reason? When?	Jeff Drazen
	Perspectives on Patient-Level Clinical Data Sharing From a Patient Advocate, Bioethicist and Industry Physician	AJ Allen
9:10 – 9:20 AM	Introduction of potential data sharing models	Michelle Mello
Session I: Rationale for Increased Clinical Trial Data Sharing		***Michelle Mello***
9:20 – 9:35 AM	Case Study: Model of Data Sharing Among Multiple Stakeholders	Martha Brumfield
9:35 – 9:50 AM	Case Study: A Review of The Project Data Sphere Initiative	Robin Jenkins
9:50 – 10: 05 AM	Rationales and Benefits of Data Sharing	Pat Teden
10:05 – 10:45 AM	Discussion Panel / Q & A	
10:45 – 10:55 AM	Break	
Session II: Safeguarding Patient Privacy, Consent Principles, and the Integrity of Data Analyses		***Mark Barnes***
10:55 – 11:10 AM	Responsible Use of Data	Mark Barnes
11:10 – 11:25 AM	Case Study: GSK Data Sharing Initiative as One Model for Increasing Data Transparency and a First Step Toward a Broader Solution	Jessica Scott
11:25 – 11:40 AM	Case Study: How Anonymous is Genetic Research Data?	Yaniv Erlich
11:40 – 12:20 PM	Discussion Panel / Q & A	
12:20 – 12:40 PM	Lunch Served	

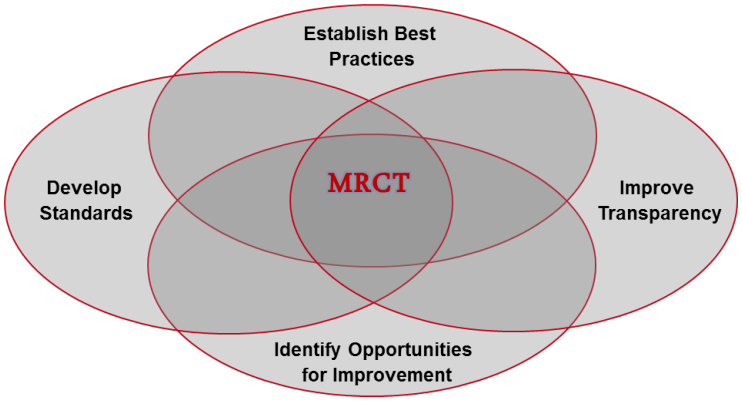
12:40 – 1:00 PM	Lessons Learned From the Implementation of FDAAA and the ClinicalTrials.gov Results Database	Deborah Zarin
1:00 – 1:10 PM	Q & A	
Session III: Balancing Companies' Intellectual Property Interests with Public Access to Data		***Justin McCarthy***
1:10 – 1:40 PM	Case Study: Legal Issues Posed by the EMA Proposed Initiative	Richard Kingham
1:40 – 1:55 PM	Maintaining Incentives to Invest in Research	Jeff Francer
1:55 – 2:35 PM	Discussion Panel / Q & A	
2:35 – 2:50 PM	Break	
Session IV: Assuming Patient-Level Data is Shared in the Public Domain, What are the Ramifications?		***Marc Wilenzick***
2:50 – 3:25 PM	European Medicines Agency Update on Clinical Trial Data Transparency	Sabine Haubenreisser
3:20 – 3:45 PM	Anticipated Impact of BMJ's Data Sharing Policy on Publications and the Scientific Process	Elizabeth Loder
3:45 – 4:00 PM	Regulatory Implications of Data Sharing	Jules Mitchel.
4:00 – 4:45 PM	Discussion Panel / Q & A	
Closing		
4:450 – 5:00 PM	Wrap Up Discussion	Mark Barnes, Barbara Bierer, Rebecca Li, Holly Lynch

INTRODUCTION

MRCT CENTER

Purpose: *To improve the design, conduct, and oversight of multi-regional clinical trials, especially trials sited in or involving the developing world; to simplify research through the use of best practices; and to foster respect for research participants, efficacy, safety and fairness in transnational, trans-cultural human subjects research.*

Objectives:



Establish Best Practices	To establish best practices of performance and to enhance the scientific and ethical conduct of multiregional research.
Develop Standards	To establish common, explicit, realistic and ethical standards for conduct of transnational research
Identify Opportunities for Improvement	To align practice with those standards and study barriers to alignment
Improve Transparency	To enhance predictability for the benefit of research participants, regulatory authorities, sponsors and researchers

EXECUTIVE SUMMARY

In light of recent decisions by the European Medicines Agency (EMA) there is likely to be a call for an increased level of data sharing including from other regulatory agencies. The objectives of this meeting were to provide a forum for discussion of the rationales and implications of this paradigm change and to begin to consider potential solutions.

KEYNOTE SPEAKERS

Access to Patient-Level Data from Clinical Trials: What Data? For Whom? For What Reason? When? **Jeffrey M. Drazen,** MD, Editor in Chief, New England Journal of Medicine, Parker B. Francis Professor of Medicine, Harvard Medical School

Dr. Drazen suggested that there is general agreement that the basic goal is to "obtain information that healthcare professionals can use to improve their ability to care for people who are sick." He discussed 3 potential uses for data sharing: to reproduce/confirm someone else's work, to apply knowledge gained from someone's else's work, to make a decision about an already formulated hypothesis, and/or to use large datasets to try to identify previously unseen patterns that may provide clues to the etiology or possible treatment of an existing disease.

There are currently two resources for patient level clinical trial data. GlaxoSmithKline (GSK) has recently established a website through which researchers can request access to anonymised patient level data from GSK clinical studies. Research proposals are reviewed by an Independent Review Panel. After approval and receipt of a signed Data Sharing Agreement, access is granted to the data through a password-protected website. Open access, participant-level, clinical trial data is also available through the Immune Tolerance Network's (ITN) clinical trials research portal TrialShare. ITN is an international clinical research consortium sponsored by the National Institute of Allergy and Infectious Diseases (NIAID), part of the National Institutes of Health (NIH).

Dr. Drazen said he believes that data sharing is possible and that it is likely to have benefit. He suggested that it is time to take the optimistic view. "Patients put themselves at risk to allow their data to be gathered – we should use it to its fullest."

Perspectives on Patient-Level Clinical Data Sharing From a Patient Advocate, Bioethicist, and Industry Physician **Albert J. Allen**, MD, PhD, Senior Medical Fellow, Bioethics & Pediatric Capabilities, Eli Lilly and Company

Speaking as a patient, Dr. Allen said that patients want to get well today (or as soon as possible), want treatments with the best possible balance of benefits to risks, want as many options for the future as possible, and want to be respected as fellow humans. Concerning data sharing he suggested that patients should have a say in whether their data are studied or there should be someone independently looking out for their interests as well as making sure the science is sound. There must also be accountability with respect to how the data are used. He suggested that many of the requirements to meet these concerns are already in place – we just need to figure out how to use them in the context of data sharing.

Session I The Rationale for Increased Clinical Trial Data Sharing **Michelle M. Mello**, JD, PhD, Professor of Law and Public Health, Harvard School of Public Health

Dr. Mello provided an overview of the four proposed models (Learned Intermediary, Data Generator, Database Query, and Open Access) and discussed some of the core principles that guided their development. (Appendix B and C)

Of primary importance among the list of core principles are respect for the diverse interests and needs of the individual participant who contributed data to the studies, use of a decision making approach that weighs the burdens that sharing of the data might pose against the benefits of the proposed research to public health, fairness, and accountability. There was agreement that requests and decisions concerning the release of data should be posted on the web and decisions about the release of data should include multiple parties (not just the data generator). In addition, for all but the open-access model, requestors should be required to

submit an analytical plan and to sign a data use agreement by which they commit that there are adequate statistical resources on the team to carry out the plan, they will not freely share the data with others, and they will not attempt to re-identify patients.

Case Study: Model of Data Sharing Among Multiple Stakeholders **Martha A. Brumfield**, PhD, CEO and President, Critical Path Institute (C-Path)

Dr. Brumfield illustrated the data-sharing model employed at C-PATH by using a recent project from The Coalition Against Major Diseases (CAMD). After deciding to pursue a research question or hypotheses, C-Path collects the data needed to accomplish the goal then applies a set of standards, which will permit the data to be pooled into a single database. A recent CAMD project on Alzheimer Disease (AD) involved evaluating whether public-private data sharing could aid in developing a tool that would help predict clinical trial design settings for investigational agents more likely to detect a difference in the treatment groups. The end result of the project was a new *in silico* modeling tool that is under review by the Food and Drug Administration (FDA) and the EMA, and a database that is now available for use by qualified scientists.

<u>Best Practices / Key insights</u>

- There is a need for a data sharing/use agreement and rules for merging and accessing data
- Legacy data conversion is resource intensive but worthwhile for specific projects
- Assurance is needed that a specific dataset will be useful in achieving research/regulatory qualification objectives
- New insights can be obtained from data converted to a common standard and aggregated to enable queries and analysis
- Addition of standardized data from other sources (prospective, retrospective) becomes simplified and expands the power and utility of an integrated data resource

Dr. Brumfield believes that shared resources can shorten the drug development time. Challenges remain but these can be overcome with collaboration, diligence, and focus.

Case Study: A Review of Project Data Sphere **Robin Jenkins, MBA**, Head, Strategic Planning North America Medical Affairs, Sanofi Pharmaceuticals

Ms. Jenkins provided an overview of Project Data Sphere, LLC, an independent initiative of the CEO Roundtable on Cancer's Life Sciences Consortium that seeks to alter the trajectory of the cancer death rate by revolutionizing cancer research through responsible sharing and analysis of oncology clinical trial datasets.

Ms. Jenkins noted that Project Data Sphere benefits research scientists by providing them with access to drug discovery data from oncology clinical trials through a single software platform. Companies and institutions benefit from the opportunity for collaboration optimization of trial designs that improve transparency, patient selection, and data analysis. Patients who have participated in clinical trials benefit by knowing that their contributions reach more patients with cancer. About 10 datasets are currently available for sharing. The expectation is to have 60+ datasets, by priority disease area available by the end of 2013.

Subgroup I Report

Patricia Teden, MBA, Principal, Teden Consulting LLC, reported the outcome of Subgroup 1's discussions of the Rationales and Benefits of Data Sharing. Consensus was reached on the following:

- Data sharing rules should apply equally to all study sponsors and data generators
- A purely voluntary regime is not sufficient to create a level playing field
- There should be global standard formats for clinical-trial data and documents to permit comparisons
- The data sharing system must be practical
- Many of the rationales/benefits require participant-level datasets
- Mechanisms for a data sharing system must ensure adequate scientific expertise among the analytical team and provide technical support sufficient to permit users to understand the data
- Some benefits are difficult to achieve in a sponsor-controlled model

- Both summary and participant-level data could be available 1 year after primary study completion, assuming an adjudicated process to obtain participant-level datasets. The evaluation of the purpose for the participant-level datasets could be different ('tighter') prior to product approval.

DISCUSSION PANEL / Q & A

The Question and Answer session centered on the following themes:

- The need for some control over access to the data
- Possible requirements for the data requestor and who will monitor for compliance
- The need for a flexible approach with respect to granting access
- Privacy and how evolving technology has and will continue to affect the ability to re-identify individual patients
- What matters to patients with respect to data sharing, what options should they be offered concerning how their data are shared, will data sharing impact trial enrollment
- How will the proposed models address some of the problems that have been known to occur in the reporting of clinical trial data in the past

Session II: *Safeguarding Patient Privacy, Consent Principles, and the Integrity of Data Analyses* **Mark Barnes, JD, LLM,** Partner, Ropes & Gray, Harvard University MRCT

Subgroup II focused on 5 questions. Mr. Barnes presented an overview of the results of the subgroup's discussions. He noted that with respect to privacy protection, there was consensus among the subgroup members that re-identification is an understandable fear of research participants particularly with respect to sensitive conditions and or genetic information. Group members noted that de-identification is not consistently defined, for example, genetic information is becoming increasingly identifiable, and this may make the Health Insurance Portability and Accountability Act (HIPAA) de-identification standards obsolete.

Overall, however, the group felt the risks associated with data sharing to be limited but dependent on the way in which data are shared, the entities or persons who receive the data,

and the specificity of the data. There may be certain individuals who may have a more compelling reason to fear re-identification and thus privacy violation.

There was general agreement that data sharing must be included as specific information in the informed consent process. Participants should be informed about how their data will be used and by whom, regardless of the method used for sharing

The group found no substantial evidence to indicate that prospective participants in general are greatly dis-incentivized from clinical trial participation out of a fear of privacy violations. Nevertheless, some participants have heightened sensitivity to privacy concerns and certain types of trials pose increased risk of privacy violations because of uniqueness of data, (e.g., pediatric trials, orphan drug trials, rare disease trials, or multi-year trials). It is also possible that IRBs may be reluctant to approve some clinical trials involving participant-level data sharing, when danger of re-identification could be significant.

With respect to liability in the case of misuse of data leading to compromised privacy, the group preferred some form of legislative and/or regulatory measure that would impose civil or criminal liability on data recipients who engage in data "misuse" or who otherwise violate conditions of data use that are set forth in agreements between a data generator or learned intermediary and the data recipient.

Case Study: GSK Data Sharing Initiative as One Model for Increasing Data Transparency and a First Step Toward a Broader Solutions **Jessica Scott, MD, JD,** Director, North America Medical Advocacy and Policy, GlaxoSmithKline

Dr. Scott discussed GSKs transparency initiatives, reiterating GSK's commitment to public disclosure of aggregated clinical research information for all human subject research studies that evaluate investigational or approved medicinal products, including phase I-IV, meta-analysis and observational studies. This public disclosure includes posting of protocol summaries at the start of studies, result summary after completion, publication of the studies along with posting of full protocols. Clinical study reports are posted after approval or

termination of the medicine on the GSK Clinical Study Register. GSK recognizes that these disclosures have limitations as to their usefulness for further scientific research. Therefore, GSK developed an approach to sharing anonymised patient level data with external researchers whereby researchers wishing to access data from clinical studies can select trials listed on a web site and submit a research proposal including an analysis plan, publication plan and details of the researcher expertise and experience. These proposals are reviewed by a panel external to GSK. After approval by the panel, a Data Sharing Agreement is put in place and, access to the data is provided via a secure website.

Case Study: How Anonymous is Genetic Research Data? **Yaniv Erlich, PhD,** Andrea and Paul Heafy Family Fellow and Principal Investigator, Whitehead Institute for Biomedical Research

Dr. Erlich discussed the anonymity of genetic research data by way of an example in which he was able to recover the identity of anonymous sequencing datasets by using only public records combined with allowable HIPAA identifiers of state and age. [Gymrek M et al. *Science* 2013;339(6117):321-324]

DISCUSSION PANEL / Q & A

Questions from the audience focused on the following themes:

- The specific patient-level data that will become available to non-GSK researchers
- The past and current legal practice and perceived obligations concerning the use of participant-level data and how those are changing
- The amount of data collected during a clinical trial and the labor (and associated expense) required to make it usable for an actionable purpose
- Whether there has been an attempt to identify privacy concerns in the developing countries
- What are the current privacy norms and what expectations exist from that perspective
- What role might there be for the IRBs/ECs in setting policy

Lessons Learned from the Implementation of FDAAA and the ClinicalTrials.gov Results Database **Deborah A. Zarin, MD,** Director, ClinicalTrials.gov

During an assessment of the registry database, it became apparent to Dr. Zarin and her staff that there is inconsistent adherence to study protocols and continued evidence of selective publication and outcomes reporting. An assessment of the results database showed a lack of clarity about who is in charge of the science and inconsistent rigor in the data analysis and reporting. In some trials the protocol may be vague, may not exist, or may not be followed. Summary data are not always readily available, even for published trials. For many trials, no one could explain the structure or analysis. In general, said Dr. Zarin, there is no objective, easy to describe route from initial participant level data to the summary data—many people and many judgments are involved. These issues will complicate any data sharing initiative if they are not addressed.

Session III: Balancing Companies' Intellectual Property Interests with Public Access to Data
Justin McCarthy, JD, Senior Vice President, Associate General Counsel, Chief Counsel Worldwide Research and Development, Pfizer

Case Study: Legal Issues Posed by the EMA Proposed Initiative **Richard Kingham, JD,** Partner, Covington and Burling, LLP

Mr. Kingham told the group that until recently, the EMA policy on disclosure of clinical trial information was consistent with that of other developed countries: full study reports were treated as confidential commercial information and not disclosed and assessment reports were disclosed, but with redactions. This policy changed in November 2010 as a direct result of a case brought to the EU Ombudsman by Danish academic researchers seeking data in the MAs for Xenical®/Orlistat.

In March of 2012, the EMA and heads of national medicines agencies (HMA) issued a joint guidance, which stated:

- "Information encompassing non-clinical and clinical development of the medicinal products … is not per se confidential" and

- "In general, the entire content of modules 4 and 5 (nonclinical and clinical study information) can be released, including "case report forms and individual patient listings, when submitted"

Two MA holders sought an annulment of this decision during 2012 and both were granted preliminary measures (a stay pending review). The EMA has announced, however, that it plans to continue releasing CSRs not directly affected by court order.

At the moment, two ongoing legal issues bear watching as they have the potential to affect patent protection and regulatory exclusivity periods. In October 2012, the EMA issued a Guidance requiring sponsors of clinical trials to submit summary reports to the European Union Drug Regulating Authorities Clinical Trials (EudraCT) database. The Guidance includes the timeframe in which this must be done. There is also a proposal before the European Parliament to create a centralized regulatory program for clinical trials in the EU that would require all clinical documents to be submitted to a single "EU portal" and incorporated in an EU database. The database will be publicly accessible unless there is an exemption and will be covered by the same principles of confidentiality as are at issue in the litigation relating to MAs.

Maintaining Incentives to Invest in Research **Jeff Francer, JD**, MPP, Assistant General PhRMA

Mr. Francer said that Biopharmaceutical companies support and are seeking responsible data sharing arrangements, which respect the privacy of research participants and the integrity of the regulatory system but there are some concerns about how forced early disclosure of data might affect a company's ability to recoup that investment. Earlier disclosure of data could require inventors to file patent applications prematurely and possibly in less complete condition, resulting in increased difficulty in prosecution of applications and uncertain patent protection in many countries. Earlier filing also could also reduce the effective patent term for a marketed pharmaceutical invention. Mandated disclosure of certain regulatory files could allow competitors to obtain regulatory approval in other countries without having invested in the research ("free-rider problem"), thus affecting incentives for investment in

biomedical research. In addition, since many data exclusivity provisions only protect data that have not been disclosed to the general public, early release of those data may pose additional negative effects on incentives to invest in biomedical research.

DISCUSSION PANEL / Q & A

Questions centered on the following themes:

- Concern about the need to protect a company's investment leading to human subject research being unnecessarily repeated
- Data sharing may also benefit industry in terms of efficiency and cost avoidance
- Are there legal methods that can be employed to eliminate/reduce the risk of the "free-rider problem"
- The potential for having different policies for safety versus other concerns
- The need to vet the data that are posted to the web to avoid to poor quality data being used to generate "bad" science
- How important is it to limit competitor access to data versus establishing a method to control that use for joint benefit

Session IV: *Assuming Research Participant Data IS Shared in the Public Domain, What are the Ramifications?* **Mark Wilenzick, JD,** Senior Advisor, Harvard University MRCT

European Medicines Agency Update on Clinical Trial Data Transparency **Sabine Haubenreisser, MSc, PhD,** European Medicines Agency Liaison, United States FDA

Dr. Haubenreisser said that the EMA wishes to enable public scrutiny and secondary analysis of clinical trials to allow others to verify/challenge the regulatory authority's position, move regulatory decision-making closer to citizens, and promote better use of medicines. There is also a belief that independent replication of clinical trial data is a legitimate scientific and societal goal. In addition, the EMA believes that access to clinical trial data in an analyzable format benefits public health. It will lead to more efficient drug development, allow developers to learn from past successes and failures, and permit the scientific community to develop new knowledge. The EMA intends to ensure that the personal information of

patients is adequately protected by developing a guarded approach to sharing patient-level data. Policies will respect the boundaries of informed consent. An EMA draft policy will be published by June 30, 2013 and open for public consideration until September 30, 2013. The final policy will be published by November 30, 2013 and will be effective January 1, 2014.

Anticipated Impact of BMJs Data Sharing Policy on Publications and the Scientific Process
Elizabeth Loder, MD, MPH, Senior Research Editor, BMJ, Brigham and Women's Hospital, Harvard Medical School

The British Medical Journal (BMJ) has recently instituted a new data sharing policy [Godlee F, Groves T. *BMJ* 2012;345:e7888] under which drug and device trials will only be considered for publication if the author(s) commit to making relevant anonymised patient level data available on reasonable request. The policy applies to drug and device trials submitted for publication beginning January 2013 regardless of when the trial was conducted and regardless of the source of funding or sponsorship. Dr. Loder anticipates that others will follow and in fact, there is a growing movement in the direction of data sharing and for requiring the registration and publication of all trials. The impact of the policy is expected to have ethical and moral, as well as practical and scientific effects – arguments that are very similar to those made for trial registration. [Krleza-Jeric K et al. *BMJ* 2005;330:956-958]

Regulatory Implications of Data Sharing
Jules T. Mitchel, PhD, MBA, President, Target Health, Inc. reported the results of Subgroup IV's discussions with respect to the following questions.

1. What are the implications of public sharing of clinical trial data for regulatory processes?
    - Will there be a regulatory mechanism for researchers conducting secondary analyses to provide their respective findings to regulators?
    - Will there be obligations imparted to regulatory bodies as a result of any secondary analyses? eg., should the information be sent to a FDA Advisory Committee?
    - What are the implications for drug or device labeling?

- What are the regulatory processes that need to be followed by sponsors?
- Drug companies and medical device manufacturers have certain reporting obligations. What will be the minimum reporting requirements to sponsoring companies for unaffiliated researchers conducting secondary analyses?
- Journals may become inundated with publications from those outside the company performing sub-studies or post hoc analyses and this may lead to second-guessing of labeling, etc.

2. Do the potential benefits of data sharing for regulatory processes outweigh the risks (eg., second-guessing regulatory agencies, premature or incorrect conclusions on risk/benefit profile of medicines)?
    - There could be serious implications if there is selective disclosure about unapproved uses of a drug or medical device that are positive but do not include the negative results.
    - Proponents and opponents of a specific drug may swiftly move to conduct subset analyses of their competitor's product, and this may inevitably lead to challenges with respect to regulatory determinations.
    - Having personal data available, researchers may be able to conduct investigations involving targeted medical treatments, which could define responder rates for subgroups within an indication. What happens then?
3. Can a move toward increased public data sharing jeopardize ongoing efforts toward improved regulatory harmonization?
    - Due to cultural, political, and regulatory differences among regions of the world, there may be conflicts between regions on the use of shared patient level data.
    - Awareness and sensitivity to these issues must be taken into consideration when initiating global clinical programs.

Mr. Wilenzick asked **Robert O'Neill, PhD**, Senior Statistical Advisor, CDER, OTS, FDA; **Deborah A. Zarin, MD**, Director, ClinicalTrials.gov, National Library of Medicine; **Evgeny Rogoff, MD, PhD, JD**, Clinical Trials Control Department, Roszdravnadzor of Russia; and **Toshiyoshi Tominaga, PhD**, Professor and Director, Food and Drug Evaluation Center,

Osaka City University Hospital, to join Drs. Haubenreisser, Loder, and Mitchel on the Panel. He asked each to comment on one aspect of the regulatory issues.

Commenting on the models, Dr. O'Neill said that without a significant upfront clean-up process or recognition that someone is going to have to pay for the cleanup, the type of data sharing under discussion would be impossible to accomplish. He expressed concern about the lack of expertise in analyzing patient-level data in regulatory agencies other than the FDA. In Dr. O'Neill's opinion – the model that works has not yet been thought of.

Dr. Klein said that since the Ottawa Agreement, there has been discussion of some type of results posting in Canada but it has never been clarified. Although there is nothing specific about data sharing in the Canadian regulations, the Minister is permitted to request whatever information he/she believes to be necessary to support a submission. That has always been interpreted to be everything that has been done in terms of studies. There is a general understanding among the Canadian regulators that data sharing needs to and will happen.

Dr. Rogoff commented that there is strict personal data protection legislation in Russia. Violators are prosecuted and can be jailed. As for data, study sponsors do not have an obligation to post their study. The data contained in the MA application is not published and neither are the results of the application review.

Dr. Tominaga noted the Pharmaceuticals and Medical Devices Agency (PMDA) in Japan do not require the submission of patient-level data. However, if there were a requirement for raw patient-level data to be published then PMDA would have an obligation to review them. There is currently a lack of resources and experience to do that but PMDA biostatisticians have begun discussions with their EMA counterparts on how this is done. Japanese regulators are in favor of transparency but they believe it will take time. At the moment, the Japanese government is undecided as to whether clinical trial data are commercial property. With respect to informed consent, they believe that to the extent that clinical data will be used for unknown purposes in the future, explicit consent from the subject is necessary.

DISCUSSION PANEL / Q & A

- Whether a voluntary or mandatory process is preferable
- The need to disclose the extent of the disclosure
- How clinicaltrials.gov might be used as either a source or link to other documents
- Is there a mechanism for the regulators to respond to issues or concerns raised by the re-analysis
- Is there mistrust concerning the regulators' ability to be the sole gatekeeper of the clinical trial data

CLOSING

Dr. Barnes closed the meeting by summarizing some of the important take-away messages.

- There is a positive to data sharing for all of the stakeholders
- Data sharing can be calibrated by the amount of data that are shared, the form in which the data are shared, the level of underlying risk to patients, and to the potential benefit
- The possibility of data sharing and the potential uses to which that data might be put should not be withheld from study participants during the informed consent process
- De-identification is not perfect and will only become more difficult to achieve as technology advances and as public databases advance. This means that we have to be even more careful in how we calibrate the disclosure/sharing of data and the terms and conditions that obligate the recipients of the data for secondary uses
- Recognizing that there are both commercial benefits and risks with data sharing
    - we need to identify those areas in which data sharing can be allowed that will have the maximum public benefit with the minimum disincentive for R&D investment
    - we need to identify which intellectual property laws could be changed to allow us to maximize data sharing while not harming the intellectual property rights the promote investment
- Transparency should extend to the secondary users of the data
- The Learned Intermediary Model appears to best accommodate the different pressures and concerns about data sharing

- Change can be incremental and it's going to be painful and take time
- As more and more data are shared and reanalyzed, there is likely to be increased friction in the regulatory environment
- Data sharing can be a positive for industry because it can lead to abbreviated processes and might also increase confidence in the clinical trial process.
- To the degree that data sharing leads to the absolute right of researcher to use data and bio specimens for downstream uses in controlled and appropriate ways that is a positive good for everyone.

MEETING SUMMARY

It appears that – even if not at the level currently proposed by the European Medicines agency (EMA) – there is likely to be a call for an increased level of data sharing from the various regulatory agencies. Such a change would affect the conduct of trials worldwide – as such, it is a subject of interest for the Multi Regional Clinical Trials Group (MRCT).

Mark Barnes, JD, LLM, Partner, Ropes & Gray, and Harvard University MRCT and **Barbara Bierer**, MD, Sr. Vice President of Research, Brigham & Women's Hospital, Harvard Medical School, and Harvard University MRCT, co-chairs of the meeting, welcomed the speakers and attendees and reviewed the specific goals of the meeting, which were to:

- Review the rationales for requirements to disclose participant-level clinical trials data
- Discuss implications of data disclosure requirements
- Review evidence from recent experiences with participant-level data disclosure
- Provide a new, multi-stakeholder perspective on potential solutions and criteria for access to participant-level data for public health and scientific research purposes
- Identify potential areas of collaboration on these issues among stakeholder groups

Dr. Bierer explained that the process for this meeting began in February 2013 with the formation of the Clinical Trial Data Sharing & Transparency Working Group, which is co-chaired by Salvatore Alesci MD, PhD, Pharmaceutical Research and Manufacturers of America (PhRMA) and Michelle Mello, JD, PhD, Harvard School Public Health. A complete list of the Working Group and Subgroup Members can be found in Appendix A.

The working group comprises 4 subgroups each charged with exploring one of the themes discussed during this meeting:

- Rationales for and Benefits of Data Sharing
- Responsible Use of Shared Data
- Innovation and Incentives
- Regulatory Implications

The specific objectives of the subgroups were to establish common criteria for data-sharing models, identify and describe reasonable data sharing models, and involve diverse stakeholders across pharma, biotech, and academia, while balancing between:

- assuring simplicity in operation, maintenance and access

- protecting participant privacy
- maintaining commercial incentives for medical product development
- meeting the public health and scientific requirements for data transparency

The meeting began with two keynote speakers.

Keynote 1

Access to Patient-Level Data from Clinical Trials: What Data? For Whom? For What Reason? When?

Jeffrey M. Drazen, MD, Editor in Chief, New England Journal of Medicine, Parker B. Francis Professor of Medicine, Harvard Medical School

Although the specific benefits of clinical trial data sharing are somewhat unclear at the moment – Dr. Drazen suggested that there appears to be general agreement that the basic goal is to "obtain information that healthcare professionals can use to improve their ability to care for people who are sick." He discussed 3 potential uses for data sharing: to reproduce/confirm someone else's work, to apply knowledge gained in someone's else's work to make a decision about an already formulated hypothesis, and/or to use large datasets to try to identify previously unseen patterns that may provide clues to the etiology or possible treatment of an existing disease.

The idea of data sharing is not new. In the area of gene research (and for gene studies conducted as part of clinical trials), most journals require the authors to provide full access to relevant data sets through a publicly accessible repository such as the Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) or ArrayExpress (http://www.ebi.ac.uk/arrayexpress/). The accessibility of these data allows researchers to pursue a hypothesis at less cost than in the lab and without putting patients at risk. Another use of large data sets is the FDA Mini-Sentinel Pilot Project (http://www.mini-sentinel.org/), a national electronic system launched by the Food and Drug Administration (FDA) in 2008 to help scientists understand potential safety issues associated with approved medical products. The database includes clinical data from 17 database US partners and data for 99 million patients, 2.9 billion prescription drug dispensings, and 2.4 billion unique medical

encounters, including 38 million acute inpatient hospital stays. There is no direct access to the data. Scientists pose a query and the individual database partners run the query against their own data, and respond. Because the personal information associated with the individual data is never released, privacy is protected.

When the discussion turns to sharing clinical trial data, the questions and the answers take some thought. It is generally accepted that the highest quality data come from randomized clinical trials (RCTs) but data generated in RCTs are expensive to get and there is a lot of it in various stages of detail from individual instrument readings, laboratory results, imaging results, and patient and clinician completed assessments to summary data from the clinical study and other reports. In a sense, the answers to the "What" "For Whom" and "Why" are inter-related.

There are currently 2 resources for patient level clinical trial data. GlaxoSmithKline has just recently established a website (https://clinicalstudydata.gsk.com/) through which researchers who wish to conduct additional research can request access to anonymised patient-level data from GSK clinical studies. Studies are listed on the site after the drug studied has been approved by regulators or terminated from development and the study has been accepted for publication. Currently listed studies (~200 as of the date of the meeting) include those completed since 2007. They may also inquire about the availability of data from clinical studies not listed on the site before they submit a research proposal. Research proposals are reviewed by an Independent Review Panel. Although the panel is appointed by GSK, GSK is only involved in the decision making process for inquiries about access to data from studies not on this site. Following approval and after receipt of a signed Data Sharing Agreement, access is granted to the password-protected website. The second resource for open access to participant-level clinical trial data is the Immune Tolerance Network's (ITN) clinical trials research portal TrialShare (https://www.itntrialshare.org/). ITN is an international clinical research consortium sponsored by the National Institute of Allergy and Infectious Diseases (NIAID), part of the National Institutes of Health (NIH).

In closing, Dr. Drazen believes that data sharing is possible and that it is likely to have benefit. He suggested that although it could also bring trouble – it is time to take the optimistic view. "Patients put themselves at risk to allow their data to be gathered – we should use it to its fullest."

Keynote 1

Perspectives on Patient-Level Clinical Data Sharing From a Patient Advocate, Bioethicist, and Industry Physician

Albert J. Allen, MD, PhD, Senior Medical Fellow, Bioethics & Pediatric Capabilities, Eli Lilly and Company

Dr. Allen's perspective on the topic of data sharing is informed by his education, his career as a physician and medical researcher, his personal life and family life, and the fact that he is currently receiving chemotherapy for multiple myeloma. For his presentation today – he spoke primarily as a patient and in that regard about what patients want. He said that patients want to get well today (or as soon as possible), want treatments with the best possible balance of benefits to risks, want as many options for the future as possible, and want to be respected as fellow humans. Concerning data sharing he suggested that patients should have a say in whether their data are studied or there should be someone independently looking out for their interests as well as making sure the science is sound. There must also be accountability with respect to how the data are used.

He also identified some potential negative consequences of data sharing: the impact of publicity, litigation, education, etc. on medical practice and public health outcomes and on future research/investment into treatments; the impact of diversion of resources from other research to examine questions raised by flawed analysis.

In closing, Dr. Allen suggested that many of the requirements to meet these concerns are already in place – we just need to figure out how to use them in the context of data sharing.

Q & A

Questions following these two presentations focused on patient attitudes toward data sharing and the unique needs of the regulators. Some members of the group suggested that patients expected their data to be shared. Others disagreed and specifically mentioned trials involving what might be considered sensitive conditions. Concerning the regulators need for transparency there was a suggestion that timing of the data release might be an issue. The benefit of clinical trial data with respect to surveillance for adverse events was mentioned as having a public health value.

Introduction To The Models

Michelle M. Mello, JD, PhD, Professor of Law and Public Health, Harvard School of Public Health opened the formal sessions with a discussion of the core principles that guided the model development and an overview of the four proposed models. (Appendix B and Table 1)

The models were developed using a set of core principles and common elements (Appendix C). Dr. Mello considers several of these to be of paramount importance, in particular respect for the diverse interests and needs of the individual participant who contributed data to the studies. Application a cost/benefit approach to each model is another important concept and there was a general consensus among the workgroup participants that some case-by-case analysis of the benefits of a particular analysis to public health ought to be weighed against the burdens that sharing of the data might pose. Fairness – that is treating all data generators equally – is also critical to acceptance of any of the models and, finally, the models should include mechanisms to ensure accountability on the part of all involved. To this end, there was agreement among the workgroup that requests and decisions concerning the release of data should be posted on the web and decisions about the release of data should include multiple parties (not just the data generator). In addition, for all but the open-access model, requestors should be required to submit an analytical plan and to sign a data use agreement by which they commit that there are adequate statistical resources on the team to carry out the plan, they will not freely share the data with others, and they will not to attempt to re-identify patients.

Table 1 Models for Review

Gatekeeper Models		
	Learned Intermediary Model	**Data Generator Model**
Decision-maker	Review Board*	Data Generator
Criteria	1. *Sound science:* Is there a reasonable scientific hypothesis, sound analytical plan, and adequate plan to disseminate findings? 2. *Benefit/risk balancing:* Do the potential public health benefits of answering the proposed question(s) outweigh the probable adverse effects on the data generator (intellectual-property interests, competitive concerns, and technical support burden) and the potential risks to research participants? 3. *Expertise:* Does the research team have expertise sufficient to carry out the proposed analyses?	
Process	• Board reviews request, collects input from data generator, decides, and publicly documents rationale for decision	• Data generator reviews request, decides, and publicly documents rationale for decision. • Denials are appealable to independent appellate board, whose decision is final.
"Black Box" / Database Query Model		
Decision-maker	Independent review board ***or*** data generator	
Criteria	Same as for Gatekeeper Model	
Process	• Requester submits a research query to the data holder • Data holder runs the query and returns results—not data	
Open Access Model		
Decision-maker	None	
Criteria	Responsible-use attestation: All requests granted if Requester attests that data will not be used inappropriately (eg., to re-identify research participants)	
Process	• Data generator routinely posts data from trials when results are publicly reported or submitted to regulator, along with documentation to facilitate use of data • Researchers can simply download the material	

* That is independent of the data-generator

Session I: Rationale for Increased Clinical Trial Data Sharing

Moderator: Michelle Mello

Case Study: Model of Data Sharing Among Multiple Stakeholders

Martha A. Brumfield, PhD, CEO and President, Critical Path Institute (C-Path)

The pace at which knowledge, data, information, and technology advance is beyond the ability of any one entity to harness – a collaborative model in which risk and cost are shared is more productive. C-Path is a non-profit institution that functions as a neutral 3rd party to

bring together experts from around the world to gain consensus on particular research objective or hypothesis. C-Path also actively works with various regulatory agencies. Seven global consortia are currently collaborating with 1,000+ scientists and 41 companies looking at biomarkers, clinical outcome assessment instruments, clinical trial simulation tools, and data standards. The premise from which C-Path approaches data-sharing has to do with accelerating the learning about disease, particularly very complex disease such as neurologic disorders, and how that can inform the development of better therapies.

Dr. Brumfield used The Coalition Against Major Diseases (CAMD) work on Alzheimer Disease (AD) to illustrate how data sharing can be leveraged to accelerate biopharmaceuticals development. After deciding to pursue a research question or hypothesis, C-Path collects the data needed to accomplish the goal then applies a set of standards, which will permit the data to be pooled into a single database. The AD project serves as a case study for public-private data sharing to develop a tool that would help predict clinical trial design settings for investigational agents more likely to detect a difference in the treatment groups. CAMD received legacy data (in various formats) for 24 AD trials from 9 companies. The data were then remapped to create a fully integrated database that could be used for queries. The end result of this effort was a new *in silico* modeling tool, which is under review by the FDA and EMA, and a database that is now available for use by qualified scientists

Best practices that emerged from this process included the need for the following governance considerations:

1. Data Sharing/Use Agreement
    - Protects subjects and owners of data
    - Up front identification of objectives and governance (how the data will be collected, what will be done with it, how it will be analyzed, published, etc.)
2. Rules for merging data
    - De-identify data to HIPAA “Safe Harbor” and European Union Data Protection Directive requirements
- High value data in a (Computer Society International Design Competition (CSIDC) standard accepted by regulatory agencies

3. Rules for accessing data
    - Obtain broadest possible data use agreement that meets regulatory requirements for secondary use of data
    - Use access controls appropriate to research objectives

Key insights

- Legacy data conversion is resource intensive but worthwhile for specific projects
- Assurance is needed that a specific dataset will be useful in achieving research/regulatory qualification objectives
- New insights can be obtained from data converted to a common standard and aggregated to enable queries and analysis
- Addition of standardized data from other sources (prospective, retrospective) becomes simplified and expands the power and utility of an integrated data resource

Dr. Brumfield believes that shared resources can shorten the drug development time. Challenges remain (eg., patient privacy/informed consent, review/approval for secondary use, respect for confidentiality and intellectual property, data sharing/pooling mechanisms, risk mitigation, cultural resistance, effective communication of why and value proposition, incentives, and funding sources) but these can be overcome with collaboration, diligence, and focus. Suggested additional reading: Sharing Clinical Research Data Workshop Summary available at http://books.nap.edu/openbook.php?record_id=18267

Case Study: A Review of Project Data Sphere

Robin Jenkins, MBA, Head, Strategic Planning North America Medical Affairs, Sanofi Pharmaceuticals

Project Data Sphere, LLC (http://ceo-lsc.org/projectdatasphere) is an independent initiative of the CEO Roundtable on Cancer's Life Sciences Consortium that seeks to alter the trajectory of the cancer death rate by revolutionizing cancer research through responsible sharing and analysis of oncology clinical trial datasets.

The group has worked to address 4 issues that have been identified as barriers to progress in the area of data sharing:

1. *Privacy* Although privacy is one of the most commonly expressed concerns in connection with data sharing, de-identifying patient-level data for research purposes is possible. Ms. Jenkins discussed one company that accomplished it by using an Expert Determination Model. Alternative methodologies and additional information are available at the Department of Health and Human Services (HHS) website: http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/guidance.html.
2. *Security* Data security cannot be overlooked, and Project Data Sphere has collaborated with the SAS Institute, a leader in data security and analytics, to develop a platform that should be publically available during summer 2013. All of the data will be hosted by SAS, and will be behind their firewall and encrypted using Secure Socket Layer (SSL) Protocols. All documents will be scanned for viruses before being made public. Users must apply for and be granted permission to access the data, which is controlled through role based permissions and passwords.
3. *Contracts* A team of legal experts working with Project Data Sphere has developed uniform legal agreements that balance the needs of both the data provider and the data user. The data provider retains ownership and all existing IP while contributing meaningful datasets. Users may not seek patent protection for research procedures or research designs that result from their research using the database and may not assert against a data provider any patent right that results from the user's use of the provider's data. There are no other restrictions or limitations on users seeking to patent inventions that result from their research using the database.
4. *Resources* In the view of Project Data Sphere, the resource commitment on the part of the data contributor is fairly minimal, relative to trial cost and benefits of sharing, when the existing legal, information technology, and biostatistics infrastructure are leveraged. A critical success factor is to identify a "Champion" of data sharing to facilitate effort within organization.

Under Project Data Sphere, the benefits of data sharing accrue to researchers by providing access to drug discovery data from oncology clinical trials through a single software platform and to companies and institutions by enabling collaboration on cancer research and

optimization of trial designs that improve transparency, patient selection, and data analysis. Benefits for trial participants include the opportunity to see their efforts reach beyond a single study to reach more patients with cancer. About 10 datasets are currently available for sharing. The expectation is to have 60+ datasets by priority disease areas available by the end of 2013.

Patricia Teden, MBA, Principal, Teden Consulting LLC, reported the discussions from Subgroup 1.

Questions considered:

1. What are the specific rationales for data sharing? What public health benefits arise from being able to share and access data?
2. What kinds of data need to be shared to reach those goals and in what format?
3. What conditions must be present to ensure that data sharing adequately achieves the identified goals

Team discussions on questions 1 & 2 led to the identification of 11 Rationales/Benefits, 3 potential data formats, and 4 potential document types that could be shared with 7 possible types of stakeholders. To refine their results, the team conducted 2 internal surveys, one to connect the stakeholders to one or more rationales/benefits and a second to map the rationales/benefits with data types and document formats. The top two stakeholder beneficiaries for each of the 11 rationales/benefits are shown in Figure 1. Table 2 contains the team consensus on the data formats and types of documents that are necessary to achieve each goal/benefit. They considered data summaries (similar to those found on clinicaltrials.gov to be important for all 11 rationales/benefits.

Figure 1 All Stakeholders Stand to Benefit

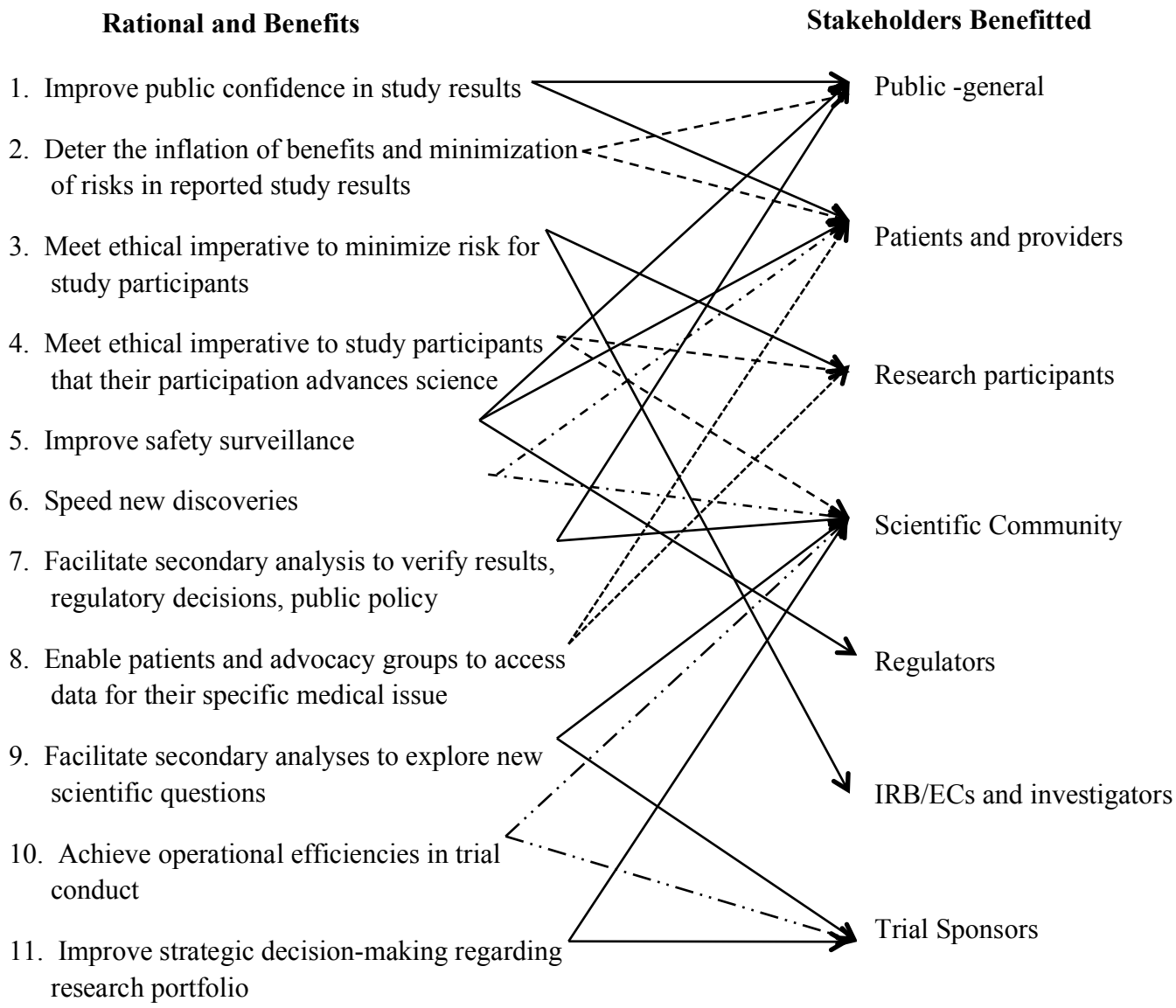


Table 2 Materials That Support the Rationales and Benefits

	Rationales and Benefits	Material That Should Be Shared
1.	Improve public confidence in study results	Data summaries, participant –level datasets Documents: Lay summaries
2.	Deter the inflation of benefits and minimization of risks in reported study results	Data summaries, participant-level datasets, all datasets Documents: Clinical Study Reports (CSRs)
3.	Meet ethical imperative to minimize risk for study participants	Data summaries Documents: scientific publications
4.	Meet ethical imperative to study participants that their participation advances science	Data summaries, participant-level datasets Documents: scientific publications
5.	Improve safety surveillance	Data summaries, participant-level datasets, all datasets Documents: CSRs
6.	Speed new discoveries	Data summaries, participant-level datasets, all datasets Documents: CSRs
7.	Facilitate secondary analysis to verify results, regulatory decisions, public policy	Data summaries, participant-level datasets, all datasets
8.	Enable patients and advocacy groups to access data for their specific medical issue	Data summaries, participant-level datasets, all datasets Documents: CSRs, lay summaries, scientific publications
9.	Facilitate secondary analyses to explore new scientific questions	Data summaries, participant-level datasets, all datasets Documents: CSRs, scientific publications
10.	Achieve operational efficiencies in trial conduct	Data summaries Documents: CSRs
11.	Improve strategic decision-making regarding research portfolio	Data summaries Documents: CSRs

The subgroup reached consensus on the following concepts:

1. Data sharing rules should apply equally to all study sponsors and data generators
2. Something beyond a purely voluntary regime is desirable to create a level playing field
3. There should be global standard formats for clinical-trial data and documents to permit comparisons
4. The rationales and benefits presuppose that initial and re-analyses of shared data will reflect sound science
    a. Data sharing systems should have mechanisms for promoting responsible use of data

    b. Accountability standards should be similar for the initial sponsor or data generator, and a researcher conducting a re-analysis
5. Data sharing system must be practical
6. Many of the rationales/benefits require participant-level datasets
7. Important mechanisms for a data sharing system must exist to:
    a. Ensure adequate scientific expertise among the analytical team
    b. Provide technical support sufficient to permit users to understand the data
8. Some benefits are difficult to achieve in a sponsor-controlled model
9. Timing of availability for both summary and participant-level data should be 1 year after primary study completion.
    a. Assuming an adjudicated process to obtain participant-level datasets, evaluation of the purpose for the participant-level datasets could be different ('tighter') prior to product approval.

Reactions to the Four Models

The Learned Intermediary model was thought to be appealing – but there were questions about who would run and pay for it. The team felt Data Generator Review was insufficient to garner public trust and ensure consistency across all sponsors/data generators and that the Database Query model was also flawed on the grounds of transparency and practicability. While they considered the Open Access model to have maximal transparency, there was some concern about the cost of ensuring scientific integrity in how data are used. The group felt they did not have enough information about the experience with the existing initiatives nor measures of their effectiveness.

DISCUSSION PANEL / Q & A

The speakers for this session were joined by **John J. Orloff, MD**, Chief Medical Officer, Senior Vice President, Global Development, Novartis Pharmaceuticals, **James Ware, MA, PhD,** Frederick Mosteller Professor of Biostatistics, Associate Dean for Clinical and Translational Science, Harvard School of Public Health and **Sally Okun RN, MMHS,** Vice President, Advocacy, Policy & Patient Safety, *PatientsLikeMe*. Dr. Orloff, Dr. Ware, and Ms.

Okun were given the opportunity to comment on the presentations prior to the audience question and answer period.

Dr. Ware discussed the labor intensiveness of preparing the data for sharing. After acknowledging the excellent work that has already been done, Dr. Orloff suggested that additional attention should be paid to the process of matching the data to the need, in particular to the level of data. Ms. Okun spoke to how important the issue of data sharing is to patients, including their right and desire to be informed at the end of the study so they can better understand what their commitment meant. With respect to the labor intensiveness of the process, she agreed with the other speakers, but suggested that it is important to balance this against the cost of not sharing. In closing, she mentioned that some effort would have to be made to determine how patients will be able to access these data.

The audience Question and Answer session centered on the following themes:

- The need for some control over access to the data to ensure an appropriate level of scientific rigor and accountability in the use of the data
- Who controls the access (eg., in the C-Path model it is the data owner not C-Path)
- Possible requirements for the data requestor, (eg., a data analysis plan, a publication plan, willingness to sign a data-use agreement) and who will monitor for compliance
- The need for a flexible approach with respect to granting access (eg., the risk of the condition, control arm data vs. active arm data, the level of data that is being requested, pre-vs. post-approval for different trial stages)
- Privacy and how evolving technology has and will continue to affect the ability to re-identify individual patients
- What matters to patients with respect to data sharing, what options should they be offered concerning how their data are shared, whether data sharing will impact trial enrollment, how will patients who do not want their data shared be handled at enrollment,
- The status of the somewhat restrictive informed consent documents used in older studies
- How well will the proposed models address some of the problems that have been known to occur in the reporting of clinical trial data in the past

Session II: Safeguarding Patient Privacy, Consent Principles, and the Integrity of Data Analyses

Moderator: Mark Barnes

Mr. Barnes reminded the group that there is already a method that investigators can use to access data from closed clinical trials. Investigators wishing to access data from closed clinical trials can apply to the appropriate Institutional Review Board (IRB) for a waiver of informed consent. The IRB then reviews each request on a case-by-case basis using criteria that essentially requires that there be minimal risk to the patients/subjects. What is problematic with this approach is that to obtain all the data for a single study would require application to many IRBs – the transaction cost for such an endeavor would be quite high.

Subgroup II focused on 5 questions. Mr. Barnes presented a summary of the group's Consensus for following each question.

1. What are the risks of clinical trial data sharing with regard to privacy protection, and how can they be balanced against the potential social benefits of data sharing?
    a. *Risks:* De-identification (eg., by employers, insurance companies, family members, or self-identification) is a fear of research participants particularly with respect to sensitive conditions and or genetic information. There are also certain discrete and insular minorities who may have a greater risk of privacy violation, even without de-identification.
    b. *Consensus:* Overall, the risks associated with data sharing are limited but they depend on the way in which data are shared, the entities or persons who receive the data, and the specificity of the data. Despite a generally overall low risk, there may be certain individuals who may have a more compelling reason to fear privacy violation. The risk suggests that researchers should disclose to subjects the full extent (or at least the range) to which collected data will be used.

2. Would privacy concerns related to clinical trial data sharing deter prospective research participants from participating in clinical studies?
    a. *Consensus:* No substantial evidence indicates that prospective participants are greatly dis-incentivized from clinical trial participation out of a fear of privacy violations. Nevertheless, some participants have heightened sensitivity to privacy concerns and certain types of trials pose increased risk of privacy violations because of uniqueness of

data, (eg., pediatric and orphan drug or multi-year trials). It is also possible that IRBs may be reluctant to approve clinical trials involving participant level data sharing.

3. Does de-identification of data solve the problem of risks to participant privacy and confidentiality?
    a. *Consensus:* De-identification is not consistently defined, and it is becoming increasingly difficult to assure as technology improves. Genetic information is becoming increasingly identifiable, which may make the HIPAA de-identification standards obsolete. The degree of de-identification is inversely related to data usefulness: the more identifiers removed, the less useful the data become to subsequent researchers

4. Should participants' consent serve as a precondition for sharing of clinical trials data?
    a. *Consensus:* Data sharing must be included as a specific part of the informed consent process and should be a precondition to participation in a clinical trial. Participants should be informed about how their data will be used and by whom, regardless of the method used for sharing.

5. What should be the consequence (eg., liability) if/when privacy is compromised as a result of increased clinical trial data sharing?
    a. *Consensus:* The range of potential consequences for privacy violations depends on a variety of factors. The first preference of the subgroup was for legislative and/or regulatory measures that impose civil or criminal liability on data recipients who engage in data "misuse," such as re-identifying participants or sharing the clinical data set with additional users without authority to do so. Alternatively, data sharing provisions should be enforced through agreements between a learned intermediary and the data user in which the data user states they will not "misuse" information and acknowledges that breaking of the pledge can lead to liability for breach of contract.

Reactions to the Models

From a privacy perspective, the subgroup felt that the Learned Intermediary and Data Generator model were similar in that in both cases there is an entity that can assess the risk of re-identification for each data set disclosed, deny or rigorously control access in situations of heightened risk, and tailor a data use agreement to the appropriate risk level. They felt the Database Query model to be the

safest from a privacy perspective because the data requester never gains access to the data set; however, there is a concern that this may be difficult to operationalize because the system must respond to a heterogeneous group of data queries. The most dangerous model, from the perspective of privacy is the Open Access model. If adopted, there must be appropriate penalties are in place to discourage misuse of data.

Jessica Scott, MD, JD, Director, North America Medical Advocacy and Policy, GlaxoSmithKline discussed GSKs transparency initiative.

GSK has committed to public disclosure of clinical research information for all human subject research studies that evaluate investigational or approved medicinal products (includes Phase 1 to 4, meta-analyses, and observational studies). Disclosure begins with the posting of the protocol summary at study start. Within ~8 to 12/18 months of study completion GSK will post a summary of results. GSK also commits to submit a manuscript for each study within 18 to 24 months of study completion and to post the full protocol and CSR on the GSK Clinical Study Register at the time of publication (in the case of the protocol) or after approval or termination of the medicine in the case of the CSR. These publicly disclosed results include summarized data with statistics to compare treatment groups but not primary data from each research participant.

GSK has also committed to make patient level data available in a manner that protects the privacy of research participants and ensures that the data are used for valid scientific investigation. Under the current model, they will provide anonymised patient level data to a GSK appointed Independent Data Custodian after completion of a project and publication of the results. Researchers wishing to access the data are required to submit scientific proposals and analysis plans to the Independent Data Custodian, who reviews the proposals and expertise of the requestor and manages any conflicts as well as privacy. Requests and responses are made and access to data is provided via a secure website (https://clinicalstudydata.gsk.com).

There is also a method through which researchers may inquire about the availability of data from GSK studies that are not listed before they submit a research proposal. These inquiries will be reviewed by GSK.

Any arrangement to share data is subject to the terms of a Data Sharing Agreement under which the data requestor agrees to:

- Only use the data for the agreed purpose
- Not to attempt to establish the identities of research participants
- Inform regulatory authorities and GSK of any safety concerns as soon as they are identified
- Provide GSK with a copy of the manuscript after submission to a peer reviewed journal
- Post and seek publication of the research
- Allow GSK to use any invention that comes out of the research and negotiate any further rights in good faith

Yaniv Erlich, PhD, Andrea and Paul Heafy Family Fellow and Principal Investigator, Whitehead Institute for Biomedical Research discussed the anonymity of genetic research data by way of an example in which he was able to recover the identity of anonymous sequencing datasets by using only public records combined with allowable HIPAA identifiers of state and age. [Gymrek M et al. *Science* 2013;339(6117):321-324]

The speakers for Session II were joined by **Kristen M. Henderson, JD**, Regulatory Affairs and Privacy Attorney, Quintiles, **Mark D. Lim, PhD**, Innovations Program Manager, FasterCures, and **Claudia I. Emerson, PhD**, Senior Scientist, Sandra Rotman Centre, University of Toronto, Assistant Professor, McMaster University for a question and answer session.

DISCUSSION PANEL / Q & A

Questions from the audience focused on the following themes:

- The specific data that will become available to researchers outside of GSK as part of the patient level data (eg., age, adverse events, etc)
- The past and current legal practice and perceived obligations concerning the use of participant level data and how those are changing
- The amount and breadth of data collected during a clinical trial and the labor (and associated expense) required to make it usable for an actionable purpose
- Whether there has been an attempt to identify privacy concerns in the developing countries, especially in light of past experiences with ethical violations and taking into consideration linguistic and cultural differences
- What are the current privacy norms and what expectations exist from that perspective
- What role might there be for the IRBs/Ethics Committees (ECs) in setting policy
- What is the process when a request from an outside researcher is denied

Lessons Learned From the Implementation of FDAAA and the ClinicalTrials.gov Results Database
Deborah A. Zarin, MD, Director, ClinicalTrials.gov

ClinicalTrials.gov is a registry (>145,000 listed trials) and results database (results from >8000 trials) of publicly and privately supported clinical studies of human participants conducted around the world. The results database requires sponsors to report (for publication) the "minimum data set" specified in the trial protocol in tabular format. The report is divided into four modules: participant flow, baseline characteristics, outcome measures, and adverse events. Dr. Zarin noted that when she and her staff assessed the registry database they identified inconsistent adherence to study protocols (more so outside of industry studies) and continued evidence of selective publication and outcomes reporting. In a similar assessment of the results database, the found a lack of clarity about who is in charge of the science and an inconsistent rigor in the data analysis and in the reporting of all subjects and data (changed outcomes, measures, structural changes such as study arms that are eliminated or added).

An underlying assumption of the results database was that the required data (number of participants starting and completing the trial and their characteristics, the prespecified outcomes and the results on those outcomes, and the number and type of adverse events) were being routinely generated after the clinical trial. There was also an assumption that the burden of reporting would be mostly associated with data entry and time requirements. What the clinicaltrial.gov staff found, however, was that these assumptions were incorrect. In some trials the protocol is vague, may not exist, or may not be followed. Summary data are not always readily available, even for published trials. For many trials, no one could explain the structure or analysis. In general, there is not an objective, easy to describe route from initial participant-level data to the summary data—many people and many judgments are involved. A comparison of 110 matched pairs of clinicaltrials.gov results entries and a corresponding publication showed that 82% had at least one important discrepancy.

Dr. Zazin noted that many documents contribute to understanding the path from un-coded participant-level data to summarized results – including the Data Safety Monitoring Board (DSMB) reports, which have not been part of the discussion. She cautioned the group, however, that the huge volume of trials that will need to be considered in any plan to share data and non-systematic data release could generate a new kind of "disclosure bias."

Session III: Balancing Companies' Intellectual Property Interests with Public Access to Data

Moderator: **Justin McCarthy, JD**, Senior Vice President, Associate General Counsel, Chief Counsel Worldwide Research and Development, Pfizer

Enhanced data sharing is accompanied by many complex issues. One of those is the need to maintain the incentives to undertake high-risk research. Those incentives take the form of intellectual property protection like patents, regulatory data protection, and trade secrets, all of which can be affected by data sharing. Patents can be compromised by the premature release of data, as can regulatory data exclusivity – particularly in countries where the regulatory protection is not that robust. Trade secrets can be affected by the release of the CSR, which contains regulatory information not directly related to patients. These intellectual property questions go to the heart of some of the issues being addressed in the EU right now, including changes to clinical trial regulation.

Case Study: Legal Issues Posed by the EMA Proposed Initiatives

Richard Kingham, JD, Partner, Covington and Burling, LLP discussed public access to nonclinical and clinical data in marketing authorization (MA) files, disclosure of information relating to clinical trials while the trials are underway, and the legal and policy implications of the recent EU transparency initiatives.

Until recently, the EMA policy on disclosure of clinical trial information was consistent with that of other developed countries. Full study reports were treated as confidential commercial information and not disclosed. Assessment reports were disclosed, but with redactions. Patient-level data has not previously been an issued in the EU since the EMA does not require the submission of case report forms and patient listings they have not previously been available to disclose. This policy changed in November 2010, however, as a direct result of a case brought to the EU Ombudsman by Danish academic researchers seeking data in the MAs for Xenical®/Orlistat. Having failed to get the parties to compromise the Ombudsman took the position that the data contained in full CSR were not confidential commercial information and that the EMA was wrong to withhold the documents. As a result, the EMA agreed to revise their disclosure policy. In March of 2012, the EMA and heads of national medicines agencies (HMA) issued a joint guidance, which stated:

- "Information encompassing non-clinical and clinical development of the medicinal products … is not per se confidential" and

- "in general, the entire content of modules 4 and 5 (nonclinical and clinical study information) can be released"

Two MA holders sought an annulment of this decision during 2012 and both were granted preliminary measures (a stay pending review) by the President of the General Court of the European Union. The EMA has announced, however, that it plans to continue releasing CSRs not directly affected by court order. They are also considering the possibility of streamlined procedures that could conceivably lead to a policy under which the agency would routinely upload some set of information from the CSRs as drugs are approved. There have also been reports that the EMA is considering a policy that would require the submission of individual patient listings or case report forms. Mr. Kingham mentioned that there have been reports indicating that many of the requests for data have been submitted by competitors or law firms and only a minority has been submitted by researchers.

At the moment, two ongoing legal issues bear watching as they have the potential to affect patent protection and regulatory exclusivity periods. In October 2012, the EMA issued a Guidance addressing the posting and publishing of result-related information relating to clinical trials conducted in the EU. The Guidance requires sponsors of clinical trials to submit summary reports to the European Union Drug Regulating Authorities Clinical Trials (EudraCT) database and outlines the timeframe in which this must be done. Under this Guidance, non-pediatric Phase 1 studies will not be released but all other trials will be made public within 15 days of posting. There is no exception for studies completed prior to receiving a MA. There is also a proposal before the European Parliament to create a centralized regulatory schema for clinical trials in the EU that would require all clinical documents to be submitted to a single "EU portal" and incorporated in an EU database. Submissions will include full clinical trials applications (CTAs), and all of the other documents relating to the trial, including correspondence, inspection reports, etc. There is also the requirement for submission of a summary report of all clinical studies conducted under the regulation to be submitted within 1 year of study completion. The database will be publicly accessible unless there is an exemption and will be covered by the same principles of confidentiality as are at issue in the litigation relating to MAs.

The patent issue is related to the timing of the data release. Product patents are almost always filed well before the initiation of clinical trials. Although use patents can be filed at any time, it is becoming increasingly common for them not to be filed until after the clinical trials for a particular condition have been completed. Public disclosure of clinical trial results prior to submission of use patents could undermine patentability by forcing submission of applications before full information is

available to support them. Data exclusivity provisions differ around the world but all rely on some period of time during which the government cannot rely on data summited by innovators to approve a copy product. These regulations were designed with the expectation that the data submitted by the innovator would not be public. Making the data public may undermine the basis for this system, which serves as an important adjunct to patent protection. Even if full reports are not used directly to support competitive applications, they can contain valuable commercial information, for which the innovator has already paid.

Mr. Kingham concluded by noting that policies on public disclosure of safety and effectiveness data should be developed in a manner that takes account of the need to preserve incentives for innovation, as well as the legitimate needs of researchers and others for information relating to the drug development and approval process.

Jeff Francer, JD, MPP, Assistant General PhRMA, discussed some of the issues that Subgroup III discussed relating to *Maintaining Incentives to Invest in Research*. The questions addressed by Group III were

1. What intellectual property rights, proprietary interests, and competitive concerns do companies have that may be adversely affected by data sharing [by either voluntary or mandated clinical data disclosure policies]?
2. Would the impingement on these interests that could accompany data sharing likely affect public and private investments in R&D and, ultimately, innovation?
3. How should these concerns be balanced against the potential benefits of data sharing?
4. What strategies might effectively address companies' legitimate concerns while maximizing the public benefit of data sharing?
5. Is imposing a "learned intermediary" between those who seek access to data and the data sources a possible approach to ease competitive concerns while still allowing reasonable access for independent researchers?

According to a 2007 economic study, it takes an average of 10 to 15 years and costs an average of $1.2 billion to bring one new drug to market. [DiMasi JA and Grabowski HG *Managerial and Decision Economics* 2007; 28: 469–479] Biopharmaceutical companies support and are seeking responsible data sharing arrangements that respect the privacy of research participants and the integrity of the regulatory system but there are some concerns about how forced early disclosure of data might affect a company's ability to recoup that investment. Earlier disclosure of data could

require inventors to file patent applications prematurely and possibly in less complete condition, resulting in increased difficulty in prosecution of applications and uncertain patent protection in many countries. Earlier filing also could also reduce the effective patent term for a marketed pharmaceutical invention. Mandated disclosure of certain regulatory files could allow competitors to obtain regulatory approval in other countries without having invested in the research ("free-rider problem"), thus affecting incentives for investment in biomedical research. In addition, since many data exclusivity provisions only protect data that have not been disclosed to the general public, early release of those data pose additional negative effect on incentives to invest in biomedical research.

The question Subgroup III addressed was how to achieve the goals of data sharing without harming the incentives to invest in biomedical research. They identified several potential strategies:

- Restricting access of data to "qualified" individuals / institutions (not competitors)
- Restricting the use of data through a contractual arrangement
    - Specify acceptable uses of data for research purposes
    - Specifying requirements for keeping data set confidential / not transferable
    - Prohibiting the filing of patent applications on inventions made from data set
    - Potential penalties for misuse
- Delay disclosure of data
- "Learned intermediary" to manage and control the data flow and use

Susan Forda, BPharm, MSc, PhD, Vice President Global Regulatory Affairs, International, Eli Lilly and Company, **Benjamin Roin, JD**, , Harvard Law School, **Ira Shoulson, MD**, Georgetown University, **Aaron S. Kesselheim, MD, JD, MPH**, Harvard Medical School, and **Sandra A. Morris, PhD, PMP**, Vice President, Strategy Realization, Johnson and Johnson joined Mr. Kingham and Mr. Francer on the panel.

Dr. Morris commented that there is a broad range of rationales for requesting access to data from verifying the outcomes of a study to using the data to move forward on a new concept. As such, it is important to match the type of data that is released with the question that is being asked and further to match the control process to the type of data. From a European perspective, Dr. Ford noted that with the new EMA process, industry is looking for 3 basic guarantees:

- that there be a robust process in place that allows the sponsor to interact with the EMA to discuss what the release will encompass,
- that there be due consideration for what should be considered confidential, and

- the relationship as to whether the data release is pre- versus post-approval.

She also mentioned that even though the data required for the new EudraCT database are the same as that for clinicaltrials.gov the upload process is completely different, meaning that everything will need to be redone for postings to both databases. It has been estimated that within 5 years the cost of this new process will be the same as that needed to bring one new molecular entity to market. She asked that the personnel from the databases come to agreement to either merge the databases or agree to the same upload process to avoid the duplicate cost.

DISCUSSION PANEL / Q & A

Questions for this panel centered on the following themes:

- Concern about the need to protect a company's investment leading to research being repeated on new study participants without full knowledge that answers may already have been found in earlier studies and/or without full knowledge of potential adverse effects
- Data sharing may also benefit industry in terms of efficiency and cost avoidance
- Are there legal methods that can be employed to eliminate/reduce the risk of the "free-rider problem"
- The potential for having different policies for safety versus other concerns
- Concern that without some vetting of the data that are posted to the web may lead to poor quality data being used to generate "bad" science
- How important is it to limit competitor access to data versus establishing a method to control that use for joint benefit
- Concern that the discussions/concerns about what *can't be done* are standing in the way of immediately doing what is easily done to address some of the really pressing medical issues we are facing

Session IV: Assuming Research Participant Data IS Shared in the Public Domain, What are the Ramifications?

Moderator: **Mark Wilenzick, JD,** Senior Advisor, Harvard University MRCT

European Medicines Agency Update on Clinical Trial Data Transparency

Sabine Haubenreisser, MSc, PhD, European Medicines Agency Liaison, United States FDA provided an EMA update on Clinical Trial Data transparency.

The new EMA position that clinical trial data should be openly available is based on several factors:

- it is necessary to build trust and confidence in the system
- there is an ethical responsibility to the patients enrolled in clinical trials
- independent (re)analysis of data broadens the knowledge base and contributes to the public health
- sharing of complex data can open new horizons and contribute to scientific progress

Most, but not all, stakeholders are very positive about the Agency's initiative. Parts of the drug industry are lobbying against full transparency and they have expressed concern that the availability of detailed clinical trial datasets will encourage flawed 'secondary analyses' that will give rise to unfounded public health scares and that such data could be used by competitor companies to produce copycat drugs in less regulated markets.

Since the November 2012 stakeholder meeting, the EMA has formed four advisory groups (Table 3) consisting of representatives from patient groups, the pharmaceutical industry, consultants, law firms, research institutes, non-governmental organizations, healthcare professionals, academia, regulators and the media. The advisory groups met between January-March 2013 and a final advice from each group was published 30 April 2013. The stakeholders expressed divergent views and suggested different options on particular aspects and these are reflected in the final advice from each group, which are posted on the EMA website. The EMA will make the final decision regarding which option to choose.

Table 3 EMA Advisory Groups/Questions of Interest

Protecting Patient Confidentiality
Question: How can the Agency ensure through its policy that patient and other personal information will be adequately protected, that patients cannot be retroactively identified when clinical trial data are released, and that applicable legislation, standards, and rules regarding personal data protection will be respected?
Clinical Trial Data Formats
Question: How can the Agency ensure through its policy that clinical trial data can be shared in the interests of public health, in a clear and understandable format that enables appropriate analyses, and with a swift implementation without undue burden to stakeholders?
Rules of Engagement

Question: Are there rules or conditions that should be in place before an external stakeholder can download clinical trial data (e.g. formal acceptance of the need to respect personal data rules, uploading of analysis plans)?
Legal Aspects
Question: Are there any legal aspects other than personal data protection that need to be addressed when drafting the Agency's policy? Are there exceptional circumstances under which data can be claimed to be commercially confidential?

The EMA wishes to enable public scrutiny and secondary analysis of clinical trials to allow others to verify/challenge the regulatory authority's position and to take regulatory decision-making closer to citizens and promote better use of medicines. There is also a belief that independent replication of clinical trial data is a legitimate scientific and societal goal. In addition, the EMA believes that access to clinical trial data in an analyzable format benefits public health. It will lead to more efficient drug development, allow developers to learn from past successes and failures, and permit the scientific community to develop new knowledge. The EMA intends to ensure that the personal information of patients is adequately protected by developing a guarded approach to sharing of patient-level data. Policies will respect the boundaries of informed consent. In the view of the EMA, patients participate to support development of particular medicine; any other use (e.g. for commercial purposes) oversteps the boundaries and should not be enabled. In order to protect the public health and regulatory decisions against claims of inappropriate secondary data analyses, EMA policy will address conflicts of interest and set quality standards. The same standard of transparency will apply to secondary analyses as was applied to the primary analysis.

An EMA draft policy will be published by June 30, 2013 and open for public consideration until September 30, 2013 for public consultation. The final policy will be published by November 30, 2013 and will be effective January 1, 2014.

Dr. Haubenreisser was asked whether the policy would apply broadly to all trial sponsors – not just industry. She responded that the policy covers all CSRs submitted to the EMA as part of a MA application for which the scientific committee has given an opinion (ie. approved or denied applications). It also includes the CSRs that were received in connection with MA applications that were withdrawn during the authorization process.

Over lunch and during the afternoon break the attendees were asked to vote for the model they most preferred. Mr. Wilenzick announced that the model with the most votes was the Learned Intermediary model.

Anticipated Impact of BMJs Data Sharing Policy on Publications and the Scientific Process
Elizabeth Loder, MD, MPH, Senior Research Editor, BMJ, Brigham and Women's Hospital, Harvard Medical School

The British Medical Journal (BMJ) has recently instituted a new data sharing policy [Godlee F, Groves T. *BMJ* 2012;345:e7888] under which drug and device trials will only be considered for publication if the author(s) commit to making relevant anonymised patient-level data available on reasonable request. The policy applies to drug and device trials submitted for publication beginning January 2013 regardless of when the trial was conducted and regardless of the source of funding or sponsorship. Trials of diagnostic tools or surgical operations or other interventions that are not drugs or devices are currently excluded from the policy. The *BMJ* also encourages researchers to seek informed consent for sharing at recruitment, even if there are no current plans to share data.

With the new policy, the Editors of the *BMJ* expect researchers who are requesting data to submit a protocol for their re-analysis to the authors and to commit to making their results public. The Editors encourage those requesting data to send a rapid response to bmj.com describing their intended use of the data. If they refuse a request for access, the authors of the paper will be asked to explain why. All of this is accomplished in a public forum. Relevant data is defined as "All anonymised data on individual patients on which the analysis, results, and conclusions reported in the paper are based."

The Editors of *BMJ* see this as an evolution of policy and as an acknowledgement that a journal article can no longer standalone as a sufficient record, especially over time, of what happened in a clinical trial. They anticipate that others will follow and in fact, there is a growing movement in the direction of data sharing and for requiring the registration and publication of all trials. The impact of the policy is expected to have ethical and moral as well as practical and scientific effects – arguments that are very similar to those made for trial registration. [Krleza-Jeric K et al. *BMJ* 2005;330:956-958]

Jules T. Mitchel, PhD, MBA, President, Target Health, Inc. presented the results of subgroup IV's discussion on the following questions.

1. What are the implications of public sharing of clinical trial data for regulatory processes?
    - As the results of analyses become publicly available, will there be any issues that the FDA or other regulatory bodies need to address?
    - Will there be obligations imparted regulatory bodies as a result of any secondary analyses? eg., should the information be sent to a FDA Advisory Committee?
    - What are the implications for drug or device labeling?
    - What are the regulatory processes that need to be followed by sponsors?
    - Will there be a regulatory mechanism for researchers conducting secondary analyses to provide their respective findings to regulators?
    - Drug companies and medical device manufacturers have certain reporting obligations (eg., adverse events or patient safety issues) to regulatory agencies- what will be the minimum reporting requirements to sponsoring companies for unaffiliated researchers conducting secondary analyses.
    - Journals may become inundated with publications from those outside the company performing sub-studies or post hoc analyses and this may lead to second-guessing of labeling, etc.
2. Do the potential benefits of data sharing for regulatory processes outweigh the risks (eg., second-guessing regulatory agencies, premature or incorrect conclusions on risk/benefit profile of medicines)?
    - There could be serious implications if there is selective disclosure about unapproved uses of a drug or medical device that are positive but do not include the negative results.
    - Proponents and opponents of a specific drug may swiftly move to conduct subset analyses of their competitor's product, and this may inevitably lead to challenges with respect to regulatory determinations.
    - Having personal data available, researchers may be able to conduct investigations involving targeted medical treatments, which could define responder rates for subgroups within an indication. What happens then?
3. Can a move toward increased public data sharing jeopardize ongoing efforts toward improved regulatory harmonization?
    - Due to cultural, political, and regulatory differences among regions of the world, there may be conflicts between regions on the use of shared patient level data.

- Awareness and sensitivity to these issues must be taken into consideration when initiating global clinical programs.

Robert O'Neill, PhD, Senior Statistical Advisor to CDER, OTS, FDA; **Toshiyoshi Tominaga, PhD**, Professor and Director, Food and Drug Evaluation Center, Osaka City University Hospital; **Agnes Klein, MD, DPH**, Director, Centre for Evaluation of Radiopharmaceuticals and Biotherapeutics, Health Canada; **Deborah A. Zarin, MD**, Director, ClinicalTrials.gov, National Library of Medicine; and **Evgeny Rogoff, MD, PhD, JD,** Clinical Trials Control Dept. Roszdravnadzor of Russia, joined Drs. Haubenreisser, Loder, and Mitchel for the panel discussion.

Before proceeding to the audience question and answer session Mr. Wilenzick asked the new members of the panel for comment on specific items. Dr. O'Neill was asked to comment on the models. Noting the enormous volume of complex data that are submitted to with FDA with a New Drug Application (NDA), Dr. O'Neill said that without a significant upfront clean-up process or recognition that someone is going to have to pay for the clean-up, the type of data sharing that is being discussed will be impossible to accomplish. He expressed concern about the lack of expertise in analyzing patient-level data in regulatory agencies other than the FDA. In Dr. O'Neill's opinion – the model that works has not yet been thought of.

Dr. Klein was then asked to comment on the impact of patient-level data sharing on the regulatory process. Although the Canadian system of review is somewhat of a hybrid between the FDA model and the model commonly understood to be used by the EMA, much like the FDA, in Canada they tend to look at the totality of the evidence. Since the Ottawa Agreement, there has been discussion of some type of results posting but what exactly that means has never been clarified. There is nothing specific about data sharing in the Canadian regulations but the regulations do permit the Minister to request whatever information he/she believes to be necessary to support a submission. That has always been interpreted to be - everything that has been done in terms of studies. That said there is a general understanding among the Canadian regulators that data sharing and the associated processes need to happen.

Dr. Rogoff was asked whether making patient level data available is an issue in Russia. He commented that there is strict personal data protection legislation in Russia. Violators are prosecuted and can be jailed. As for data, study sponsors do not have an obligation to post their study. The data

contained in the MA application is also not published and neither are the results of the application review.

Dr. Tominaga, was asked whether access to patient-level data is a concern in Japan. At the moment, the Pharmaceuticals and Medical Devices Agency (PMDA) in Japan do not require the submission of patient-level data. However, if there is a requirement for raw patient level data to be published then PMDA would have to review them and there is a lack of resources and experience to do that. PMDA biostatisticians have just begun discussions with their EMA counterparts on how this is done. Japanese regulators are in favor of transparency and believe it is right to open up the data but it will take time. At the moment, the Japanese government is undecided as to whether clinical trial data are commercial property and with respect to informed consent they believe that to the effect that clinical data will be used for unknown purposes in the future, explicit consent from the subject is necessary.

DISCUSSION PANEL / Q & A

Questions for the panel centered on the following themes:

- Whether a voluntary or mandatory process is preferable
- The need to disclose the extent of the disclosure, for example a list of all possible clinical trial documents with check marks next to those that are being disclosed for a particular trial and a list of all related trials with an indication of which are being disclosed
- How clinicaltrials.gov might be used as either a source or link to other documents
- Is there a mechanism for the regulators to respond to issues or concerns raised by the re-analysis
- Is there mistrust concerning the regulators' ability to be the sole gatekeeper of the clinical trial data

Meeting Summary and Close

Mr. Barnes closed the meeting by summarizing some of the important take-away messages from the meeting.

- There are significant benefits of data sharing, for all of the stakeholders.
- Data sharing can be calibrated by the amount of data that are shared, the form in which the data are shared, the level of underlying risk to patients, and to the potential benefit.
- The possibility of data sharing and the potential uses to which that data might be put should be told to study participants during the informed consent process.

- De-identification is not perfect and will only become more difficult to achieve as technology advances and as public databases advance. This means that we have to be even more careful in how we calibrate the disclosure/sharing of data and the terms and conditions that obligate the recipients of the data for secondary uses
- Recognizing that there are both commercial benefits and risks with data sharing:
    - we need to identify those areas in which data sharing can be allowed that will have the maximum public benefit but with minimum disincentive for R&D investment; and
    - we need to identify which intellectual property laws could be changed to allow us to maximize data sharing while not harming the intellectual property rights that promote R & D investment.
- Transparency should extend to the secondary users of the data.
- The learned intermediary model appears best to accommodate the different pressures and concerns about data sharing, although there are logistical and practical issues with relying solely on that model.
- Change can be incremental, and it will take time to get data sharing "right."
- As more and more clinical trials data are shared and reanalyzed, there is likely to be increased friction in the regulatory environment, especially concerning second guessing regulatory agency decisions.
- Data sharing can be a positive for industry because it can lead to abbreviated processes and might also increase confidence in the clinical trial process overall.
- To the degree that data sharing leads to the researcher's use of data and bio-specimens for downstream uses in controlled and appropriate ways, it is a positive good for everyone.

Dr. Bierer agreed with Mr. Barnes adding that the models are not yet sufficiently nuanced. Additional work is need to understand the kinds of data, the kinds of uses, the kind of intermediary or perhaps no intermediary, and how we want to understand data. In particular, data integrity is critical. We need to think about who is going to bear the cost of doing the work that needs to be done to validate the date and perform the analysis to make the posted summary data reliable not only for the primary posting but for the secondary study. Increased transparency will improve public trust over time. She suggested once a decision has been made concerning the posting of clinical trial data there is an obligation to go back to the IRBs for the studies in which the informed consent documents were not specific about future use to obtain advice and consent about where and what to post.

Holly Fernandez Lynch, JD, MBE, Executive Director, Petrie-Flom Center, Harvard Law School thanked the staff of the MRCT for their cooperation in putting this conference together. She invited the attendees to visit the Petrie-Flom website (http://www.law.harvard.edu/programs/petrie-flom/) to read more about their programs.

Appendices

Appendix A: Clinical Trial Data Sharing & Transparency Working Group

Appendix B: MRCT Consolidated Models Summary Table

Appendix C: Common Criteria Summary

APPENDIX A

Clinical Trial Data Sharing & Transparency Working Group

Launched February 15, 2013

Co-chairs Salvatore Alesci (PhRMA)/Michelle Mello (Harvard School Public Health)

AJ Allen (Lilly)
Mark Barnes (Ropes & Gray / MRCT)
Barbara Bierer (Partners Healthcare / MRCT)
Melissa Binz (Novartis)
Karen Craun (Sanofi)
Amy Davis (Public Responsibility in Medicine and Research)
David Dorsey (Johnson & Johnson)
Jeffrey Francer (PhRMA)
Kate Heffernan (KGH Advisors)
Julie Kaberry (Harvard)
Marcia Levenstein (Pfizer)
Jennifer Miller (Harvard)
Jules Mitchel (Target Health)
Sandra Morris (Johnson & Johnson)
Pearl O'Rourke (Partners Healthcare)
Mercy Osakpawan (Harvard School of Public Health)
David Peloquin (Ropes & Gray)
Thomas Peppard (Gates)
Roshni Persaud (MRCT)
Patricia Teden (Teden Consulting)
Fabio Thiers (Vis Research)
Marc Wilenzick (MRCT)

Subgroups

Subgroup 1: Rationales for and Benefits of Data Sharing *Lead: Patricia Teden*

AJ Allen	Eli Lilly & SACHRP
Karen Craun	Sanofi Oncology
Amy Davis	Public Responsibility in Medicine and Research
Mercy Imahiyerobo	Harvard School of Public Health
Michelle Mello	Harvard School of Public Health
Jennifer Miller	Safra Center, Harvard University
David Peloquin	Ropes & Gray
Tom Peppard	Global Health Program, Gates Foundation
Pat Teden	Teden Consulting
Fabio Thiers	ViS Research Institute

Subgroup II: Responsible Use of Data Lead: Mark Barnes / Roshni Persaud

Mark Barnes	MRCT, Harvard Law School, Ropes & Gray LLP
Melissa Binz	Novartis Pharmaceuticals
Jeffrey Francer	PhRMA
Kate Gallin Hefferman	KGH Advisors
Michael Hughes	Harvard School of Public Health
Mercy Imahiyerobo	Harvard School of Public Health
Julie Kaberry	Harvard School of Public Health
Marcia Levenstein	Pfizer
Pearl O'Rourke	Partners HealthCare
David Peloquin	Ropes & Gray LLP
Roshni Persaud	MRCT

Subgroup III: Innovation and Incentives Lead: Salvo Alesci / Jeff Francer

Salvatore Alesci	PhRMA
Melissa Binz	Novartis
Jeffrey Francer	PhRMA
Rebecca Li	MRCT
Justin McCarthy	Pfizer
Sandra Morris	Johnson & Johnson
David Peloquin	Ropes & Gray
Roshni Persaud	MRCT
Fabio Thiers	ViS Research

Subgroup IV: Regulatory Implications of Data Sharing Lead: Jules Mitchel

Barbara Bierer	MRCT / Brigham and Women's Hospital
David Dorsey	Janssen Research & Development
Rebecca Li	MRCT
Jules Mitchel	Target Health
Sandra Morris	Johnson & Johnson
David Peloquin	Ropes & Gray
Roshni Persaud	MRCT
Marc Wilenzick	MRCT

APPENDIX B

MRCT Consolidated Models Summary

1. Learned Intermediary Model

Who: In this model, a learned intermediary body ("Review Board") would review requests for access to the clinical trial data. The members of the Review Board would have no financial ties or other relationship with the Data Generating companies.

What: Applications to the Review Board for release of data must provide information sufficient for the Review Board to make the following three determinations:

1. Is there a reasonable scientific hypothesis, sound analytical plan, and adequate plan to disseminate findings?
2. Do the potential public health benefits of answering the proposed question(s) outweigh the probable adverse effects on the Data Generator, considering intellectual-property interests, competitive concerns, technical-support burden, and the potential risks to research participants?
3. Does the researcher and research team have expertise sufficient to carry out the proposed analyses?

How:

- Requests would be submitted to the Review Board via a website.
- Data Generators would be given 3 weeks to submit information detailing any concerns they have about the request, and estimating the technical-support burden involved.
- The Review Board would publicly document the rationale for each decision.
- If the request is approved, the Review Board would also oversee the creation of a standard template data user agreement to be signed by the requester and Data Generator for each release.

When: The Review Board would ordinarily respond to requests within 60 days.

2. Data Generator Review Model:

Who: Requests would be reviewed by the Data Generator, whose decision would be subject to appeal through an external review process.

What: Applications to the Data Generator for release of data must provide information sufficient for the Data Generator to make the following three determinations:

1. Is there a reasonable scientific hypothesis, sound analytical plan, and adequate plan to disseminate findings?
2. Do the potential public health benefits of answering the proposed question(s) outweigh the probable adverse effects on the Data Generator, considering intellectual-property interests, competitive concerns, technical-support burden, and the potential risks to research participants?
3. Does the researcher and research team have expertise sufficient to carry out the proposed analyses?

How:

- The data requester would submit a written request via website to the Data Generator which would apply the three criteria specified above to determine whether or not to grant each request.
- The Data Generator would publicly document the rationale for each decision.
- For denied requests, the requester could appeal to an Appellate Board that is independent of the Data Generator.
- The Data Generator would be given an opportunity to submit further explanation of its decision to the Appellate Board.
- The decision of the Appellate Board would be final.

When: The Sponsor would ordinarily respond to requests within 60 days. For denied requests, an appeal should be received within two weeks. The Appellate Board should make a decision within 30 days of the appeal request.

3. Black Box / Database Query Model:

Who: A learned intermediary or the Data Generator could serve as the decision maker. This party (the "data holder") would keep the dataset(s), which would be available for queries; however, the data would not be directly accessed or downloaded by others. Rather, the data holder would implement the query and return the results to the requester.

What: Applications for release of data must provide information sufficient to make the following three determinations:

1. Is there a reasonable scientific hypothesis, sound analytical plan, and adequate plan to disseminate findings?
2. Do the potential public health benefits of answering the proposed question(s) outweigh the probable adverse effects on the Data Generator, considering intellectual-property interests, competitive concerns, technical-support burden, and the potential risks to research participants?
3. Does the researcher and research team have expertise sufficient to carry out the proposed analyses?

Due to the fact that only results, not research participant level data, are released, privacy concerns assume less importance in the benefit/risk weighing for this model than in other models.

How: The requester would send a research question to the data holder. A software program would then run on the remote dataset(s) and return an answer to the researcher. For complex queries that exceed the capabilities of the software, in-house programmers may be needed.

When: A timeframe for responding to such requests would need to be established. This time frame will depend in part on the complexity of the query and the amount of programming expertise required beyond automated software algorithms.

4. Open Access Model

Who: Data Generators would post clinical trial data on a central portal maintained by an independent organization. Researchers would have open access to the data. Data Generators could petition an independent board for approval to omit certain studies from the portal due to national security concerns, heightened risk to research participants' confidentiality, or compelling intellectual property issues.

What: Going forward, research participants' identifiable information might be included, but at least for existing data sets, the data are likely to require redaction or de-identification.

How: When a trial has been completed and the results are publicly reported or submitted to a regulatory agency, the Data Generator would post in the public domain: the study protocol for the published trial, trial amendments made for specific jurisdictions, and underlying clinical trial data (with research participants' information removed). Once data are made available, researchers would then be able to download the data set via the web. All requests would be granted so long as requester attests that data will not be used for inappropriate purposes (e.g., attempts to re-identify patients).

When: A timeframe for posting such data, after approval of the drug or, in the case of an approved drug, completion of the trial, would need to be established.

APPENDIX C

Common Criteria Summary

Core Principles Used in the Development of the Data Sharing Models

It is recommended that any model ultimately adopted for Data Sharing should promote, to the maximum extent possible, the following principles:

- Protect human research participants
- Advance biomedical innovation and public health
- Balance risks with benefits to be gained from data sharing
- Treat all data generators equally, independent of their affiliation, (eg., industry, government, academia)
- Make data disclosure practicable by avoiding undue burdens on data generators and requesters
- Provide timely access to data
- Transparency
- Accountability

Common Elements

It is recommended that each of the data sharing models contain the following common elements (with a few notable exceptions)

1. **Use of a web interface where summary information about each data request is publicly viewable**, including the identity of the requestor, the proposed uses, the outcome of the request; and for denied requests a comprehensible explanation of the reason for denial. The rationale for this element is to maximize transparency and accountability for data requestors, data generators, and any intermediary bodies
2. **Pre-commitment by requesters for a rigorous analytical plan** (*may not be applicable to the "Open Access" Model). There are three rationales for this element: (1) to promote adherence to sound scientific methods, including the notion that analyses should be tailored to address a reasonable, prospectively posed hypothesis; (2) to enable data generators or intermediaries evaluating requests for data to weigh the risks and benefits associated with particular proposed releases; and (3) to minimize risks to research subjects whose data are released

A. as a precondition for receiving data, the data requester must register the following information on the website:

- The identity of the requester and all persons who will have access to the data
- The scientific hypothesis to be explored using the data
- A short statement of the analytical plan, sufficient to permit a conclusion to be drawn about the requester's ability to address the stated hypotheses by following the stated methods
- The minimum necessary information to carry out the proposed analyses

B. As a precondition for receiving data, the data requester must sign a data use agreement in which they attest that:

- Data will not be shared with others outside the team that is registered on the website. Additional team members will be added to the registered team before the data are shared with them
- No attempt will be made to re-identify research participants
- The team includes appropriate expertise adequate to carry out the proposed analyses
- Prior to pursuing any analyses outside the scope of the registered plan, the team will register a supplemental analysis plan on the website and obtain approval to proceed.

C. Models should specify some mechanism for enforcing or promoting compliance with these requirements

3. **Participation by both the data generator and other parties in any decisions made about data releases.** The rational for this element is to help ensure that while the data generators' views are represented, data generators do not inappropriately deny requests for data, and in doing so, public trust is maintained in the data-sharing system