TITLE: A Method for Optimizing Clinical Data Standardization

INVENTOR: Kalyan Gopalakrishnan

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application is related to, and claims priority from, U.S. Provisional Patent application no. 60/982,737 filed on October 26, 2007, by Kalyan Gopalakrishnan entitled "Process for Optimization for Clinical Data Standardization", the contents of which are hereby incorporated by reference.

FIELD OF THE INVENTION

[0002] The present invention relates to methods for gathering and storing data, and more particularly, to methods for transforming data gathered in clinical trials into regulatory compliant formats.

BACKGROUND OF THE INVENTION

[0003] In the US, and most other industrial countries, pharmaceuticals are strictly regulated by government agencies such as the US Food and Drug Administration (FDA). Part of this regulation requires that pharmaceutical and medical companies conduct extensive clinical research in order to demonstrate the safety and efficacy of new drugs and devices before they can be approved for commercial use.

[0004] These clinical trials are typically costly and time consuming and usually involve significant amounts of data. In order to speed up, and reduce the

cost of, clinical trials and their evaluation, the FDA is attempting to implement standardization of the data that is electronically submitted to them as regulatory submissions for evaluation.

[0005] In particular, they are attempting to standardize on the Clinical Data Interchange Standards Consortium (CDISC)'s Study Data Tabulation Model (SDTM).

[0006] A problem facing the pharmaceutical and medical industry is that currently clinical, non-clinical, lab and healthcare data exists in multiple physical formats, with datasets typically having heterogeneous syntax and semantics. The data may also be located across multiple geographical locations and organizations, and may exist at different stage of the clinical research life cycle. All of these factors make it difficult to assemble and aggregate such data for regulatory submissions.

SUMMARY OF THE INVENTION

[0007] Briefly described, the invention provides a method of integrating clinical trial data into a form required by a standards based data format.

[0008] In a preferred embodiment, the method includes using a case report form (CRF) designed for a clinical trial. Typically such a CRF is designed to optimize data collection in a particular clinical trial, including not collecting unnecessary data while collecting all necessary data. The CRF is also typically designed to make the data collection as easy and accurate as possible for personal

administering the clinical trial and may, therefore, include syntax and semantics that are not directly compliant with the standards requirements.

[0009] The trial CRF may be annotated with one or more annotations that map report form data to standards compliant data. For instance, a CRF in a particular clinical trial may have as one of the options for the outcome "Death". The standard may, however, stipulate that such an outcome be recorded as "Fatal". The annotation, in such a case, may be a metadata entry that automatically transforms any occurrence of that variable as "death" to "fatal".

[0010] The annotations may then be stored in, for instance, a metadata repository or other storage module. These stored annotations may then be applied to data that was captured in a clinical trial data using that particular case report form to produce standards compliant data for that trial.

[0011] One exemplary data standard that may be mandated by the US Food and Drug Administration for electronic data submission is the Clinical Data Interchange Standards Consortium (CDISC)'s Study Data Tabulation Model (SDTM).

[0012] In a further preferred embodiment, the data sets produced are validated as being standards compliant.

[0013] These and other features of the invention will be more fully understood by references to the following drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0014] FIG. 1 is a schematic flow diagram of selected steps in an exemplary clinical trial.

[0015] FIG. 2 is a schematic view of an exemplary embodiment of the present invention of integrating clinical trial data into a format suitable for submission to a regulatory agency.

[0016] FIG. 3 is a schematic view of a further exemplary embodiment of the present invention of integrating clinical trial data into a format suitable for submission to a regulatory agency.

DETAILED DESCRIPTION

[0017] A preferred embodiment of the invention will now be described in detail by reference to the accompanying drawings in which, as far as possible, like elements are designated by like numbers.

[0018] Although every reasonable attempt is made in the accompanying drawings to represent the various elements of the embodiments in relative scale, it is not always possible to do so with the limitations of two-dimensional paper. Accordingly, in order to properly represent the relationships of various features among each other in the depicted embodiments and to properly demonstrate the invention in a reasonably simplified fashion, it is necessary at times to deviate from absolute scale in the attached drawings. However, one of ordinary skill in the art would fully appreciate and acknowledge any such scale deviations as not limiting the enablement of the disclosed embodiments.

[0019] FIG. 1 is a schematic flow diagram of selected steps in an exemplary clinical trial 10. In a typical first step 12, a clinical trial is designed. This may, for instance, be initiated by a sponsor or investigator that has identified a medication or device to be tested, typically termed the new agent, in the clinical trial. Usually, one or more pilot experiments are conducted prior to the clinical trial 10 to gain insights for the design of the clinical trial 10. The sponsor typically first decides what to compare the new agent to. They may, for instance, decide to compare the new agent to one or more existing treatments or a placebo or some combination thereof. Decisions are also made as to what kind of patients might benefit from the medication/device that is the new agent. These decisions are typically made in coordination with a panel of expert investigators such as, but not limited to, physicians well-known for their publications and clinical experience.

[0020] Once the clinical trial protocol has been designed, a Case Report Form (CRF) is designed in step 14. The CRF may be used in the clinical trial to collect the specific data needed in order to test the efficacy of the new agent in the clinical trial. The CRF may be a paper or electronic research questionnaire, or some combination thereof. A well designed CRF may capture all the data necessary for the purpose of the clinical trial without additional data. Clinical trials are expensive, so collecting all the necessary data that may enable a successful clinical trial is typically a goal when designing a CRF. Collecting unnecessary data, or data that may not be analyzed in assessing the clinical trial, may slow down the clinical trial by unnecessary data processing. CRF's may,

therefore, be a document that is unique to a specific clinical trial, and may, for instance, contain terminology specific to a particular area of medicine or particular treatment. A CRF may range in size and complexity from a handwritten one-time 'snapshot' of a patient's physical condition to hundreds of pages of electronically captured data obtained over a period of weeks or months.

[0021] Once a CRF has been successfully designed and produced, a typical next step in a case study is step 16 of recruiting clinical subjects. If a sponsor cannot obtain enough patients with the specific disease or condition relevant to the case study at one location, then investigators at other locations who can obtain the same kind of patients to receive the treatment may be recruited into the study.

[0022] Having recruited suitable clinical subjects, a clinical trial typically then takes step 18 of performing the study. This typically takes the form of administering the treatment(s), i.e., providing the new agent to the clinical subjects in the manner laid out in the clinical trial protocol.

[0023] In step 20, the data is collected typically using the CRF. These data may, for instance, include measurements like vital signs, amount of study drug in the blood, and whether the patient's health gets better or not.

[0024] In step 22 the data is analyzed. This is typically done by the sponsor of the clinical trial. Once the data has been analyzed and organized, it may be submitted to the appropriate regulatory authority in step 24.

[0025] Regulatory authorities such as, but not limited to, the US Food and Drug Administration (FDA) are the bodies whose responsibilities typically

include protecting the public health by assuring the safety and efficacy of human and veterinary drugs and medical devices. The regulatory authorities typically access the data from the clinical trials and may use them to make decisions on whether or not to allow the drug to be used by the medical community and the public and may also decide on conditions of that use. In order to make these decisions more efficiently, regulatory authorities are typically requiring that the data submitted to them in step 24 is in a prescribed, standardized data format.

[0026] The US FDA has, for instance, begun mandating that clinical trial data submitted to them conform to the Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model (SDTM). The CDISC is a non-profit organization, whose mission is to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare. Their CDISC SDTM is written in XML Schema.

[0027] SDTM is built around the concept of observations collected about subjects who participated in a clinical study. Each observation can be described by a series of variables, corresponding to a row in a dataset or table. Each variable can be classified according to its Role. A Role determines the type of information conveyed by the variable about each distinct observation and how it can be used. Variables may be classified into five major roles:

1. Identifier variables, which identify, for instance, the study, subject of the observation, the domain, and the sequence number of the record.

2. Topic variables, which specify the focus of the observation, such as, but not limited to, the name of a lab test.

3. Timing variables, which describe the timing of the observation, such as, but not limited to, start date and end date.

4. Qualifier variables, which include additional illustrative text, or numeric values that describe the results or additional traits of the observation , such as, but not limited to, units or descriptive adjectives.

5. A Rule that may express an algorithm or executable method to, for instance, define start, end, or looping conditions in the Trial Design model.

[0028] FIG. 2 is a schematic view of an exemplary method of the present invention of integrating clinical trial data into a format suitable for submission to a regulatory agency 31. Integrating clinical trial data into a format suitable for submission to a regulatory agency 31 may be accomplished using a CRF optimized for a specific clinical trial 30. The CRF optimized for a specific clinical trial 30 may be a paper form or an electronic form, or some combination thereof.

[0029] The CRF optimized for a specific clinical trial 30 may be converted into a CRF annotated to conform to a standard 34 by a process of annotation 32. The standard that the CRF annotated to conform to a standard 34 conforms to may, for instance, be, but is not limited to, the CDISC SDTM standard model supported by the FDA's electronic submission process. In the annotation 32 process, both the format and terminology of the CRF optimized for a specific clinical trial 30 may be conformed to the CDISC SDTM standard

model. For instance, the term "DIED" may be used to describe a trial outcome where the accepted CDISC SDTM standard model terminology is "FATAL".

[0030] The annotation 32 may, for instance, be accomplished using an electronic or e-CRF version of the CRF optimized for a specific clinical trial 30. The annotation 32 may be done by one or more skilled personal, or by one or more skilled personal augmented by suitable software, including, but not limited to, third party drag and drop tools and a resident standards metadata library that may be used for validation of the annotations. The software may, for instance, run on a general purpose server.

[0031] The annotations used to accomplish the annotation process may be stored in an annotation store 36. The annotation store 36 may, for instance, be a data storage device that is internal to the general purpose server running the annotation support software. The data stored in the annotation store 36 may, for instance, be metadata or XML metadata related to the CRF optimized for a specific clinical trial 30.

[0032] The CRF optimized for a specific clinical trial 30 may be used in a clinical trial 38. The data collected in the clinical trial 38 may be stored in a trial data store 40. The trial data store 40 may, for instance, be any suitable data storage device.

[0033] Once there is data in the trial data store 40, it may be mapped into data suitable for submission to a regulatory agency. A mapping of trial data to submission data process 42 may be done by suitable software running on a server that has access to the annotation data stored in the annotation store 36. Once trial

data has been mapped or transformed into data suitable for submission to the regulatory agency, the data may be stored in a submission data store 44.

[0034] FIG. 3 is schematic view of a further exemplary embodiment of the present invention 51 of integrating clinical trial data into a format suitable for submission to a regulatory agency. The further exemplary embodiment of the present invention 51 includes a main database 50, a mapping module 52 and a validation module 54.

[0035] The mapping module 52 includes a eCRF annotation module 56, a metadata capture module 58, a mapping and transformation rule capture module 60, a metadata repository 62, a dataset generation module 64, a data mapping module 66 and a rules engine 68.

[0036] The validation module 54 includes a SDTM compliance module 70 and an SDTM IG compliance module 72.

[0037] The main database 50 may store raw data from current or historical clinical trials. The mapping module 52 may include one or more software programs running on one or more general purpose computers may be designed to take the raw data from the clinical data and convert the syntax and or semantics associated with the raw data into a form that is in compliance with a standard form such as, but not limited to, the CDISC STDM standard or the ADaM standard. The eCRF annotation module 56 facilitates and captures annotations to the CRF used to gather data in the clinical trial dataset being analyzed. This may involve computerized augmentation of an operator's interaction with an electronic version of the original CRF used in the clinical

trial. The metadata capture module 58 may, for instance, capture the metadata associated with such annotations. The mapping and transformation rule capture module 60 facilitates and captures mapping and transformation rules for converting the data set captured in the clinical trial into a standards compliant format, syntax and semantics.

[0038] The captured data may be validated in real-time using, for instance, a metadata library that may, for instance, be stored in the metadata repository 62.

[0039] The metadata repository 62 may be a general purpose, electronic data storage device that may be used to store the mapping and transformation rules captured by the mapping and transformation rule capture module 60, as well as the metadata captured by the metadata capture module 58 and the annotations captured by the eCRF annotation module 56.

[0040] The data mapping module 66 uses the mapping and transformation rules, metadata and annotations stored in the metadata repository 62 to map the raw trial data into standards compliant data. For data to be submitted electronically to the US FDA this standard is typically the CDISC STDM.

[0041] The dataset generation module 64 takes the data that has been source-to-target mapped by the data mapping module 66 and generates the data sets in standard complaint form.

[0042] The interaction of the various modules of the mapping module 52 may be managed and controlled by a rules engine 68.

[0043] The validation module 54 typically validates the generated data for compliance with the required standards. For the CDISC STDM, this may include validation against the standards model by a SDTM compliance module 70 as well as against the standards Implementation Guidelines (IG) by a SDTM IG compliance module 72.

[0044] The output standardized and validated data may become interim data for analysis 76 or final submission and report data 74. The data may also be feed back by a data feed back path 78 to the mapping module 52. The feed back data may, for instance, be used to further refine the annotations, metadata and mapping rules.

[0045] Although the invention has been described in language specific to structural features and/or methodological acts, it is to be understood that the invention defined in the appended claims is not necessarily limited to the specific features or acts described. Rather, the specific features and acts are disclosed as exemplary forms of implementing the claimed invention. Modifications may readily be devised by those ordinarily skilled in the art without departing from the spirit or scope of the present invention.

What is claimed is:

1. A method of integrating clinical trial data into a form required by a standard, said method comprising:

providing a case report form designed for a clinical trial;

annotating said case report form with one or more annotations comprising a transformation rule that maps report form data to standards compliant data;

storing said annotation; and

applying said annotation to trial data captured using said case report form to produce standards compliant.

2. The method of claim 1 wherein said report form data further comprises metadata.

3. The method of claim 2 wherein said standards compliant data and metadata comprises an XML data format.

4. The method of claim 2 wherein said standards compliant data and metadata comprises a Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model (SDTM).

5. The method of claim 1 wherein said standard produces data suitable for submission to a regulatory agency.

6. The method of claim 5 wherein said regulatory agency is the US Food and Drug Administration (FDA).

7. The method of claim 1 further comprising the step of verifying said annotations using a standards model.

8. The method of claim 7 wherein said standards model is a Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model (SDTM).

9. The method of claim 1 further comprising the step of verifying said annotations using a standards interpretation guide.

10. The method of claim 6 wherein said standards interpretation guide is a Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model Interpretation Guide (SDTMIG).

11. A computer-readable medium, comprising instructions for:

augmenting annotation of a case report form designed for a clinical trial with one or more annotations comprising a transformation rule that maps report form data to standards compliant data;

storing said annotation; and

applying said annotation to trial data captured using said case report form to produce standards compliant.

12. A computing device comprising: a computer-readable medium comprising instructions for:

augmenting annotation of a case report form designed for a clinical trial with one or more annotations comprising a transformation rule that maps report form data to standards compliant data;

storing said annotation; and

applying said annotation to trial data captured using said case report form to produce standards compliant.