Discussion based on Cluster Meeting 6/24/2005  
Present: Ed Stanek, Wenjun Li, George Reed

Note taker: Ed Stanek

1. Silvina, Julio, and Ed’s Simulation Paper. 
   Ed described the work that Julio and Silvina have been doing, and where he thought it stood. In this regard, Ed has worked on it some more- revising the programs to include two different estimators for the SS and RP model (based on estimated shrinkage constants). He also create a 'mega' program that will run the 504 simulations for case 1 (N=10, M=5 with equal within cluster variances), and create an output data set with the results. The program names are listed below, including the mega program, resulting data set, and two programs that produce plots (one for theoretical plots, and one for empirical plots).

His first impression of this is that although there are some patterns, it is complicated to explain.

ceds05pred_equal23.sas (mega program for N=10, M=5 Normal Normal that creates summary of simulations)
ceds05pred_equal24.sas (corrects some output to create data set for plotting summary)
ceds05pred_equal25.sas (creates plots comparing theoretical MSE (from 504 simulations)
r10_snnv2.sas7bdat Data set with results from simulations
ceds05pred_equal26.sas (creates plots comparing empirical MSE)

After doing this, Ed looked at how the shrinkage constants relate between the different predictors. One setting was selected, and scatter plots were constructed of the shrinkage constants. This appeared interesting, but somewhat overwhelming since there are so many settings. One ad hoc idea is to use the comparison to compare the best predictor with the RP predictor, and then see where the shrinkage constants differ. It may be possible to alter the shrinkage constant so that the RP predictor is better. This may be future research rather than needed work for Silvina’s manuscript.

Ed is not sure what to say about the simulations now. He feels like we may want to complete the paper similar to the way it was written, and expand the discussion to highlight the additional work needed. Julio and Silvina's thoughts on this are appreciated. Ed would only like to be a contributor in this respect, and not the lead for the paper- so Julio and Silvina can call the shots here. We can use where we end up as a starting point for future research.

2. Unbalanced Cluster Sampling
   Ed need to complete the unbalanced paper, and so doesn't want to focus on anything else (apart from participating in discussions of other papers).

3. Covariate Paper
   Wenjun has a draft paper (which Ed’s reading and making comments) from his dissertation. This is on adding a covariable in the model (like 'farm size’ in a sampling context investigating
total yield). His objective is to get the paper submitted, and Ed thinks this should not be too long a process. After this paper, the next agenda for Wenjun is to write an application paper related to standardization in Epi. Then, to expand this to a 2 stage sample setting. (See unpublished results.)

4. Multi-dimensional Graphical Analysis
George was reading Wenjun's paper, and otherwise has not committed himself to an area. He is interested in exploring other graphical methods for summarizing Silvina’s simulation results (for an exploratory data analysis kind of paper). Julio and Silvina have already done quite a bit on this, but it is a challenging multi-dimensional problem. Ed thought this would be a place for George to start. It is clear that the simulation paper is with Silvina, Julio, and Ed (so that George looking at it won't slow the progress down, but may provide some insights).

5. Additional Areas to Pursue

There are several areas to pursue. Ed is happy to assign topics to students. In this way, more can be accomplished. On the other hand, Ed is concerned over "saving topics" for students if such a step slows down the progress of the research. Timely progress is needed. If progress isn’t at a sufficient pace (based on Ed and Julio’s assessment), others may step in and do the work. We need a clear understanding of this with students.

1. Developing predictors with 3-stage sampling (at first simply assuming we have multiple measures (representing response error) on sampled units). This can be done with equal numbers of measures on units, and with unequal numbers of measures. This should be done first, since it is straightforward.

2. With two measures on a unit corresponding to two times. This is in some ways a trivial problem, but in other ways, an important building block for additional methods. There can be many variations of the population setting: each subject measured at age 8 and 10 with no response error; each subject measured at two known (in the population) times with no response error; each subject measured at age 8 and 10, but response observed with error; each subject measured at two known (in the population) times with response observed with error.

3. Extend 2 to a setting where there are 3 possible times per subject, there is lack of fit for a linear model, and a sample of 2 measures are to be selected.

…