Opportunities through data sharing

Andrew Freeman
Head of Medical Policy
GSK

Disclosures and Potential Conflicts of Interest

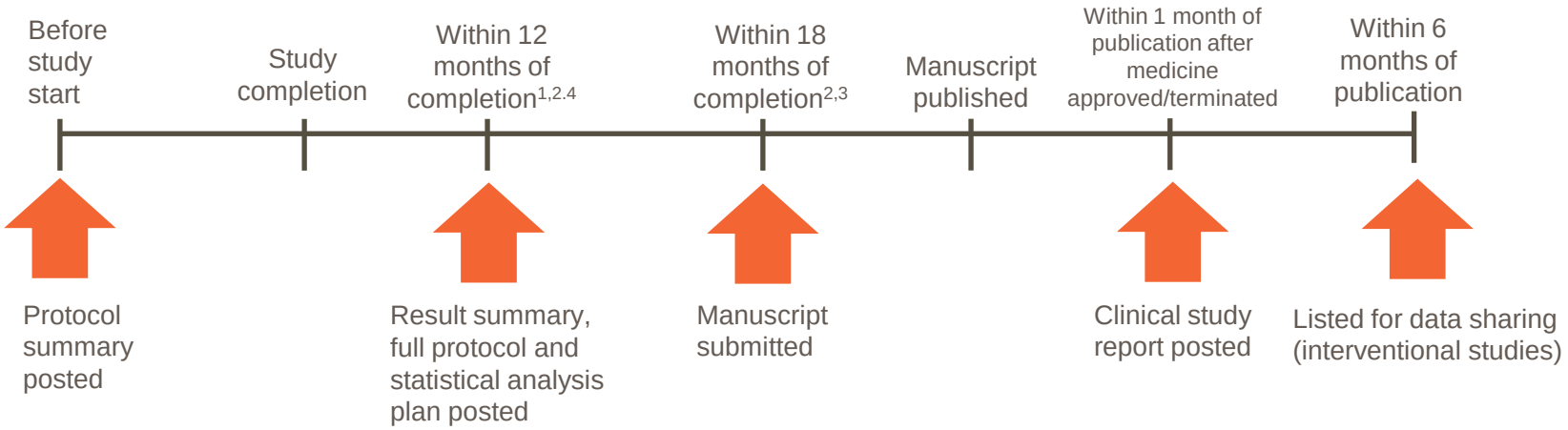
I am a full time employee of GSK and hold stocks and stock options

The views presented in these slides are my own and not necessarily those of GSK

What I’m going to talk about...

- I’m old ---I’ll say a little bit about the history of clinical trial data sharing at GSK
- I’ll show you some data on re-use of clinical trial data and some examples
- Along the way I’ll discuss some of the issues and opportunities
  - Fragmentation
  - Impact
  - Using different types of data

What is our disclosure Policy for clinical studies on our products?



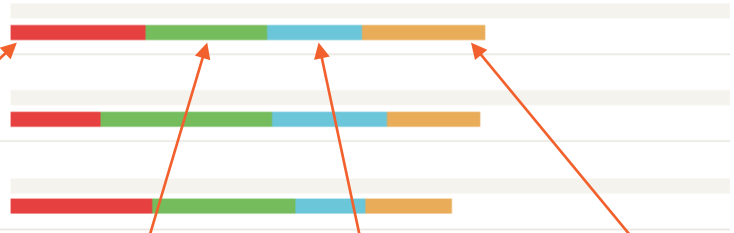
[1] From primary completion date for interventional studies.
[2] From completion of analysis for non-interventional studies.
[3] From last subject last visit for interventional studies
[4] Or earlier if required by regulation

[4] Disclosure of clinical studies that evaluate the clinical efficacy, safety, or effectiveness of a product in humans. Additional analyses and other types of human subject research, where results provide important scientific knowledge or are relevant for patient care are submitted for manuscript publication or, where manuscript publication is not practical, submitted for a citable abstract publication or a results summary is posted

All Trials Transparency Index

Company ranking

RANK	NAME	SCORE	BREAKDOWN
1	**GlaxoSmithKline**	1041	
2	**LEO Pharma**	1030	
3	**Takeda**	969	



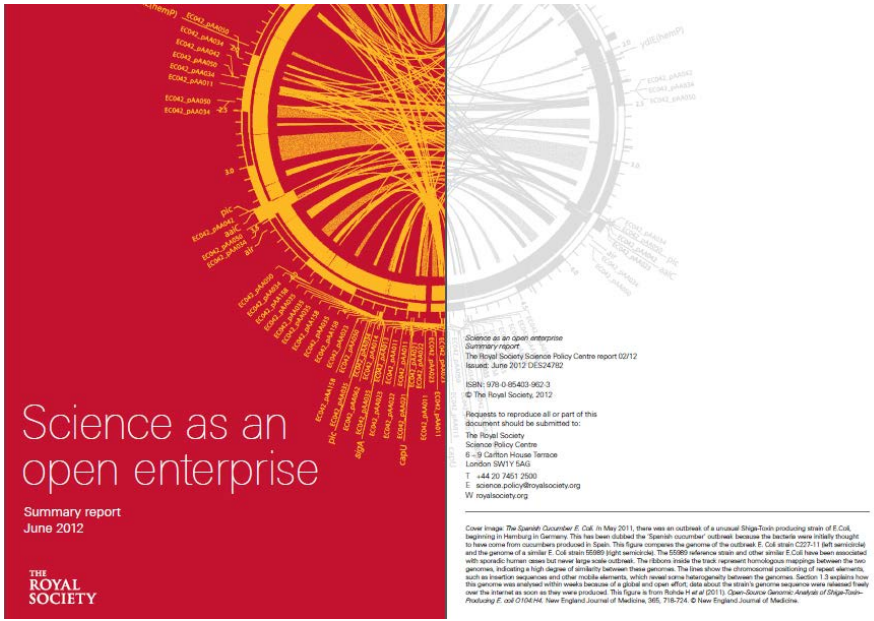
Protocol Registration (>4,000)

Summary results (>6,000)

Clinical study reports (>2,500)

Individual patient data (>2,000)

UK Royal Society 2012



Science as an open enterprise
Summary report
The Royal Society Science Policy Centre report 02/12
Issued: June 2012 DES24782

ISBN: 978-0-85403-962-3

The Royal Society
Science Policy Centre
6 – 9 Carlton House Terrace
London SW1Y 5AG
T +44 20 7451 2500
E science.policy@royalsociety.org
W royalsociety.org

Cover image: *The Spanish Cucumber E. Coli.* In May 2011, there was an outbreak of a unusual Shiga-Toxin producing strain of E.Coli, beginning in Hamburg in Germany. This has been dubbed the 'Spanish cucumber' outbreak because the bacteria were initially thought to have come from cucumbers produced in Spain. This figure compares the genome of the outbreak E. Coli strain C227-11 (left semicircle) and the genome of a similar E. Coli strain 55989 (right semicircle). The 55989 reference strain and other similar E.Coli have been associated with sporadic human cases but never large scale outbreak. The ribbons inside the track represent homologous mappings between the two genomes, indicating a high degree of similarity between these genomes. The lines show the chromosomal positioning of repeat elements, such as insertion sequences and other mobile elements, which reveal some heterogeneity between the genomes. Section 1.3 explains how this genome was analysed within weeks because of a global and open effort; data about the strain's genome sequence were released freely over the internet as soon as they were produced. This figure is from Rohde H *et al* (2011). *Open-Source Genomic Analysis of Shiga-Toxin-Producing E. coli O104:H4.* New England Journal of Medicine, 365, 718-724. © New England Journal of Medicine.

Scientists should communicate the data they collect and the models they create, to allow free and open access, and in ways that are intelligible, assessable and usable for other specialists in the same or linked fields wherever they are in the world. ”

Benefits and risks

MAIN BENEFITS

- ✓ Verify publications/ enhances reproducibility
- ✓ Ask new research questions
- ✓ Helps avoid waste
- ✓ Helps ensure the data provided by research participants are used to maximum effect in the creation of knowledge and understanding.

MAIN RISKS

- ❖ Protecting the privacy of research participants.
- ❖ Ensuring the data are used for valid scientific investigation.
- ❖ Ensuring the data are usable.

Senior management endorsement

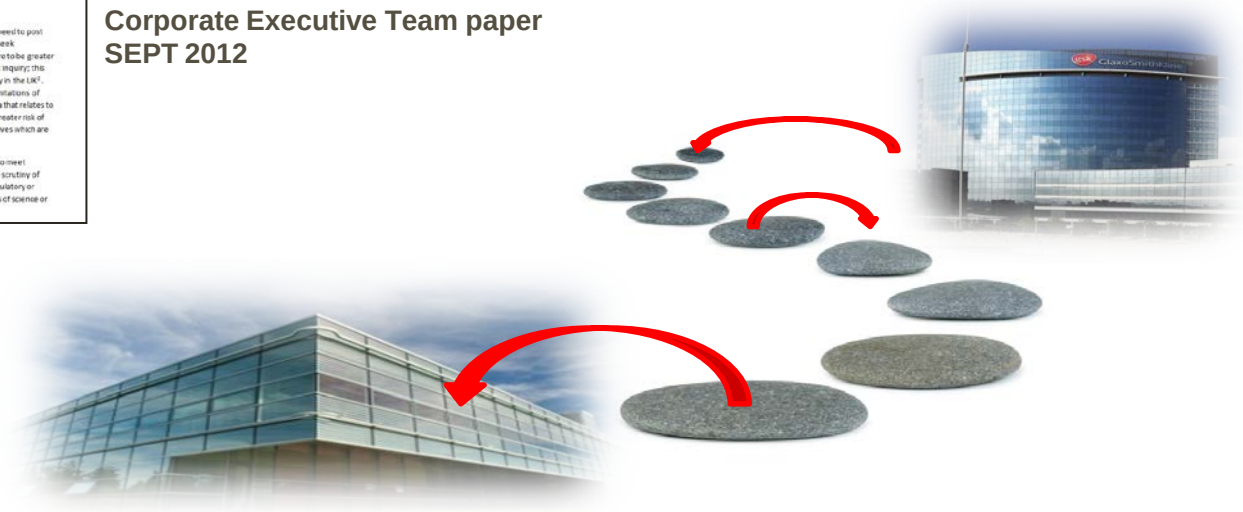
Sharing Patient Level Data from GSK Sponsored Clinical Trials

1. Background

Clinical trial disclosure requirements have evolved over the last ten years to include the need to post protocol and result summaries of our studies on internet registers and commitments to seek publication of studies in the scientific literature. However there are ongoing calls for there to be greater access to underlying (or patient level) data to verify findings and re-use data for scientific inquiry; this view has support from European regulators[1] and other groups including the Royal Society in the UK[2]. There is a good scientific rationale for providing greater access to these data given the limitations of analyses based on published results and the perceived need for greater openness of data that relates to the efficacy and safety of medicines. Nonetheless, access to these data may result in a greater risk of litigation and there are risks that third parties may access the data with nefarious objectives which are not in the best interests of science or patient care.

Finding an acceptable mechanism for sharing patient level data may become necessary to meet external expectations, build trust and provide assurance that industry data can stand the scrutiny of independent scientific review. There is also a likelihood that there may be legislative, regulatory or external policy requirements in this area and a risk these may not be in the best interests of science or the development of medicines for patients.

Corporate Executive Team paper
SEPT 2012



"It is proposed that the ultimate objective is a model where data from all research sponsors (including industry and academe) is made available for research (STEP 2), however the first step would be for GSK to adopt a leadership role to demonstrate our commitment and the feasibility of the approach (STEP 1)."

Public commitment

Press Release Oct 2012

Expanding further on its commitments to openness and transparency, GSK also announced today that the company will create a system that will enable researchers to access the detailed anonymised patient-level data that sit behind the results of clinical trials of its approved medicines and discontinued investigational medicines.

This initiative is a step towards the ultimate aim of the clinical research community developing a broader system where researchers will be able to access data from clinical trials conducted by different sponsors. GSK hopes the experience gained through this initiative will be of value in developing and catalysing this wider approach.

PharmaTimes online

Search news,

Magazine Web Exclusives News Competitions Appointments Jobs Bus

Free news subscriptions Free RSS feeds

GSK outlines plan to share patient data online

GlaxoSmithKline opens the door on clinical data sharing

Support The Guardian

Contribute → Subscribe →

Search jobs Dating Sign in Search The Guardian UK edition

News Opinion Sport Culture Lifestyle More

UK World Business Football UK politics Environment Education Society Science Tech Global development Cities Obituaries

Health

GlaxoSmithKline opens door on data

Business Report

Big Pharma Opens Up Its Big Data

BUSINESS DAY

Glaxo Opens Door to Data on Research

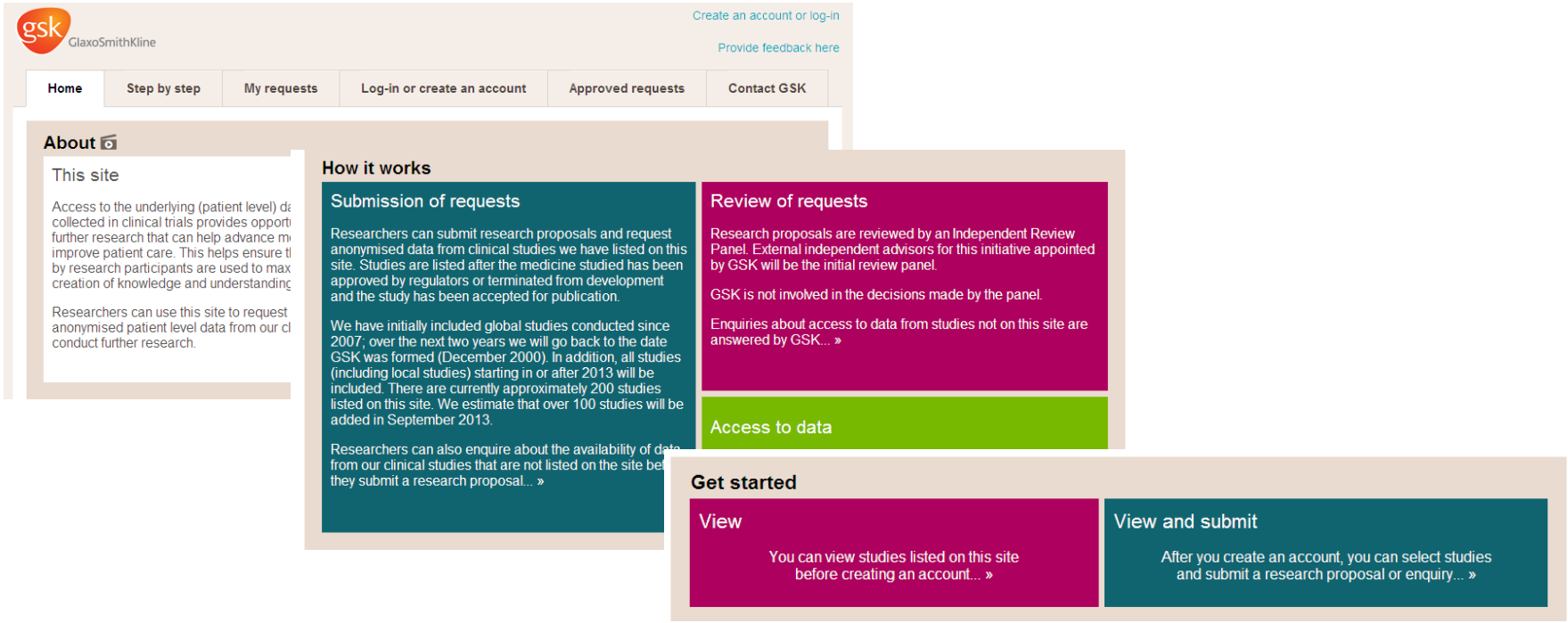
By KATIE THOMAS OCT. 11, 2012

News

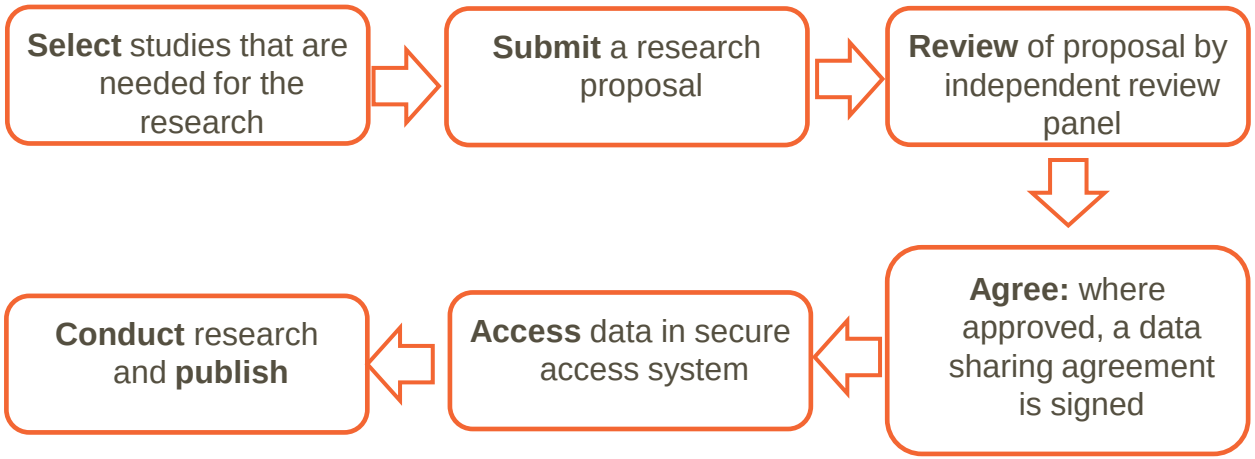
GlaxoSmithKline grants researchers access to clinical trial data

BMJ 2012 ; 345 doi: https://doi.org/10.1136/bmj.e6909 (Published 12 October 2012)
Cite this as: *BMJ* 2012;345:e6909

GSK request site – May 2013



How it works…


Select studies that are needed for the research
Submit a research proposal
Review of proposal by independent review panel
Agree: where approved, a data sharing agreement is signed
Access data in secure access system
Conduct research and **publish**

Study Sponsor: GSK

Study Title
A single-blinded randomized, placebo-controlled, staggered-parallel, escalating-dose study to investigate the safety, tolerability, pharmacokinetics and pharmacodynamics of subcutaneous injections of GSK 716155 in Subjects with type 2 Diabetes Mellitus

Medicine or Vaccine (generic name)
albiglutide

Sponsor Identification Number
GLP106073

Trial Registry Identification Number(#'s)
NCT00354536

Medical Condition
Diabetes Mellitus, Type 2

Phase
Phase 2

Link to study details on the GSK Clinical Study Register
http://www.gsk-clinicalstudyregister.com/study/GLP106073

Link to study details on other Trial Registries (if available)
http://clinicaltrials.gov/show/NCT00354536

Datasets and Documents Available for this Study

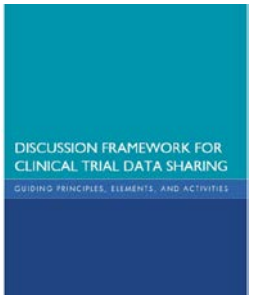
✔Analysis-ready dataset ✔Reporting and analysis plan ✔Clinical study report ✔Raw dataset
✔Annotated case report form ✔Dataset specifications ✔Protocol with any amendments

Additional information about the data and documents available for this study

Date Added to this Site
June 2014

…some significant things happened

Industry commitment July 2013



Institute of Medicine guiding principles 2015

Recommendation 1: Stakeholders in clinical trials should foster a culture in which data sharing is the expected norm, and should commit to responsible strategies aimed at maximizing the benefits, minimizing the risks, and overcoming the challenges of sharing clinical trial data for all parties.

Journal editors' proposal April 2016

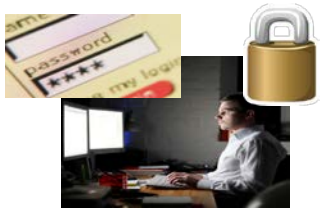

EDITORIAL

Sharing Clinical Trial Data
A Proposal From the International Committee of Medical Journal Editors

Collaboration -
Site reconfigured and owned/used by multiple sponsors

The challenge of fragmentation..

Secure access system used by Data provider 1



Secure access system used by Data provider 2

Secure access system used by Data provider 3

Combining data from different systems and using software not provided in these systems is problematic

The challenge of fragmentation..

Perspective

Data Sharing at a Crossroads

Frank Rockhold, Ph.D., Perry Nisen, M.D., Ph.D., and Andrew Freeman, B.Sc.

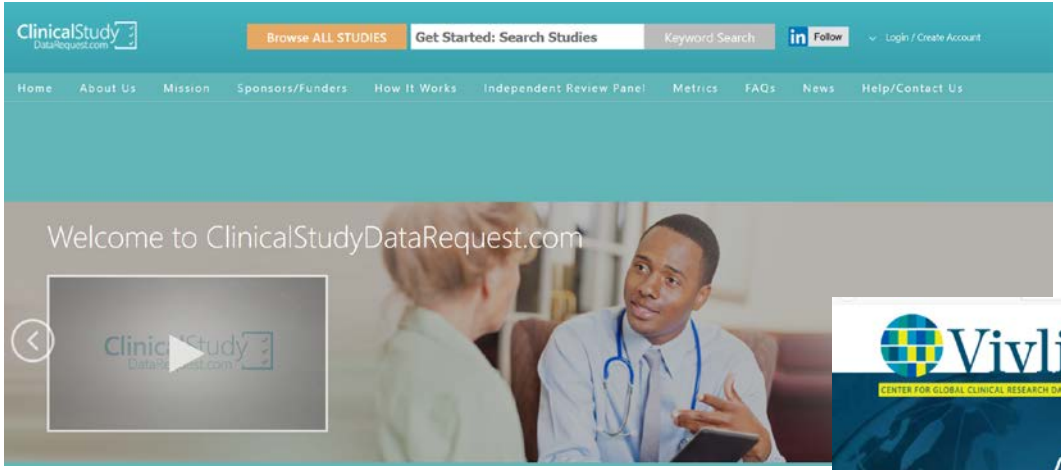
September 22, 2016
N Engl J Med 2016; 375:1115-1117
DOI: 10.1056/NEJMp1608086

of data availability and expertise in using data from clinical trials compromising independence. The Wellcome Trust agreed to accom-

Our vision of a simple single system may be challenging to achieve. But there's a risk that if myriad systems emerge, the benefits will be limited by the complexity of obtaining data.

in pooled and meta-analyses are likely to increase use. Since the modate this approach to enable more sponsors to share data

This is not a theoretical issue…

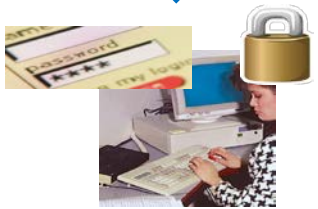


An alternative model – the researcher provides the secure access system

Secure download by Data Provider 1

Secure download by Data Provider 2

Secure download by Data Provider 3



- Contractual IT and data governance requirements
- Subject to audit
- Accreditation of researchers?

Has our data sharing enabled more science?

Has our data sharing enabled more science?

Number of Research Proposals submitted up to **30 September 2019**		543
Requirements check	In process	32
	Withdrawn by the requestor	49
	No response received	24
	Did not meet requirements (further details)	23
	Potential conflict of interest or an actual or potential competitive risk	0
	Met requirements	415

Has our data sharing enabled more science?

IRP review	In process	6
	Withdrawn by the requestor	4
	No response received	2
	Rejected or advised to re-submit	60
	Approved or approved with conditions	343

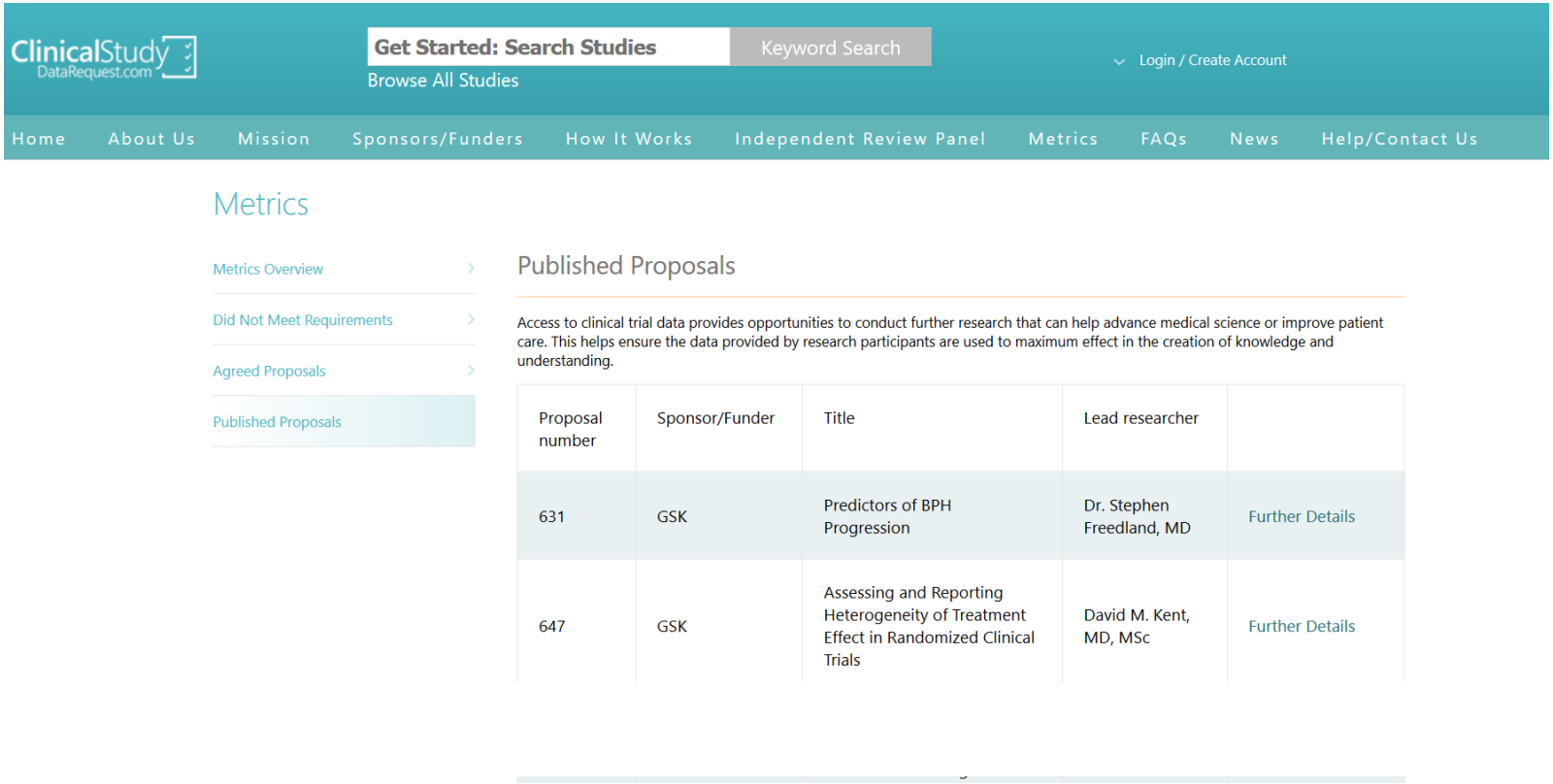
Has our data sharing enabled more science?

Data Sharing Agreement	In process	36
	Withdrawn by the requestor	18
	No response received	10
	Not agreed by requestor	1
	Agreed (signed) View details of these Research Proposals	277

Has our data sharing enabled more science?

Data Access	Access provided	142
	Withdrawn by the requestor	2
	Access closed	122
Publication	Publication anticipated	35
	Published View details of these Research Proposals	53
	Withdrawn by requestor	13
	No response received	20
	Publication not received after 18 months of Access Closed	9

Examples of published research



What has been the impact?

Data Sharing — Is the Juice Worth the Squeeze?

Brian L. Strom, M.D., M.P.H., Marc E. Buyse, Sc.D., John Hughes, B.Sc., and Bartha M. Knoppers, Ph.D.

Making trial data broadly available is ethically imperative and scientifically justified and has the potential to increase public understanding of and support for clinical research. But it seems critical to find ways to improve the use and output of data-sharing projects before the clinical research community invests the substantial effort and resources required to broaden the effort to include academic and other noncommercial investigators.

October 27, 2016
N Engl J Med 2016; 375:1608-1609
DOI: 10.1056/NEJMp161033

Awareness and expertise may limit use…

FIRST OPINION

Sharing clinical trial data: lessons from the YODA Project

By JOSEPH S. ROSS, JOANNE WALDSTREICHER, *and* HARLAN M. KRUMHOLZ / NOVEMBER 18, 2019

But the largest challenges to achieving the promise of data sharing are broadening awareness of data availability and fostering expertise in using data from clinical trials. The skills needed for analyzing clinical trial data, particularly for meta-analysis, are considerable. The primary reason that requests for data that have been approved by the YODA Project do not result in a completed project is external investigators' realization that their expertise was insufficient to perform the proposed analyses.

https://www.statnews.com/2019/11/18/data-sharing-clinical-trials-lessons-yoda-project/?utm_source=STAT+Newsletters&utm_campaign=ff05bd5e58-Pharmalot&utm_medium=email&utm_term=0_8cab1d7961-ff05bd5e58-135106209

Can we do more to enable use of different types of data…?

Sex and BMI Alter the Benefits and Risks of Sulfonylureas and Thiazolidinediones in Type 2 Diabetes: A Framework for Evaluating Stratification Using Routine Clinical and Individual Trial Data

John M. Dennis[1], William E. Henley[1], Michael N. Weedon[2], Mike Lonergan[3], Lauren R. Rodgers[1], Angus G. Jones[4,5], William T. Hamilton[2], Naveed Sattar[6], Salim Janmohamed[7], Rury R. Holman[8,9], Ewan R. Pearson[3], Beverley M. Shields[4] **and** Andrew T. Hattersley[4,5], **on behalf of the MASTERMIND Consortium***

Corresponding authors: Beverley M. Shields, b.shields@exeter.ac.uk, and Andrew T. Hattersley, a.t.hattersley@exeter.ac.uk.

Author Affiliations

Diabetes Care 2018 Sep; 41(9): 1844-1853.
https://doi.org/10.2337/dc18-0344

CONCLUSIONS Patient subgroups defined by sex and BMI have different patterns of benefits and risks on thiazolidinedione and sulfonylurea therapy. Subgroup-specific estimates can inform discussion about the choice of therapy after metformin for an individual patient. Our approach using routine and shared trial data provides a framework for future stratification research in type 2 diabetes.

Can we do more to enable the use of AI and machine learning…?

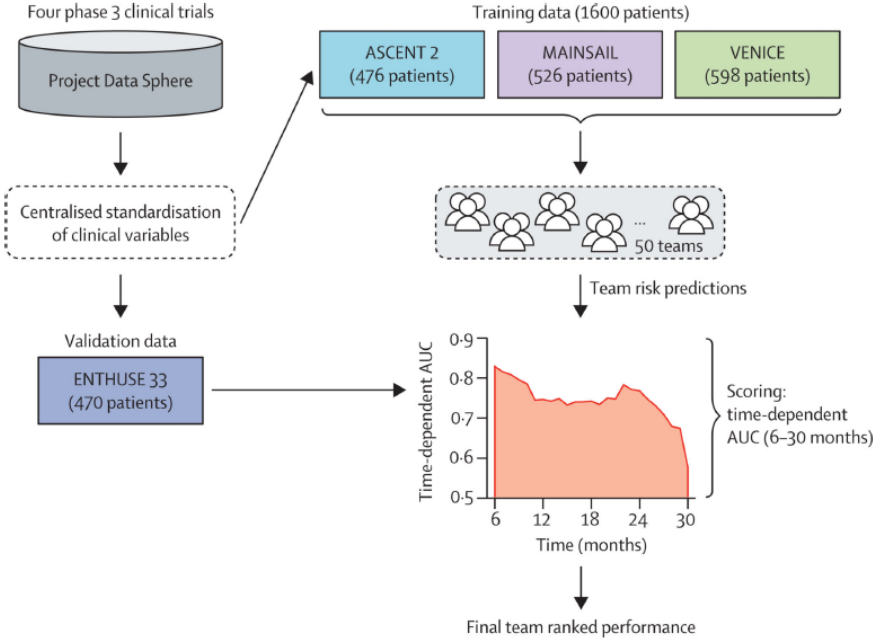
Predicting corticosteroid-free endoscopic remission with vedolizumab in ulcerative colitis

A. K. Waljee[1,2] | B. Liu[3] | K. Sauder[2] | J. Zhu[3] | S. M. Govani[2] | R. W. Stidham[2] | P. D. R. Higgins[2]

Aliment Pharmacol Ther. 2018;47:763–772.

Conclusions: A machine learning algorithm using laboratory data through week 6 of vedolizumab therapy was able to accurately identify which UC patients would achieve corticosteroid-free endoscopic remission on vedolizumab at week 52. Application of this algorithm could have significant implications for clinical decisions on whom to continue on this costly medication when the benefits of the vedolizumab are not clinically apparent in the first 6 weeks of therapy.

Can we incentivise data re-use with Challenges?
Prostate Cancer Challenge



Study design

Data were acquired from Project Data Sphere and curated centrally by the organising team to provide a harmonised dataset across the four studies. Three studies were provided as training data (ASCENT2, MAINSAIL, and VENICE) and the fourth (ENTHUSE 33) was the validation dataset. Teams submitted risk scores for ENTHUSE 33, then their predictions were scored and ranked using an integrated time-dependent area under the curve (AUC) metric.

Guinney J et al Lancet Oncol. 2017 18(1) 132-142

Summary

- We have been sharing anonymised patient level data from our clinical trials since 2013
- These data are used in further research
- The use of centralised secure data repositories could limit use of these data
- Lack of awareness/capabilities may limit use
- Can we do more to enable use of these data using machine learning or with other types of data?
- Are data challenges a good way to encourage use?