JOHNS HOPKINS
UNIVERSITY

Center for Clinical Trials

Department of Biostatistics
Department of Epidemiology
Department of International Health

Department of Medicine
Department of Ophthalmology
Oncology Center

25 June 2012

Memorandum

To: Trialists

Fr: Curtis Meinert

Re: The story of the duplicate LSOCA patient

LSOCA (Longitudinal Study of Ocular Complications of AIDS) is a long-term multicenter cohort followup study of persons with AIDS and ocular disease. The study is funded by the National Eye Institute and has been ongoing since 1998. Presently, LSOCA has 13 study clinics, two in New York city, Memorial Sloan-Kettering (MSK) and New York University (NYU).

In December 2011, the MSK coordinator, in discussion with the person's primary care physician, learned that the person was also enrolled at the NYU clinic. The coordinator alerted the NYU coordinator and us to the problem. Checks indicated that the person was enrolled on 15 September 2008 at the NYU clinic and on 24 October 2008 at the MSK clinic.

We informed our IRB of the problem. It requested that we "*revise study procedures in the research plan to require that the two NYC clinics confer with each other if presented with a duplicate Id*". I informed the IRB chair that Id numbers are, by definition, unique and that we do not have the means of preventing duplicate enrollments since we do not receive personal identifiers. I added that one of the reasons for our not having such information was because of concerns of the IRB regarding risk of breaches of confidentiality.

The enrollment raised questions as detailed below.

Question: Which coordinator should speak to the patient to determine the clinic from which to be withdrawn?
Resolution: After consultation with the LSOCA Policy and Data Monitoring Board and the two clinic coordinators, the decision was for the coordinator at the clinic next scheduled to see the patient to pose the question – the MSK coordinator. The patient elected to remain in the MSK clinic and withdraw from the NYU clinic.

Question: What IRBs are informed?
Resolution: The IRB of the coordinating center and those of NYU and MSK.

Question: Which data and banked specimens does the coordinating center purge?
Resolution: Data and specimens collected at the NYU clinic.

Question: How many publications include data on the duplicate enrollment?
Resolution: Four.

Question: Did the data affect published results and conclusions?
Resolution: No.

Question: Should the editors of the published papers be informed of the problem?
Resolution: Yes (being done).

Observations

The duplicate enrollment was discovered by chance. That reality raised questions as to whether we, at the coordinating center, should have had systems in place to reduce the likelihood of such occurrence. One obvious check would have been inclusion of a question on the enrollment form to screen for duplicate enrollments even if of questionable utility; utility depends on persons answering honestly and knowing the study is the same as the one already in.

We have no idea why the person opted for duplicate enrollment. While it is true that persons enrolled receive payments for visits (about $50), it is uncertain if money was a motive for duplicate enrollment. It is not even clear the person knew the study being considered was the same as the one already in.

If a coordinating center collects personal identifiers it, at least in theory, has means of checking for duplicates. That option was precluded in LSOCA because the coordinating center does not receive such information. But even if we did, use of that information is not likely to be particularly useful because names and addresses change and people are not required to disclose their social security numbers.

Bottom line: There is no guarantee against duplicate enrollments.

A lesson, known to anyone who has tried to purge bad data, is that it is well nigh onto impossible to get bad data out of datasets once incorporated. We have spent considerable time identifying errant datasets and still searching.

There were several datasets created for manuscripts with cutoff dates of 24 October 2008 or later – the date of the duplicate enrollment. We identified four publications including the duplicate data. Other datasets created after 24 October 2008 for publications were purged of the duplicate data before manuscripts were finished.

We are still tracking specimens to determine if shipments included specimens for the duplicate patient.

One side benefit is that the event has provided an opportunity to compare data collected on the same person at two clinics. The data are remarkably concordant.

Last. One might say the effort to identify datasets containing the duplicate enrollment was unnecessary because a single duplicate would not affect published results. But editors do not have senses of humor about errors known to authors not revealed to them. In any case, coordinating centers fear the slippery road embarked upon by ignoring errors, even if inconsequential.