



Center for Clinical Trials

*Department of Biostatistics  
Department of Epidemiology  
Department of International Health*

*Department of Medicine  
Department of Ophthalmology  
Oncology Center*

Wednesday, 16 November 2005

**Memorandum**

**To:** Center for Clinical Trials Students, Staff, and Faculty

**Fr:** Curtis Meinert

**Re:** Tables 101: Data cutoffs and freezes

I wrote previously on cross checking. You cannot produce a coherent set of tables without a frozen dataset. Cross checking is futile if the database changes over the course of production.

A term best deleted from one's vocabulary is "final" in regard to datasets. Largely, at least in trials, there is no such thing. "Final", even when a trial is finished and "mothballed" is more a declaration of hope than of reality. The data collection and keying processes are error prone. The likelihood of finding errors and the need to change things in the dataset pertains as long as data are being analyzed, possibly years after the trial has finished.

Basically, to freeze the dataset, one has to declare a cutoff date beyond which additions or changes are barred. The more "final" the dataset, the harder it is to live with such cutoffs. Living with arbitrary "cutoffs" is relatively easy for interim looks for the purpose of treatment effects monitoring but becomes increasingly difficult as one approaches immortalized datasets created at the "end" of trials.

Typically, the frozen dataset created during the trial includes "dirty" data, ie, data with outstanding edits. Clearly, finished datasets, prepared at the end of the trial, should be with edits resolved and, hence, free of "dirty" data.

A dilemma faced in "cut offs" is what to do when errors are discovered. They are easily ignored when trivial, but increasingly difficult to ignore when of possible consequence in analyses. For example, what does one do when a person, counted as alive in the analysis, is discovered to have died? The temptation is to change the dataset to reflect the change but those changes, unless tracked back to the basic underlying dataset of the trial, will not be permanent.

Basically, the best policy, at least for interim analyses, is to avoid such changes unless one is willing to re-freeze and regenerate all tables in a report. The notion that the effect of a change effects only "one table" is usually errant. Mixing tables with different databases, even when involving only "trivial" changes, will almost always lead to tables that do not "cross check".

---

The tendency, too often, is to freeze unreasonably close to the report due date. The tendency is born of naiveté or a "macho" attitude in the coordinating center to demonstrate its analytic prowess. Save a nickel spend a dollar!

Another temptation to be resisted is providing up-to-date counts and analyses for key outcome measures. Nine chances out of ten the difference in denominators for the "updated" counts versus those for the other tables in the report will lead to confusion.

---

(Sat 6:52am) 23 Jul 05

VTables.101\Freeze.WPD

### Distribution

Debbie Amend-Libercci	Rosetta Jackson	Dave Shade
Ming-Wen An	Jennifer Jones	Anne Shanklin
Jeannette Beasley	Charlene Levine	Charles Shiflett
Pat Belt	Tianjing Li	Jackie Smith
Elena Blasco-Colemanares	Simon Liu	Michael Smith
Cathy Bosley	Hope Livingston	Paul Smith
Rob Casper	Nancy Maldeis	Alice Sternberg
Hui-Ming Chung	Barbara Martin	Jennifer Thorne
Betty Collison	Reena Masih	Andrea Tibbs
Ryan Colvin	Curtis Meinert	Ada Tieman
Christine Costantino	Jill Meinert	James Tonascia
Kay Dickersin	Vinnette Morrison	Susan Tonascia
John Dodge	Wai Ping Ng	Aynur Ünalp-Arida
Michele Donithan	Deborah Nowakowski	Mark Van Natta
Lea Drye	Kapreena Owens	Swaroop Vedula
Ann Ervin	Bonnie Piantadosi	Margie Wild
Ingrid Friberg	Steven Piantadosi	Laura Wilson
Julia Gage	Nancy Prusakowski	Robert Wise
Judy Harle	Chrissy Riggs	Claudine Woo Shinoff
Janet Holbrook	Linda Roberts	Kathy Yates
Rosemary Hollick	Karen Robinson	Tables Notebook
Milana Isaacson	Bobbi Scherer	Chronologic file