# Points to consider for statistical analysis using the database for Aggregate Analysis of ClinicalTrials.gov

Introduction

This document suggests points for investigators to consider when planning a statistical analysis of the database for Aggregate Analysis of ClinicalTrials.gov (AACT). It is not intended to be a comprehensive guide for using or analyzing data in AACT.

The current version of AACT contains content that was downloaded on 27 March 2016 and includes studies that were registered at ClinicalTrials.gov and publicly released by 24 March 2016. The 2016 March release of AACT includes both the study registration fields and the basic results and adverse events reporting fields.

Population: studies likely to be represented

Virtually any clinical study may be registered at ClinicalTrials.gov and therefore be included in AACT. However, the registry is more likely to include certain types of study relative to others. These biases are summarized by Zarin et al. [1]:

“… [T]here are undoubtedly trials that are not registered in ClinicalTrials.gov or any other publicly accessible registry. Coverage in ClinicalTrials.gov is likely to be most complete for trials of drugs or devices that are sponsored by U.S.-based or multinational organizations (e.g., major pharmaceutical companies).”

The ClinicalTrials.gov trial registry was released for the registration of studies on February 29, 2000. The database downloaded by the Clinical Trials Transformation Initiative (CTTI) and the Duke Clinical Research Institute (DCRI) on March 27, 2016 includes 211,437 studies. Of these, 170,688 are interventional studies in which participants are assigned according to a research protocol to receive specific interventions. The registration of studies and reporting of results and adverse events has been mandated to a large extent by requirements (both legal and institutional) implemented as part of the Food and Drug Administration Amendments Act (FDAAA), as well as by requirements introduced by the International Committee of Medical Journal Editors (ICMJE) and the European Medicines Agency (EMA) regarding registration of clinical studies. Table 1 describes the scope of these requirements.

**Table 1:** Scope of Interventional Studies Covered by Major Reporting Policies\*

|  |  |  |
| --- | --- | --- |
| **Policy Requirements** | **Registration & results reporting**  **Requirements** | **Effective**  **date** |
| NCI Access Policy | The National Cancer Institute (NCI), NIH, issued its [Policy Ensuring Public Availability of Results from NCI-supported Clinical Trials](https://grants.nih.gov/grants/guide/notice-files/NOT-CA-15-011.html). Generally, for "all initiated or commenced NCI-Supported Interventional Clinical Trials whether extramural or intramural" (Covered Trials), "Final Trial Results are expected to be reported in a publicly accessible manner within twelve (12) months of the Trial's Primary Completion Date regardless of whether the clinical trial was completed as planned or terminated earlier." This policy "will be incorporated as a Term and Condition of the award." | January, 2015 |
| FDAAA[2] | The following must be registered in ClinicalTrials.gov:   * Interventional studies of drugs, biologics, or devices (whether or not approved for marketing) * Studies phases 2 through 4 * Studies with at least 1 U.S. site or   conducted under IND/IDE  Results and adverse event reporting is required for studies that meet the above registration requirements if they study drugs, biologics, or devices that are approved, licensed, or cleared by the FDA. | **September 27, 2007.** Studies initiated after this date, or with a completion date later than 12/25/2007 are subject to FDAAA requirements. Registration is required no later than 21 days after first patient is enrolled. Results and adverse events must be reported for these studies (if required) within 1 year of completing data collection for the pre-specified primary outcome.  **September 2008.** Results reporting launched with optional adverse event reporting.  **September 2009.** Adverse event information became required. |
| ICMJE[3] | The following must be registered in ClinicalTrials.gov or other approved registry:   * Interventional studies of any intervention type, phase, or geographical location   No results reporting requirements | **July 1, 2005.** Studies initiated after this date must be registered before first patient enrolled; studies initiated before this date must be retrospectively registered to be considered for publication. |
| EMA[4,5] | The following must be registered in ClinicalTrials.gov or other approved registry:   * Interventional studies of drugs or biologics (whether or not approved for marketing) * Phase 1 studies (pediatrics only); * Studies in phases 2 through 4 * Studies taking place in at least 1 European Union site   Results reporting required for all studies that meet registration requirements. | **May 1, 2004.** EMA launched EudraCT  https://eudract.ema.europa.eu/  **March 22, 2011.** The [EU Clinical Trials Register](https://www.clinicaltrialsregister.eu/) was launched by the European Medicines Agency (EMA) |

*\* Adapted from [1]. For complete descriptions of policy requirements, see the references cited.*

*EMA denotes European Medicines Agency; FDAAA, Food and Drug Administration Amendments Act; ICMJE, International Committee of Medical Journal Editors; IDE, investigational device exemption; IND, investigational new drug application.*

Based on these requirements, the following are examples of characteristics that may influence the likelihood that a study is included in the ClinicalTrials.gov registry:

* Interventional studies are more likely to be registered than observational studies.
* Studies that began before the ICMJE requirement in July 2005 are less likely to be registered, especially if their results are unpublished (e.g., negative studies).
* Studies with drug, biological, or device interventions are more likely to be registered than studies of other interventions.
* Studies with at least one site in the United States or European Union are more likely to be registered than studies with no such sites.
* Studies involving a drug or device that is manufactured in the United States are more likely to be registered than studies involving a drug or device manufactured outside of the United States.
* Studies subject to an IND or IDE are more likely to be registered (i.e., if the study is intended to support approval for marketing in the United States).
* Phase 1 adult drug studies or small feasibility studies of devices are less likely to be registered.
* Studies in pediatric populations may be more likely to be registered.

Duplicate records

Studies registered at ClinicalTrials.gov are identified by a unique identifier, the NCT\_ID. Because of the quality assurance measures applied by ClinicalTrials.gov staff on registration entries, we can be reasonably certain that each study (i.e., NCT\_ID) entered in ClinialTrials.gov refers to a unique clinical study. However there may be a small number of duplicate records within the database [6].

Types of questions that can be investigated using the ClinicalTrials.gov data

The version of the ClinicalTrials.gov database that has been made publicly available through the CTTI and the DCRI contains study registration and results records. The registration records describe the study characteristics, including sponsor, disease condition, type of intervention, participant eligibility, anticipated enrollment, study design, locations, and outcome measures. Summary results and adverse events data are included in the current version of the AACT database. The article by Tse et al [8] may be helpful in understanding the components of the basic results that are reported at ClinicalTrials.gov.

*Using the study registration data:*

We anticipate that investigators will use the current database to explore the characteristics of selected subsets of clinical studies (e.g., typical enrollment for a phase 3 study in breast cancer patients), and to compare and contrast these characteristics across different subgroups of studies (e.g., sponsor; device versus drug intervention; or prevention versus treatment).

*Using the study results and adverse events data:*

Researchers may be able to use the basic results and adverse events summary data reported at ClinicalTrials.gov for meta-analysis or systematic review (e.g., to compare the efficacy and safety of different types of diabetes therapies). However because only a small subset of studies registered in ClinicalTrials.gov are required to report results, the results data from ClinicalTrials.gov will most likely be a useful supplement to traditional data sources used for meta-analysis or systematic review, such as published and unpublished manuscripts and abstracts. Standard techniques for valid meta-analysis or systematic review (e.g., PRISMA statement[7]) should be used when determining how to appropriately identify and aggregate summary data gleaned from ClinicalTrials.gov and/or literature.

Interpretation of variables

When interpreting the study characteristics collected for a study registered with ClinicalTrials.gov, investigators are encouraged to refer to 2 NLM sites that provide authoritative definitions for the data element: [study registration data elements](http://prsinfo.clinicaltrials.gov/definitions.html) and [basic results data elements](http://prsinfo.clinicaltrials.gov/results_definitions.html). Interpretation of a variable may depend on:

* **How the question was phrased.** For example, the definition of “Sponsor” does not necessarily imply that the sponsor is the agency paying for the clinical study, as might be expected from the common use of the term.
* **Whether the respondent can enter a free-text answer to a specific question, or is restricted to a fixed set of possible responses.** 
  + Note that the definition of a data element and the available responses may have changed over time. The most recent data element definitions are available at <http://prsinfo.clinicaltrials.gov/definitions.html> (study data) and <http://prsinfo.clinicaltrials.gov/results_definitions.html> (results data).  A history of changes through September 2011 for the study variable definitions can be viewed in the Comprehensive Data Dictionary 2011 available at http://ctti-clinicaltrials.org/files/documents/AACTcomprehensiveDataDictionaryV3\_2011.xlsx.
* **Whether there is dependence between fields.** Certain data elements need to be interpreted together with other data elements. For example, data elements such as enrollment date and completion date have a companion data element that indicates whether the value in the first field is an anticipated or actual value.
  + Note that the study record may be updated by the owner of the record at any time. Fields such as enrollment type may be changed from anticipated to actual, indicating that the value entered now reflects the actual rather than the planned enrollment. When data are downloaded, the result is a snapshot of the database at that particular time point, and the history of changes made to the field is lost.

Data completeness and accuracy

The presence of a record in a dataset indicates that information was submitted to ClinicalTrials.gov for at least one variable in that data set before the data were downloaded from ClinicalTrials.gov. Some data elements are more/less likely than others to have missing information, depending on several known factors. For example:

* **The data element being required by the FDAAA and/or the ClinicalTrials.gov website.** Refer to data element definitions and the comprehensive data dictionary for specifics regarding these requirements. Requirements may have changed over the history of the ClincalTrials.gov database.
* **The date when the data element was introduced.** Not all data elements were included in the database at the time of its launch in 2000, but were added later. Studies registered after FDAAA when into effect must meet more requirements than studies registered earlier in the life of ClinicalTrials.gov.
* **The branching structure of questions.** The availability of certain questions to the person submitting data depends on answers to previous questions. For example, questions about bio-specimen retention are only available for observational studies. Therefore, interventional studies should be excluded when analyzing data elements pertaining to bio-specimens.
* **The list of possible answers for data elements with a fixed set of responses.** For example, questions that include “N/A” as a possible response are likely to have fewer missing values than questions that do not provide a “N/A” response.

“Missingness” of data may also depend on other unknown factors. However, regardless of the cause of missing data, users of ClinicalTrials.gov datasets are encouraged to specify clearly how missing values and “N/A” values are handled in their analysis. For example, are studies with missing values excluded from statistics summarizing that data element, or are they included? In some cases, missing values may be imputed based on other fields (e.g., if a study has a single arm, it cannot employ a randomized design). In other cases, a sensitivity analysis may be appropriate for exploring the effect of different assumptions about the missing values on analysis results.

Although the FDAAA and other requirements do not apply to all fields in the database, users might consider including only studies registered post-FDAAA (September 2007), or studies with a primary completion date after December 2007. This will help to limit the number of missing values across many data elements. Users could also consider annotating data elements used in analysis according to whether or not they are FDAAA-required fields, if the user believes this might affect the extent of missing data.

Even when the data elements for a particular study are complete, users are cautioned to have modest expectations about the accuracy of the data. In particular, results data posted at ClinicalTrials.gov may not be subject to the same level of critical scrutiny as results published in a peer-reviewed journal. As described by Zarin and colleagues [1], ClinicalTrials.gov has implemented several measures to assure data quality. For example, staff applies automated business rules that alert providers when required data are missing or are internally inconsistent. In addition, some manual review is performed, and a record may be returned to the data provider if revision is required. However, ClinicalTrials.gov staff cannot always validate the accuracy of submitted data (e.g., against an independent source). As Zarin et al. note, “… individual record review has inherent limitations, and posting does not guarantee that the record is fully compliant with either ClinialTrials.gov or legal requirements” [1].

During our own analysis of the ClinicalTrials.gov database, several potentially unrealistic values for numeric data elements were encountered, such as an anticipated enrollment of several million subjects. When aggregate summaries of numeric data are provided, analysts are encouraged to use measures that are robust to outliers, such as medians and interquartile ranges, rather than measures such as means ± SD, which could be strongly influenced by unusually large or small values. Users may also wish to run their own consistency checks (e.g., to compare whether the number of arm descriptions provided for the study matches the data element that quantifies the number of arms in the study design).

AACT is a snapshot at one time point

The data downloaded from ClinicalTrials.gov and stored in the database for Aggregate Analysis of ClinicalTrials.gov (AACT) is a snapshot of the information that was publicly available at ClinicalTrials.gov on the download date. Data submitters may update their ClinicalTrials.gov study record at any time but a particular version of AACT (e.g., the version downloaded on 27 March 2016) only captures the information present at one point in time. Although changes to a study are stored in an archive history at ClinicalTrials.gov, these changes are not captured in a particular version of AACT. A user may find that information contained in AACT differs from information currently listed on ClinicalTrials.gov. For example, after the data were downloaded and loaded into AACT, a study may have completed enrollment and updated the enrollment status and enrollment values at ClinicalTrials.gov.

Use of appropriate statistical inference

If the AACT results data are to be used to support a meta-analysis or systematic review of the safety or efficacy of a particular intervention, then standard methods of meta-analysis or systematic review (e.g., the PRISMA statement [7]) should be used to appropriately account for study-to-study variability and other sources of uncertainty or bias. We recommend that authors consider the following points when deciding whether to report p-values, confidence intervals, or other probability-based inference when performing aggregate analysis of the ClinicalTrials.gov database.

1. **Is the data-generating mechanism random?**

Methods of statistical inference such as p-values and 95% confidence intervals are most appropriate when used to quantify the uncertainty of estimates or comparisons due to a random process that generates the data. Examples of such processes include selection of a random sample of subjects from a broader population, randomly assigning a treatment to a cohort of subjects, or a coin toss about which we aim to predict future results.

In the following examples, we recommend against reporting p-values and 95% confidence intervals because the data generating mechanism is not random.

Example 1: Descriptive analysis of studies registered in the ClinicalTrials.gov database.

In this case, the “sample” equals the “population” (i.e., the group about which we are making conclusions) and there is no role for statistical inference because there is no sample-vs-population uncertainty to be quantified.

Example 2: Descriptive analysis of the “clinical trials enterprise” as characterized by the studies registered in ClinicalTrials.gov.

Despite mandates for study registration (Table 1), it may be that some studies that are required to be registered are not. In this case the sample (studies registered in ClinicalTrials.gov) may not equal the population (clinical trials enterprise). However, it is likely that those studies not registered are not excluded at random, and therefore neither p-values nor confidence intervals are helpful to support extrapolation from the sample to the population. To support such extrapolation, we recommend careful consideration of the studies that are highly likely to be registered (see section above on Population), and to limit inference to this population so that sample-vs-population uncertainty is minimal.

1. **How can I objectively identify important differences?**

In practice, p-values and confidence intervals are often employed even when there is no random data generating process in order to help highlight differences that are larger than “noise” (e.g., authors may want to highlight differences with a p-value < .001). While this practice may not have a strong foundation in statistical philosophy, we acknowledge that many audiences (e.g., journal peer reviewers) may demand p-values because they appear to provide objective criteria for identifying larger-than-expected signals in the data. While we don’t encourage reporting of p-values for this purpose, we do encourage analysts to specify objective criteria for evaluating signals in the data. For example,

1. Prior to examining the data, specify comparisons of major interest, or quantities to be estimated.
2. Determine the magnitude of differences that would have practical significance. E.g., a 25% difference in source of funding between studies of two pediatric conditions, or a difference in enrollment of 100 participants.
3. Determine appropriate formulas for quantifying differences between groups or summarizing population variability. This quantification could take account of the observed difference, variability in the data, and the number of observations. For example,
   * When summarizing a continuous characteristic such as enrollment, the analyst might choose to report the median and 5th to 95th percentiles.
   * To quantify signal to noise, the analyst could calculate a t-statistic or a chi-squared statistic (without the p-value) and rank differences between two groups based on these values. The analyst might pre-specify a threshold (e.g., absolute value of 3) to flag notable differences.

Specific tips for working with the AACT database

1. ***Data formats and supporting documents:***

The AACT database extracts are available for download in three formats:

* Oracle dmp file(s) for users who plan to import the data extracts into an Oracle database. A “readme” document is provided with instructions and troubleshooting tips.
* SAS transport files for users who plan to work with the data extracts in SAS software (version 9 or later). A “readme” document is provided with instructions for reading the transport files in SAS.
* Pipe-delimited text files for users who plan to work with the data extracts in other software. Three “readme” documents are provided, the first providing general guidance for using the text files, the second providing suggestions for reading these files in the R statistical computing environment, and the third providing suggestions for reading these files with SAS software. SAS users will probably find the SAS transport files more convenient to work with, but may need to use the text files if unable to read the transport files (e.g., if using a SAS version preceding version 9).

Past users have reported some difficulties using the text files, partly due to the presence of some fields with very long text values which span multiple lines, may include embedded soft returns (line feed or LF characters), or may include isolated or unmatched double quotes. These characteristics may result in incorrect parsing of fields and/or records. Suggestions for addressing these specific issues are provided in the three “readme” documents (README\_GENERAL, README\_R, and README\_SAS) provided with the text files. Users are encouraged to use the file CLINICAL\_STUDY\_NOCOB.txt file in place of CLINICAL\_STUDY.txt, unless they need the information contained within the long descriptive fields (for example BRIEF\_SUMMARY, CRITERIA, DETAILED\_DESCRIPTION).

1. ***Schema diagram, Record identifiers & NCT\_IDs, and MESH\_THESAURUS data set:***

* Users are encouraged to refer to the Schema diagram in the Comprehensive Data Dictionary to determine the relationships between the different data sets that comprise AACT. These relationships determine how data sets may be merged together.
* Each data set in AACT contains a unique record identifier which is referred to as the primary key in the Comprehensive Data dictionary. For example, the NCT\_ID is the primary key in the CLINICAL\_STUDY data set. This NCT\_ID also identifies unique studies in AACT. Some data sets also contain an identifier referred to as a foreign key. For example, in the CONDITIONS data set CONDITION\_ID is the primary key and NCT\_ID is the foreign key. This data set may contain multiple records per NCT\_ID. The CONDITIONS data set may be merged with the CLINICAL\_STUDY data set by the NCT\_ID in order to obtain the disease conditions that are associated with each study.
* Some data sets do not contain the NCT\_ID identifier. To associate records in such data sets with a particular study, the user will need to merge these data sets with the data set or data sets that are higher in the schema hierarchy in order to obtain the NCT\_ID. For example the data set RESULTS\_BASELINE\_MEASURES can be merged with RESULTS\_BASELINE by the identifier RSLTS\_BASELINE\_ID. This merged data can then be merged with CLINICAL\_STUDY by the NCT\_ID.
* The MESH\_THESAURUS data set does not have a formal relationship with other data sets in AACT. The MESH\_THESAURUS is provided so that users may look up a particular MeSH term (MESH\_TERM) in the thesaurus, for example, to determine the position of this term in the MeSH trees hierarchy (MESH\_ID). MESH\_THESAURUS was created from the most recently available (2015) MeSH trees information available from [NLM’s MeSH Site](http://www.nlm.nih.gov/mesh/meshhome.html). For more information on MeSH, please refer to the section below titled ‘MeSH terms in CONDITION\_BROWSE and INTERVENTION\_BROWSE’ and to [NLM’s MeSH Site](http://www.nlm.nih.gov/mesh/meshhome.html).

1. ***Use of the ARM\_GROUPS data set***

The ARM\_GROUPS data set contains records that describe the arms or groups in the clinical study. For example a two arm study may assign participants to one of two study arms: the first in which participants receive placebo drug in addition to standard of care, and the second in which participants receive an experimental drug in addition to standard of care. Information about arms or groups is provided by the data submitter when a study is first registered with ClinicalTrials.gov. Once the study is completed, results data may be submitted to ClinicalTrials.gov. At that time results will be described for each arm or group (and potentially also for the overall study population), along with a description of the arm or group. Thus a study may have multiple records in the ARM\_GROUPS data set, some coming from the information that was submitted when the study was registered, and some coming from the information that was submitted when results and adverse events were reported. The following values in the ARM\_GROUP\_TYPE field indicate that this record in the ARM\_GROUPS data set relates to a specific component of the reported results:

* Participant Flow
* Baseline
* Results Outcome
* Reported Event

Users may wish to merge records from the ARM\_GROUPS data set with records in one or more of the results data sets, for example to see the age distribution of participants in the Placebo arm. This can be achieved by merging the required data sets (e.g., RESULTS\_BASELINE\_MEASURE\_CATGY with ARM\_GROUPS) by the ARM\_GROUP\_ID identifier. When merging data sets with ARM\_GROUPS, users are cautioned against overwriting variables named DESCRIPTION that are present in various results data sets with the variable named DESCRIPTION in the ARM\_GROUPS data set.

1. ***Information about trial sites (FACILITIES, REMOVED\_COUNTRIES, and LOCATION\_COUNTRIES)***

Information about organizations (trials sites) where the protocol is/was being conducted is stored in the FACILITIES dataset. This is a snapshot of the facility information that was included in the study record on the date when the data were downloaded from ClinicalTrials.gov. Sites that were initiated after the data were downloaded are not included.

In some cases, data submitters remove sites from the study record. If a site was removed before the data were downloaded then it is not included in the FACILITIES dataset. If all sites in a country have been removed from the study record for a particular NCT\_ID then an algorithm run at the National Library of Medicine (NLM) creates a record for that country and NCT\_ID in the REMOVED\_COUNTRIES dataset. The reasons why sites are removed from the FACILITIES dataset are varied but unknown. For example, a site may have been removed because it was never initiated or because it was entered with incorrect information. The recommended action for sites that have completed or terminated enrollment is to change the enrollment status to “Completed” or “Terminated”, however, such sites are sometimes deleted from the study record by the responsible party. Data analysts may consider using information from the REMOVED\_COUNTRIES dataset to supplement the information about trial locations that is contained in FACILITIES, particularly for studies that have completed enrollment and have no records in FACILITIES.

The LOCATION\_COUNTRIES dataset is created by a National Library of Medicine (NLM) algorithm from the FACILITIES dataset. This algorithm identifies the unique facility countries associated with each NCT\_ID in the FACILITIES dataset and stores these countries in the LOCATION\_COUNTRIES dataset for each NCT\_ID. In the 2016 March release, all countries listed in the FACILITIES data set for a particular NCT\_ID were also listed in the LOCATION\_COUNTRIES data set for that NCT\_ID, and vice versa. For this release, users who are interested in identifying countries where participants are being/were enrolled may use either the FACILITIES or the LOCATION\_COUNTRIES datasets with equivalent results.

1. ***MeSH terms in CONDITION\_BROWSE and INTERVENTION\_BROWSE***

When data submitters provide information about studies at ClinicalTrials.gov they are encouraged to use Medical Subject Heading (MeSH) terminology in fields such as interventions, conditions, and keywords. The CONDITION\_BROWSE and INTERVENTION\_BROWSE data sets contain relevant MeSH terms generated by an NLM algorithm that may be used to supplement the information about study conditions and interventions supplied by the data submitter. The CONDITION\_BROWSE and INTERVENTION\_BROWSE datasets contain Medical Subject Heading (MeSH) terms generated by an algorithm run by the National Library of Medicine (NLM). The NLM algorithm is re-run nightly on all studies in the ClinicalTrials.gov database, and sources the most up-to-date information in the study record, the latest version of the algorithm, and the version of the MeSH thesaurus in use at that time.

1. ***“Delayed Results” data elements available in AACT***

A responsible party may delay the deadline for submitting results information to ClinicalTrials.gov if one of the following two certification conditions applies to the trial:

* Initial approval: trial completed before a drug, biologic or device studied in the trial is initially approved, licensed or cleared by the FDA for any use.
* New use: the manufacturer of a drug, biologic or device is the sponsor of the trial and has filed or will file within one year, an application seeking FDA approval, licensure, or clearance of the new use studied in the trial.

A responsible party may also request, for good cause, an extension of the deadline for the submission of results.

The date when the first certification or extension request for a trial was received by ClinicalTrials.gov is public information, but is not included among the data elements that can be downloaded from ClinicalTrials.gov. The National Library of Medicine sent a list of studies that had submitted a certification or extension request before March 27, 2016, and the date of the first certification or extension request (FIRSTRECEIVED\_DISPOSITION\_DATE) to CTTI/DCRI for incorporation into the 2016 March update of AACT. This information is not currently included in the Oracle dmp file, but is included in the following formats:

* Pipe delimited text output: File is named ‘delayed\_results.txt’
* SAS CPORT transport: Data set is named DELAYED\_RESULTS and is included in the SAS transport file named “aact201603\_study\_data\_1.xpt”.

1. ***Pipe (‘|’) characters within character fields replaced with slash (‘/’)***

In general, the content that is contained in the AACT database preserves the content in the source XML files that are downloaded from ClinicalTrials.gov. However in the March 27, 2016 and other recent downloads we found that some character fields contained a vertical pipe (‘|’) within the field. Since these embedded pipes would create problems for users attempting to read the pipe-delimited text file extract format of AACT, and no other character string could be identified to use as an alternative delimiter, it was decided to remove embedded pipes from within character fields and replace them with a slash (‘/’).

**References**

1. Zarin, D. A., Tse, T. T., Williams, R. J., Califf, R. M., and Ide, N. C. (2011). The ClinicalTrials.gov results database – update and key issues. *N Engl J Med* 364: 852–60.
2. Food and Drug Administration Amendments Act of 2007. Public Law 110-95.
3. Laine, C., Horton R., DeAngelis C.D., et al. Clinical trial registration – looking back and moving ahead. *N Engl J Med* 356: 2734–6.
4. Communication from the Commission regarding the guideline on the data fields contained in the clinical trials database provided for in Article 11 of Directive 2001/20/EC to be included in the database on medicinal products provided for in Article 57 or Regulation (EC) No 726/2004. In: European Commission, ed. Official Journal of the European Union, 2008. (2008/C 168/02.)
5. Guidance on the information concerning paediatric clinical trials to be entered into the EU Database on Clinical Trials (EudraCT) and on the information to be made public by the European Medicines Agency (EMEA), in accordance with Article 41 of Regulation (EC) No 1901/2006. In: European Commission, ed. Official Journal of the European Union, 2009. (2009/C 28/01.)
6. Zarin, D. A., Ide, N. C., Tse, T. et al. (2007). Issues in the registration of clinical trials. *JAMA* 297: 2112—2120.
7. Moher, D., Liberati, A., Tetzlaff, J. and Altman, D. G. (for the PRISMA Group)(2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement, *BMJ* 339: 332—336.
8. Tse, T., Williams, R. J., Zarin, D. A. (2009). Reporting “basic results” in ClinicalTrials.gov. *CHEST* 136: 295—303.