Avoiding excessive enrollment of a single cohort in a clinical trial is prudent in order to avoid imbalances and to prevent one cohort from having a disproportionate influence on the results of a trial and perhaps even negating positive findings of the clinical trial. Numerical criteria are provided here to evaluate the expected influence of a large cohort as a function of both its size and the relative effect of interventions, in comparison to those of other groups. Measures of expected influence are obtained as a function of the parameters of the distribution of statistics measuring influence. Calculated numerical criteria for the binomial, continuous and time-to-event contexts are presented. Details of the application of this method and sensitivity analyses conducted during the planning stages of a multiple myeloma clinical trial are provided. Numerical criteria are derived under asymptotic conditions and thus results hold for large cohorts. The numerical criteria are easy to compute and are useful tools to assess possible detrimental effects of large cohorts during the design of a study or during enrollment prior to any un-blinding. The numerical criteria allow for a-priori sensitivity analyses of the likely influence of large cohorts under varying conditions.

Keywords: Influence of large sites, large strata, large cohorts, scaled inflation in influence, binomial, time-to-event, continuous, expected influence.

Introduction

The methods presented here address concerns of the negative impact that a large coherent but discrepant
patient group can have on the statistical integrity of the overall results of a randomized clinical trial when two arms are being compared. We refer to such a patient group as a strata or a cohort and define these as sub-populations whose subjects share the same attribute on a statistical factor (factors include investigator site, geographical location, demographic groups, genetic marker groups etc.). Many factors can lead to the occurrence of a large cohort including competitive enrollment [1, 2], a large research facility with extended geographical coverage, and as a consequence of authorship incentives for recruiting [3]. Uncontrolled competitive enrollment, can lead to imbalances through a large number of small sites as well as a few very large good enrollers. This is a common occurrence. Pierre [4] reported that as many as 50% of independent inpatient and outpatient sites in clinical trials enrolled only one or no patients, leaving fewer sites to carry the patient load. The National Institute for Mental Health [5] notes the following in its points to consider about recruitment and retention while preparing a clinical research study:

“As a study progresses, sites tend to fall into the categories of strong and weak enrollers. As enrollment goals slip behind, study leaders often decide to allow the strong enrollers to “over-recruit” in order to meet the overall sample size goal. However, because part of the site selection involves balance of geography, type of clinic, and racial/ethnic diversity, the post-design imbalance in enrollment across sites often impacts statistical analysis in a manner rarely considered.”

Large geographical locations in multi-regional trials can also result in a large strata or cohort [6]. The nature of a stratum can influence the efficacy one might expect at the strata. Determinants of differing efficacy include the use of community versus academic sites, differing countries with a differing standard of care and differing demographic and genetic profiles. Under these circumstances, the credibility and generalizability of the results of a clinical trial may be suspect due to the influence of a few large cohorts. An a-priori method to assess the potential influence of a site and thereby take action to prevent this from occurring is the subject of this report.

The consequence of and relevant analysis to account for differences across patient strata are a subject in popular texts such as Milliken and Johnson [7] for continuous data, Agresti [8] for binary data and Collett [9] for time-to-event data. A practical guide to the analysis of stratified data in this context is in Dmitrienko et. al. [10]. Multiple approaches are provided for the analysis of stratified data. For instance Dmitrienko et al. note that Type I analysis of continuous data requires equal weighting of observations while a Type III analysis, in the presence of differing treatment effects across strata, would involve equal weighting of strata. Dmitrienko et al. [10] provides some guidelines on appropriate methods but notes a lack of consensus. We approach this issue with the objective of pre-empting large discrepant strata effects when designing a clinical trial or during trial enrollment, rather than providing further options for analyses once the trial is fully-enrolled.

**Methods: Expected Influence of Large Strata in Clinical Trials**

Measures of influence have largely been used retrospectively as a post-hoc statistic. We intend to use our measures of influence prospectively, as an a-priori function of parameters, by obtaining a measure which characterizes how influential a stratum is likely to be in the analysis which will eventually be conducted, based on stratum size and an assessment of the degree to which results in the stratum are likely to differ from those outside the stratum. Measures of influence to be used in this report will include Cook’s [11] influence measure applied to continuous and binary data and the homogeneity of strata statistic for time-to-event data [12]. These influence statistics, in our context, are shown or noted to have chi-squared distributions with a non-centrality
parameter $\lambda$. For a chi-square with $n$ d.f and non-centrality parameter $\lambda$, the expected value is $n+2\lambda$ and the variance is $2(n+4\lambda)$. When there is no influence, $\lambda = 0$ and the expected value is $n$ and the expected variance is $2n$. These expected values provide useful parameters to gauge the likely influence of a large stratum if it differs from the rest of the trial. One particularly interesting parameter, a scaled inflation in influence, is obtained as the inflation in influence divided by the square root of the variance, or standard deviation, under no influence. This measure equals $2\lambda / \sqrt{2n}$. The influence chi-square statistics for continuous and binary data have 2 degrees of freedom (d.f) while that for time-to-event data has 1 d.f and hence the scaled inflation in influence are given by $\lambda$ and $\sqrt{2\lambda}$ respectively. Analogous to the frequent use of half a standard deviation as clinically significant [13] and the classification by Cohen [14] of effect sizes as low, medium and high for values of 0.2, 0.5 and 0.8 respectively, a good rule would consider influence as likely when the expected scaled inflation in influence is between 0.2 and 0.5, impactful when between 0.5 and 0.8 and serious when $>0.8$. Other measures of influence are the percent inflation in influence given by 100$\lambda$ and 200$\lambda$ for the statistics with 2 and 1 d.f., and the percent inflation in the variance of influence given by 200$\lambda$ and 400$\lambda$ respectively.

Scenarios are presented where a calculated expected inflation in influence is provided for continuous, Binomial and time-to-event endpoints. The derivations assume independence across strata and treatment, and asymptotic distributional assumptions.

**Binomial Case**

For the binomial case we look at the distribution of the following Cook’s influence statistic applied to proportions

$$
(\hat{p} - \hat{p}_{-i})' \hat{V}^{-1} (\hat{p} - \hat{p}_{-i}) = \begin{bmatrix}
\hat{p}_1 - \hat{p}_{1(-i)} \\
\hat{p}_2 - \hat{p}_{2(-i)}
\end{bmatrix} \begin{bmatrix}
\hat{V}_{11} & \hat{V}_{12} \\
\hat{V}_{21} & \hat{V}_{22}
\end{bmatrix}^{-1} \begin{bmatrix}
\hat{p}_1 - \hat{p}_{1(-i)} \\
\hat{p}_2 - \hat{p}_{2(-i)}
\end{bmatrix},
$$

where $p$ and $V$ are the vector of proportions and the variance covariance matrix respectively, $\hat{p}_1$ is the proportion of responders in the first arm and $\hat{p}_{1(-i)}$ is the proportion of responders in the first arm after excluding the $i^{th}$ stratum. $\hat{V}_{ii}$ is the variance $\text{Var}(\hat{p}_1 - \hat{p}_{1(-i)})$. Similarly for the second arm we have $\hat{p}_2$, $\hat{p}_{2(-i)}$ and $\hat{V}_{22}$. The parameters $\hat{V}_{21}$ and $\hat{V}_{12}$, the covariances, are equal to zero through the independence of observations in the two arms.

Let $n_{1(-i)}$ and $n_{ii}$ be the number of responders in the first arm for those outside stratum $i$ and those in stratum $i$ respectively. These can be considered the sum of successes in $N_{1(-i)}$ and $N_{ii}$ Bernoulli trials. For the sufficiently large $N_{1(-i)}$ and $N_{ii}$ being considered here $n_{1(-i)}$ and $n_{ii}$ can be considered asymptotically normal by the central limit theorem. The difference $\hat{p}_1 - \hat{p}_{1(-i)}$ can be written as a sum of linear functions of $n_{1(-i)}$ and $n_{ii}$ as follows

$$
\hat{p}_1 - \hat{p}_{1(-i)} = \frac{n_{1(-i)} + n_{ii}}{N_{1(-i)} + N_{ii}} - \frac{n_{1(-i)}}{N_{1(-i)}},
$$

hence this difference can be assumed to approximately normal. Note that the distributional properties of the single deletion Cook’s statistic are intractable as we cannot assume asymptotically normality of the difference $\hat{p}_1 - \hat{p}_{1(-i)}$. Here we exclude a large number of observations through the exclusion of an entire large stratum.

We determine the distributional properties of the multi-deletion Cook’s statistic in the discussion the follows using the approximate normality of the difference $\hat{p}_1 - \hat{p}_{1(-i)}$ due to the large number of observations deleted.
The variance $V_{11}$ is obtained as

$$V_{11} = \text{Var}(\hat{p}_1 - \hat{p}_{1(-i)}) = \text{Var}\left(\frac{n_{1(-i)}}{N_{1(-i)} + N_{li}} - \frac{n_{1(-i)}}{N_{1(-i)} + N_{li}}\right) = \frac{1}{(N_{1(-i)} + N_{li})^2} \text{Var}\left(\frac{n_{li} - n_{1(-i)}}{N_{1(-i)}}\right)$$

where $p_{li}$ and $\hat{p}_{1(-i)}$ are the proportions expected during the planning of the study in the large stratum $i$ and in other strata excluding the large stratum. $V_{22}$ can be obtained similarly.

The expected value of $\hat{p}_1 - \hat{p}_{1(-i)}$ can be obtained as

$$E(\hat{p}_1 - \hat{p}_{1(-i)}) = E\left(\frac{n_{1(-i)} + n_{li}}{N_{1(-i)} + N_{li}} - \frac{n_{1(-i)}}{N_{1(-i)}}\right) = \frac{E(n_{1(-i)}) + E(n_{li})}{N_{1(-i)} + N_{li}} - \frac{E(n_{1(-i)})}{N_{1(-i)}}$$

where

$$E = \frac{N_{1(-i)}p_{1(-i)} + N_{li}p_{li}}{N_{1(-i)} + N_{li}} - \frac{N_{1(-i)}p_{1(-i)}}{N_{1(-i)} + N_{li}} = \frac{N_{li}(p_{li} - p_{1(-i)})}{N_{1(-i)} + N_{li}}$$

The expected value of $\hat{p}_2 - \hat{p}_{2(-i)}$ can be obtained similarly.

Graybill [15] notes the following general results: For an $n$-dimensional random vector $Y$ distributed normally with mean vector $\mu$ and variance covariance matrix $\Sigma$, where $\Sigma$ has rank $n$, the random variable $U = Y^T \Sigma^{-1} Y$ is a chi-squared variable with $n$ d.f. and non-centrality factor $\lambda = 1/2 \mu' \Sigma^{-1} \mu$. Furthermore the mean of this chi-squared distribution is $n + 2\lambda$ and the variance is $2(n + 4\lambda)$.

We have an approximately normally distributed vector $Y = \hat{p} - \hat{p}_-$ for which the mean vector and variance covariance matrix have been evaluated. We can obtain the distribution of the influence statistic as an approximate chi-squared distribution ($\alpha = 2$ d.f.) with the mean and variance given through a non-centrality parameter $\lambda$ as $2 + 2\lambda$ and $2(2 + 4\lambda)$ respectively. The Graybill [15] result continues to hold under asymptotic conditions due to a result in Billingsley [16]. $\lambda$ can be computed as

$$\lambda = 0.5 * [E(\hat{p}_1 - \hat{p}_{1(-i)})] V_{11}^{-1} E(\hat{p}_1 - \hat{p}_{1(-i)}) = 0.5 * \left[\frac{(E(\hat{p}_1 - \hat{p}_{1(-i)}))^2}{V_{11}} + \frac{(E(\hat{p}_2 - \hat{p}_{2(-i)}))^2}{V_{22}}\right]$$

**Binomial Numerical Example**

Consider for example, a randomized two-arm study seeking to compare response rates of 75% and 60% between treatment and control with 80% power and a two sided significance level of 0.05. This requires a study with 152 subjects per arm. We would like to see when a stratum has a high scaled expected inflation in influence so that we can restrict the maximal enrollment in that stratum. The scaled increase in influence is in Table 1 under various scenarios. Scenarios 1, 3 and 5 have zero calculated scaled increase in influence as the
Measures of Expected Influence Provide Useful Constraints to Enrollment in Randomized Multi-Center Clinical Trials for Binomial, Continuous and Time-to-Event Endpoints

sites being evaluated have identical treatment and control response rates to those in the rest of the clinical trial. The other scenarios have a potentially influential stratum with anticipated response rates differing from those at other strata. In two of the scenarios (2 and 4) these strata are clearly influential with scaled increases in Influence exceeding 0.5. Scenario 6 is close to the 0.5 cut-off.

Table 1

<table>
<thead>
<tr>
<th>Scenario</th>
<th>% of Patients</th>
<th>N Per Arm</th>
<th>Treatment</th>
<th>Control</th>
<th>Scaled Inflation (Lambda)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A Non-influential Stratum</td>
<td>15</td>
<td>23</td>
<td>75</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>All Other Strata</td>
<td>85</td>
<td>129</td>
<td>75</td>
<td>60</td>
</tr>
<tr>
<td>2</td>
<td>Potentially Influential Stratum</td>
<td>15</td>
<td>23</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>All Other Strata</td>
<td>85</td>
<td>129</td>
<td>75</td>
<td>60</td>
</tr>
<tr>
<td>3</td>
<td>A Non-Influential Stratum</td>
<td>25</td>
<td>38</td>
<td>75</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>All Other Strata</td>
<td>75</td>
<td>114</td>
<td>75</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>Potentially Influential Stratum</td>
<td>25</td>
<td>38</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>All Other Strata</td>
<td>75</td>
<td>114</td>
<td>75</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td>A Non-Influential Stratum</td>
<td>35</td>
<td>53</td>
<td>75</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>All Other Strata</td>
<td>65</td>
<td>99</td>
<td>75</td>
<td>60</td>
</tr>
<tr>
<td>6</td>
<td>Potentially Influential Stratum</td>
<td>35</td>
<td>53</td>
<td>70</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>All Other Strata</td>
<td>65</td>
<td>99</td>
<td>75</td>
<td>60</td>
</tr>
</tbody>
</table>

The Continuous Case

For the continuous case consider the distribution of the Cook’s statistic

\[
(\bar{Y} - \bar{Y}_i)^' \hat{V}^{-1}(\bar{Y} - \bar{Y}_i) = \begin{bmatrix} \bar{Y}_1 - \bar{Y}_{1(i(-i))} \\ \bar{Y}_2 - \bar{Y}_{2(i(-i))} \end{bmatrix}' \begin{bmatrix} \hat{\nu}_{11} & \hat{\nu}_{12} \\ \hat{\nu}_{21} & \hat{\nu}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \bar{Y}_1 - \bar{Y}_{1(i(-i))} \\ \bar{Y}_2 - \bar{Y}_{2(i(-i))} \end{bmatrix}
\]

where \( \bar{Y}_i \) is the mean response in the first arm and \( \bar{Y}_{1(i(-i))} \) is the proportion of responders in the first arm after excluding the \( i^{th} \) stratum. \( V_{11} \) is the variance \( \text{Var}(\bar{Y}_1 - \bar{Y}_{1(i(-i))}) \). Similarly for the second arm we have \( \bar{Y}_2 \), \( \bar{Y}_{2(i(-i))} \) and \( V_{22} \). The parameters \( V_{21} \) and \( V_{12} \), the covariances, are as in the binomial case equal to zero through the independence of observations in the two arms.

Partition and order the \( N_1 \) observations \( Y_{ij} \) in the first arm into observations in the \( N_{1(i)} \) observations outside the potentially influential stratum \( i \) followed by the \( N_i \) observations in the stratum. Let the mean and variance of the \( N_{1(i)} \) observations be \( \mu_{1(i)} \) and \( \sigma^2_{1(i)} \) and that for the \( N_{ij} \) observations be \( \mu_i \) and \( \sigma^2_i \). Then the sums \( \sum_{j=1}^{N_{1(i)}} Y_{ij} \) and \( \sum_{j=N_{1(i)+1}}^{N_i} Y_{ij} \) under these assumptions of first and second order moments are can be considered approximately normal for the sufficiently large \( N_{1(i)} \) and \( N_i \) being considered on applying the central limit theorem. The difference \( \bar{Y}_1 - \bar{Y}_{1(i(-i))} \) is then approximately normal as it can be written as a sum of
Measures of Expected Influence Provide Useful Constraints to Enrollment in Randomized Multi-Center Clinical Trials for Binomial, Continuous and Time-to-Event Endpoints

linear functions of $\sum_{j=1}^{N_{l(-i)}} Y_{ij}$ and $\sum_{j=N_{l(-i)}}^{N_l} Y_{ij}$ as in the following

$$\bar{Y}_i - \bar{Y}_{l(-i)} = \frac{\sum_{j=1}^{N_{l(-i)}} Y_{ij}}{N_{l(-i)} + N_{li}} - \frac{\sum_{j=N_{l(-i)}}^{N_l} Y_{ij}}{N_{l(-i)} + N_{li}} = \left( \frac{\sum_{j=1}^{N_{l(-i)}} Y_{ij}}{N_{l(-i)} + N_{li}} - \frac{\sum_{j=1}^{N_{l(-i)}} Y_{ij}}{N_{l(-i)} + N_{li}} \right) + \left( \frac{\sum_{j=N_{l(-i)}}^{N_l} Y_{ij}}{N_{l(-i)} + N_{li}} \right)$$

Hence this difference can be assumed to be approximately normal. As in the binomial case, the multi-deletion Cook’s statistic’s distribution can be derived as the difference $\bar{Y}_i - \bar{Y}_{l(-i)}$ achieves an approximate normality due to the large number of observations. The variance $V_{11}$ is obtained as

$$V_{11} = \text{Var}(\bar{Y}_i - \bar{Y}_{l(-i)}) = \text{Var}\left( \frac{\sum_{j=1}^{N_{l(-i)}} Y_{ij}}{N_{l(-i)} + N_{li}} - \frac{\sum_{j=1}^{N_{l(-i)}} Y_{ij}}{N_{l(-i)} + N_{li}} \right) + \text{Var}\left( \frac{\sum_{j=N_{l(-i)}}^{N_l} Y_{ij}}{N_{l(-i)} + N_{li}} \right)$$

$$= \text{Var}\left( \frac{\sum_{j=1}^{N_l} Y_{ij}}{N_{l(-i)} + N_{li}} - \frac{N_{li} \sum_{j=1}^{N_{l(-i)}} Y_{ij}}{(N_{l(-i)} + N_{li})^2} \right) = \frac{1}{(N_{l(-i)} + N_{li})^2} \text{Var}\left( \frac{\sum_{j=1}^{N_l} Y_{ij}}{N_{l(-i)} + N_{li}} \right)$$

where $\sigma_{l(-i)}^2$ and $\sigma_{li}^2$ are the variances expected outside and within the potentially influential stratum during the planning of the study. $V_{22}$ can be obtained similarly.

The expected value of $\bar{Y}_i - \bar{Y}_{l(-i)}$ can be obtained as

$$E(\bar{Y}_i - \bar{Y}_{l(-i)}) = E\left( \frac{\sum_{j=1}^{N_{l(-i)}} Y_{ij}}{N_{l(-i)} + N_{li}} - \frac{\sum_{j=1}^{N_{l(-i)}} Y_{ij}}{N_{l(-i)} + N_{li}} \right) + E\left( \frac{\sum_{j=N_{l(-i)}}^{N_l} Y_{ij}}{N_{l(-i)} + N_{li}} \right)$$

$$= \frac{N_{l(-i)} \mu_{l(-i)}}{N_{l(-i)} + N_{li}} - \frac{N_{li} \mu_{li}}{N_{l(-i)} + N_{li}} = \frac{N_{li} (\mu_{li} - \mu_{l(-i)})}{N_{l(-i)} + N_{li}}$$
Measures of Expected Influence Provide Useful Constraints to Enrollment in Randomized Multi-Center Clinical Trials for Binomial, Continuous and Time-to-Event Endpoints

The expected value of \( \bar{Y}_2 - \bar{Y}_{2(-i)} \) can be obtained similarly.

Using the Graybill [15] result noted for the binomial case we can obtain the distribution of the influence statistic as an approximate chi-squared distribution \((n = 2 \text{ d.f.})\) with the mean and variance given through a non-centrality parameter \( \lambda \) as \(2+2\lambda \) and \(2(2+4\lambda)\) respectively. Since \( Y = \bar{Y} - \bar{Y}_i \) is an approximately normal vector with the mean vector and variance covariance matrix evaluated, \( \lambda \) can be computed as

\[
\lambda = 0.5 * E\left( (\bar{Y} - \bar{Y}_i)' \right) \left( \Sigma^{-1} \right) E\left( (\bar{Y} - \bar{Y}_i) \right) = 0.5 * \left[ \frac{\left( E\left( \bar{Y}_1 - \bar{Y}_{1(-i)} \right) \right)^2}{V_{11}} + \frac{\left( E\left( \bar{Y}_2 - \bar{Y}_{2(-i)} \right) \right)^2}{V_{22}} \right]
\]

Continuous Numerical Example

Consider a randomized two arm study seeking to compare a mean response of 86 and 90 in treatment and control with 80% power and a two sided significance level of 0.05. This requires a study with 143 subjects per arm for a common standard deviation of 12. We would like to see when a stratum has a high scaled expected inflation in influence so that we can restrict the maximal enrollment in that stratum. The scaled increase in Influence is in Table 2 under various scenarios.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>% of Patients</th>
<th>N Per Arm</th>
<th>Mean Response</th>
<th>STD of Response</th>
<th>Inflation in Influence (Lambda)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A Non-influential Site</td>
<td>15</td>
<td>21</td>
<td>86</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>All Other Sites</td>
<td>85</td>
<td>122</td>
<td>86</td>
<td>90</td>
</tr>
<tr>
<td>2</td>
<td>Potentially Influential Site</td>
<td>15</td>
<td>21</td>
<td>88</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>All Other Sites</td>
<td>85</td>
<td>122</td>
<td>86</td>
<td>90</td>
</tr>
<tr>
<td>3</td>
<td>Potentially Influential Site</td>
<td>15</td>
<td>21</td>
<td>88</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>All Other Sites</td>
<td>85</td>
<td>122</td>
<td>86</td>
<td>90</td>
</tr>
<tr>
<td>4</td>
<td>A Non-Influential Site</td>
<td>25</td>
<td>36</td>
<td>86</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>All Other Sites</td>
<td>75</td>
<td>107</td>
<td>86</td>
<td>90</td>
</tr>
<tr>
<td>5</td>
<td>Potentially Influential Site</td>
<td>25</td>
<td>36</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>All Other Sites</td>
<td>75</td>
<td>107</td>
<td>86</td>
<td>90</td>
</tr>
<tr>
<td>6</td>
<td>A Non-Influential Site</td>
<td>35</td>
<td>50</td>
<td>86</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>All Other Sites</td>
<td>65</td>
<td>93</td>
<td>86</td>
<td>90</td>
</tr>
<tr>
<td>7</td>
<td>Potentially Influential Site</td>
<td>35</td>
<td>50</td>
<td>88</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>All Other Sites</td>
<td>65</td>
<td>93</td>
<td>86</td>
<td>90</td>
</tr>
</tbody>
</table>

Scenarios 1, 4 and 6 have zero calculated scaled increase in influence as the sites being evaluated have identical treatment and control response means as the rest of the clinical trial. The other scenarios have potentially influential sites with anticipated response means and/or standard deviations differing from those at other sites. Scenario 5 clearly has an influential site as the effect size in the potentially influential site is zero. Scenarios 2 and 3 are an interesting contrast. The influential site in scenario 2 has large standard deviations and therefore adds little information to the pooled analysis and is therefore not as influential as the
more precise information provided by the potentially influential site in scenario 3. Scenario 7 is close to the 0.5 cut-off.

The Time-to-Event Case

Let \( U_{Li} \) be the log rank statistic given by

\[
U_{Li} = \sum_{j=1}^{r} \left( d_{1ji} - e_{1ji} \right),
\]

where \( r \) is the number of distinct event times, \( d_{1ji} \) is the number of deaths in the first arm at the \( j \text{th} \) event time in the potentially influential \( i \text{th} \) stratum, and \( e_{1ji} \) is the expected value of the number of deaths in the first arm at the \( j \text{th} \) event time in the \( i \text{th} \) stratum. Similarly \( U_{L(-i)} = \sum_{j=1}^{r} \left( d_{1(j-i)} - e_{1(j-i)} \right) \), where \((-i)\) denotes the deletion of the \( i \text{th} \) stratum (the influential stratum) when computing the number of events and the expected number of events. Let the log hazard ratio in the \( i \text{th} \) and \( i \)-deleted strata be denoted by \( \hat{\xi}_i \) and \( \hat{\xi}_{-i} \). Then the homogeneity of strata statistic

\[
Q = \text{var}(U_{Li}) (\hat{\xi}_i - \bar{\xi})^2 + \text{var}(U_{L(-i)}) (\hat{\xi}_{-i} - \bar{\xi})^2
\]

for

\[
\bar{\xi} = \text{var}(U_{Li}) \hat{\xi}_i + \text{var}(U_{L(-i)}) \hat{\xi}_{-i} + \left( \text{var}(U_{Li}) + \text{var}(U_{L(-i)}) \right).
\]

follows a chi-squared distribution with 1 d.f.[12]. A non-centrality parameter is obtained on substituting parameters for the estimates in the expression for \( Q \) as

\[
\lambda = 0.5 \left( V_{Li} (\hat{\xi}_i - \bar{\xi})^2 + V_{L(-i)} (\hat{\xi}_{-i} - \bar{\xi})^2 \right)
\]

with

\[
\bar{\xi} = V_{Li} \hat{\xi}_i + V_{L(-i)} \hat{\xi}_{-i} + \left( V_{Li} + V_{L(-i)} \right).
\]

The variances \( V_{Li} \) and \( V_{L(-i)} \) can be obtained approximately as the product \( r^*(1-r) \) of the randomization fractions \( r \) and \((1-r)\) multiplied by the expected number of events in the \( i \text{th} \) and \( i \)-deleted strata [17]. The variances \( V_{Li} \) is given as in the following using expressions for the expected number of events from [18,19].

\[
V_{Li} = r(1-r) \left[ \frac{a_{1i} \lambda_{1i}}{\lambda_{1i} + \gamma} \left[ A - \frac{e^{-(\lambda_{1i} + \gamma)A}}{\lambda_{1i} + \gamma} \right] \right] + \frac{a_{2i} \lambda_{2i}}{\lambda_{2i} + \gamma} \left[ A - \frac{e^{-(\lambda_{2i} + \gamma)A}}{\lambda_{2i} + \gamma} \right] - \lambda_{1i} \gamma A - \lambda_{2i} \gamma A
\]

for variable follow up. For the first and second arm in the \( i \text{th} \) site, \( a_{1i} \) and \( a_{2i} \) are the uniform accrual rates and \( \lambda_{1i} \) and \( \lambda_{2i} \) are the hazard rates. \( A \) is the accrual period and \( \gamma \) is the drop-out hazard rate and \( l \) is the calendar time \((A+\text{follow-up})\). For fixed follow-up the expression for \( V_{Li} \) can be obtained on taking the limit as \( A \) approaches zero as

\[
V_{Li} = r(1-r) \left[ \frac{N_{1i} \lambda_{1i}}{\lambda_{1i} + \gamma} \left[ e^{-(\lambda_{1i} + \gamma)l} \right] + \frac{N_{2i} \lambda_{2i}}{\lambda_{2i} + \gamma} \left[ e^{-(\lambda_{2i} + \gamma)l} \right] \right].
\]

Similar expressions can be obtained for the \( V_{L(-i)} \). The scaled inflation in influence is given by \( \text{SQRT}(2)\lambda \), the percent inflation in influence is 200\( \lambda \) and the percent inflation in variance is given by 400\( \lambda \).
Measures of Expected Influence Provide Useful Constraints to Enrollment in Randomized Multi-Center Clinical Trials for Binomial, Continuous and Time-to-Event Endpoints

**Time to Event Numerical Example**

Consider a randomized study seeking to compare treatment and control arms with median time-to-events of 10 and 14 months with 80% power and a two sided significance level of 0.05. This requires a study with 217 subjects per arm for a total of 277 events in a fixed follow-up study of 18 months. We would like to see when a stratum has a high scaled expected inflation in influence so that we can restrict the enrollment in that stratum. The scaled increase in Influence is included in Table 3 below, under various scenarios. Scenarios 1, 3 and 5 have zero calculated scaled increase in influence as the strata being evaluated have identical treatment and control response medians as the rest of the clinical trial. The other scenarios have potentially influential strata with anticipated median time-to-events differing from those at other sites. Scenario 6 clearly has an influential stratum as the hazard ratio in the potentially influential stratum is 1.09 compared to 1.4 in other strata. Scenarios 2 will not have a substantial qualitative influence in the eventual analysis of such a trial. Scenario 4 is close to the 0.5 cut-off.

**Table 3**

<table>
<thead>
<tr>
<th>Scenario</th>
<th>% of Patients</th>
<th>N Per Arm</th>
<th>Median Time to Event</th>
<th>Hazard Ratio</th>
<th>Inflation in Influence (sqrt(2)*Lambda)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A Non-influential Stratum</td>
<td>15</td>
<td>33</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>All Other Strata</td>
<td>85</td>
<td>184</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>Potentially Influential Stratum</td>
<td>15</td>
<td>33</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>All Other Sites</td>
<td>85</td>
<td>184</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>A Non-Influential Stratum</td>
<td>25</td>
<td>54</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>All Other Strata</td>
<td>75</td>
<td>163</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>Potentially Influential Stratum</td>
<td>25</td>
<td>54</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>All Other Strata</td>
<td>75</td>
<td>163</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>A Non-Influential Stratum</td>
<td>35</td>
<td>76</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>All Other Strata</td>
<td>65</td>
<td>141</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>Potentially Influential Stratum</td>
<td>35</td>
<td>76</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>All Other Strata</td>
<td>65</td>
<td>141</td>
<td>14</td>
<td>10</td>
</tr>
</tbody>
</table>

**Results: Application to a planned clinical study**

This real world case study involves a planned randomized clinical trial involving Lenalidomide maintenance therapy versus observation for patients with newly diagnosed multiple myeloma who have had 6 to 9 cycles of induction therapy (ClinicalTrials.gov identifier NCT02155634). The primary endpoint was progression free survival (PFS) post induction. A sample size of 172 subjects with 103 events were needed for 90% power, a two-sided significance level of 0.05, assuming 18 month accrual, 12 months of follow-up post accrual, a 15% censoring drop-out rate (corresponding to a drop out hazard rate of 0.01354), 15% non-evaluable rate for patients, and 10 and 20 months of median PFS in observation and maintenance arms respectively. Should enrollment be slower than expected in US sites, the study team was considering expanding enrollment to Canadian sites and expanding enrollment in Canada to 30% of total enrollment. Canadian sites are expected to be homogeneous across Canada but may differ from US sites due to the possibility of differing standard of care (other than study interventions) at Canadian sites due to the differing health care systems.
Sensitivity analyses were conducted under the assumptions of increased median survival in Canada of 23 months in the maintenance arm and 14, 15 and 16 months in the observation arm. For a 14 months median PFS in the observation arm, using the expressions in the time to event section above, the scaled inflation is expected influence was 0.126 and the percent inflation in expected influence was 17.8%. For 15 month median PFS these compute to 0.226 and 32% respectively and for 16 month median PFS these compute to 0.343 and 48.5%. Given the general clinical impression of the implausibility of the higher median PFS in the observation arm and the lower expected inflation of influence in the considered scenarios, the sensitivity analysis concluded that the impact of adding Canadian sites would not be detrimental. A similar sensitivity analysis, presuming lower outcomes in the Canadian cohort, using a 15 month PFS for treatment and a 9 month PFS for control leads to a scaled inflation of influence of 0.129 and a percent inflation of influence of 18.2%.

Discussion

This report provides tools to compute the expected inflation in influence due to large strata in the context of randomized clinical trials under various conditions, with comparisons between cohorts on binomial, time-to-event and continuous endpoints. Instead of a focus on the empirical values of an influence statistic once data has been collected and analyzed, this report looks at the distribution of the influence statistic while designing a study or during enrollment and computes measures of influence based on the parameters of such distributions. Similar non-central distributions have been obtained for continuous normally distributed data in the regression context supporting the derivations in this article [20]. The maximal strata size in a clinical trial can be determined for anticipated deviations of such strata from the norm at other strata in the clinical trial. Easily computable expressions are provided for the scaled inflation in expected influence. These expressions allow for a-priori sensitivity analyses of the likely influence of large cohorts under varying conditions. Derivations assume asymptotic results and the expressions provided would need to be validated for smaller strata sizes. An EXCEL calculator implementing calculations of the measures of inflated influence is available at http://resourcetepee.com/free-statistical-calculators/large-site-or-strata-influence-calculator/.

References

[12]. Marubini E and Valsecchi MG. Analyzing Survival Data from Clinical Trials and Observational Studies. Chichester, West


