



Preparing individual patient data from clinical trials for sharing: the GlaxoSmithKline approach

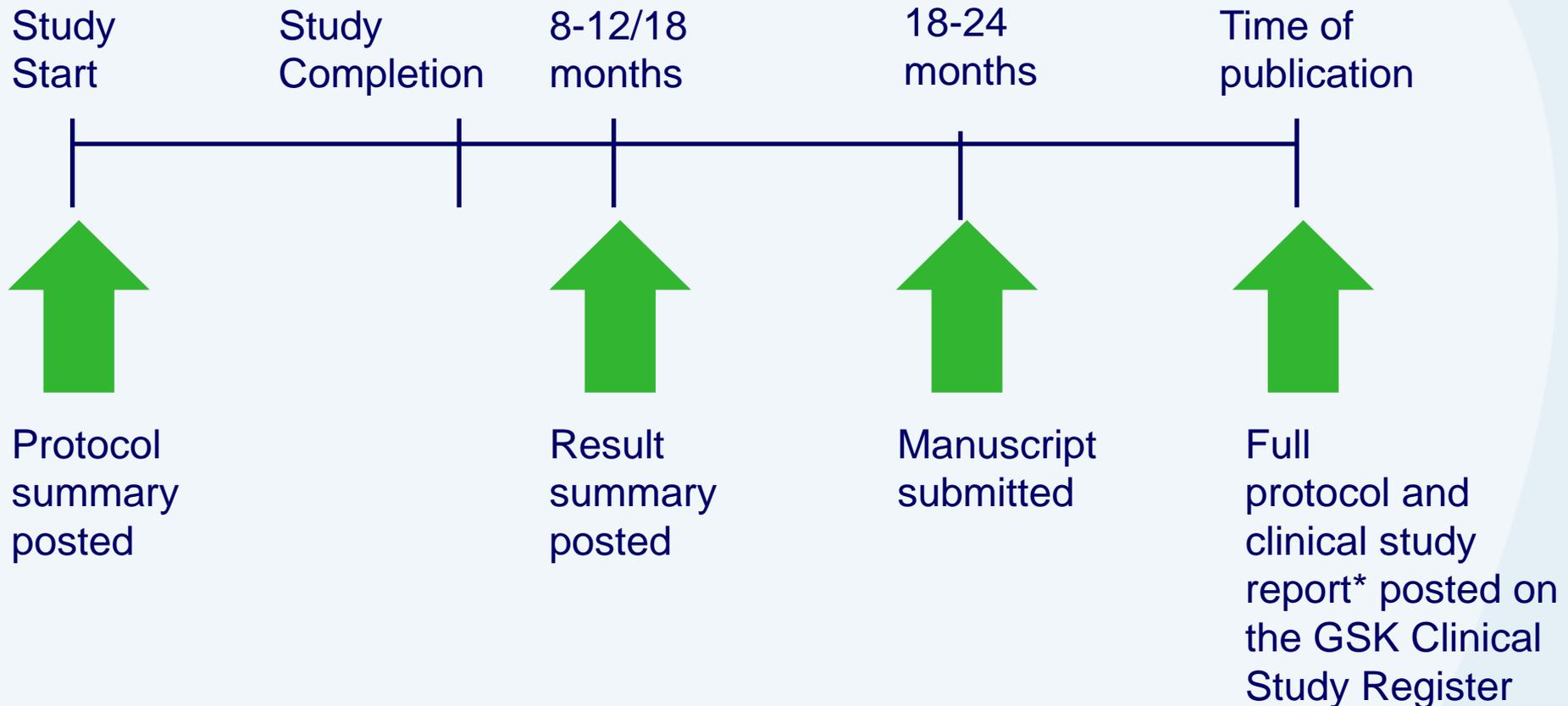
Presenter: Sara Hughes, VP & Head Clinical Statistics, GSK

Co-authors: Karen Wells, Paul McSorley, Andrew Freeman, GSK

Pharmaceutical Statistics 2014; 13: 179-183.

GSK publicly discloses clinical research information

All human subject research studies that evaluate investigational or approved medicinal products – (phase I-IV, meta-analyses, observational studies)



Result summaries and publications have limitations

Publicly disclosed results:

- Summarise data from the study population with statistics to compare treatment groups
- Do not include the primary data from each research participant

Primary Efficacy Results: Total Population			
Day 1 To Day 5	Dose 1	Dose 2	Dose 3
Treatment Response, n (%)			
Complete (0 Episodes)	7 (19)	8 (22)	10 (31)
Major (1-2 Episodes)	10 (28)	14 (39)	10 (31)
Minor (3-5 Episodes)	0	1 (3)	0
Failure (>5 Episodes/Rescued)	19 (53)	13 (36)	12 (38)
p-value (stratified for centre)			
Dose 2 vs Dose 1	0.848		
Dose 2 vs Dose 3	0.467		

Benefits of greater access to patient level data

- Enables the identification of trends and associations that may provide greater insight or help develop hypotheses and theories for further research
- Enables the review of results from clinical trials to validate the results
- Helps ensure the data provided by research participants are used to maximum effect in the creation of knowledge and understanding
- Strengthens trust in clinical research through enhanced openness and transparency

Main issues

- Protecting the privacy and confidentiality of research participants
- Ensuring the data are used for valid scientific investigation

A solution

Research sponsors



Anonymised patient level data provided after completion of the project and publication

Independent Data Custodian



Undertakes scientific review of proposals

Reviews expertise and management of any conflicts of interest

Manages privacy

Researchers



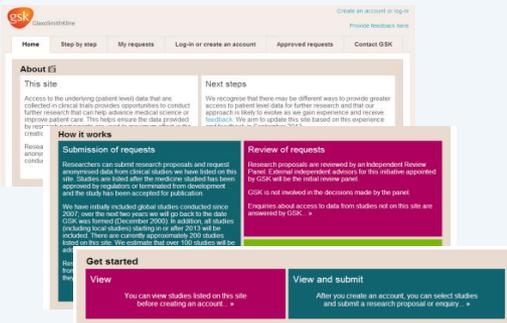
Submits scientific proposals and analysis plans

Agreements to protect privacy, and publish the results

GSK took a first step in May 2013

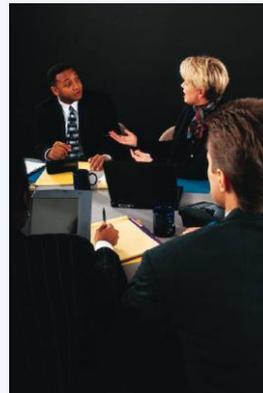
In May 2013 GSK launched a system with three components to provide greater access to patient level data from our clinical trials

1



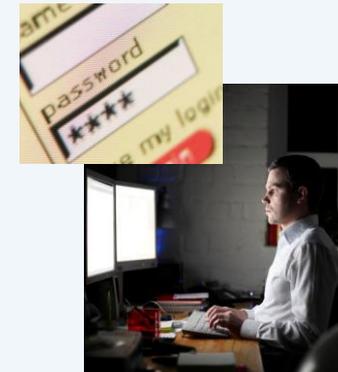
GSK Request Site

2



Independent Review Panel

3



GSK Access system Panel

A first step

When GSK launched this system the aim was to help realise a broader solution with an Independent Data Custodian to allow access to data from clinical trials conducted by multiple companies and organisations.

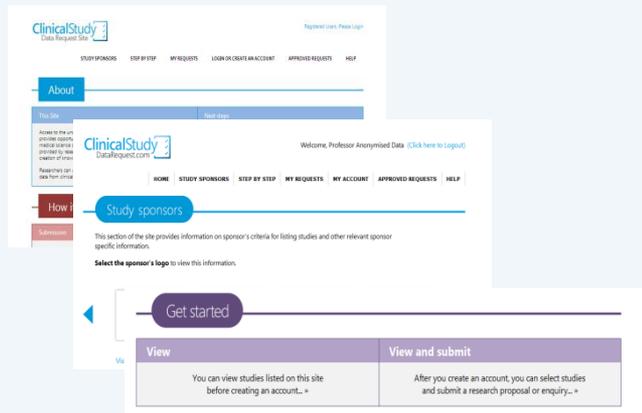


January 2014 – An important step

January 2014:

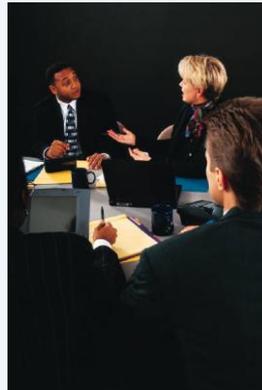
- A new multi-sponsor request site clinicalstudydatarequest.com made available
- A multi-sponsor access system under development

1



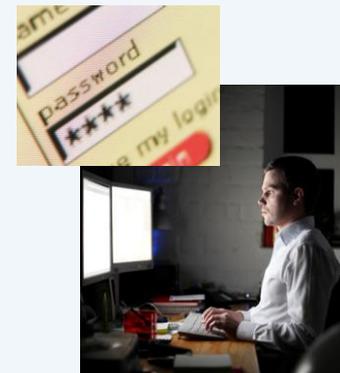
Multi-sponsor Request Site

2



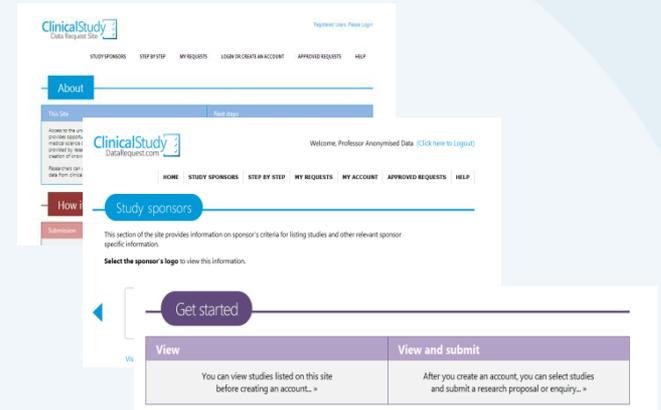
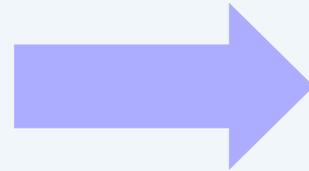
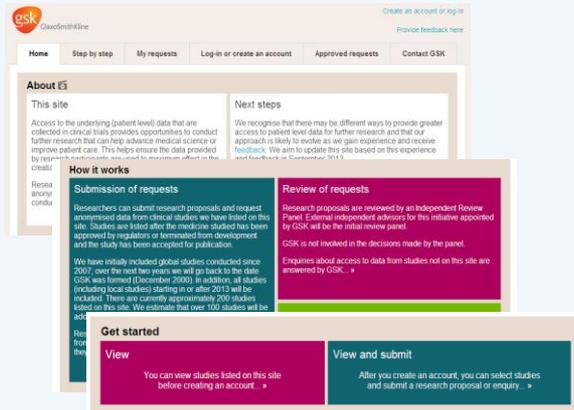
Independent Review Panel (same panel)

3



Multi-sponsor Access System (available Q1,2014)

GSK has transferred studies to the new site



- Our commitment to providing access to these data is unchanged.
- Over 900 GSK studies are now listed on the site.

We will regularly update this list to include global studies going back to the formation of GSK in December 2000 and all our studies (including local studies) started from 2013 after the medicine studied has been approved by regulators or terminated from development and the study has been accepted for publication.

Next steps

We hope that:

1. Other industry and academic study sponsors will join
2. An independent body will come on board to administer the initiative and manage the independent review of research proposals.

But how to anonymise the data?

- Minimise breaches to patient confidentiality
- Adhere to privacy laws and regulatory guidance
- Maximise scientific usefulness of data

Consider published proposed approaches to clinical trial data anonymisation:

- One publication relates to the specific situation where data will be published and therefore freely available [Hrynaszkiewicz et al, 2010]
- Another proposed by an academic research organisation in support of their focus to share knowledge and data when possible [Shostak et al, 2006]

High level summary of approach

- For both SDTM and ADaM (or equivalent) datasets:
 1. Remove Personally Identifiable Information (PII) → de-identification
 2. Destroy link between old and new datasets → anonymisation; store separately
- SAS macros automate some of process
- Additional programming & QC required for each study
- Manual review of de-identified datasets by independent reviewer to verify correctness of anonymisation

De-identification approach

- USA federal privacy law (HIPAA): methods for de-identification of health information – requires removal of 18 specific identifiers
- GSK approach follows HIPAA methodology (where applicable)
- Any other PII that may be present is also removed

De-identification in detail – step 1

1. Recoding identifiers (or code numbers)

- A new subject identifier (or code number) for each research participant
- A new investigator identifier (or code number) for each investigator. The investigator name is set to blank.
- A new laboratory identifier for each laboratory
- A new centre identifier for each centre

De-identification in detail – step 2

2. **Removing free text verbatim terms.** Free text verbatim terms are set to 'blank', including the following:

- Adverse events
- Medications
- Other, for example, medical history
- Other specific verbatim free text

De-identification in detail – steps 3 & 4

3. Replacing date of birth with year of birth

- with the exception of ages older than 89 years, which are aggregated into a single category of '90 or older'

4. Replacing all original dates relating to a research participant

- A random offset is generated for each research participant and applied to all dates for that research participant
- All original dates are replaced with the new dummy dates so that relative times of all events/ observations/interventions etc for each research participant are retained

De-identification in detail – step 5

5. Reviewing and removing other PII. Eg,

- Any names and initials (e.g. any investigator names and any subject names and initials)
- Information from variable names (e.g. laboratory names may contain location information)
- Other geographic information such as place of work (e.g. if socioeconomic data is collected)
- Investigator comments which may identify a subject.
- Genetic data that would enable a direct trace back to an individual subject
- Kit numbers and device numbers (e.g. container numbers and lab sample numbers)

Summary

- A multi-sponsor system to provide greater access to patient level data is an important step on the road to a fully independent system where data from multiple companies and organisations are available for research.
- GSK has transferred studies to this new system. Our commitment remains unchanged. There are currently over 450 GSK studies on clinicalstudydatarequest.com.
- We hope that other industry and academic study sponsors will join and that an independent body will come on board to administer the initiative and manage the independent review of research proposals.
- **A standardised approach to data anonymisation is another important step in order to simplify pooling of data across sponsors**