



**Biostatistics Core Meeting
September 25, Lusaka, Zambia**

MINUTES: Biostatistics Breakout Meeting

This document mainly serves to remind us of our topics of discussion, more than to document decisions made.

Baobab Room

In Attendance: Corbett, Crawford, Fielding, Golub, Hayes, Lawrence, Lewis, Moulton, Pacheco, Pan, Russell, Samandari, Schaap, Sismanidis; later, joined by most of THRio.

DATA MANAGEMENT

Trish Crawford brought up an important but difficult topic—how to decide how much data cleaning is enough. With millions of data items, there are thousands of queries, which can take some time to resolve. There are several levels of criticality. For example, data on process measures and data on data are not as critical. TB and safety data are most important. Post-coding, usually of open fields or fields with multiple possible responses, should be done centrally, not by site/cluster.

Kathy Lawrence explained how MS SQL Server is sufficient for handling their size of data system and needs.

Antonio Pacheco discussed cause-of-death algorithms, and how CoDe from Denmark has been adapted—puts all sources of data into standardized form, which is then reviewed by committee and assigned levels of certainty (definite, probable, possible).

With causes, can do competing risks analyses. Competing with AIDS are non-AIDS, and Unknown. Would be good to get mortality of patients on IPT, especially TB-related mortality.

In Thibela TB, use Human Resources data; basic, pretty much only know if it was from a mine accident or not. Can get data on those on IPT, as they have consented; but that of course is only in the intervention arm.

May be useful to check HR mortality data to see if additional resources are needed to investigate differentials in mortality (if they exist).

Mortality reporting may vary across mine companies, but the companies are fairly well distributed across the two trial arms.

Ab Schaap:

1. He is “grumpy” about TB registers. How will they do quality control? Can track data entry. All TB patients should have an outcome—death, success, failure of treatment.
2. Linking of data sets—TB registers. Household intervention—if new patient, goes thru process. Link register with household intervention. Use TB Register ID, names, gender, age, maybe date of notification/intervention.

3. TB register data—this year, some sites in Capetown have new format of TB register, there are three versions now. Sometimes don't know if treatment has been completed; coding of pulmonary TB, extrapulmonary etc. has changed.

In Zambia, registers introduced in 2006 are much better. ECF clusters—decreased notification, but expected greater notifications. But perhaps denominators are different across ECF status of arms. ECF is smallest, so may make sense, 10% smaller.

4. Sharing data between SA, ZA, and UK. Platform for sharing of data? Need some server, can give access over internet, for sharing these.

JOINT WITH MODELLING BREAKOUT: ORGANIZATIONAL CONSIDERATIONS—Liz Corbett

Peter Dodd has been at Imperial since August. Mainly will be working on case-finding—ZAMSTAR and DetectTB.

Will be looking at TST, ELISPOT, HIV data from our studies, looking at existing models in the literature. Will need household composition, and distribution of Mtb and HIV.

Brian Williams will retire from WHO and come to Imperial this fall, will help orient Peter regarding TB modeling.

There will be a pre-start-up meeting in November in London.

Emilia Vynnycky will focus on Thibela TB. She's at the Health Protection Agency. Funding is available for a postdoc for 20 months to work on modelling the Bethel and Thibela data. The postdoc will be based at HPA

The Steering Committee is quite big: most PIs and data people from Thibela TB and ZAMSTAR on it. And, there will be a Technical Management group mainly of modelers.

One question is how to deal with requests for extra work?

E.g. a request from Lois regarding how the recruitment rate might affect the intervention effect? ZAMSTAR to provide baseline prevalence data, anonymised HIV and TB data. Babis will transfer these data.

Thibela—there will be periodic transfer of data, at shaft level. Mixing data collected at baseline.

Other data—TST/IGRA/HIV

Prevalence of Mtb by site

GIS data

Household data

A Draft Memorandum of Understanding was circulated—internal, so people will know their communications, confidentiality, publication, etc.

DETERMINISTIC MODELING OF TB/HIV IN RIO (JOINT WITH MODELLING BREAKOUT)

Antonio Pacheco presented some possible approaches to modeling the Rio epidemic, augmented with THRio data. Wants to look at interactions between HIV+ and HIV-. Liz Corbett mentioned looking at the interaction between IPT and infection control among patients. Is the clinic the source of much TB infection?

MDRTB AND PREVIOUS TB—James Lewis

James commented on a paper by Andy Hall on risk factors for MDRTB in Europe. There is an odds ratio of 10 relating prior TB to MDRTB; mainly people failing therapy. But James notes that 1/3 with MDR have no prior TB history, there is a lot more going on.

ANALYSIS PLAN FOR THRIO, REVISITED—Larry Moulton

The addition of an extra year of follow up means revisiting the analysis, may well need to go “horizontal” instead of “vertical.” Decision will depend on simulation of data and applying the different methods.

Need to change the primary endpoint from pure ITT to a restriction to those “eligible,” will be more powerful and also more applicable to other situations.

PLANS FOR GIS DATA COLLECTION AND ANALYSIS

Bill Pan described how the GIS data are collected, the use of enumeration areas, and testing of the polygon method. The plan is to hire some Zambian students to obtain the data and create maps, in 6-9 months. Data from the 2000 census, clinics and schools will also be put into the data base.

There was a question about the need to map 100% of the enumeration areas, or perhaps just 20-30% of them, but it appears that for very few additional resources all of them can be covered, so may as well do so.

Mapping will facilitate the final prevalence survey that requires taking samples of 5000 from the “yolks” of the intervention areas. Need some backup areas in case don’t get the full 5000. Mapping will include more landmarks, etc.

DSMB REPORTING

Want to revise THRIO analysis plan to give to them.

Thibela TB may expand their analysis plan. May change case definitions, depending on availability of lab data.

ZAMSTAR--analysis plan needs to get fleshed out, go into details.
Otherwise, more of the same, with whatever manuscripts in progress there are.

OTHER ISSUES

PREVALENCE SURVEYS

For ZAMSTAR, there was some discussion of the cluster sampling methodology—clusters in the clusters. An ICC of 0.0014 has been calculated. There are about 500 per each of 10 enumeration areas. Gives deff of 4? Ab will go to London to sit down with Babis and Richard and Katherine to work out the cluster/cluster sample size issues, perhaps in November.

For Thibela TB and the 750 sampling:

This size represents 7-44% of each cluster size, most under 20%. The question is how much this will affect long-term dynamics.

One of several alternatives: Taking 19% of each increases total sample by 40%, keeps same power.

MID-YEAR MEETING

Perhaps in Centurion/Johannesburg in March 2009.

Agenda:

Data sharing/ public data repository/ reproducible epidemiology

Other items?

ADJOURNED TO THE CHIT-CHAT CAFÉ FOR A RELAXING DINNER