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: Data collection good practice policies and procedures (GPPP)

Definitions

baseline period *n* - [general] A **period** of **time** that is used to perform procedures needed to assess the suitability and eligibility of a **study candidate** for **enrollment** into a **study**, to collect required **baseline data**, and to carry out **consent processes**. [**trials**] 1. The period for a person or **treatment unit** defined by the start of the first **data collection visit** and ending with **assignment** to **treatment** or start of treatment. 2. Such a period ending with a **visit** or time point occurring shortly after **assignment** to **treatment** or the start of treatment. 3. A period of time during the course of treatment or **followup** of a person or treatment unit, usually marked by some **event**, process, or procedure, in which new **measurements** or **observations** are made to serve as a base for gauging subsequent change. 4. **enrollment period** rt: **treatment period**, **followup period**, **treatment and followup period**, **close-out period**, **baseline observation period**, **lead-in period**, **run-in period** ***Usage note***: Avoid in the sense of defn 2 without qualification (see note for **baseline** *adj*) and in the sense of defn 4. Provide qualifying detail for uses in the sense of defn 3. Traditionally, the point defining the end of the **baseline period** in trials is marked by the **treatment assignment** or initiation of **treatment**. The tendency to "stretch" the baseline period as in defn 2 arises from a desire to reduce missing **baseline data**. Clearly, the utility of an observation as a baseline, in the strict sense of usage, is diminished if there is any possibility of the observation being influenced by treatment. Hence, the practice is not recommended, even if the time interval following treatment assignment or initiation of treatment is small and even if the likelihood of treatment having had an effect on the **variable** being observed within that interval is small. See also notes for **baseline** *adj* and for **baseline** *n*.

contiguous time window *n* - A **time window** constructed to adjoin but not overlap the preceding or following time window. rt: **disjoint time window**, **ideal time window**, **overlapping time window**, **permissible time window**, **time window**

data collection *v* - 1. The act of observing, recording, or assembling **data** for some defined purpose, such as in a **research project**. 2. The act of observing and completing **data forms** on **observation units** of a **study**, as in relation to **patients** in a **trial** at a **clinic**; the act of mailing and receiving forms in the case of **surveys** or **followup** done by mail.

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

---

data collection compliance *n* - The nature, extent, or degree of **compliance** to prescribed **data collection** procedures. rt: **followup compliance**, **protocol compliance**, **treatment compliance**

data collection compliance measure *n* - Any of various **measures** of **data collection compliance** including **edit query**, **missed data**, **dropout**, **out of range**, and **outlier** rates, as well as counts indicative of failure to comply to required data collection procedures (eg, counts of ineligible persons **enrolled** and of missing **baseline** or **followup data items**). rt: **followup compliance measure**, **protocol compliance measure**, **treatment compliance measure**

data collection design *n* - **Design** relating to **data collection**, especially that relating to mode of collection, schedule of collection, and classes of data to be collected as expressed in the **data collection protocol** and by the **data collection schedule**.

data collection form *n* - A **data form** used for **data collection**.

data collection instrument *n* - **data collection form**

disjoint time window *n* - A **time window** neither adjoining or overlapping a preceding or following time window. rt: **contiguous time window**, **ideal time window**, **overlapping time window**, **permissible time window**, **time window**

followup period *n* - [**trials**] 1. The **period** of **time** from **enrollment** to termination of **followup** of a person or **treatment unit**. 2. The period of time in the course of a trial defined by the start of followup of the first observation unit **enrolled** and by the end of followup of the last unit enrolled; **followup stage**. rt: **baseline period**, **treatment period**, **treatment and followup period**, **close-out period**

ideal time window *n* - A **time interval** in a **permissible time window** within which an activity or procedure is ideally performed, eg, a 14 day interval centered at the **ideal visit time** within a permissible time window of 56 days similarly centered. rt: **permissible time window**

interim followup visit *n* - [**trials**] Any **followup visit** that takes place after **enrollment** of a person into a trial that is not part of the required sequence of **followup visits** and that is initiated because of some problem or concern; ordinarily not counted as a **required followup visit** unless it takes place within the specified time period for a required visit and all the required procedures for that visit are carried out as part of the interim visit; **nonrequired followup visit**. ant: **required followup visit**

observation variable *n* - A **variable** associated with the **observation unit** of a **study** and that is to be or has been observed at one or more points in time in the study, eg, gender as recorded upon **enrollment** or blood pressure as measured upon entry and at specified points in time during **followup**.

overlapping time window *n* - A **time window** overlapping the preceding or following time window. rt: **contiguous time window**, **disjoint time window**, **ideal time window**, **permissible time window**, **time window**

permissible time window *n* - The allowable **time interval** for performing a specified activity or procedure; usually centered at the **ideal visit time** and usually **contiguous** to adjacent time windows for **visit schedules**; **time window** in the absence of **ideal time window**. rt: **ideal time window**

repeated measure *adj* - Of or relating to a **design** or process involving **repeated measures**. ***Usage note***: See **repeated measures design**.

repeated measures design *n* - 1. A general class of **designs** having one or more **outcome measures** amenable to repeated measurement. 2. That subset of designs defined in defn 1 in which the **primary outcome measure** is amenable to repeated measurement. ***Usage note***: Note that defn 1 includes **crossover trials** and most trials having **parallel treatment designs** to the extent that the latter have at least one outcome measure amenable to repeated measurement. Defn 2 is limited to the set of designs in which the primary outcome measure (usually the **design variable**) is amenable to repeated measurement. Hence, in this sense, a trial having a parallel treatment design with mortality as the design variable is not a member of the class because the variable, death, is not amenable to repeated measurement. However, such trials, to the extent they involve other outcome measures amenable to repeated measurements (eg, a trial with mortality as the primary outcome measure and blood pressure change as a secondary outcome measure, as observed at various points over the course of treatment) are members of the class as defined by defn 1.

scheduled followup visit *n* - 1. **Required followup visit**; not an **interim followup visit**. syn: **required followup visit** ant: **unscheduled followup visit** 2. Any **visit** to a **study site** after **enrollment** that is **scheduled** (ie, as distinct from unscheduled, as with emergency visits), whether or not part of the sequence of followup visits specified in the **study protocol**. ***Usage note***: Usage in the sense of defn 2 is not recommended when there is a potential for confusion with use in the sense of defn 1, or when used interchangeably without distinction. Whether or not a visit is scheduled, in the sense of defn 2, in studies involving specified periods of followup, such as in most **clinical trials**, is of less importance than whether or not the **data** collected and procedures performed at the visit are sufficient to satisfy the **data collection** requirements for a specified **followup period**; if so, the visit may be considered part of the required set of visits, even if it was not scheduled in advance.

screening variable *n* - A **variable** used for **screening** (defn 2). rt: **baseline variable**

time window *n* - The **time interval** for performing a specified activity or procedure. In **trials** and other **followup studies**, usually the **window** for performing a specified **examination** or type

of **data collection**, such as for a **baseline** or **followup visit**. rt: **contiguous time window**, **disjoint time window**, **ideal time window**, **overlapping time window**, **permissible time window**, **time interval**, **time measure**, **time period**, **time point** ***Usage note***: See **time measure**.

unscheduled followup visit *n* - **interim followup visit**, also **nonrequired followup visit** ant: **scheduled followup visit**

P&P 1: Design data collection forms consistent with *Forms design good practice policies and procedures* (GPPP).

P&P 2: Establish data collection schedules consistent with the following principles:
- That persons are not randomized until baseline data and data establishing eligibility have been collected and recorded on study forms
- That the schedule for followup data collection be independent of treatment assignment and adherence to treatment
- That time windows for scheduled visits be constructed to be centered on the ideal time points for visits and that visits not made within the specified time windows be counted as missed

P&P 3: Strive for data collection schedules that are the same for all treatment groups, ie, independent of treatment assignment.

Comment

Ideally, the examination schedule should be the same for all study subjects. That must be in the case in masked trials, but may not be in unmasked trials, eg, as in trials involving different modes of treatment (eg, surgery vs medical treatment).

Schedules that differ by treatment group may produce artifactual differences. The adage that "the more you look the more you find" applies to observations in trial. One can expect that clinic personnel will observe and report more side effects and morbid events for persons seen more frequently (see, eg, morbidity results reported from the Hypertension Detection and Followup Program where persons assigned to Stepped Care were seen more frequently than persons assigned to Usual Care).

The "workaround" when dealing with differential schedules is censoring. Censoring may reduce artifactual differences, but is not likely to eliminate them (because observations made at visits are not independent of previous observations).

P&P 4: Divide the clinic visit/examination schedule into the following periods:
- Baseline
- Treatment
- Followup
- Close-out

Comment

The end of the baseline period, usually marked by issue of the treatment assignment, also marks the start of the treatment and followup periods. The treatment period ends with the last application of treatment. The followup period ends with the last contact for data collection. In many cases, the endpoint for the two periods is the same – the case in most trials involving treatment of chronic conditions, eg, as in the CDP. The two periods are different in trials where treatment is applied over a short period of time and where followup extends over a longer period of time.

The factors to be considered when determining the schedule for the different periods are as outlined below.

Baseline period

- Time and number of visits needed for diagnosis and determining eligibility
- Number of procedures needed to establish diagnosis, eligibility, and baseline for followup
- Number of visits required to obtain necessary baseline data
- Time needed to solicit consent and for study candidate to consider options
- Time needed to issue treatment assignment and to inform study subject of assignment

Treatment period

- Period of treatment
- Place of treatment: out-patient or in-patient
- Need for documentation of the fact of administration (eg, need for observed administration by clinic personnel)
- Need for treatment adjustments (eg, dosage change depending on observed effect)
- Likelihood of treatment-related side effects
- Need for treatment "touch-up" or reapplication of treatment
- Need for measuring compliance

Followup period

- Underlying event rate
- Likely rate of change for measures used to assess outcome
- Number of different procedures that can be reasonably performed at any given visit
- Maximal allowable separation between visits consistent with care requirements

Close-out period

- Time required to safely withdraw treatment
- Time needed to collect desired data on persons prior to separation
- Time needed for deconsent and for orderly transfer of care (when indicated)

P&P 5: Choose a time unit (eg, hour, day, week, month, or year) for constructing the data collection schedule; use the same time unit for the baseline and followup periods of observation.

Comment

The most common time unit for long-term trials is month. The difficulty with month is that months vary in length. The variation can be problematic when constructing the visit schedule and time windows for visits; use week to avoid the problem.

P&P 6: Establish a date format convention and maintain across forms and throughout the trial.

Comment

The 6 digit numeric format of the form __ __/__ __/__ __ or __ __-__ __-__ __ is least desirable because of the potential for confusion (month/day/year vs day/moth/year). The format should not be used without subscripts to indicate fields for day, month, and year. The recommended format is day (2 digit, numeric), month (3 alphabetic characters), year (2 or 4 digit, numeric).

P&P 7: Establish convention for recording time of day; maintain consistency across forms and throughout the trial.

Comment

The options are 24-hour or 12-hour clocks. The 24-hour clock has the advantage of avoiding the confusion of am and pm, but is difficult for people to understand. The 12-hour clock is preferred in most settings.

P&P 8: Specify extant units of measurements for continuous variables; opt in favor of units in use in participating clinics.

P&P 9: Measurements should be recorded in the unit obtained.

Comment

Data are not improved by conversion to other unit systems, even when the conversion amounts to nothing more complicated than multiplying or dividing by 100. If scales for measuring weight and height are calibrated in pounds and inches, weights and heights should be recorded in pounds and inches. Conversion to kilograms and centimeters, if required for publication, should when the manuscript is readied for submission.

P&P 10: Ideally, the unit of measurement for a variable should be invariant across forms and over time in the trial.

P&P 11: In multicenter trials, choose the unit of measurement common to the majority of clinics if units vary across clinics; in such cases establish data forms to allow for variation in observation units (eg, by having the unit of measurement as a variable to be recorded or by having different versions of forms, depending on clinic).

P&P 12: Calculations and analyses should be carried out using data as reported with as few conversions as possible. If values are to be reported in international scientific units in manuscripts (as required in journals adherent to the international scientific vocabulary), those conversions should be done when a manuscript is submitted for publication or after it has been accepted for publication. The manuscript should include a footnote indicating that the conversion was done using finished results based on a conventional unit system.

P&P 13: List observation variables; maintain and update over the course of the trial; for each variable list visits or contacts at which observed, unit of measurement (when applicable), state requirement(s) (when applicable, eg, no food or drink 12 hours prior to test), and intended use (eg, stratification, assessing eligibility, followup).

P&P 14: Establish permissible and ideal time windows for baseline and followup visits.

Comment

The schedules for data collection are idealized. The actual schedules will vary around the idealized schedule, hence, designers have to set allowable limits for variation. That limits are necessary is obvious. Personnel at clinics will need rules for deciding when they can start seeing patients in relation to a specified visit. People responsible for performance monitoring or for analysis of study data will need rules for counting visits as completed or missed.

The usual approach to dealing with the variation is to construct time windows about the idealized time points. The normal method of construction for followup visits is to center the window on the idealized time of the visit and to make them contiguous. For example, if followup visits are to be at 4 week intervals following randomization, then the maximum allowable intervals are 4 weeks, centered at 4 weeks, 8, weeks, etc. The window for the first visit opens at the start of week 3 and closes at the end of week 6. The window for the 2nd followup visit opens at the start of week 7 and closes at the end of week 10, etc.

The advantage of having contiguous time windows is that the next window opens when the preceding one closes. Use of contiguous windows avoids "dark periods" where data are not admissible.

A difficulty with the contiguous method of construction is that there is no minimum for separation of visits. Technically, therefore, it is possible to perform followup visits for two time periods on adjoining days. The solution, if such possibilities are to be precluded, is to impose a minimum time separation. The minimum imposed has the effect of causing a window to remain closed until a specified time has lapsed following completion of the visit in the preceding time window.

P&P 15: In constructing the visit schedule for baseline periods (including the randomization visit), construct windows to indicate permissible intervals; specify allowable overlap and minimum separation for adjoining visits; specify maximum permissible time for the entire baseline period.

Comment
Minimum separations are required when procedures or visits are to be performed on different dates and where they must be separated in time.

A time window for the entire set of baseline visits is needed to ensure that data used for establishing baselines and for assessing eligibility are collected within a reasonably short period of time prior to randomization. The usual practice, when the window is exceeded, is to deny randomization until or unless the patient is "recycled" through a new period of observation and evaluation.

P&P 16: Except for observation variables that are invariant (eg, sex), construct the followup data collection schedule to provide for repeat observation of variables in assessing treatment effects.
Comment
Observation variables, checked as being needed to establish baselines (P&P 13) should be slated for observation at indicated followup visits.

P&P 17: Except for trials done in emergency settings, construct the visit schedule for the baseline period of observation to involve at least two visits, separated in time by 24-hours or more.
Comment
The time and separation is useful in ensuring quality informed consent.

P&P 18: In regard to a variable that is used to screen for eligibility and that is also to be tracked over time, make certain that the baseline value for the variable is independent of screening; ie, do not use the screening value for baseline.

P&P 19: List procedures and tests to be performed; maintain and update over the course of the trial; list in tabular fashion with columns indicating procedure, visits at which performed, place preformed, time of day preformed, person(s) preforming the procedure, and state of patient when presenting for the procedure (eg, overnight fast).

P&P 20: Separate and schedule for different visits, procedures that are incompatible with being performed on the same visit, that together lead to unreasonably long visits, or that, when done together, are considered to be too stressful or fatiguing for patients.

P&P 21: Arrange procedures to be conducted during a visit in an order that minimizes interference and in descending order of importance in the trial.
Comment
Easier said than done because "interference" may be subtle and not evident until well into the trial; sometimes only after a review of distributions by visits.

P&P 22: Establish a hierarchy of procedures so that, in cases in which, patients, are not willing to submit to all procedures, those most important to the trial are done before those of less importance.

P&P 23: Prior to implementation, perform a walk-through of the data collection procedures.

P&P 24: Design the data collection system to distinguish between scheduled and unscheduled (interim) visits.

Comment

At a minimum, it is necessary to have a data collection system capable of providing counts of interim visits because of the adage noted in P&P 2. Differential interim visit rates can lead to artifactual differences among the treatment groups.

P&P 25: In trials where different modes of data collection can be used to satisfy data collection requirements (eg, use of home visits in lieu of clinic visits), design so as to be able to identify the mode of data collection.

Comment

A good many observations in trial are setting dependent. Mixing observations taken in different settings usually increases variation. For example, blood pressures taken at home are generally different from those taken in clinics. Designing the data collection procedures and data system to include codes to indicate where the visit was done allows analysts to restrict analyses, when desired, to measurements made in a particular setting.

P&P 26: When designing data collection schedules, distinguish between visits and data needed to care for a patient and those needed for the trial; limit the contact schedule and data collection to that needed for the trial.