JOHNS HOPKINS
U N I V E R S I T Y

Center for Clinical Trials

Department of Biostatistics
Department of Epidemiology
Department of International Health

Department of Medicine
Department of Ophthalmology
Oncology Center

(Thursday 9:59pm) 15 June 2000

Memorandum

To: Center for Clinical Trials faculty and staff

Fr: Curt Meinert

Re: Analysis good practice policy and procedures (GPPP)

Definitions

adjustment *n* - 1. The act or process of **adjusting**; the state of being **adjusted**. 2. The **control** of extraneous **sources of variation** during **data analysis** that affect, or are believed to affect, some **comparison** by use of an **adjustment procedure**. Those procedures for **trials** traditionally involve the use of **baseline variables**, especially those considered to have **distributions** that differ by **treatment group** and that influence, or are suspected of influencing, **treatment** or **outcome**.

adjustment procedure *n* - Any of a variety of procedures intended to remove the effect of one or more extraneous **sources of variation** that could affect, or are believed to affect, a particular **result**. Procedures include **direct** and **indirect rate adjustment**, **subgroup analysis**, **analysis of covariance**, and **linear** and **nonlinear regression analysis**. rt: **standardization**, **stratification**

adjustment variable *n* - A **variable**, such as age or gender, used for **adjustment** via some **analysis procedure**; in **trials**, usually a **baseline variable** or a demographic characteristic such as sex, race, or age on entry.

ad hoc subgroup *n* - A **subgroup** (defn 2) identified by **data analysis** (not identified by other means prior to data analysis).

ad hoc subgroup comparison *n* - A **comparison** based on an **ad hoc subgroup**. ant: **designed subgroup comparison**

analysis by administered treatment *n* - **Data analysis** in which **tabulations** and summaries are by the **administered treatment** (as opposed to the **assigned treatment**), eg, done by grouping **results** for **patients** who were assigned to the **test treatment** but refused the treatment with those for patients assigned to and receiving the **control treatment** in the case of a **placebo-controlled trial**. Not recommended as the primary method of analysis (see **analysis by assigned treatment**). ant: **analysis by assigned treatment** rt: **analysis by level of treatment compliance**

Room 5010 • 615 North Wolfe Street • Baltimore, Maryland 21205 • 410 955-8175 • Fax: 410 955-0932
E-mail: cctmail@jhsph.edu • Web site: http://www.jhsph.edu/Research/Centers/CCT

analysis by assigned treatment *n* - **Data analysis** in which **tabulations** and summaries are by **assigned treatment** regardless of **administered treatment**. Recommended primary method of analysis. syn: analysis by intention to treat ant: **analysis by administered treatment**

Bayesian *adj* - Being or relating to a school of thought in which a **prior probability** distribution is assigned to **parameters** (**hypotheses**) fashioned from **observed data** by application of **Bayes' theorem**. The resulting **posterior probabilities** can be viewed as measures of existing evidence and prior opinion, a result of logical reasoning, or subjective degree of belief. rt: **frequentist**, **likelihoodist**

Bayesian *n* - One who subscribes to the Bayesian school of thought for the **analysis** and interpretation of **data**. rt: **frequentist**, **likelihoodist**

composite event *n* - An **event** that is considered to have occurred if any one of several different events or **outcomes** are observed (eg, occurrence of an attack of angina pectoris, transient ischemic attack, or myocardial infarction). rt: **composite outcome**

composite outcome *n* - An **outcome** comprised of any of several different outcomes (eg, prolonged attack of angina pectoris, elevated bilirubin, or abnormal ECG tracing in a cardiovascular **trial**). rt: **composite event**

counting rule *n* - [**trials**] A **rule** related to the counting of persons (**observation units**) **enrolled** into a trial or **events** observed in relation to the **primary analysis**. The counting rules, discussed by Meinert and Tonascia [1986], include: 1) All persons enrolled (**assigned** to **treatment**) should be counted in the **denominator** for the primary analysis. 2) All events should be counted regardless of when they occur after enrollment. 3) Events should be counted in the **treatment group** to which a person (observation unit) was assigned, regardless of degree of **compliance** to the assigned treatment. 4) Counts of subsets of events (eg, deaths due to cardiovascular causes) should not be used for analyses until counts and analyses of the higher order events (eg, deaths, regardless of cause) have been performed. rt: **analysis principle**, **analysis rule**

data dredging *v* - **Data analyses** done on an **ad hoc** basis, without benefit of prior stated **hypotheses**, especially those done with the aim or intent of trying to find noteworthy differences within or among different **subgroups**; **exploratory data analysis**; see **dredge**. ***Usage note***: Often used in a pejorative sense, especially in reference to analyses in which it appears that only large differences are presented and where the number of **comparisons** made is not specified. Not to be confused with **subgroup analysis**; see usage note for that term.

data freeze *n* - **Data** held in a fixed state, especially such a state imposed on an active **database** or **data file** in order to complete some task requiring a stable, nonchanging, database or data file (eg, as required for preparation of a **treatment effects monitoring report**). rt: **data snapshot**

designed subgroup comparison *n* - A **subgroup comparison** in a **subgroup** enrolled to a **quota**, especially one based on a **sample size** calculated to yield a specified level of **precision**. ant: **ad hoc subgroup comparison**

frequentist *adj* - Being of or relating to a school of thought in which **inferences** about a particular **dataset** depend on the **probability distribution** for particular **parameter** values based on the hypothetical notion of a **study** being repeated many times under the same conditions; also referred to as **sampling** theory approach. rt: **Bayesian**, **likelihoodist**

frequentist *n* - One who subscribes to the frequentist school of thought for the **analysis** and interpretation of **data**. rt: **Neyman-Pearson theory**, **Bayesian**, **likelihoodist**

interim data analysis *n* - [**trials**] 1. **Data analysis** carried out during a **trial** for the purpose of **treatment effects monitoring**. 2. Any data analysis done before data collection is completed, for whatever reason, but usually concerned with assessments of **treatment effects**. rt: **sequential data analysis** ***Usage note***: There are subtle distinctions between **interim data analysis** and **sequential data analysis** (see usage notes for **sequential** and **sequential data analysis**) that should be preserved in usage. Strictly speaking, the term, **interim data analysis**, applies to any trial, (**fixed sample size design** or a **sequential design**) in which interim analyses are done. However, the recommended convention is to reserve the term for fixed sample size designs and to use the term, **sequential data analysis**, when referring to the analyses required for sequential designs. See also note for **sequential data analysis**.

frozen data *n* - **Data** held in a fixed state by virtue of a **data freeze**.

frozen dataset *n* - A **dataset** created by virtue of a **data freeze**.

intention to treat *n* - [**trials**] A philosophy in which there is an **intent** to account for all **observation units** enrolled and to perform **analyses by assigned treatment**, regardless of observed course of **treatment**. ***Usage note***: Use only in the presence of language detailing the operational implications of the intent, as in **analysis by assigned treatment**. See also usage note for **intention**. The term as an **analysis principle** arises from the essence of the **pragmatic trial**, as discussed by Schwartz and Lellouch (1967). Armitage (1979) relates **intent to treat** and **analysis**. Sackett and Gent (1979) discuss the elements of the principle in their paper on counting and attributing events in clinical trials. Pocock (1983) uses the term in his book on clinical trials.

likelihoodist *adj* - Being or relating to a school of thought for **analysis** and interpretation of **data** based on the **likelihood principle**. rt: **frequentist**, **Bayesian**

likelihoodist *n* - One who subscribes to the likelihood school of thought for the **analysis** and interpretation of **data**. rt: **frequentist**, **Bayesian**

---

subgroup analysis *n* - 1. Any **data analysis** focused on a selected **subgroup** (defn 2). 2. Analysis aimed at characterizing **observed differences** among different subgroups, eg, **comparison** of **treatment differences** in a **trial** for different subgroups of **patients** defined by sex, age at entry, and other **baseline characteristics**. 3. A form of **exploratory data analysis** aimed at trying to identify a subgroup of persons that account for an observed difference, eg, such an analysis in a trial to determine whether or not an observed **treatment difference** can be accounted for by some **subgroup**. See also **data dredging**. ***Usage note***: Not to be used interchangeably with **data dredging**. Data dredging is **value-laden** and pejorative. **Subgroup analysis** is neutral in connotation and is descriptive of a process. Analysis involving subgroups formed using entry **demographic** and other **baseline characteristics** is an essential part of the analysis process for a trial. The analyses are done to determine whether or not it is reasonable to regard the **treatment effect** observed as being **homogeneous** (ie, **independent** of entry and other important baseline characteristics). The analysis has bearing on conclusions reached from the trial. Evidence of **qualitative** or **quantitative** treatment by baseline characteristic **interaction** obligates the trialist to temper or qualify the conclusion accordingly. A treatment effect cannot be assumed to be homogeneous across subgroups absent analyses aimed at addressing the question of homogeneity of treatment effect. Subgroup analyses become forms of data dredging if results of such analyses are used to identify "**significant**" differences without regard to the number of subgroups studied or when the results are presented so as to suggest that the difference is the result of clinical insight regarding an underlying **disease** process.

P&P 1: Specify the primary treatment comparison when the trial is designed; specify in the study protocol and handbook.

P&P 2: Commit to primary analysis by assigned treatment; specify commitment in the study protocol.

P&P 3: Detail counting rules for defining numerators and denominators for treatment comparisons during the design phase of the trial; commit to writing in the study protocol or study handbook.

P&P 4: Adhere to the following counting rules:

1. Count a person as randomized when the assignment is revealed to clinic personnel.
2. Count all persons randomized to the treatment group to which assigned, regardless of course of treatment.
3. Count all events occurring after randomization, regardless of time of occurrence.
4. Count events to the assigned treatment group, regardless of degree of compliance to the assigned treatment at the time of the event.

Comment

The rules are necessary to meet strict requirements for *intention to treat analyses*.

P&P 5: Do not report analyses for lower order events in the absence of analyses for higher order events.

Comment

A treatment difference based on a subset of events is interpretable only if consistent with results for the entire set of events. For example, suppose a difference in cardiovascular mortality favoring the test treatment in a heart trial. To make anything of that difference, the difference in overall mortality has to be in the same direction. If the difference for overall mortality is nil or in the opposite direction, the difference in cardiovascular mortality is meaningless because it is offset by increased mortality in the test-treated group for deaths from non-cardiovascular causes.

P&P 6: Do not base analyses on a composite measure without first analyzing and reporting results for the individual measures represented in the composite measure.

Comment

Similar to the rationale for P&P 5.

P&P 7: Replicate crucial counts, such as number of assignments by treatment group, number of deaths by treatment group, and number of other key events by treatment group, prior to stopping a trial or publishing results from the trial.

Comment

The time to find mistakes is before action is taken and publication. Mistakes contained in publications are embarrassing and indelible.

The preferred method of checking is to have two people, independent of each other, make essential counts.

P&P 8: Replicate key analyses prior to publication.

Comment

Do not assume that analysis programs are tried and true (even if they have run flawlessly for years), or that the results produced by an analysis program are reproducible. Most analyses involve dozens of assumptions, eg, in regard to how missing values are handled; in regard to how outlier values are dealt with; etc. Different programmers will make different assumptions. The only way to determine whether an analysis procedure is reproducible is to have two people perform the same analysis using the same program or using different programs.

P&P 9: Check for errors in results reported in manuscripts prior to submission for publication.

P&P 10: Do not mask data analysts.

Comment

Masked data analysis is done to increase objectivity in the analysis process. However, the added objectivity is achieved at the expense of competency and is, therefore, not advised.

P&P 11: Assess baseline comparability of the treatment groups; summarize and report in primary study publications; be cautious in interpretation; avoid characterizations of differences as being indicative of "breakdowns" in randomization merely because of differences yielding p-values of ≤ 0.05.

P&P 12: Calculated adjusted treatment differences; do so even if treatment groups are regarded as "comparable" as assessed under P&P 11.

Comment

Generally, the crude and adjusted treatment difference is the same, but there is no way to be certain that that will be the case without undertaking the effort to adjust.

P&P 13: Report crude and adjusted treatment differences in publications if adjustment makes a difference; otherwise report crude or adjusted differences along with a statement indicating that the crude and adjusted are the same.

P&P 14: Carry out analyses on frozen datasets.

Comment

Freezing is necessary to allow for production of internally consistent analyses and to allow analysts to check analyses.

P&P 15: Establish policy and rules for data freezes during the start-up phase of the trial.

Comment

The rules will be different depending on the reason for the freeze (eg, whether for production of an interim treatment effects or performance monitoring report, a publication, or archiving a final dataset). The rules should indicate the time of the freeze relative to when finished analyses are required and whether the frozen dataset is to include data under query. In regard to final datasets, the policy should be written to indicate the date beyond which changes to the dataset may not be made or accepted.

P&P 16: Characterize the analysis philosophy of the CC; characterize at the outset; indicate the extent to which philosophy departs from the traditional frequentist philosophy.

P&P 17: Indicate philosophy regarding use of p-values for characterizing treatment differences and whether values reported are nominal or "adjusted" (for multiple looks or multiple outcomes).

P&P 18: When presenting results, avoid binary characterization of results as being significant or nonsignificant; report observed p-values.

P&P 19: In trials with reasonably large sample sizes (say 100 or more per treatment group), before reporting a treatment effect, perform selected subgroup analyses to determine whether the observed treatment effect is homogeneous across subgroups.

Comment

One should not represent a treatment effect as being homogeneous (ie, the same across the spectrum of persons represented in the trial) in the absence of evidence to support the supposition.

The utility of subgroup analyses diminishes with diminishing size of the trial. The smaller the sample size the more improbable it is that one will discover heterogeneity of effect, even if present.

P&P 20: Limit subgroup analyses to subgroups defined by demographic and baseline entry characteristics.

Comment

Variables observed during followup are confounded by treatment effect and are suitable for use in adjustment or for subgrouping.

P&P 21: If a subgroup result is reported in a publication, indicate whether it is for a designed or ad hoc subgroup comparison.

Comment

The distinction is important for frequentists interested in p-values associated with the difference.

P&P 22: In large trials (see P&P 19), give results in finished reports by gender, ethnic origin, and age, or provide a statement indicating such analyses were done and that the differences observed were nil.

Comment

Much has been made in recent years of gender representation in trials and of the fact that treatments may work differently in women than in men. The best way to address this concern is to provide estimates of treatment effects by gender and other selected demographic variables. If there are no differences, or if editors "ax" the results because of space limitations, include a statement in the finished report indicating that such analyses were performed and that no differences were found.

P&P 23: Indicate in finished publications whether subgroup analyses were performed and the subgroups examined.

P&P 24: Do not report dredged results (see definitions for **data dredging** and **subgroup analysis**).

P&P 25: For publications, set up and maintain archives for associated dataset and programs used to produce them.