PhUSE 2009

Paper RG09

ClinicalTrials.gov - a programmer's perspective

Ralf Minkenberg, Boehringer Ingelheim Pharma GmbH & Co. KG, Ingelheim, Germany

ABSTRACT

Since September 2008, public law requires the reporting of results from all FDA-approved trials or cleared drugs and devices on the website www.ClinicalTrials.gov. Basic results, such as baseline characteristics, primary and secondary outcomes and the corresponding statistical analyses have to be disclosed. Starting in September 2009 adverse event reporting will also be required. An expansion of results by rulemaking is foreseen in September 2010. Collecting and entering the needed data in the appropriate forms for the ClinicalTrials.gov website could be done in several ways and at several time points during the analysis and reporting of a clinical trial. Most information needed to be disclosed to the ClinicalTrials.gov such as the pre-specified tables, listings and graphs are already produced during the creation of the clinical study report. Therefore, simplifying the review and automating the upload of data to ClinicalTrials.gov during the creation of the clinical study report should reduce the additional workload for all involved tasks. In addition, any manually entered data should be avoided as much as possible during the collection and formatting of such data.

INTRODUCTION

After the "basic result" database was launched in September 2008, result reporting is legally required and could be done through a Web-based Protocol Registration System (PRS) on the website ClinicalTrials.gov. No later than one year after the estimated or actual completion date of a trial, whichever is earlier, all baseline characteristics as well as primary and secondary outcomes from all FDA-approved or cleared drugs and devices trials have to be uploaded to the ClinicalTrials.gov website.

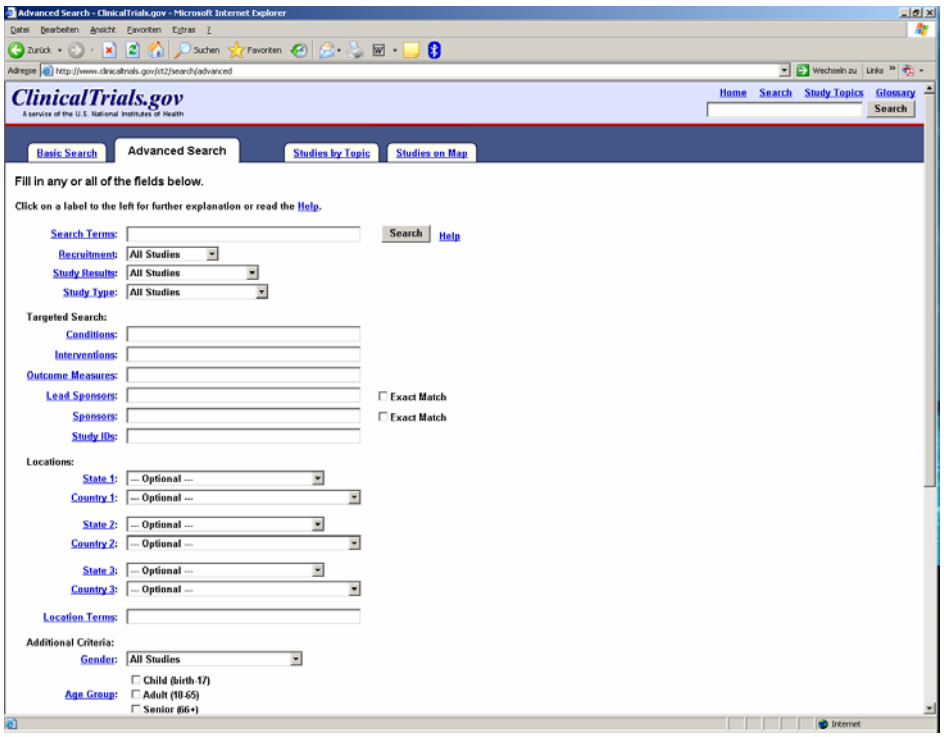

Figure 1: Advanced searching interface on ClinicalTrials.gov

All data uploaded to ClinicalTrials.gov must be in a specific format. The ClinicalTrials.gov required format differs from formats which are currently used when the clinical study report is created. Thus, we have to either enter the data manually into the ClinicalTrials.gov website or transfer our data into the ClinicalTrials.gov required format. This paper describes the process of collecting the needed information as early as they are created during the course of a clinical study. Additionally, the process should minimize data errors by reducing manually entered data or copying and pasting of data as much as possible. Since the results needed for public disclosure are available at the same time the Clinical Trial Report is final, therefore timelines will not be of an issue in this case. The responsible statistician, programmer, medical writer and physician are all involved in the process of collecting and saving the required information for the ClinicalTrials.gov website. As the medical staff mainly review and validate the final result data, the statistician and the programmer are responsible for the specification, collection and storage of the results.
The process should add minimal workload to all involved persons.

RESULTS DISCLOSURE ON CLINICALTRIALS.GOV

The website ClinicalTrials.gov launched in February 2000 to capture key summary protocol information before and during a trial. In the following years, calls for increased transparency of clinical trials were announced e.g. in Maine State Law and from the International Committee of Medical Journal Editors. With the FDA Amendment Act (FDAAA) 801 in 2007 the registry of ClinicalTrials.gov was expanded and a results database was added. The results section was enabled on September 27, 2007. Starting in December 2007 public law requires trial registration of Phase II-IV drug and device trials for all diseases. Starting September 2008 results reporting is required for all trials of FDA-approved or cleared drugs and devices. The “basic” results which have to be disclosed are participant flow, baseline characteristics (age, sex), primary and secondary outcomes (at least all outcomes which were registered), and statistical analyses. Adverse events have to be disclosed from September 2009. An “expansion” of results by rulemaking is foreseen in September 2010.

ClinicalTrials.gov organizes information for each registered study as an integrated unit, displaying the study protocol record and, if available, the corresponding results record under different tabs on the same page to facilitate cross-referencing.

If you are non-compliant, many enforcement mechanisms could be carried out. Non-compliance could be noticed publicly, and civil monetary penalties up to $10,000 per day could be declared. Withholding of NIH funds and further FDA sanctions are also possible.

In September 2009 there are 78,145 studies registered on ClinicalTrials.gov. 65,103 (83%) are interventional studies, 12,727 (16%) are observational studies. Trials done by industry are 31,643 (40%), US Federal institutions (incl. NIH) provide 18,703 (24%) studies, and 40,113 (51%) studies are provided by universities and other organizations (multiple answers are possible). Today about 300-350 trials per week are newly registered.

For 711 studies results are submitted on ClinicalTrials.gov (September 2009). 611 studies (86%) with results are from industry, 24 (3%) from US Federal institutions (incl. NIH), and 93 (13%) from universities/other organizations.

Currently there are about 40 new results from trials per week, from ClinicalTrials.gov about 150 per week are anticipated.

BASIC RESULT ITEMS ON CLINICALTRIALS.GOV

The ClinicalTrials.gov Results Database allows data providers to report summary results of registered clinical trials and observational studies. Results are reported in a standard, tabular format. The basic results part of ClinicalTrials.gov is divided into six main sections, which are described briefly in the following paragraphs.

1. Participant flow – a table is displayed including the number of patients who dropped out of the clinical trial and the number of patients excluded from the analysis, if any. Therefore, for every defined milestone during a trial the number of patients started, completed and not completed are required, and reasons for not completing the trial (or a part of the trial) must be given.

2. Baseline measures – tables of the demographic and baseline data collected overall and for each arm of the clinical trial. As default required measures age and gender are necessary. Additional user-specified measures (e.g. baseline measures used in the statistical analyses) could be added.

3. Outcome measures – tables of values for each of the primary and secondary outcome measures for each arm of the clinical trial. A measure name and description should be given for every outcome measure. Also a time frame and if the outcome measure is a safety issue are required. Additionally the reporting groups and the measured values have to be entered.

4. Statistical analysis – tables of values for each of the primary and secondary outcome measures, including the results of scientifically appropriate tests of the statistical significance of such outcome measures. In addition to the reporting groups and measured values for the statistical analysis estimate, confidence interval and p-value could be given. Additional information about the analysis, such as null hypothesis, power calculation, definition of non-inferiority margin, and other key parameters could be entered.

5. Serious adverse events – a table of anticipated and unanticipated serious adverse events grouped by organ system, with number and frequency of such event in each arm of the clinical trial.

6. Frequent adverse events – a table of anticipated and unanticipated adverse events that are not included in the serious adverse event table, that exceed a frequency of 5 percent within any arm of the clinical trial, grouped by organ system, with number and frequency of such event in each arm of the clinical trial. For both adverse event sections the

total number of affected participants and for every preferred term in every organ system the number of participants at risk, the number of events and the number of participants affected are required.

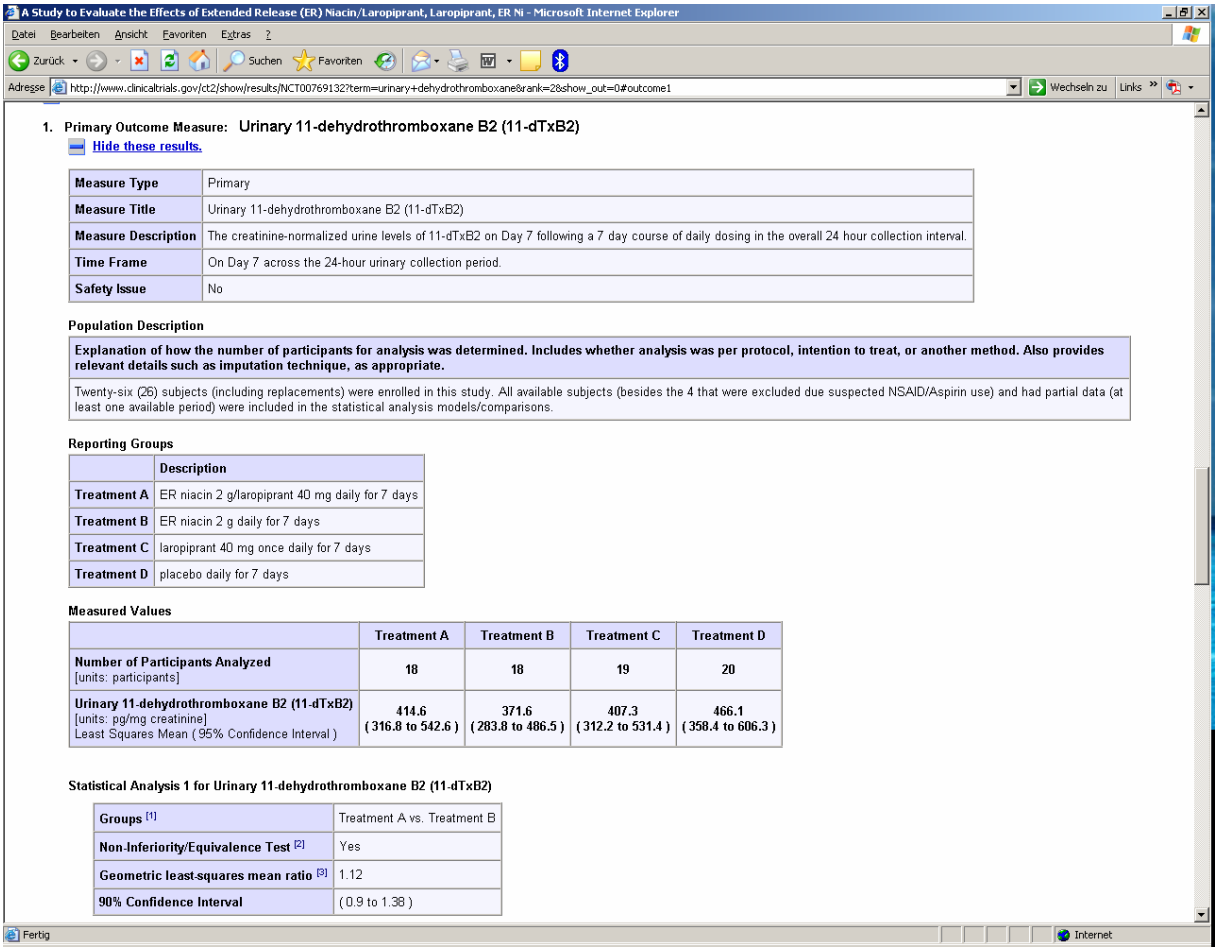

Figure 2: An example of statistical analysis results disclosed on Clinical Trials.gov

BASIC RESULT ENTRY ON CLINICALTRIALS.GOV

A Web-based Protocol Registration System (PRS) on the website ClinicalTrials.gov enables the user to enter results directly via an form based interface. Entries are checked and incorrect values will be highlighted. This kind of manual data entry is only feasible for a very small number of trials with a reasonable number of necessary entries.

As the data on ClincialTrials.gov are saved in a xml format, it is possible to upload the complete result file to ClincialTrials.gov instead of entering data manually. The contents of such a xml file is defined by a “basic result” data element definition (draft version), which defines and explains every possible entry of the result database. A xml schema for the results portion of clinical trial data to be uploaded into the Protocol Registration System (PRS) is available. In this schema generally the tag names and comments within this document relate to fields on the data entry screens and those have links into the data element definitions document for full definitions. This schema could be changed without notice from the operators whenever it is needed.

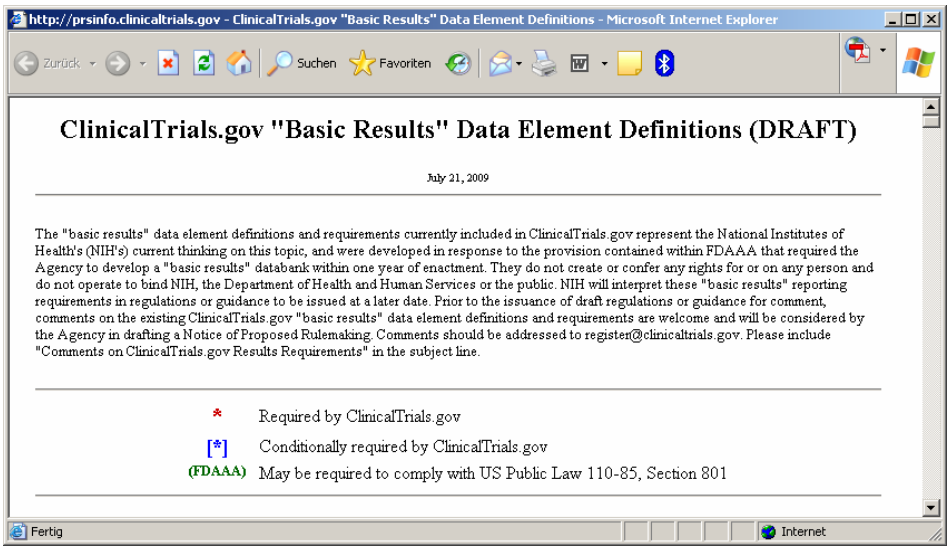

Figure 3: Top of the “Basic Results” Data Element Definitions website

COLLECTING AND UPLOADING THE RESULTS

The collection and upload of the information needed for clinical trial disclosure on ClinicalTrials.gov could be done in different ways. The needed information for results disclosure should entirely exist in the Clinical Trial Report. But extraction of the necessary information from the report requires a great amount of manual work, which is time consuming and vulnerable to entry errors. Therefore, additional validation and quality checks have to be done.

CURRENT PROCESS

Nevertheless the current, first temporary process used at Boehringer Ingelheim to collect data for ClinicalTrials.gov can’t avoid these manual work. After identifying the trials, for which results have to be published, the trial team will be contacted and an excel based template will be delivered. The trial team will have to fill this excel template until a pre-specified deadline. After a first formal quality check the information from the excel template will be entered into the preview web form from ClinicalTrials.gov. At last this file will be reviewed and uploaded to ClinicalTrials.gov.

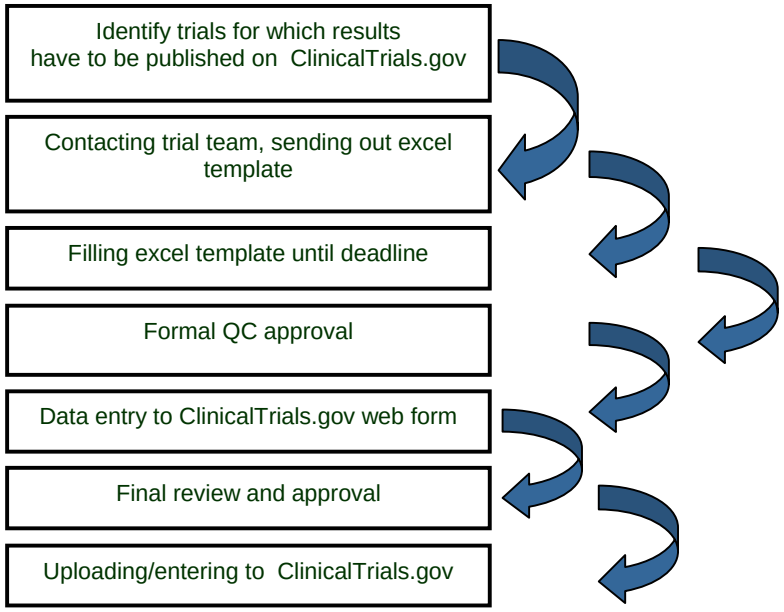

Figure 4: Current process description for collecting and entering data to ClinicalTrials.gov

FUTURE PROCESS

In future a more automatic process should be established, in which the programming department will play a major role.

The basic idea is to save the information needed for ClinicalTrials.gov at that point during a clinical trial, at which these data are produced anyway. This would be the time, when the tables for the Clinical trials Report are created. Therefore, the programmers could save the relevant information at the same time, when they produce a table program for the Trial Report. If this could be achieved, no manual entry of study results is needed anymore and the correctness of the result data could be established during normal program validation of study programs.

In a first step all needed information have to be specified in advance using the table templates. This should ensure that it is uniquely defined which information comes from which table. Therefore, during programming the extraction of the relevant information could easily be inserted in the corresponding table program. Using a validated SAS macro the programmer can save all relevant data during table creation into a dataset for ClinicalTrials.gov usage. The final output dataset for CliniclaTrials.gov will be available at the same time the final tables for the Clinical Trial Report are finalized.

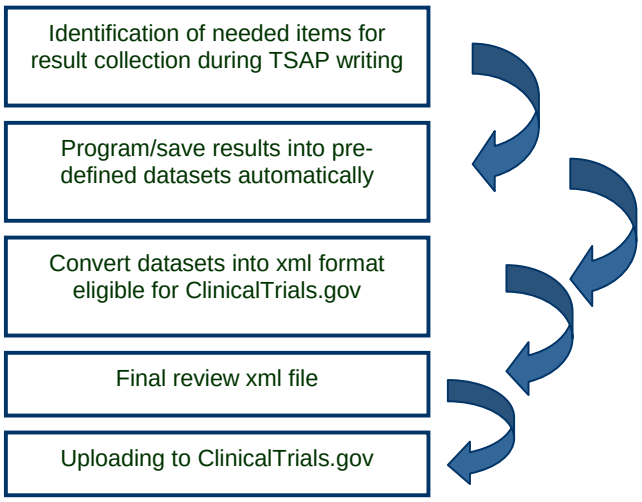

Figure 5: Final planned process description for collecting and entering data to ClinicalTrials.gov

THE PROGRAMMER'S RESPONSIBILITIES

In the planned future process of data collection for ClinicalTrials.gov the statistical programmers will fill a major task. All needed data will be derived and saved during the standard programming process for the tables, listings and figures needed for the Clinical Trial Report. The additional programming workload should not be very high. The extraction of necessary information for the dataset, where the disclosure information will be stored, will be done using a standardized, validated macro. Therefore, only the information, which values from which variables in which (table) datasets will be collected for which item in the ClinicalTrials.gov file must be specified during programming. The underlying specifications which values are necessary will be delivered by the responsible statistician during creation of the Statistical Analysis Plan. The macro used for extracting table data for Clinicaltrials.gov will have to cover all possible entries which could be entered on the website and also will have to formal validate the saved data.

The programmer will have to ensure that the correct table numbers are chosen according to the specifications. If tables which contain information for the result dataset will change during table programming the macro call will have to be updated accordingly. During validation of the table programs the correct usage of macro calls will be verified.

CONCLUSION

Clinical trials result disclosure on the website ClinicalTrials.gov is mandatory since September 2008 and causes additional work to enter relevant data or upload results in a format eligible for the website, respectively. A first preliminary process to comply with the legal requirements was a mainly manual process. Data entry and copy were the main tasks to get the needed information. In a planned process most of the data collection should be done automatically. The data collection will be performed during programming of the tables for the Clinical Trial Report using standardized macro calls. Because the data are produced nevertheless at this point the additional workload will be minimized. Further there will be less efforts needed to check and validate every single number in the result dataset for ClinicalTrials.gov. This process, therefore, should give a solution, where additional workload is small and the accuracy of the delivered information is high. The data from the final result dataset could easily be transformed into the xml format required by ClinicalTrials.gov.

REFERENCES

- ClinicalTrials.gov "Basic results" Data Element Definitions (DRAFT), July 21, 2009 – http://prsinfo.ClinicalTrials.gov/results_definitions.html
- U.S. Public Law 110-85 (Food and Drug Administration Amendments Act of 2007 or FDAAA), Sept 27, 2007 – http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=110_cong_public_laws&docid=f:publ085.110.pdf
- Tony Tse, Rebecca J. Williams, Deborah A. Zarin – Reporting "Basic Results" in ClinicalTrials.gov, Chest 2009 Jul; 136(1):295-303
- Tony Tse, Rebecca J. Williams, Deborah A. Zarin – Update on Registration of Clinical Trials in ClinicalTrials.gov, Chest 2009 Jul; 136(1):304-305
- Ellen Strahlman, Frank Rockhold, Andrew Freeman – Public Disclosure of Clinical Research, Lancet 2009 Apr; 373(9672):1319-1320
- Tony Tse, Rebecca J. Williams – ClinicalTrials.gov to Include Basic Results Data – NLM Tech Bull 2008 Sep; 364:e15
- Maine State Law §2700-A. Prohibitions and Required Disclosures, March 2007
- Catherine De Angelis, Jeffrey M. Drazen, Frank A. Frizelle et al. – Clinical Trial Registration: a Statement from the International Committee of Medical Journal Editors, N Engl J Med 2004; 351:1250-1251

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Ralf Minkenberg
Boehringer Ingelheim Pharma GmbH & Co. KG
Binger Str. 173
55216 Ingelheim
Work Phone: +49-6132-7797209
Fax: +49-6132-7297209
Email: ralf.minkenberg@boehringer-ingelheim.com