



Center for Clinical Trials

Department of Biostatistics
Department of Epidemiology
Department of International Health

Department of Medicine
Department of Ophthalmology
Oncology Center

(Mon) 25 July 2011

Memorandum

To: Trialists

Fr: Curt Meinert

Re: Estimating costs for public use data deposits

A little help please: e-mail me your guesstimate of coordinating center dollar cost associated with preparing datasets for deposit under the NIH data sharing policy. To be useful your guesstimate should be independent of others, so please do your guesstimating on your own. Horseback estimates are OK or you can do your guesstimating by guessing at the number of person hours involved in the different aspects of deposit.

In theory, CC cost is a function of the nature and size of the dataset, but most of the effort involved is largely the same regardless of size. In making your guesstimate assume the data deposit involves:

- Staff mtgs to devise the deposit plan and to identify variables to be re-coded, redacted, and collapsed
- Programming to re-code, redact, and collapse and to create the de-identified dataset
- Verification of the dataset to ensure proper coding, stripping, and collapsing
- Preparation of a data dictionary and accompanying documentation
- Checking accuracy of the data dictionary and accompanying documentation
- Testing the data file for robustness of user instructions
- Challenges to de-identification (to see if persons can be re-identified)
- Preparation and submission of IRB submission detailing nature and extent of data sharing to CC IRB
- Preparation of memo to heads of data collection and generation sites informing them of intention to deposit de-identified data and instruction to inform local IRBs of impending deposit
- Collection of local IRBs approvals of data deposit (essential when consents do not include mention of intent to data share; see *Story of the fellow who did*; previously distributed)
- Revision of the dataset based on interaction with CC IRB and data collection and generation sites and their respective IRBs, including stripping data from sites where deposit is not approved by local IRBs
- Interaction with funding agency and repository custodian regarding deposit
- Transmission of dataset to repository

To give uniformity to the guesstimates assume the dollar cost per hour of effort is \$132, derived as below.

Median FTE salary of persons involved in preparing dataset	\$100,000
Employee fringe benefit cost (35% of salary)	\$35,000
Supplies and equipment costs per FTE (10% of salary)	\$10,000
Total direct cost (TDC)	\$145,000

Estimating costs for public use data deposits
(Mon) 25 July 2011

F&A cost (62% of TDC).	\$89,900
Total cost per FTE of effort.	\$234,900
Work hours per yr (52wks x 40hrs).	2,080 hrs
Less vacations (3wks x 40 hrs).	-120 hrs
Less holidays (12 days).	-96 hrs
Less sick days and family leaves (5 days).	-40 hrs
Less time for professional development (5 days).	-40 hrs
Net work hours per yr.	1,784 hrs
Hourly rate (\$234,900/1,784 hrs).	\$132.00
