Longitudinal designs often change at critical times based on available funding, staffing, scientific opportunities, and subjects. This paper presents three levels of investigation into missingness by design in a partially completed longitudinal study: missingness that is completely at random (MCAR), at random (MAR), and nonignorable (MN). We first derive new expressions for the asymptotic variance and power based on multivariate normal data which are either MCAR or are missing by design (MAR). These formulae allow for any and all patterns of missing data to occur. The special case of a two stage longitudinal design is described in detail. We then present a general design and analytical strategy for protecting against nonignorably missing data midway into a longitudinal study. The new design involves stratified sampling for follow-up based on the pattern of missing data already obtained, and the corresponding estimator is based on an approximate likelihood. The methodology for MCAR, MAR, and MN are in turn applied to the design of a follow-up study to examine the effect of lifetime lead exposure on neuropsychological measures of attention. Our conclusion in this example is that a design exists that has sufficiently high statistical power to address the main scientific questions and also provides protection against a broad class of nonignorably missing data.

KEY WORDS: Design for follow-up; Missing data pattern; Missing data mechanism; selection model; pattern mixture model; nonignorably missing data.
1. INTRODUCTION

Missing data are the rule rather than the exception in longitudinal studies. Most of the time we consider missing data as a nuisance which necessitates special handling (Rubin, 1987; Little and Rubin, 1987; Schafer 1997). However, our focus in this paper is primarily on data missing by design, sometimes called planned missingness. A study design can select which cases will be incomplete in order to reduce the cost of a study. If successful, this planned missingness will have little effect on the precision of the parameter estimates of most concern. The most common example of such planned missingness is two-stage designs. The first stage measure, which is given to everyone, consists of an inexpensive instrument while the second stage measure, being more expensive to administer, is given to a subset of the original sample. This second stage sample can be a random subset of the entire sample or a stratified random sample with selection probabilities dependent on the first stage measure. The correlation between the two measures and selection criteria for the second stage sample provide the information needed to make full-sample inferences about the distribution of the second stage measure.

When data are randomly sampled for additional follow-up, the data are said to be missing completely at random, MCAR (Rubin, 1976). When the probability of a subject being followed up depends on observed variables, data are said to be missing at random (MAR). Missing nonignorably (MN) occurs when the probability of being missing depends on values which are missing. Several general methodologies, most notable the EM algorithm (Dempster, Laird, and Rubin 1977) and multiple imputations (Rubin and Schenker, 1986; Rubin, 1987; Schafer, 1997) exist for obtaining estimates from incomplete datasets under a MAR assumption. Despite the general methodology available for MAR estimation, confidence intervals, and testing, less is known about the statistical power for alternative design decisions which are under the control of the researchers. Most of this literature on power exists in the survey sampling literature (Rubin, 1987) or with binary outcome variables (Shrout and Newman, 1989). This paper presents a methodology and provides new formulae for asymptotic variances and statistical power for different missing by design choices.

Multistage designs with complete follow-up of selected cases lead to datasets which are MAR. In two-stage designs, one uses stratified sampling to select subjects for additional assessment on the more expensive second stage variables based on their scores on first stage variables. In two-stage designs we observe two patterns in our data, one with a complete set of observations and the other with first stage measures only. Multistage designs continue this strategy of using different probabilities of selection based on the outcomes of previous stage measures. There are many varieties of this general planned missingness strategy. Fuller (1987) discussed validation studies where the inexpensive measure, called a surrogate, is used to provide missing information on a more valid measure. Barron (1977) provided the basic

Two-stage designs, in which the response on the first level assessment determines the probability of assessment on the second, are often used in psychiatric epidemiology studies (Hay and Fairburn, 1998; Kisely and Goldberg, 1997; Lavigne et al., 1996). A variant of a two-stage design is the deliberate oversampling of hard to reach populations as was used in the National Comorbidity Survey (Kessler et al., 1995); Tarter and Hagedus (1991) used a multi-stage design for drug screening.

In medicine, multiple tests are sometimes obtained to decide whether a biopsy or other definitive but invasive test should be performed to obtain a diagnosis. Baker (1995) discusses the use of three assessment methods to aid in the determination of prostate cancer.

It is even possible to plan for missingness on both independent and dependent variables simultaneously (Tosteson and Ware, 1990). Buxton et al. (1991) use a theoretical mediating variable to determine the probability of being missing on both predictor and response variables.

Researchers generally understand the value of such designs with planned missingness. Nevertheless, few methods exist to help researchers decide how many subjects with each pattern of missing data are needed and what the stratification criteria should be. Complex design decisions are often based on guesses. Graham, Hofer, and MacKinnon (1996), for example, succeeded in designing a preventive trial with only one-third of the sample measured on a key drug resistance measure. Moke (1993) and Brown and Liao (1998) describe an unsuccessful two-stage design where an unanticipated large false negative rate resulted in imprecise estimates of a conduct disorder diagnosis. This paper responds to this void by providing asymptotic variance formulae and power calculations for a large class of problems where data are missing by design.

The example described in this paper refers to designing the follow-up of a prospective longitudinal study midway into the study. Midway into a study is not ordinarily the optimal time to design the follow-up of future measures, but there are many circumstances where it is impossible to begin with a complete design for follow-up or where the existing design must be changed. For example, NIH funded grants are nearly always limited to five year funding intervals, so longitudinal studies which span decades (Crum et al., in press) must continually update their follow-up designs. Secondly, budget limitations and other logistical considerations often require major design revisions. Under such circumstances the power calculations in this paper are most useful.
The example we describe in this paper involves an 11 year follow-up of students from second grade (age 8) to age 19. At the time of redesign, data have already been collected at ages 8, 10, and 13. A major aim of this follow-up study is to correlate lifetime lead exposure with attention measures taken throughout the child’s school years. Lifetime lead exposure ($Pb$) is measured by a special type of X-ray of the bone. In this project, it will be measured on a subset of the sample when they reach age 19. Because it is an expensive measure to obtain, it cannot be obtained on every subject. Our design problem is to select subjects for follow-up based on their observed data and current missing data pattern.

Attention is measured with a neuropsychological laboratory test called the continuous performance test (CPT). This was measured in grade 2 when the child was about 8 years old ($CPT_8$), at age 10 ($CPT_{10}$), and 13 ($CPT_{13}$), and also will be measured at follow-up at age 19 ($CPT_{19}$). This variable is also costly to obtain, and there are various levels of incompleteness across time. Among a total of 1196 subjects, 435 were randomly selected in second grade to receive $CPT_8$. Less money was available in later years; consequently only 216 of the 435 subjects were assessed at both ages 10 and 13. Standard tests on means revealed that there were no significant differences, nor even substantively meaningful differences, on means of observed variables over the different patterns of missing data. This suggests but does not confirm that a MCAR/MAR assumption may not be too far from the truth. Our handling of missing data mechanisms in this paper will begin with plausibly correct MCAR, move to the more general MAR, and end with a class of MN mechanisms.

With renewed funding of this longitudinal study, there is now an opportunity to obtain measures of $CPT_{19}$ and Pb on a portion of the subjects. Given existing resources, we wanted to choose 200 subjects for follow-up who provide the most information needed to address the specific research questions. The design questions pertaining to this study will be used as a motivating example.

In the next section we provide the general formulae on MAR and apply them to the example just mentioned. We discuss missingness mechanisms in sequence. Designs for MCAR are discussed in section 3, followed by MAR in section 4, and MN in section 5. Under MCAR we tentatively decide on one design choice that provides sufficient power in our example. From this we examine how various two-stage designs which oversample high-risk subjects, affect the power to address our scientific questions. We then consider a stratified design which protects against a wide class of MN mechanisms. Sufficient conditions for identifiability under this class of MN mechanisms are presented, and we also introduce an approximate likelihood estimator and show it is asymptotically unbiased. Finally, the last section discusses how these different issues can be combined to make the most informed design choice.
2. GENERAL ASYMPTOTIC VARIANCE EXPRESSIONS FOR MISSING AT RANDOM

We first develop the basic formulae for the asymptotic variance under a general missing at random case. These formulae involve fixed coefficients which are the same for any MAR design ($g$ and $h$ in equations (??) and (??)), and second, third, and fourth order moments which do depend on the design. The following two sections apply these formulae to MCAR and two stage designs.

Let $X$ be a $q$-dimensional random vector with distribution $N(\mu, \Sigma)$. Let $R$ be the $q$-dimensional response variable signifying which of the $q$ variables are observed. We assume first that $(X_i, R_i), i = 1, \ldots, n$ all have the same distribution and are independent across subjects. Secondly, in this section we assume that the data are MAR. Let $X_{obs}$ denote the observed variables for observation $i$, and let $n_r$ denote the number of subjects with pattern $r$. We introduce the MCAR marginal distribution $X_{obs} \sim N(\mu_{obs}, \Sigma_{obs})$.

The log likelihood under MAR is given by (Rubin, 1976)

$$
\ell(\mu, \Sigma) = -1/2 \sum_{i=1}^{n} \left( X_{obs} - \mu_{obs} \right)^\prime \Sigma_{obs}^{-1} \left( X_{obs} - \mu_{obs} \right) - 1/2 \log(|\Sigma_{obs}|) + \text{constant}.
$$

We now derive general formulae for the asymptotic variance of the maximum likelihood estimator for $\theta = (\mu, \Sigma)$ under MAR as $n \to \infty$ and the proportions of subjects with each pattern converge to a constant. To obtain the score functions for $\mu$ and $\Sigma$, we introduce the following notation. For any pattern $r$ let $p = p(r) = \{ j : r_j = 1 \}$ and $m = m(r) = \{ j : r_j = 0 \}$ be the observed and missing components respectively. Let $\bar{X}^{(r)}$ be a $q$ dimensional vector which contains the sample means over all observations with pattern $r$. Specifically,

$$
\bar{X}_{j}^{(r)} = \begin{cases} 
\sum_{i:R_i=r} X_{ij} / n_r & \text{for } j \in p(r) \\
0 & \text{otherwise}
\end{cases}
$$

and similarly define $\mu^{(r)}$ as the $q$ dimensional population mean vector with nonzero values only for observed variables $p(r)$. Let $\Sigma_{pp}^*$ be the $q \times q$ matrix with nonzero elements only in the $p(r)$ rows and columns, and zero elsewhere. Similarly let $V_{pp}$ be the $q \times q$ matrix with elements $(V_{pp})_{ij} = (\Sigma_{pp})_{ij}^{-1} = \sigma_{ij:m}^{p}$ for $i, j \in p$ and zero otherwise. Also, for each pattern $r$ let $S^{(r)}$ be a $q \times q$ matrix whose nonzero elements are in the submatrix determined by observed variables with elements equal to $\sum_{i:R_i=r} (X_{ip} - \mu_p)(X_{ip} - \mu_p)^\prime$.

Then the score function for the mean vector is

$$
\frac{\partial \ell}{\partial \mu} = \sum_{r} n_r V_{pp}(\bar{X}^{(r)} - \mu^{(r)}).
$$
or

\[
\frac{\partial \ell}{\partial \mu_v} = \sum_{w=1}^{q} \sum_{v,w=1}^{q} g_{r,wv}(\bar{X}_w^{(r)} - \mu_w)
\]  \hspace{1cm} (1)

where \(g_{r,wv} = n_r \sigma_{uw,m}^{v} \) if \((v, w) \in p(r)\) and zero otherwise.

To obtain the score function for \(\Sigma\), note that

\[
d\ell = -1/2 \sum_r \text{tr} S^{(r)} dV_{pp} - (n_r/2)d \log |\Sigma_{pp}|.
\]

Using the relationships \(dV_{pp} = -V_{pp} d\Sigma_{pp}^* V_{pp}\) and \(d \log |\Sigma_{pp}| = \text{tr} V_{pp} d\Sigma_{pp}^*\), we obtain

\[
\frac{\partial \ell}{\partial \sigma_{ij}} = \sum_{(v,w)} \sum_{(r)(i,j,v,w) \in p(r)} h_{r,vw}^{(ij)}(S_{vw}^{(r)} - n_r \sigma_{vw})
\]

where

\[
h_{r,vw}^{(ij)} = \begin{cases} 
1/2 \sigma_{iv.m}^{v} \sigma_{iw.m}^{w} & \text{if } (i, v, w) \in p(r), i = j \\
1/2 \sigma_{iv.m}^{v} \sigma_{jw.m}^{w} + 1/2 \sigma_{iw.m}^{v} \sigma_{jv.m}^{w} & \text{if } (i, j, v, w) \in p(r), i \neq j \\
0 & \text{otherwise}
\end{cases}
\] \hspace{1cm} (2)

Equation (2) can then be used to compute the \((a, b)\) element of the information matrix involving elements of \(\mu\),

\[
G_{ab} = E \frac{\partial^2 \ell}{\partial \mu_a \partial \mu_b} = E \left\{ \sum_r \sum_{v=1}^q g_{r,av}(\bar{X}_v^{(r)} - \mu_v) \right\} \left\{ \sum_r \sum_{w=1}^q g_{r,bw}(\bar{X}_w^{(r)} - \mu_w) \right\}
\]

\[
= \sum_r \sum_r' \sum_{v=1}^q \sum_{w=1}^q \kappa_{r,r'}^{av,bw} E \left\{ (\bar{X}_v^{(r)} - \mu_v)(\bar{X}_w^{(r)} - \mu_w) \right\}.
\] \hspace{1cm} (3)

where \(\kappa_{r,r'}^{av,bw} = g_{r,av} g_{r',bw}\).

To find the \((ij), (st)\) element in the information matrix for \(\sigma\), we note from equation (2) that

\[
H_{ij, st} = E \frac{\partial^2 \ell}{\partial \sigma_{ij} \partial \sigma_{st}} = E \left\{ \sum_r \sum_{v=1}^q h_{r,vw}^{(ij)} (S_{vw}^{(r)} - n_r \sigma_{vw}) \right\} \left\{ \sum_{r'} \sum_{x,y=1}^q h_{r',xy}^{(st)} (S_{xy}^{(r')} - n_{r'} \sigma_{xy}) \right\}
\]

\[
= \sum_r \sum_{r'} \sum_{v=1}^q \sum_{w=1}^q \sum_{x,y=1}^q \psi_{r,r'}^{(vw),(ij),(xy),(st)} \left\{ (S_{vw}^{(r)} - n_r \sigma_{vw})(S_{xy}^{(r')} - n_{r'} \sigma_{xy}) \right\},
\] \hspace{1cm} (4)

where \(\psi_{r,r'}^{(vw),(ij),(xy),(st)} = h_{r,vw}^{(ij)} h_{r',xy}^{(st)}\).
To obtain the $(a), (ij)$ element of the information matrix for $\mu_a$ and $\sigma_{ij}$, we have a similar expression.

\[
J_{a,ij} = E \left[ \frac{\partial^2 \ell}{\partial \mu_a \partial \sigma_{ij}} \right] = E \left\{ \sum_r \sum_{v=1}^q g_{r,av} (\bar{X}_v^{(r)} - \mu_v) \left\{ \sum_{r'} \sum_{w=1}^q n_{r',w} h_{r',xy}^{(ij)} (\nu_{r'} - n_{r'} \sigma_{xy}) \right\} \right\}
\]

where $\eta_{r,r'}^{(av),(xy),(ij)} = g_{r,av} h_{r',xy}^{(ij)}$.

3. ASYMPTOTIC VARIANCE EXPRESSIONS FOR MCAR

Under MCAR $G, H,$ and $J$ from expressions ?? - ?? can be simplified considerably: $G_{ab} = \sum_r n_r \sigma_{ab,m(r)}$; similarly, $H_{ij,st} = \sum_r \sum_{v,w,x,y=1}^q n_{r,vw} h_{r,xy}^{(ij)} (\sigma_{vx} \sigma_{wy} + \sigma_{vy} \sigma_{wx})$, and as usual, the terms off the block diagonals, the $J$ matrix, have zero expectation under MCAR. Asymptotic variances for means, variances, and covariances can be computed using the delta method. Approximations to the power for hypotheses tests can then be constructed using normal approximations.

3.1 Missingness on Existing Attention Measures

We present an example to illustrate how these power calculations can be used to decide the number of subjects to be assessed at follow-up in a longitudinal study. The data in this example address whether during childhood and adolescence there are critical times that attention may be affected by lifetime exposure to lead.

As in any longitudinal design, there are multiple reasons why different parts of the data are missing. We describe these in some detail below. This study began as a two cohort study. In 1985 the first cohort of 1196 children attended first grade. The parents of these children and the children themselves were asked to give their consent to the first round of assessments; consent was obtained for 1084 or 90.6%. By second grade, when researchers added the continuous performance test ($CPT_8$), 302 had moved out of the schools or classrooms under study, leaving 782 (72%) eligible for assessment. Thirty-three (4%) refused to take part leaving 749 who consented at both times. Because of limited time and funding, not all of these children could be assessed on $CPT_8$. To reduce the potential for any bias during data collection, the 749 children were divided into 5 matched replicates based on urban area, and data collection began on the first replicate then proceeded to each succeeding replicate until the school year ended. $CPT_8$ was obtained on 435 of the 749 (58%) available children. Forty (5%) of the 749 were missing because of absence on the day of testing. No differences in distribution were found between the original sample of 1196 and the 435 with $CPT_8$ measurements on
race/ethnicity or gender. Even though the 435 represents only one-third of the sample, these and other data suggest that selection bias is probably minimal.

$CPT_{10}$ and $CPT_{13}$ were collected on 289 (66%) and 269 (62%) of the 435 with $CPT_8$ measures. Again replicate sampling was used to reduce attrition bias across time. A second cohort of 1107 children, measured on the first grade measures, were not assessed on CPT at any of these times. Combined with the 761 children in Cohort 1, 1868 children who were not assessed on any CPT measure. The number of children in the two cohorts was 2303. A complete description of the missing data pattern is given in Table ??.

3.2 An Example of Determining a Follow-up Design Based on MCAR

With renewed funding in hand, sufficient for follow-up of about 200 subjects, these researchers had to decide which of the 2303 children should be assessed on $CPT_{19}$ and $Pb$. Our primary scientific question is to determine the time of maximum correlation between CPT and lifetime exposure to lead. The MCAR assumption seems at first glance reasonably justified by the initial replicate sampling and the lack of any apparent attrition bias.

We have altogether ten patterns of missing data to consider in the selection of subjects to follow up. These patterns are shown in Table ??, Pattern 1, for example, indicates that $n_1$ subjects have no missing data. The $n_2$ pattern 2 subjects have complete data up to but not including the follow-up at 19 years of age. The total number of subjects with patterns 1 or 2 is 216, the number of children who so far have been measured at ages 8, 10, and 13. One design decision then is to determine the value of $n_1$. Similar groupings occur for patterns 3 and 4, 5 and 6, 7 and 8, and 9 and 10. The number of subjects in patterns 1-8, 435, matches the number of subjects who have a $CPT_8$ measure. 1868 subjects with pattern 9 or 10 were either in Cohort 1 and did not receive $CPT_8$ or were in Cohort 2.

We have obtained trial values of $\mu$ and $\Sigma$ from two sources (Table 2). For those measures before age 19, sample estimates are used. For the remaining parameter values, we considered a range of correlation values from 0.05 to 0.30. This range is believed to cover plausible alternatives. Variables are standardized in the example since this does not affect the asymptotic variance of correlations.

In Table ??, the numbers of subjects in each of the odd numbered patterns is given in the first five columns. The numbers of subjects in the even numbered patterns are excluded for readability; they can be obtained by subtraction in Table 1. For comparison purposes, we have either selected all, none, or half of the subjects in each pattern category. The percent cost increase for follow-up in the middle column is proportional to the number of subjects followed up. The precision associated with each correlation estimator involving a $CPT$ measure and $Pb$ is given in the first row while all other rows are scaled as increases in precision. For the range of correlations between $Pb$ and $CPT$ we considered
(0.05 to 0.30), the percent improvement in asymptotic variance was nearly constant, so we have chosen to describe in Table ?? the results using a midrange value of 0.15.

Doubling the number of \( n_1 \) subjects with complete data to follow-up, shown in the second row of Table ??, produces the expected 100% increase in both cost and precision. When we begin selecting from patterns that are not complete, the cost increases much more quickly than the precision does. Row 6 corresponds to a design which follows up all subjects with measures on at least 2 of the first 3 CPT measures. The design efficiency for estimating all correlations is at least 118% compared to those in the first design. Including 934 subjects in the last group helps minimally with the other measures; it only makes a sizable improvement on the \( Pb \) correlations with \( CPT_{13} \) and \( CPT_{19} \). There is, however, some improvement in precision as we add subjects with less complete data to be followed up.

In Table ?? we present power calculations for testing the correlation between \( CPT_8 \) and \( Pb \) for different patterns of missing data and different correlations (assumed equal across all CPT measures with \( Pb \)). Comparing Tables ?? and ?? we see that maximum gain in precision depends most strongly on the number of observations selected in Patterns 1 and 3 i.e. on \( n_1 \) and \( n_3 \). When \( \rho = 0.15 \), we obtain (Row 6 of Table 4) sufficient power (79.5%) from a design that completely follows up all 216 subjects from Pattern 1, all 73 from Pattern 3 and all 53 from Pattern 5. Thus from a design efficiency (Table 3) and power perspective (Table 4) we find that following up all subjects with the three most complete patterns (patterns 1, 3, and 5), has good design efficiency. This is achieved by spending a little more than twice the cost of following up the complete cases only.

4. ASYMPTOTIC VARIANCE CALCULATIONS FOR THE MISSING AT RANDOM CASE

Under MAR, the distributions of \( \bar{X}^{(r)} \) and \( S^{(r)} \) are not necessarily the same as the marginal distributions based on the full multivariate normal distribution. To handle MAR cases, we need the moment structure for each pattern of observed data. The calculations below, which give the multivariate second moments of the sample means and variances, are similar in spirit to those in Kendall and Stuart (1979).

Define \( \alpha_i^{(r)} = E\{X_i|R = r\} \) for \( i \in p(r) \), 0 otherwise, and similarly, \( \beta_{ij}^{(r)} = E\{X_iX_j|R = r\} \) for \( i,j \in p(r) \), \( \gamma_{ijk}^{(r)} = E\{X_iX_jX_k|R = r\} \) for \( i,j,k \in p(r) \), and \( \delta_{ijkl}^{(r)} = E\{X_iX_jX_kX_l|R = r\} \) for \( i,j,k,l \in p(r) \). Higher order moments can be written in terms of these basic quantities. For example, the cross product of two means calculated over the same pattern, \( E(\bar{X}_u^{(r)} - \mu_u)(\bar{X}_v^{(r)} - \mu_v) = \frac{n_v-1}{n_v}\alpha_u\alpha_v - \mu_u\mu_v \). Expressions for other moments are lengthy and can be found on http://www.biostat.coph.usf.edu/research/psmg/Msgdesign/appx.html. The particular
Table 1. Pattern of Missing Data

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Sample Size</th>
<th>CPT8</th>
<th>CPT10</th>
<th>CPT13</th>
<th>CPT19</th>
<th>Pb</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>n1</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>2</td>
<td>n2</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>n3</td>
<td>x</td>
<td>x</td>
<td>0</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>4</td>
<td>n4</td>
<td>x</td>
<td>x</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>n5</td>
<td>x</td>
<td>0</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>6</td>
<td>n6</td>
<td>x</td>
<td>0</td>
<td>x</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>n7</td>
<td>x</td>
<td>0</td>
<td>0</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>8</td>
<td>n8</td>
<td>x</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>n9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>10</td>
<td>n10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>2303</td>
<td>435</td>
<td>289</td>
<td>269</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: x=observed, 0=missing

Table 2. Means and Covariances for CPT and Pb

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Means</th>
<th>Variances and Covariances</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CPT8</td>
</tr>
<tr>
<td>CPT8</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>CPT10</td>
<td>0</td>
<td>.44</td>
</tr>
<tr>
<td>CPT13</td>
<td>0</td>
<td>.30</td>
</tr>
<tr>
<td>CPT19</td>
<td>0</td>
<td>.30</td>
</tr>
<tr>
<td>Pb</td>
<td>0</td>
<td>.05</td>
</tr>
</tbody>
</table>

Table 3. Cost and Design Efficiency for Differing Patterns of Data Under MCAR

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Percent</th>
<th>Cost Increase</th>
<th>Precision for Correlations with Pb</th>
</tr>
</thead>
<tbody>
<tr>
<td>n1</td>
<td>n3</td>
<td>n5</td>
<td>n7</td>
</tr>
<tr>
<td>198</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>216</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>216</td>
<td>36</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>216</td>
<td>73</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>216</td>
<td>73</td>
<td>26</td>
<td>0</td>
</tr>
<tr>
<td>216</td>
<td>73</td>
<td>53</td>
<td>0</td>
</tr>
<tr>
<td>216</td>
<td>73</td>
<td>53</td>
<td>46</td>
</tr>
<tr>
<td>216</td>
<td>73</td>
<td>53</td>
<td>93</td>
</tr>
<tr>
<td>216</td>
<td>73</td>
<td>53</td>
<td>93</td>
</tr>
</tbody>
</table>

NOTE: Percent Cost increase is computed as the ratio of the total number of subjects for the row to 108 (total subjects in Row 1). The first row under precisions for correlation with Pb lists the asymptotic standard deviations between the CPT measures and Pb (when \( \rho = .15 \) in the last row of Table 2). Subsequent rows in the Precision column list the Percent improvement in precision over the corresponding values in row 1.
Table 4. Power Calculations for Correlation Coefficient of CPT₈ with Pb: Data MCAR

<table>
<thead>
<tr>
<th>Patterns</th>
<th>n₁</th>
<th>n₃</th>
<th>n₅</th>
<th>n₇</th>
<th>n₉</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ρ = .05</td>
</tr>
<tr>
<td>108</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.0739</td>
</tr>
<tr>
<td>216</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.1094</td>
</tr>
<tr>
<td>216</td>
<td>36</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.1209</td>
</tr>
<tr>
<td>216</td>
<td>73</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.1327</td>
</tr>
<tr>
<td>216</td>
<td>73</td>
<td>26</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.1409</td>
</tr>
<tr>
<td>216</td>
<td>73</td>
<td>53</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.1495</td>
</tr>
<tr>
<td>216</td>
<td>73</td>
<td>53</td>
<td>46</td>
<td>0</td>
<td>0</td>
<td>.1609</td>
</tr>
<tr>
<td>216</td>
<td>73</td>
<td>53</td>
<td>93</td>
<td>0</td>
<td>0</td>
<td>.1788</td>
</tr>
<tr>
<td>216</td>
<td>73</td>
<td>53</td>
<td>93</td>
<td>92</td>
<td>0</td>
<td>.4533</td>
</tr>
</tbody>
</table>

MAR design which is used will require calculating \((\alpha, \beta, \gamma, \delta)\), using these in equations and substituting the resulting expressions into formulae (?? - ??).

4.1 Asymptotic Variance Calculations for Two Stage Designs

In two stage designs one variable, say \(X_u\), can be used to determine the probabilities of observing the second stage measures. Let \(r = 1\) correspond to pattern with complete observations and \(r = 0\) the pattern with incomplete observations. Because we have assumed MAR, \(X_u\) must always be observed. We focus on the most common situation where the probabilities are based on a cut point \(d\) for \(X_u\), i.e., \(Pr\{R = 1|X_u > d\} = \omega^d\) and \(Pr\{R = 1|X_u < d\} = \omega_d\). Then conditional probabilities, means, and higher moments given the pattern can be calculated as follows. For any variables \(i, j, k, l = 1, \ldots q\)

\[
Pr\{R = 1\} = \omega^d(1 - \Phi\left(\frac{d - \mu_u}{\sigma_u}\right)) + \omega_d\Phi\left(\frac{d - \mu_u}{\sigma_u}\right) \\
\alpha_i^{(1)} = E(X_i|R = 1) = \omega^dE(X_i|X_u > d) + \omega_dE(X_i|X_u < d) \\
\beta_{ij}^{(1)} = \omega^dE(X_iX_j|X_u > d) + \omega_dE(X_iX_j|X_u < d) \\
\gamma_{ijk}^{(1)} = \omega^dE(X_iX_jX_k|X_u > d) + \omega_dE(X_iX_jX_k|X_u < d) \\
\delta_{ijkl}^{(1)} = \omega^dE(X_iX_jX_kX_l|X_u > d) + \omega_dE(X_iX_jX_kX_l|X_u < d)
\]

(6)

The expressions for \(r = 0\), i.e., \(\alpha_i^{(0)}\) are identical to those above except with weights \((1-\omega^d)\) and \((1-\omega_d)\) for the respective expectations. The moments are tedious to calculate even under the assumption of multivariate normality. We used Mathematica (Wolfram, 1990) to do these calculations which are available on the web (http://www.biostat.coph.usf.edu/research/psmg/).

We note that a possible limitation of our procedure could be its dependency on multivariate normality. Indeed, it is possible that our procedure may be robust under certain conditions but we shall not focus on robustness related issues in this article.

4.2 Example Continued for Missing at Random

In many longitudinal studies it is important to examine the developmental trajectories of the high-risk or non-normative subjects. Oversampling those who initially have abnormally high attention measures is appropriate for such an aim,
but it is intuitively counterproductive in examining correlations since it results in less variation. Consequently, we investigate how much power we lose in testing correlations between \( CPT \) and \( Pb \) by oversampling from high risk groups. We define subjects who are above the median in their \( CPT_8 \) score measures as belonging to a high risk group for possible adverse neurologic outcomes. Starting from the MCAR recommendation in the last section of a design where \( n_1 = 216, n_3 = 73, \) and \( n_5 = 53, \) we perturbed this design to examine several two-stage alternatives. All of these perturbations maintain the same \( n_3 \) and \( n_5 \) but vary \( n_1. \)

From these calculations, we conclude that oversampling of high-risk subjects can have a substantial effect on the power for testing correlations. This situation is, however, not a general result, as can be seen in the first two rows of Table ???. Here we compare two designs, both having half of the most complete cases selected for follow-up. In the first row the \( n_1 = 108 \) are distributed with equal probabilities among the high and low risk group. In the second row all the \( n_1 = 108 \) are high risk and none of the low risk group are selected. The power in the latter design, over this range of \( \rho, \) is reduced by less than 0.03, a minimal change. Thus if the amount of available money will support only 200 subjects for follow-up, we can afford to oversample the higher risk group with little effect on power.

On the other hand, if we have sufficient money to follow-up 300 or more subjects, there can be little oversampling of high risk subjects. Our recommendation from the last MCAR section was to take as many subjects with complete \( CPT \) measures as possible (\( n_1 = 216 \)). If we take all the 108 high risk subjects and sequentially increase the proportion of low risk subjects selected for follow-up from 0 to 108, as shown in rows 2-5 of Table ???, we find a strong effect on power, particularly at \( \rho = 0.15. \) The effects of stratifying on \( CPT_8 \) for patterns other than 1 or 2 is minimal because there is less information available on these more incomplete cases. Thus when 300 or more subjects can be followed up, a two-stage sample is not very efficient. It is possible to affect power adversely by oversampling high risk subjects. However, this only happens in our example when the subjects are chosen from those with mostly missing data rather than complete data. With 300 subjects the only designs with sufficient power would be ones that selected most or all of those with patterns 1-2 and also most of those in patterns 3-6 as well. Those with patterns 7-10 provide too little information to be useful. No matter how we selected the 300 from the 352 subjects with patterns 1-6, the fact that 300 is fully 85do much oversampling of those above the median on \( CPT_8 \). Thua if we oversample from the low risk subjects we see a minimal effect on power.

5. A GENERAL FOLLOW-UP DESIGN TO PROTECT AGAINST MN MECHANISMS
Table 5. Power Calculations for Correlation Coefficient of CPT with Lead: Missing at Random Case

<table>
<thead>
<tr>
<th>Patterns</th>
<th>Power</th>
<th>n1</th>
<th>n3</th>
<th>n5</th>
<th>( \rho = .05 )</th>
<th>( \rho = .10 )</th>
<th>( \rho = .15 )</th>
<th>( \rho = .30 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>above median</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.115</td>
<td>.332</td>
<td>.632</td>
<td>.997</td>
</tr>
<tr>
<td>below median</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.098</td>
<td>.299</td>
<td>.621</td>
<td>.996</td>
</tr>
<tr>
<td></td>
<td></td>
<td>73</td>
<td>53</td>
<td></td>
<td>.126</td>
<td>.372</td>
<td>.693</td>
<td>.999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>73</td>
<td>53</td>
<td></td>
<td>.132</td>
<td>.395</td>
<td>.723</td>
<td>.999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>73</td>
<td>53</td>
<td></td>
<td>.149</td>
<td>.455</td>
<td>.795</td>
<td>.999</td>
</tr>
</tbody>
</table>

So far we have discussed only designs to handle ignorably missing data. There are two reasons that non-ignorability may be a more serious issue. First, one rarely has direct information that non-ignorable missingness has occurred, and second, its primary impact is in bias not variance. We introduce a stratified sampling procedure for follow-up that provides protection midway into a longitudinal study against certain types of MN data. Protective estimators are defined as having good statistical properties, such as asymptotic unbiasedness, under a large class of MN mechanisms (Brown, 1990). Thus our approach is to find an estimator which is robust against specific classes of missing data. Such procedures other exist (Brown, 1990). For example, consider estimating \( \mu_2 \) based on samples drawn from a bivariate normal distribution with incomplete observations on the second variable. Suppose missingness is nonignorable and dependent exclusively on the second variable. The standard ML under ignorability produces a biased estimate. However, a calibration-based estimator of \( \mu_2 \) formed by regressing \( X_1 \) on \( X_2 \) is asymptotically unbiased regardless of the actual nonignorable mechanism (Brown, 1990).

For the problem we consider, we are midway into a longitudinal study. The time to observe some of the variables has already occurred, leading to different patterns of missing data, and the time to observe the remaining variables has not yet occurred. Our design choices involve which subjects to follow up in the remaining part of the study. We will find that certain designs allow us to obtain consistent estimators of all the parameters, even those involving the distribution of variables already subject to nonignorable missingness. The essential requirement for consistency is identifiability of the parameters under a class of MN mechanisms. Theorem 1 provides a sufficient condition for identifiability based on the patterns which are observed. Once identifiability has been established, we search for protective estimators based on weighted combinations of regression estimates obtained from the different patterns of missing data.

We divide the variables into two vectors, \( X = (Y, Z) \), where \( Y \) consists of all \( q_y \) variables whose time of measurement has already occurred and \( Z \) consists of all \( q_z \) variables yet to be measured. \( Y \) is allowed to be missing nonignorably. We define \( \theta_z = (\mu_z, \Sigma_{zz}) \) as the parameters of the marginal distribution of \( Z \) and \( \theta_{z|y} = (\alpha_{z|y}, \beta_{z|y}, \Sigma_{zz,y}) \) as the conditional intercept, slope and residual variance. Let \( R^y \) and \( R^z \) be the random patterns of missing data for \( Y \) and \( Z \) respectively. In the case of MN data \( R^y \) and \( Y \) will not be independent.
Similar to Little (1993), we transform our selection problem into a pattern mixture model. We begin with the selection model, \( \omega^y(r^y|y, z) = \Pr\{R^y = r^y | Y = y, Z = z\} \) which specifies the MN mechanism for \( Y \). Also, \( \omega^z(r^z|y, z) = \Pr\{R^z = r^z | R^y = r^y, Y = y, Z = z\} \) is the missing data mechanism for \( Z \). Then the pattern mixture model is expressed in terms of \( \omega^y(r^y, r^z) = \Pr\{R^y = r^y, R^z = r^z\} \) and \( f(y, z|r^y, r^z) \), the distribution of \((Y, Z)\) given the pattern \( R^y = r^y, R^z = r^z \).

**Theorem 1.** Let \((Y, Z)\) have multivariate normal density \( f(y; \theta_y)f(z|y; \theta_{z|y}) \). Suppose the missing mechanism for \( Y \) does not depend on \( Z \), i.e., for every observable pattern \( r^y \), \( \omega^y(r^y|y, z) = \omega^y(r^y|y) \). Also assume that for every observable pattern \((r^y, r^z)\), the mechanism for \( Z \) does not depend on \((Y, Z)\), \( \omega^z(r^z|r^y, y, z) = \omega^z(r^z|r^y) \). Suppose the following conditions hold.

1. The system of equations in \( \theta_{z|y} \), \( \gamma(y_{obs}) \), \( \delta(y_{obs}) \) over all patterns \((r^y, r^z)\)

\[
E(Z_{obs}|y_{obs}, r^y, r^z) = \alpha_{zobs,y} + \beta_{zobs,y} \left( y_{obs} \gamma(y_{obs}) \right)
\]

\[
Var(Z_{obs}|y_{obs}, r^y, r^z) = \Sigma_{zobs,y} \beta_{zobs,y} \begin{pmatrix} 0 \\ \delta(y_{obs}) \end{pmatrix} \beta_{zobs,y}^T
\]

(7)

admits a single solution \( \theta_{z|y} \).

2. The system of linear equations in \( E(Y_{mix}|r^y), Var(Y_{mix}|r^y), Cov(Y_{mix}, Y_{obs}|r^y) \) over all patterns \((r^y, r^z)\)

\[
E(Z_{obs}|r^y, r^z) = \alpha_{zobs}|y + \beta_{zobs}|y \left( E(Y_{obs}|r^y) \right)
\]

\[
Var(Z_{obs}|r^y, r^z) = \Sigma_{zobs,y} \beta_{zobs,y} \left( \begin{pmatrix} Var(Y_{obs}|r^y) \\ Cov(Y_{mix}, Y_{obs}|r^y) \end{pmatrix} \begin{pmatrix} Cov(Y_{obs}, Y_{mix}|r^y) \\ Var(Y_{mix}|r^y) \end{pmatrix} \right) \beta_{zobs,y}^T
\]

\[
Cov(Z_{obs}, Y_{obs}|r^y, r^z) = \beta_{zobs,y} \left( \begin{pmatrix} Var(Y_{obs}|r^y) \\ Cov(Y_{mix}, Y_{obs}|r^y) \end{pmatrix} \begin{pmatrix} Cov(Y_{obs}, Y_{mix}|r^y) \\ Var(Y_{mix}|r^y) \end{pmatrix} \right)
\]

(8)

admits a single solution for fixed \( \theta_{z|y} \) and fixed \((E(Z_{obs}|r^y, r^z), E(Y_{obs}|r^y, r^z), Var(Z_{obs}, Y_{obs}|r^y, r^z))\). Then \( \theta \) is identifiable.

**Proof:** First note that \( Z \) and the patterns \((R^y, R^z)\) are independent given \( Y \):

\[
f(z|y, r^y, r^z) = f(y)f(z|y)\omega^y(r^y|y)\omega^z(r^z|r^y)dz = f(z|y).
\]

Also, \( f(y|r^y, r^z) = f(y)\int f(z|y)\omega^y(r^y|y)\omega^z(r^z|r^y)dz / \int f(y)\int f(z|y)\omega^y(r^y|y)\omega^z(r^z|r^y)dzdy' = f(y|r^y) \).

Letting \( \gamma(y_{obs}) = E(Y_{mix}|y_{obs}, r^y) \) and \( \delta(y_{obs}) = Var(Y_{mix}|y_{obs}, r^y) \) be the unknown conditional means and variances, equations (7) are the correct conditional moments of \( Z_{obs}|Y_{obs} \). Since the conditional moments are observable, solvability of equations (7) for \( \theta_{z|y} \) means these parameters are identifiable. Similarly, if the first and second moments of \( Y_{mix} \) given the pattern
\((r^y, r^z)\) are solvable from equations (??), we can calculate the marginal moments of \(Y\) as 
\[
\mu_y = \sum \omega_y(r^y)E(Y|r^y)
\]
and 
\[
\Sigma_{yy} = \sum \omega_y(r^y)[Var(Y|r^y) + (E(Y|r^y) - \mu_y)(E(Y|r^y) - \mu_y)]
\]. From these we obtain 
\[
\mu_z = \alpha_{z|y} + \beta_{z|y}\mu_y,
\]
\[
\Sigma_{zz} = \Sigma_{zz,y} + \beta_{z|y}\Sigma_{yy}\beta_{z|y}^\prime,
\]
\[
\Sigma_{zy} = \beta_{z|y}\Sigma_{yy}\bullet
\].

Equations (??) and (??) can be checked rather easily for solvability. There are also many sets of patterns that produce identifiability. A simple one is given below.

**Corollary 1.** Suppose one variable \(Y_1\) is never missing or is missing completely at random while other \((Y_2, \ldots, Y_q)\) variables may be missing nonignorably. Suppose that the probability that \(Y\) is completely observed is nonzero, and the probability that \(Z\) is completely observed is nonzero for every observed pattern \(r^y\). Finally suppose \(q_z \geq q_y - 1\). Then \(\theta\) is identifiable.

**Proof:** First, \(\theta_{z|y}\) is solvable from the conditional distributions (??) given complete \(Y\) and \(Z\). Secondly, consider for every \(r^y\) where at least one variable is observed, the linear equations (??) in \(E(Y_{mis}|r^y)\) and \(\text{Cov}(Y_{mis}, Y_{obs}|r^y), \text{Var}(Y_{mis}|r^y))\), evaluated for complete data on \(r^z\). Because \(q_z \geq q_y - 1\), these yield at least as many linear equations as unknowns, hence the second set of equations is also solvable. Finally, we consider the remaining case of no observed \(Y\). Because \(Y_1\) is MCAR, the distribution of \(Y|r^y = 0\) is the same as the distribution of \(Y|r^y \neq 0\). The means and variances of \(Y|r^y \neq 0\), being based on at least one observed \(Y_1\), are all identifiable. Thus \((\mu_y, \Sigma_{yy})\) are identifiable \bullet

5.1 Example Continued for Nonignorably Missing Data

In our example, we consider stratified random sampling for follow-up based on the pattern of missing data \(r^y\) already observed in the \(Y = (CPT_8, CPT_{10}, CPT_{13})\) data and then average appropriately over the different patterns of missing data. Here we do not assume that follow-up data on \(Z = (CPT_{19}, Pb)\) will be complete. To ensure that missingness on \(Z\) depends almost completely on \(r^y\), we recommend forming replicate samples, using intensive follow-up procedures, and masking the interview team to \(Y\) values. In our case, \(Y_1 = CPT_8\) is considered essentially missing completely at random because subjects were randomly chosen for \(CPT_8\) at the start of the study and a minimal 5% were missing. No assumption need be made about missingness on \((CPT_{10}, CPT_{13})\). Thus this situation satisfies or at least closely approximates all the conditions of corollary 1.

We now complete the description of the sampling process and introduce a new consistent estimator. For definiteness, let \(S\) be an \(n\)-dimensional indicator variable that identifies those subjects who are to be sampled for follow-up. Let 
\[
\omega^*(s|y, z, r^y, r^z) = Pr\{S = s|Y = y, Z = z, R^y = r^y, R^z = r^z\}
\]
be the general sampling for follow-up mechanism. Since we will select subjects for follow-up based only on the pattern \(r^y\), 
\[
\omega^*(s|y, z, r^y, r^z) = \omega^*(s|r^y).
\]
Define \(n_{ry}\) as the number in our sample with pattern \(ry\). Let \(m_{r\cdot y}\) be the number with pattern \(ry\) who are selected for follow-up. And let \(m_{r\cdot y\cdot r}\) be the number selected with pattern \((r^y, r^z)\).

Let \((\omega^y, \omega^z, \mu_{y|r^y}, \Sigma_{yy|r^y}, \theta_{z|y})\) maximize \(\ell_{r^y, r^z} + \sum_{r^y} \ell_{y|r^y} + \sum_{r^z} \ell_{z_{obs}|y_{obs}} + \ell_{z_{obs}|y_{obs}}|r^y, r^z\) where

\[
\ell_{r^y, r^z}(\omega; n_{r^y}, m_{r^y}, m_{r^y\cdot r^z}) = \frac{\log \omega^y(r^y)/m_{r^y\cdot r^z}}{n_{r^y} - n_{r^y\cdot r^z}} - \frac{1}{2} \log |\Sigma_{obs|y_{obs}|r^y}| - \frac{1}{2} \Sigma_{i=1}^2 \Sigma_{i=1}^2 \log |f(z_{obs}, y_{obs})| ^{\frac{1}{2}} \Sigma_{z_{obs}|y_{obs}|r^y} + \beta_{z_{obs}}|y_{obs}| \Sigma_{yy|r^y}\]

subject to \(\mu_y = \sum_{r^y} \omega^y(r^y)\mu_{y|r^y}\) \(\Sigma_{yy} = \sum_{r^y} \omega^y(r^y)\Sigma_{yy|r^y} + (\mu_{y|r^y} - \mu_y)(\mu_{y|r^y} - \mu_y)^T\)

**Corollary 2.** Given the assumptions of Corollary 1, the estimating procedure based on (9) yields consistent estimates for \(\theta\) as \(n \to \infty\).

**Proof:** By Corollary 1 all parameters are identified. To show consistency, note that the first log likelihoods can be maximized directly: \(\hat{\omega}^y(r^y) = n_{r^y}/n\), and \(\hat{\omega}^z(r^z|r^y) = m_{r^y\cdot r^z}/m_{r^y}\). Both of these estimators are consistent. For the second through fourth log likelihoods, we rely on showing that the resulting estimating equations are unbiased. Because the derivatives of the second through fourth log likelihoods with respect to the parameters yield expressions that are linear and quadratic in the data, unbiasedness is easily verified.

### 5.2 Example of Follow-up to Protect Against MN

Trading some protection against bias for a small increase in variance is a useful strategy. We investigated different choices for follow-up of 204 subjects, the minimal number of subjects we anticipated assessing given the budget. Considering many alternative designs, we arrived at the following MN protective design in the last rows of Table ???. In keeping with our assumption that missingness on \(CPT_8\) is nearly MCAR, no subjects were taken if they were missing all variables. For all other \(r^y\) patterns at least 24 subjects were selected. In accordance with Tukey’s guide for obtaining a variance estimate which is likely to be within one-tenth of the true value (Tukey, 1986), this size sample affords some opportunity to estimate means and variances with each pattern. The number of subjects selected with complete data on \(Y\) is higher in keeping with more information being available from these cases.

We compared the efficiency and power of this design involving a follow-up of 204 subjects to the best design assuming MCAR and the best two-stage design, each with the same size sample (Table ??). The best MCAR design takes \(n_1 = 204\) and all other odd \(n\) to be zero (first rows). The best two-stage design with a median split selects all 108 subjects with complete \(Y\) values that are above the median on \(CPT_8\) and 96 of the 108 who are below the median (second row).
### Table 6. Efficiency and Power of Three Follow-up Designs Under MAR Conditions

<table>
<thead>
<tr>
<th>Design Category</th>
<th>Number of Subjects</th>
<th>Asymptotic Standard Errors (Power) of Correlations with Pb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CPT\textsubscript{8}</td>
</tr>
<tr>
<td><strong>MCAR</strong></td>
<td>n\textsubscript{1} = 204</td>
<td>n\textsubscript{2} = 12</td>
</tr>
<tr>
<td></td>
<td>n\textsubscript{3} = 0</td>
<td>n\textsubscript{4} = 73</td>
</tr>
<tr>
<td></td>
<td>n\textsubscript{5} = 0</td>
<td>n\textsubscript{6} = 53</td>
</tr>
<tr>
<td></td>
<td>n\textsubscript{7} = 0</td>
<td>n\textsubscript{8} = 93</td>
</tr>
<tr>
<td></td>
<td>n\textsubscript{9} = 0</td>
<td>n\textsubscript{10} = 1068</td>
</tr>
<tr>
<td><strong>MAR</strong></td>
<td>Above Median</td>
<td>Below Median</td>
</tr>
<tr>
<td></td>
<td>n\textsubscript{1} = 108</td>
<td>n\textsubscript{2} = 96</td>
</tr>
<tr>
<td></td>
<td>n\textsubscript{3} = 0</td>
<td>n\textsubscript{4} = 73</td>
</tr>
<tr>
<td></td>
<td>n\textsubscript{5} = 0</td>
<td>n\textsubscript{6} = 53</td>
</tr>
<tr>
<td></td>
<td>n\textsubscript{7} = 0</td>
<td>n\textsubscript{8} = 93</td>
</tr>
<tr>
<td></td>
<td>n\textsubscript{9} = 0</td>
<td>n\textsubscript{10} = 1068</td>
</tr>
<tr>
<td><strong>MN</strong></td>
<td>m\textsubscript{1} = 132</td>
<td>m\textsubscript{2} = 72</td>
</tr>
<tr>
<td></td>
<td>m\textsubscript{3} = 24</td>
<td>m\textsubscript{4} = 49</td>
</tr>
<tr>
<td></td>
<td>m\textsubscript{5} = 24</td>
<td>m\textsubscript{6} = 29</td>
</tr>
<tr>
<td></td>
<td>m\textsubscript{7} = 24</td>
<td>m\textsubscript{8} = 69</td>
</tr>
<tr>
<td></td>
<td>m\textsubscript{9} = 0</td>
<td>m\textsubscript{10} = 1068</td>
</tr>
</tbody>
</table>

The asymptotic variances of the correlations with Pb based on the three designs are compared with $\rho = 0.15$, and the assumption that everyone selected for follow-up has complete data on $Z = (CPT\textsubscript{19}, Pb)$. All three designs provide close to 80% power for the four correlations. Since the MN design provides an opportunity to deal with nonignorably missing data in $Y = (CPT\textsubscript{8}, CPT\textsubscript{10}, CPT\textsubscript{13})$, this design is preferable and was the one on which we decided. When the data collection on $Z$ is complete, we plan to calculate the maximum likelihood estimates of $\theta$ under both MCAR/MAR and MN assumptions. Similarities in the estimates would give us greater confidence in the appropriateness of the ignorable solution, while differences in the estimates would lead us to examine alternative explanations of how missingness would affect our results.

### 6. DISCUSSION

Selecting the sample to follow-up midway through a longitudinal study involves much more complex sample size calculations than those done at the beginning of a study. The first sample size calculations for a study generally involve very little data so educated guesses and oversimplifications, such as a fixed rate of missing data across the study panels, are assumed. By the middle of a longitudinal study we not only have some data that can point towards MAR or nonignorably missing data, but the observed patterns of missing data often differ substantially from what was assumed in earlier power calculations. Also by the middle of a study many more hypotheses are derived from the data so a design which has high power to answer one important question may not do so well in answering other important questions.
Cochran (1983) recommended that the sample size be calculated for each question and the final sample size be chosen sufficiently large to answer most of the important questions.

In contrast, determining which and how much of the sample to follow-up midway through the study often results in design choices which are in competition. A design which maximizes information on a population average value, such as the correlation between $CPT_{10}$ and $Pb$, will generally have relatively low power in detecting differences in $Pb$ levels across different identified risk groups unless the low frequency subgroups are specifically oversampled.

The MN results in this paper hold with obvious changes if we select subjects for follow-up based not just on $r^y$ but also on any fixed, completely observed covariate. In our specific example, several options for stratification on completely observed covariates for follow-up were considered: gender, classroom behavior and achievement, intervention status, birthdate of the child, and the pattern of missing data. Gender, behavior and achievement, and intervention status were related to important scientific questions. Stratifying on birthdate would allow us to control for age at time of observation, and stratifying by pattern of missing data provides protection against nonrandom missingness as described in the last section. From this potential list of stratifying factors, we finally decided to stratify on birthdate, pattern of missing data, gender, and intervention status. In each of the categories of birthdate (2), gender (2), intervention status (3), and pattern of missing data (4) we randomly chose to follow up at least 2 subjects, with oversampling of all those strata where $CPT$ measures were completed at ages 8, 10, and 13 (11 subjects in each cell). This oversampling of subjects with no missing data automatically provided more statistical power than a design which selected subjects with incomplete data. In those few cells where there were fewer than 11 subjects available, we filled out the numbers by selecting from cells with nearly complete data on $CPT$. The total number selected for follow-up in the 48 strata was 204. To simplify the presentation in this paper, our asymptotic variance calculations have ignored these completely observed covariates. Finally, an alternative approach to handling missing data in longitudinal studies is to model the various mechanisms directly. In this study, analyses of the yearly classroom-based measures suggests at least two distinct mechanisms are taking place. In early elementary school the more aggressive children (based on teachers’ ratings) are more likely to be missing data compared to other children while less aggressive children are more likely to be missing in the transition to middle school. Our explanation of the findings posits different mobility patterns based on poverty level. Families with low incomes are both more mobile and likely to be absent in the early elementary grades, and their children often receive higher ratings of aggressive behavior. The later mobility of better behaving children in middle school may reflect the general mobility of higher SES families to suburban schools. It would be possible to include poverty status and behavioral ratings into a model that examines mobility across time. By conditioning on these time dependent variables, a more elaborate ignorable missing data model could be developed. Nevertheless, these analytic
refinements of missing data mechanisms are likely to be too poorly known in advance to allow for improvements in the power calculations beyond those reported in the previous sections.
APPENDIX. MOMENT CALCULATIONS FOR TWO-STAGE DESIGNS UNDER MAR

Higher order moments can be written in terms of $\alpha$, $\beta$, and $\gamma$ defined in section 4. For cross products of two means calculated over the same pattern,

$$E(X_u^{(r)} - \mu_u)(X_v^{(r)} - \mu_v) = n_{r}^{-2} \sum_{i \neq j} E(X_{iu}^{(r)} - \mu_u)(X_{jv}^{(r)} - \mu_v) + n_{r}^{-2} \sum_{i} E(X_{iu}^{(r)} - \mu_u)(X_{iv}^{(r)} - \mu_v)$$

$$= n_{r}^{-1}(\alpha_u - \mu_u)(\alpha_v - \mu_v) + n_{r}^{-1}(\beta_{uv} - \alpha_u\alpha_v).$$

(10)

and for $r \neq r'$,

$$E(X_u^{(r)} - \mu_u)(X_v^{(r')} - \mu_v) = n_{r}^{-1}n_{r'}^{-1} E \sum_{i}(X_{iu}^{(r)} - \mu_u) \sum_{j}(X_{jv}^{(r')} - \mu_v)$$

$$= n_{r}^{-1}n_{r'}^{-1}(\alpha_u - \mu_u)(\alpha_v - \mu_v)$$

(11)

For cross products involving means and covariances,

$$E(X_u^{(r)} - \mu_u)(S_{uv}^{(r)} - n_r\sigma_{uv}) = n_{r}^{-2} \sum_{i \neq j} E(X_{iu}^{(r)} - \mu_u)[(X_{jv}^{(r)} - \mu_v)(X_{jw} - \mu_w) - \sigma_{vw}]$$

$$+ n_{r}^{-2} \sum_{i} E(X_{iu}^{(r)} - \mu_u)[(X_{iv}^{(r)} - \mu_v)(X_{iw} - \mu_w) - \sigma_{uv}]$$

$$= n_{r}^{-1}(\alpha_u - \mu_u)[\beta_{uv} - \mu_w\alpha_w - \mu_w\alpha_v + \mu_v\mu_w - \sigma_{uv}$$

$$+ \gamma_{uvw} - \mu_u[\beta_{uw} - \gamma_{uvw} - \alpha_u\alpha_v - \alpha_u\mu_w + \mu_u\mu_w]$$

$$- \mu_u[\beta_{uw} - \gamma_{uvw} - \alpha_u\alpha_v - \alpha_u\mu_w + \mu_u\mu_w]$$

$$+ \mu_u\mu_v(\alpha_v - \mu_v) + \mu_u\mu_w(\alpha_v - \mu_v) + \mu_v\mu_w(\alpha_u - \mu_u) - (\alpha_u - \mu_u)\sigma_{uw}$$

(12)

and for $r \neq r'$,

$$E(X_u^{(r)} - \mu_u)(S_{uv}^{(r')} - n_r'\sigma_{uv}) = n_{r}^{-1}n_{r'}^{-1} E \sum_{i}(X_{iu}^{(r)} - \mu_u) \sum_{j}(X_{jv}^{(r')} - \mu_v)(X_{jw} - \mu_w) - \sigma_{uv}$$

$$= (\alpha_u - \mu_u)[\beta_{uv} - \mu_w\alpha_w - \mu_w\alpha_v + \mu_v\mu_w - \sigma_{uv}]$$

(13)

For cross products of two covariances over the same pattern,

$$E(S_{uv}^{(r)} - n_r\sigma_{uv})(S_{uw}^{(r)} - n_r\sigma_{uw})$$

$$= \sum_{i \neq j} E[(X_{iu}^{(r)} - \mu_u)(X_{iv}^{(r)} - \mu_v) - \sigma_{uv}][(X_{jw}^{(r)} - \mu_v)(X_{jw}^{(r)} - \mu_t) - \sigma_{wt}]$$

$$+ \sum_{i} E[(X_{iu}^{(r)} - \mu_u)(X_{iv}^{(r)} - \mu_v) - \sigma_{uv}][(X_{iw}^{(r)} - \mu_u)(X_{iv}^{(r)} - \mu_v) - \sigma_{uv}]$$

$$= n_r(n_r - 1)[\beta_{uv} - \alpha_u\mu_t - \mu_u\alpha_v + \mu_u\mu_v - \sigma_{uv}][\beta_{wt} - \alpha_w\mu_t - \mu_w\alpha_t + \mu_w\mu_t - \sigma_{wt}]$$
is used to stratify for the second stage are given below. These have been computed using Mathematica.

\[ E(X^{(r)}_{iu} - \mu_u)(X^{(r)}_{iv} - \mu_v)(X^{(r)}_{it} - \mu_t) \]

\[ -n_r \sigma_{uv}[\beta_{wt} - \mu_w \alpha_t - \alpha_w \mu_t + \mu_w \mu_t] \]

\[ -n_r \sigma_{wt}[\beta_{uw} - \mu_u \alpha_v - \alpha_u \mu_v + \mu_u \mu_v] + n_r \sigma_{uv} \sigma_{wt}. \quad (14) \]

Also,

\[ E(X^{(r)}_{iu} - \mu_u)(X^{(r)}_{iv} - \mu_v) (X^{(r)}_{iw} - \mu_w)(X^{(r)}_{it} - \mu_t) = \delta_{uvw} \gamma_{wvt} - \mu_v \gamma_{uw} - \mu_w \gamma_{vt} - \mu_t \gamma_{uwv} + \mu_u \mu_v \beta_{wt} + \mu_u \mu_w \beta_{ut} + \mu_v \mu_w \beta_{ut} + \mu_v \mu_t \beta_{uw} + \mu_w \mu_t \beta_{uvw} - \mu_u \mu_v \mu_w \alpha_t - \mu_v \mu_w \mu_t \alpha_u - \mu_u \mu_w \mu_t \alpha_v - \mu_v \mu_w \mu_t \alpha_v + \mu_u \mu_v \mu_w \mu_t. \quad (15) \]

Finally, for the cross product of two covariances over different patterns \( r \) and \( r' \),

\[ E(S^{(r)}_{uv} - n_r \sigma_{uv})(S^{(r')}_{uw} - n_{r'} \sigma_{uw}) \]

\[ = \sum_i E[(X^{(r)}_{iu} - \mu_u)(X^{(r')}_{iw} - \mu_w) - \sigma_{uv} | \sum_j E[(X^{(r')}_{jw} - \mu_w)(X^{(r')}_{jt} - \mu_t) - \sigma_{wt}] \]

\[ = n_r n_{r'}[\beta_{uw} - \alpha_u \mu_t - \mu_u \alpha_v + \mu_u \mu_v - \sigma_{uv}][\beta_{uw} - \alpha_u \mu_t - \mu_u \alpha_v + \mu_u \mu_t - \sigma_{wt}], \quad (16) \]

The particular MAR design which is used will require calculating \((\alpha, \beta, \gamma, \delta)\), and using these in the equations above to obtain the required information matrix. The required moment calculations for two-stage designs where a cut-point \( d \) is used to stratify for the second stage are given below. These have been computed using Mathematica.

\[ E\{X_u | X_u > d\} = \mu_u - \frac{\phi\left(\frac{d - \mu_u}{\sigma_u}\right) \sigma_u}{\Phi\left(\frac{d - \mu_u}{\sigma_u}\right)}, \]

\[ E\{X_v | X_u > d\} = \mu_u - \frac{\phi\left(\frac{d - \mu_u}{\sigma_u}\right) \sigma_{vu}}{\Phi\left(\frac{d - \mu_u}{\sigma_u}\right) \sigma_u}, v \neq u \]

\[ E\{X_u^2 | X_u > d\} = \mu_u^2 + \sigma_u^2 - (d \sigma_u - \mu_u) \frac{\phi\left(\frac{d - \mu_u}{\sigma_u}\right)}{\Phi\left(\frac{d - \mu_u}{\sigma_u}\right)} \]

\[ E\{X_u X_v | X_u > d\} = \{\mu_u \mu_v + \sigma_{uv}\} - (\mu_u \sigma_u - d \sigma_{uv}) \frac{\phi\left(\frac{d - \mu_u}{\sigma_u}\right)}{\Phi\left(\frac{d - \mu_u}{\sigma_u}\right)} \]

\[ E\{X_v X_w | X_u > d\} = \]

\[-\mu_v^2 + \mu_u \mu_w + \mu_v \mu_w - \mu_u \mu_w \sigma_{uw} \sigma_{vu}^{-1} + \mu_u^2 \sigma_{uv}^2 \sigma_{uw}^2 - 2\]

\[ D R A F T \]
\[
\begin{align*}
&+ \left\{ -\mu_w \sigma_u - 2 \mu_v \sigma_u^2 \sigma_{uw} + 2 \mu_v \sigma_u \sigma_{uw}^2 \sigma_{uv} + \sigma_u \sigma_{uw} \sigma_{vw} + 3 \mu_u \sigma_u \sigma_{uw} \sigma_{vw} \\
&- 2 \mu_v \sigma_u \sigma_{uw} \sigma_{vw} - d \sigma_{uw}^2 \sigma_{uw} + \sigma_u \sigma_{uw} \sigma_{uw} - \mu_u \sigma_u \sigma_{uw} \right\} \frac{\phi\left( \frac{d-\mu_u}{\sigma_u} \right)}{\Phi\left( \frac{d-\mu_u}{\sigma_u} \right)} \\
&+ \mu_u \sigma_u \sigma_{uw} \sigma_{uw} + d \sigma_{uw} \sigma_v \sigma_{uw} + \mu_u \sigma_u \sigma_{uw} \sigma_{uw} - \mu_v \sigma_u^2 \sigma_{uw} \\
&+ \left( 2 \mu_u^2 \sigma_u^2 \sigma_{uw} - \mu_u \mu_v \sigma_u^2 \sigma_{uw} - \sigma_v^3 \sigma_{uw} + 2 \mu_u^2 \sigma_u^2 \sigma_{uv} - 2 \mu_u^2 \sigma_u \sigma_{uw} \sigma_{uv} \right) \sigma_u^{-1} \\
&+ \sigma_u \sigma_{uw} \sigma_v + \sigma_u \sigma_{uw}^2 \sigma_{uw} + \mu_u \sigma_u \sigma_{uw} \sigma_{uw} - 3 \mu_u^2 \sigma_u \sigma_{uw} \sigma_{uw} + \mu_u \mu_v \sigma_u \sigma_{uw} \sigma_{uw} - \mu_u \mu_v \sigma_u \sigma_{uw} \sigma_{uw} \\
&- \mu_v^2 \sigma_u \sigma_{uw} \sigma_{uw} - \sigma_u \sigma_v \sigma_{uw} - \mu_u \sigma_v \sigma_{uw} + \mu_u \mu_v \sigma_u \sigma_{uw} + \mu_u \mu_v \sigma_u \sigma_{uw} \\
&- \sigma_u \sigma_{uw} \sigma_v \sigma_{uw} + \mu_u \mu_v \sigma_u \sigma_{uw} + \mu_v^2 \sigma_u \sigma_{uw} - \mu_u \mu_v \sigma_u \sigma_{uw} \sigma_{uw} \right\} \left( \sigma_{uw}^2 - \sigma_u \sigma_v \right)^{-1}.
\end{align*}
\]

\[
E\{X_u^3|X_u > d\} = \mu_u^3 + 3 \mu_u \sigma_u^2 - \left( d^2 \sigma_u - d \mu_u \sigma_u - \mu_u^2 \sigma_u - 2 \sigma_u \right) \frac{\phi\left( \frac{d-\mu_u}{\sigma_u} \right)}{\Phi\left( \frac{d-\mu_u}{\sigma_u} \right)}.
\]

\[
E\{X_u^2X_v|X_u > d\} = \mu_u^2 \mu_v + \mu_v \sigma_u^2 + 2 \mu_u \sigma_u \sigma_{uw} - \left( d \mu_v \sigma_u - \mu_u \mu_v \sigma_u - d^2 \sigma_{uw} - 2 \sigma_u \right) \frac{\phi\left( \frac{d-\mu_u}{\sigma_u} \right)}{\Phi\left( \frac{d-\mu_u}{\sigma_u} \right)}.
\]

\[
E\{X_uX_vX_w|X_u > d\} =
\]

\[
- \left( \mu_u \mu_v^2 \right) + \mu_u^2 \mu_v + \mu_u \mu_v \mu_w + \mu_w \sigma_u^2 - \frac{\mu_u^2 \mu_w \sigma_{uw}}{\sigma_u} + \frac{\mu_u^3 \sigma_{uw}^2}{\sigma_u^2}
\]

\[
+ \left( \mu_v^2 \sigma_u - d \mu_w \sigma_u - \mu_u \mu_w \sigma_u - \mu_v \mu_w \sigma_u + \mu_u \mu_w \sigma_{uw} - \mu_v \sigma_{uw} \right) \frac{\phi\left( \frac{d-\mu_u}{\sigma_u} \right)}{\Phi\left( \frac{d-\mu_u}{\sigma_u} \right)}
\]

\[
+ \left( \left[ -\mu_v^2 \mu_u \sigma_{uw} \sigma_{uw} - \mu_u \mu_v \sigma_{uw}^2 \sigma_{uw} + 2 \mu_u \sigma_u \sigma_{uw}^2 \sigma_{uw} - 2 \mu_v \sigma_u \sigma_{uw}^2 \sigma_{uw} + \mu_u \sigma_u \sigma_{uw} \sigma_v + \mu_v \sigma_u^3 \sigma_{uw} \right] \\
- \mu_v^2 \mu_u \sigma_{uw} \sigma_v + \sigma_u \sigma_{uw}^2 \sigma_{uw} - 3 \mu_u \sigma_u \sigma_{uw}^2 \sigma_{uw} + 2 \mu_v \sigma_u \sigma_{uw}^2 \sigma_{uw} + \mu_u \sigma_u \sigma_{uw} \sigma_v + 3 \mu_u \sigma_u \sigma_{uw}^2 \sigma_{uw} \\
- \sigma_u \sigma_v \sigma_{uw} + \mu_u \sigma_u \sigma_{uw} + \mu_v \sigma_u \sigma_{uw} - 3 \mu_u \sigma_u \sigma_{uw} \sigma_{uw} - \mu_u \sigma_v \sigma_{uw} + \mu_u \sigma_v \sigma_{uw} \sigma_{uw} - \mu_u \sigma_v \sigma_{uw} \sigma_{uw} \\
+ \mu_u \sigma_v \sigma_u \sigma_{uw} + \mu_v \sigma_v \sigma_u \sigma_{uw} - 3 \mu_u \sigma_u \sigma_{uw} \sigma_{uw} - \mu_u \sigma_v \sigma_{uw} \right] \\
+ \left[ 2 \mu_u^3 \sigma_u \sigma_{uw} - \mu_u \mu_v \sigma_u \sigma_{uw}^2 \sigma_{uw} - \mu_u \sigma_u \sigma_{uw} \sigma_{uw} + 2 \mu_u^3 \sigma_u \sigma_{uw} \sigma_{uw} + 2 \mu_u^3 \sigma_u \sigma_{uw} \sigma_{uw} \right] \sigma_u^{-1}
\]

\[
- \left( \sigma_u \sigma_u \sigma_{uw} \sigma_v - d \sigma_u \sigma_u \sigma_{uw} \sigma_v - \mu_u \sigma_u \sigma_{uw} \sigma_{uw} + 3 d \mu_u \sigma_u \sigma_{uw} \sigma_{uw} \\
+ 3 \mu_u \sigma_u \sigma_{uw} \sigma_{uw} - 2 \mu_v \sigma_u \sigma_{uw} \sigma_{uw} - \mu_u \sigma_u \sigma_{uw} \sigma_{uw} \right) \\
- \sigma_u \sigma_v \sigma_{uw} \sigma_{uw} - d \mu_u \sigma_u \sigma_{uw} \sigma_{uw} - 2 \mu_u^2 \sigma_u \sigma_{uw}^2 \sigma_{uw} - \sigma_u \sigma_u \sigma_{uw}^2 \sigma_{uw} - 2 \sigma_u^2 \sigma_u \sigma_{uw}^2 \sigma_{uw}
\]
\[ + \sigma_u^2 \sigma_v \sigma_{vuv} + d \sigma_u \sigma_v \sigma_{uw} + \mu_u \sigma_u \sigma_v \sigma_{uw} - d \mu_u \sigma_u \sigma_v \sigma_{uw} - \mu_u^2 \sigma_u \sigma_v \sigma_{uw} - \mu_v \mu_u \sigma_u \sigma_v \sigma_{uw} \\
+ d^2 \sigma_{uv} \sigma_v \sigma_{uw} + d \mu_u \sigma_{uv} \sigma_v \sigma_{uw} + 2 \mu_u^2 \sigma_{uv} \sigma_v \sigma_{uw} + 2 \sigma_u^2 \sigma_{uv} \sigma_v \sigma_{uw} - d \mu_v \sigma_u^2 \sigma_{uw} \\
- \mu_u \mu_v \sigma_u^2 \sigma_{uv} - \mu_v^2 \sigma_u^2 \sigma_{uw} + \mu_u \mu_v \sigma_u \sigma_v \sigma_{uw} + d \mu_v \sigma_u \sigma_v \sigma_{uw} + \mu_u \mu_v \sigma_u \sigma_v \sigma_{uw} \\
+ \mu_v^2 \sigma_u \sigma_{uv} \sigma_{uw} - 2 d \mu_u \sigma_{uv}^2 \sigma_{uw} - 2 \mu_u^2 \sigma_{uv}^2 \sigma_{uw} \\
+ 2 d \mu_v \sigma_{uv}^2 \sigma_{uw} + \mu_u \mu_v \sigma_{uv}^2 \sigma_{uw} + \sigma_{uv}^3 \sigma_{uw} \left\{ \frac{\phi'(d-\mu_u)}{\phi(d-\mu_u)} \right\} \left( \sigma_{uw}^2 - \sigma_u \sigma_v \right)^{-1}. \] 

(17)
\[ E\{X_u^4 \mid X_u > d\} = \mu_u^4 + 6 \mu_u^2 \sigma_u^2 + 3 \sigma_u^4 - (d^3 \sigma_u - d^2 \mu_u \sigma_u - d \mu_u^2 \sigma_u - \mu_u^3 \sigma_u - 3 d \sigma_u^3 - 5 \mu_u \sigma_u^3) \frac{\phi(\frac{d-\mu_u}{\sigma_u})}{\Phi(\frac{d-\mu_u}{\sigma_u})}. \]

\[ E\{X_u^3 X_v \mid X_u > d\} = \mu_u^3 \mu_v + 3 \mu_u^2 \sigma_u \sigma_{uv} + 3 \mu_u \sigma_{uv} \sigma_u + 3 \sigma_u^3 \sigma_{uv} - (d^2 \mu_v - d \mu_u \sigma_{uv} - \mu_u^2 \sigma_{uv} - 2 \mu_v \sigma_u^3 - 3 d \sigma_u^2 \sigma_{uv} - 3 d \sigma_u^2 \sigma_{uv} - 3 \mu_u \sigma_u^2 \sigma_{uv}) \frac{\phi(\frac{d-\mu_u}{\sigma_u})}{\Phi(\frac{d-\mu_u}{\sigma_u})}. \]

\[ E\{X_u^2 X_v X_w \mid X_u > d\} = \]

\[ -\mu_u^3 \mu_w \sigma_{uw} \sigma_{uv} \sigma_u^{-1} - \mu_u \mu_w \sigma_{uw} \sigma_{uv} + \mu_u^2 \sigma_{uw}^2 \sigma_u^{-2} - \mu_u^2 \mu_v^2 \]

\[ + \mu_u^3 \mu_v + \mu_u \mu_v \mu_w - \mu_v^2 \sigma_u^2 + 3 \mu_u \mu_v \sigma_u^2 + \mu_v \mu_w \sigma_u^2 \]

\[ + \left[ d \mu_v^2 \sigma_u + \mu_u \mu_v^2 \sigma_u - d^2 \mu_v \sigma_u - d \mu_v \mu_w \sigma_u - \mu_v^2 \mu_w \sigma_u - d \mu_v \mu_w \sigma_u - d \mu_v \mu_w \sigma_u - \mu_v \mu_w \sigma_u \right] \frac{\phi(\frac{d-\mu_u}{\sigma_u})}{\Phi(\frac{d-\mu_u}{\sigma_u})} \]

\[ -2 \mu_v \sigma_u^3 + \mu_v \mu_w \sigma_u + \mu_v \mu_w \sigma_u - d \mu_v \mu_w \sigma_u - d \mu_v \mu_w \sigma_u - \mu_v \mu_w \sigma_u - \mu_v \mu_w \sigma_u \]

\[ - \mu_u^3 \mu_v \sigma_{uw} \sigma_{uv}^{-1} - \mu_u^3 \mu_v \sigma_{uw} \sigma_{uv} - \mu_u^2 \mu_v \sigma_{uw} \sigma_{uv} - 3 \mu_u \mu_v \sigma_{uv}^2 \sigma_u \sigma_{uw} - \mu_v^2 \sigma_{uw} \sigma_{uv} \]

\[ + \mu_u^3 \mu_v \sigma_{uw} \sigma_{uv}^{-1} + 6 \mu_u^2 \sigma_u \sigma_{uw}^2 \sigma_u^{-1} - 5 \mu_u \sigma_{uv} \sigma_u \sigma_{uw} - \mu_u^2 \sigma_{uw} \sigma_u \sigma_{uw} - \mu_u^3 \sigma_{uw} \sigma_u^{-2} - \sigma_u \sigma_{uw} \sigma_u \]

\[ + \mu_v \sigma_{uw} \sigma_{uv} \sigma_u + \sigma_u \sigma_{uw} \sigma_{uv} \sigma_u + \mu_v \sigma_{uw} \sigma_u \sigma_{uw} - 3 \mu_v \sigma_{uw} \sigma_u \sigma_{uw} + \mu_u^3 \sigma_{uw} \sigma_u \sigma_{uw} + 3 \mu_u \sigma_{uw} \sigma_u \sigma_{uw} \]

\[ -9 \mu_u \sigma_{uw} \sigma_u \sigma_{uw} - 5 \mu_u \sigma_{uw} \sigma_u \sigma_{uw} + \mu_v \sigma_{uw} \sigma_u \sigma_{uw} + 2 \mu_u^3 \sigma_{uw} \sigma_u \sigma_{uw} + 7 \mu_u^2 \sigma_u \sigma_{uw} \sigma_u \]

\[ + \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} + 3 \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} - \mu_u \sigma_u \sigma_{uw} \sigma_u - \sigma_u \sigma_{uw} \sigma_u - \mu_u \sigma_u \sigma_{uw} + \mu_v \sigma_{uw} \sigma_u \]

\[ + \mu_u \mu_v \sigma_u \sigma_{uw} - \mu_v \sigma_u \sigma_{uw} + 3 \mu_v \sigma_u \sigma_{uw} + \mu_v \sigma_u \sigma_{uw} - 2 \mu_u \sigma_u \sigma_{uw} \sigma_u^{-1} + \]

\[ 3 \mu_v \sigma_u \sigma_{uw} + \mu_v \sigma_u \sigma_{uw} - \mu_v \sigma_u \sigma_{uw} - \mu_v \sigma_u \sigma_{uw} + \mu_v \sigma_u \sigma_{uw} + \mu_v \sigma_u \sigma_{uw} + \mu_v \sigma_u \sigma_{uw} \]

\[ -3 \sigma_u \sigma_u \sigma_u \sigma_{uw} - 7 \mu_u \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} \sigma_u^{-1} \]

\[ (d^2 \mu_v - d \mu_u \sigma_{uv} - \mu_u^2 \sigma_{uv} - 2 \mu_u \sigma_u^3 - 3 d \sigma_u^2 \sigma_{uv} - 3 d \sigma_u^2 \sigma_{uv} - 3 \mu_u \sigma_u^2 \sigma_{uv}) \frac{\phi(\frac{d-\mu_u}{\sigma_u})}{\Phi(\frac{d-\mu_u}{\sigma_u})} \]

\[ - d \mu_u \sigma_u^2 \sigma_{uw} - d \mu_v \sigma_u^2 \sigma_{uv} + 5 \mu_u \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} + 3 d \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} + 2 \mu_u \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} \]

\[ + 2 \mu_u \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} + 2 \mu_u \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} - 2 \mu_u \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} + 2 \mu_u \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} - \mu_v \mu_u \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} \]

\[ - d \mu_u \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} - \mu_u \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} - d \mu_u \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} + \mu_u \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} + d \mu_u \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} \]

\[ + d \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} + d \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} - 5 \mu_u \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} - \mu_u \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} - 3 d \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} \]

\[ - d \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} - 2 \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} + 6 \mu_u \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} - 4 \mu_u \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} - 2 \mu_u \sigma_u \sigma_{uw} \sigma_u \sigma_{uw} \]
\[-d^2 \mu_u \sigma_{uv}^2 \sigma_{vw} - 2 d \mu_u^2 \sigma_{uv}^2 \sigma_{vw} - d^3 \sigma_{uv}^2 \sigma_{vw} + 3 d \mu_u^2 \sigma_u \sigma_{uv} \sigma_{vw} + 3 \mu_u^3 \sigma_u \sigma_{uv} \sigma_{vw}\]

\[-2 d^2 \mu_v \sigma_u \sigma_{uv} \sigma_{vw} - d \mu_u \mu_v \sigma_u \sigma_{uv} \sigma_{vw} - d^2 \sigma_u \sigma_{uv} \sigma_{vw} - d \mu_u \sigma_u \sigma_{uv} \sigma_{vw} + 3 d^2 \mu_u \sigma_u \sigma_{uv} \sigma_{vw}\]

\[-\mu_u^2 \sigma_u \sigma_{uv} \sigma_{vw} - \mu_u^2 \mu_v \sigma_u \sigma_{uv} \sigma_{vw} - d \sigma_u \sigma_{uv} \sigma_{uw} \sigma_v - d \mu_u \sigma_u \sigma_{uv} \sigma_{uw} \sigma_v - 4 \mu_u \sigma_u^2 \sigma_{uw}^2 \sigma_{uv}\]

\[+4 \mu_v \sigma_u^2 \sigma_{uw}^2 \sigma_{uv} + d \sigma_u^3 \sigma_{uw} + \mu_v \sigma_u^3 \sigma_{uw} + 2 \mu_v \sigma_u^3 \sigma_{uv} \sigma_{uw} - 2 d^2 \mu_u \sigma_u^2 \sigma_{uw}^2 - 2 \delta \mu_u^2 \sigma_{uv}^2 \sigma_{uw}\]

\[-2 \mu_u^3 \sigma_{uw}^2 \sigma_{uv} + 2 d^2 \mu_v \sigma_{uv}^2 \sigma_{uw} + d \mu_u \mu_v \sigma_{uv}^2 \sigma_{uw} + \mu_u^2 \mu_v \sigma_{uv}^2 \sigma_{uw} + d^2 \mu_v \sigma_u \sigma_{uv} \sigma_{uw}\]

\[+d \mu_u \mu_v \sigma_u \sigma_{uv} \sigma_{uw} + \mu_u^2 \mu_v \sigma_u \sigma_{uv} \sigma_{uw} + d \mu_v^2 \sigma_u \sigma_{uv} \sigma_{uw} + \mu_v \mu_u^2 \sigma_u \sigma_{uv} \sigma_{uw}\]

\[\phi\left(\frac{d - \mu_u}{\sigma_u}\right) \phi\left(\frac{d - \mu_v}{\sigma_u}\right) \left(\sigma_{uw}^2 - \sigma_u \sigma_v\right)^{-1}\]