Ideas for Research with Recai

1. Start with a simple two stage problem:

   2 stage sampling
   Equal size clusters
   Equal size sample per cluster
   Response error equal for all units

   \( Y_{stk} = \mu_{stk} + W_{stk} \)
   \( = \mu + \beta_s + \varepsilon_{stk} + W_{stk} \)
   Assume \( \text{var}_i (E_{stk}) = \sigma_{R}^2 \)

   Representation of random variables:
   \( Y_{ijk} = \mu + B_i + E_{ij} + W_{ijk} \)

   Representation of Sample random variables:

   \[
   \begin{align*}
   \text{PSU}(i) & \quad \text{SSU}(j) & \text{Potential Obs} & \text{Actually Obs (0/1)} \\
   1 & j = 1, \ldots, m & \begin{pmatrix} Y_{11k} \\ Y_{12k} \\ \vdots \\ Y_{1mk} \end{pmatrix} & \begin{pmatrix} R_{11k} \\ R_{12k} \\ \vdots \\ R_{1mk} \end{pmatrix} \\
   2 & j = 1, \ldots, m & \begin{pmatrix} Y_{21k} \\ Y_{22k} \\ \vdots \\ Y_{2mk} \end{pmatrix} & \begin{pmatrix} R_{21k} \\ R_{22k} \\ \vdots \\ R_{2mk} \end{pmatrix} \\
   \vdots & & & \vdots \\
   n & j = 1, \ldots, m & \begin{pmatrix} Y_{n1k} \\ Y_{n2k} \\ \vdots \\ Y_{nmk} \end{pmatrix} & \begin{pmatrix} R_{n1k} \\ R_{n2k} \\ \vdots \\ R_{nmk} \end{pmatrix}
   \end{align*}
   \]

2. Start with simpler single stage problem
Traditional Survey sampling methods use inclusion probabilities, $\pi_i$, rather than the random variables $Y_i$.

**Imputation with One variable.**

Suppose we observe household medical care costs per person and gross household income (at two levels, high vs low) in a simple random sample of households. Also, suppose that response for the income question is missing 50% of the time given a household has high income, and 10% of the time given a person has low income. Other variables, such as household members may be available on all subjects. There is interest in analyzing these data, and there may be a variety of analysis questions. How should it be done?

One strategy is multiple imputation. Can we identify how this will occur under a model for missing at random.