Fallacies of Last Observation Carried Forward

John M. Lachin

The Biostatistics Center, The George Washington University, Rockville, MD, USA

Abstract

Background—Last Observation Carried Forward (*LOCF*) is a common statistical approach to the analysis of longitudinal repeated measures data where some follow-up observations may be missing. In a *LOCF* analysis, a missing follow-up visit value is replaced by (imputed as) that subject's previously observed value, i.e. the last observation is carried forward. The combination of the observed and imputed data are then analyzed as though there were no missing data.

Purpose—There have been numerous statistical demonstrations of faults of this approach. In 2012 the National Research Council's Panel on Handling Missing Data in Clinical Trials issued a report that raised concerns with the use of *LOCF*, and described alternate methods that offer greater statistical validity. Nevertheless, the method persists and its use is rampant. A search of the key word "LOCF" using Google Scholar yielded "about 1360" published citations during 2014 alone, the overwhelming majority presenting the results of scientific studies. However, there has not been a simple explanation of the statistical deficiencies of *LOCF*. Such a description is presented herein.

Results—A simple repeated measures model is described for quantitative observations at two times (e.g. 1 and 2-years), with complete values at 1-year that are used to impute by *LOCF* the missing values at 2-years under the missing completely at random (MCAR) assumption. This results in a mixture distribution of observed and imputed values at 2-years with mean and variance that are a function of the mixture of the 1 and 2-year distributions. The expressions show that *LOCF* is only unbiased when the distribution of the observed values at 1-year is exactly equal to the distribution of the missing values at 2-years, the latter of course being unknown.

Limitations—When the values at 2-years are not randomly missing, no simple expressions for the mean and variance of the mixture distribution are possible without additional unverifiable assumptions.

Conclusion—All analyses using *LOCF* are of questionable veracity, if not being outright specious (def: appearing to be true but actually false). It is hoped that future studies will make a more vigorous attempt to minimize the amount of missing data, and that more valid statistical analyses will be employed in cases where missing data occurs. *LOCF* should not be employed in any analyses.

Author for correspondence: John M. Lachin, The Biostatistics Center, The George Washington University, Rockville, MD, USA. jml@bsc.gwu.edu.

Introduction

Every statistical analysis attempts to provide an unbiased (i.e. reliable) assessment of study results. However, various factors may yield results that are biased (distorted) in some way, perhaps the most common being missing data, i.e. measurements that were potentially measurable or expected under the study design, but were not measured for some reason. The resulting bias arises not just because some data are missing, but more importantly, the reason that the values are missing, or the missing data mechanism.

Little and Rubin[1] review the types of missing data and the underlying mechanism. Data that are *missing completely at random* do not confer any bias, meaning that the data are missing purely by chance. The simplest case is data that are missing by design or administratively, e.g. when a subject who entered a 5 year study in the second year is administratively missing a year 5 measurement. *Missing at random* refers to data where the probability of being missing may may depend on other observed data, and adjusting for that data can provide an unbiased result. However, data that are *missing not at random* will introduce a bias in the analysis results regardless of statistical adjustment for other measured covariates.

A National Research Council panel reviewed the numerous statistical methods that have been proposed to address the problem of non-randomly missing data.[2–5] The report concludes that every method of analysis makes various assumptions that are technically unverifiable, and that the best approach is to minimize the extent of missing data.

Last Observation Carried Forward (LOCF) is among the simplest methods to account for missing data in a longitudinal analysis of repeated measures over time. Historically, its origins are obscure. There was no single statistical article that originally proposed the approach. In fact, there is not a single peer-reviewed statistical publication that describes general conditions under which *LOCF* provides a statistically unbiased result. Rather, many authors have been critical of *LOCF*.[6–8] Nevertheless, *LOCF* remains pervasive. A search for "LOCF" using Google Scholar yielded "about 1360" published citations during 2014 alone, the overwhelming majority presenting the results of scientific studies, all of which are of questionable veracity, if not being outright specious.

Herein I employ the well-established properties of a mixture of observations drawn from two separate distributions to compute the bias that can (will) be introduced by *LOCF* under a range of conditions.

One-sample Biases

While *LOCF* is most frequently applied to the analysis of two or more treatment groups, it is instructive to first consider the biases that can be introduced within a single group for data that are missing completely at random, and then for non-randomly missing data.

Missing completely at random data

Assume that the outcome is measured at 1 and 2-years in a single sample of subjects, such as an analysis of the change from baseline after 1 and 2-years of follow-up. The primary analysis consists of an estimate of the mean change at 2-years, with a 95% confidence interval and a one-sample test of significance of the null hypothesis that the mean change is zero using perhaps a paired t-test, or as herein, a simple large sample Z-test.

Assume that the measures at 1-year (X_1) and 2-years (X_2) are randomly drawn from distributions with respective means μ_1 and μ_2 and variances σ_1^2 and σ_2^2. Also assume that all of the X_1 measures are observed but some of the X_2 measures are missing completely at random; meaning that having a missing X_2 is not influenced by the value of X_1 or the true value of X_2. Technically X_1 and X_2 are drawn from a bivariate distribution, such as a bivariate normal, with correlation ρ_{12}. Let τ denote the fraction of X_2 values that are missing and imputed by the X_1 values, where τ is fixed. The Appendix shows that identical results apply when the fraction missing is random.

The set of observed and imputed 2-year values, denoted as $\tilde{X}_2$, then consists of a mixture of samples from two distributions with expected mean[9]

$$E(\tilde{X}_2)=\tilde{\mu}_2=\tau\mu_1+(1-\tau)\mu_2. \quad (1)$$

The expression for the variance of the observed/imputed values, $\tilde{\sigma}_2^2=V(\tilde{X}_2)$, is presented in the Appendix. Note that since the missing observations are completely random, the mean and variance of $\tilde{X}_2$ do not depend on the correlation of X_1 and X_2.

For a sample of n observed/imputed values with mean $\overline{\tilde{X}}_2$ and sample variance $\hat{V}(\tilde{X}_2)$, the one-sample Z-test of the null hypothesis $\mu_2 = 0$ is computed as $\tilde{z}=\overline{\tilde{X}}_2/\left[\hat{V}(\tilde{X}_2/n)\right]^{-1/2}$, with significance (one-sided) at level α if $z \tilde{} Z_{1-\alpha}$. For specified values of the mean μ_2 and variance $\tilde{\sigma}_2^2$ of the distribution of observed/imputed values, the probability of a statistically significant test result can be obtained as

$$P(\tilde{z} \geq Z_{1-\alpha})=1-\Phi\left[Z_{1-\alpha}-\frac{\tilde{\mu}_2\sqrt{n}}{\tilde{\sigma}_2}\right]. \quad (2)$$

A like expression provides the probability of significance for a Z-test with complete data for a given mean and variance μ_2 and σ_2^2.

While the distribution of the observed/imputed values $\tilde{X}_2$ depends on the variances of X_1 and X_2, the possible bias due to the *LOCF* imputation is principally a function of the mean value $\tilde{\mu}_2$. If $\tilde{\mu}_2 \lesseqgtr \mu_2$ then then analysis of the observed/imputed values is biased, leading to a probability of a significant test $\gtreqless$ the true probability.

The following computations assume $\tau = 0.30$ and $n = 100$.

The null hypothesis—First consider cases where the true mean change from baseline at 2-years is zero, or $E(X_2) = \mu_2 = 0$, for which the probability of significance is the desired level α, 0.05 one-sided herein.

Table 1 then shows the mean of the observed/imputed values ($\tilde{\mu_2}$, 30% imputed) and the corresponding variance $\tilde{\sigma}_2^2$. The first three cases assume that $\sigma_1^2 = \sigma_2^2 = 20$, i.e. the variance of the imputed values equals that of the true (but missing) values.

In case 1, $E(X_1) = \mu_1 = 0$ and $V(X_1) = 20$, the same as for X_2, and $\tilde{\mu_2} = 0$ and $\tilde{\sigma}_2^2 = 20$ so that no bias is introduced and the probability of a significant result is not affected. In other words, the value 0.05 is the theoretical type I error probability with no missing data, and no *LOCF*.

In case 2, there is a decrease from baseline at 1-year ($\mu_1 = -1$) that yields a decrease in the observed/imputed values at 2-years, $\tilde{\mu_2} = -0.3$ resulting from the 70% observed values having mean zero and the 30% imputed values having mean -1. While the 1 and 2-year values have the same variance ($\sigma_1^2 = \sigma_2^2 = 20$), the variance of the observed/imputed values is increased slightly to $\tilde{\sigma}_2^2 = 20.21$ because the distribution is now bimodal. Accordingly, the probability of a significant increasing value (one-sided) is decreased to only 0.010.

In case 3, there is an increase at 1-year ($\mu_1 = 1$) that yields an increase in the observed/ imputed values ($\tilde{\mu_2} = 0.3$) resulting from the 30% imputed values having mean 1. Again the variance is increased slightly owing to the bimodal distribution. Accordingly, the probability of a significant increase from baseline (one-sided) is inflated to 0.164. That is, an observed *LOCF* p-value of exactly 0.05 corresponds to a true false positive probability of 0.164, over 3 times the nominal *p*-value.

Case 4 is the same as case 3 except that the X_1 variance is also higher ($\sigma_1^2 = 30$) than the X_2 variance, so that the variance of the observed/imputed values is increased further to $\tilde{\sigma}_2^2 = 32.21$. The resulting inflation in the false positive probability is slightly less at 0.153, but still severely biased. If the variance at 1-year was lower than at 2-years ($\sigma_1^2 = 10$) the variance of the observed/imputed values is less than the true variance, $\tilde{\sigma}_2^2 = 17.21$, and the false positive probability is increased to 0.178.

By construction, the test will have type I error probability $\alpha = 0.05$ whenever $\tilde{\mu_2} = 0$, regardless of the value of the variance $\tilde{\sigma}_2^2$. Thus, under the null hypothesis of no change from baseline at 2-years, *LOCF* imputation will only be unbiased when the mean value at 1-year (μ_1) is also zero.

The alternative hypothesis—Now assume that the true change from baseline at 2-years is $\mu_2 = 1$ with variance $\sigma_2^2 = 20$ and $n = 100$, as above, in which case the probability of a statistically significant one-sided test at p 0.05 is 0.723, or the power of the study with no missing data.

In case 5, the mean and variance at 1-year are the same as at 2-years, $E(X_1) = \mu_1 = 1$ and $V(X_1) = 20$, so that the mean and variance of the observed/imputed values is unaffected, $\tilde{\mu}_2 = 1$ and $\tilde{\sigma}_2^2 = 20$, and no bias is introduced. Accordingly, the probability of a significant result is not affected, which in this case is the power of the original study (0.723) with no missing data, and no *LOCF*.

In case 6, there is a decrease from baseline at 1-year ($\mu_1 = -1$). When the 30% imputed values from this distribution are mixed with the 70% observed values with mean $\mu_2 = 1$, the mean change in the observed/imputed values is diluted so that $\mu e_2 = 0.4$, resulting in a markedly reduced probability of a significant result of 0.221. Owing to the bimodal distribution, the variance of the observed/imputed values is increased slightly to $\tilde{\sigma}_2^2 = 20.84$ that has a trivial effect on the probability of a significant result.

In case 7, there is no change from baseline at 1-year ($\mu_1 = 0$) so that again the mean of the observed/imputed values is diluted, $\tilde{\mu}_2 = 0.7$ and the variance increased slightly, so that the probability of a significant result is reduced to 0.465.

In case 8, there is a greater change from baseline at 1-year ($\mu_1 = 2$) than at 2-years so that the mean of the mixture of observed/imputed values is now greater than that of the true 2-year values, $\tilde{\mu}_2 = 1.3$, and the variance increased slightly, so that the probability of a significant result is substantially increased to 0.893. In case 9 the variance of the 1-year values ($\sigma_1^2 = 30$) is also greater than that of the true 2-year values that dampens slightly the increased probability of a significant result to 0.854.

Note that these latter probabilities have not been labeled "power", since they will not correspond to the theoretical probability of rejecting the null hypothesis under the parametric settings of the original study design before the missing data occurs. For the assumptions herein, $\mu_2 = 1$, $\sigma_2^2 = 20$ and $n = 100$, the power is 0.723 as shown above. Thus, under case 8, the proper interpretation is that the bias in the observed/imputed data introduced by *LOCF* increases the probability of a significant result to 0.893. Clearly some of the inflation/deflation in the probability of significance under the alternative is a result of the inflation/deflation in the type 1 error probability under the null, but not all.

Figure 1 shows the expected mean value of the observed/imputed values ($\tilde{\mu}_2$) under the null and alternative hypotheses ($\mu_2 = 0$, $\mu_2 = 1$, respectively) over a range of mean values at 1-year (μ_1) and variances $\sigma_1^2 = (10, 20, 30)$. From (1), the expectation at 2-years ($\tilde{\mu}_2$) is a simple linear function of the mean value at 1-year (μ_1). Figure 2 displays the variance of the observed/imputed values ($\tilde{\sigma}_2^2$) under the null and alternative hypotheses for 1-year variances $\sigma_1^2 = (10, 20, 30)$ and 2-year variance $\sigma_2^2 = 20$ over the range of 1-year mean values. While the variance expression (A.1) is also a function of the mean values, the Figure shows that the variance of X_2 is principally a function of the variance of the 1-year values used for the imputation. Figure 3 then shows the probability that a one-sample test of the observed/imputed values at 2-years would be statistically significant (nominally) under the null and alternative hypotheses (lower and upper curves, respectively) as a function of the mean and

variance of the 1-year values. Under the null and alternative hypotheses, differences in the variances have little effect on the bias introduced by *LOCF*.

Under the null hypothesis $\mu_2 = 0$, *LOCF* imputation only yields an unbiased result when the mean value at 1-year (μ_1) is also zero (and the variances equal), in which case the probability of a falsely significant result (α) is the desired 0.05 (one-sided). If $\mu_1 < 0$ then α decreases, and if $\mu_1 > 0$ then α increases, possibly substantially.

Under the alternative hypothesis that there truly is a positive change from baseline at 2-years ($\mu_2 = 1$, the upper set of curves), *LOCF* yields an unbiased result only when the mean value at 1-year (μ_1) is also 1 (and the variances equal), in which case the probability of a significant result equals the power of the initial study design (0.723, indicated by the dot). *LOCF* introduces a negative bias when there is a lesser effect at 1-year than at 2-years ($\mu_1 < 1$), and a positive bias when if there is a greater effect ($\mu_1 > 1$), so that the expected mean difference at 2-years is an under- or overestimate, respectively, of the true change, and the probability of a significant result is less or greater than that provided by the original design.

Non-randomly missing data

Non-randomly missing data occur when the likelihood that a given datum may be missing, and the true but unobserved value depends on other information (data) that may not have been observed, including the missing X_2 value itself. In this case, the properties of an analysis of the simple change from baseline are more complicated to assess and describe. In terms of the above example, under the alternative hypothesis the true mean value of $\mu_2 = 1.0$ would apply to the 70% observed values, but yet some other mean would apply to those non-randomly missing, say μ_2^*, because the m missing observations are not a random subset of the true X_2 values. Likewise, owing to the correlation of the 1 and 2-year values, the subset of 1-year values used for imputation are a non-random subset of the X_1 values with some unknown mean μ_1^*. Then from the above equation, the mean of the observed/imputed values becomes

$$E(\tilde{X}_2) = \tilde{\mu}_2 = \tau \mu_1^* + (1 - \tau) \mu_2^*. \quad (3)$$

Even when the true 1-year values have the same mean as the true 2-year values ($\mu_1 = \mu_2 = 1$) then $\tilde{\mu}_2 \neq 1$ and the study results are biased.

Two Sample Biases

The most common application of *LOCF* is to the comparison of two or more groups at a given point in time, such as at 2-years herein. Again consider the simplest case where the 2-year missing values are missing completely at random within the two groups labeled as *A* and *B*. We also assume that the fraction missing (τ) is the same in the two groups. The values in the *i*th group at the *j*th time are then distributed with mean values μ_{ij} and variances σ_{ij}^2, with mean difference at the *j*th time $\triangle_j = \mu_{aj} - \mu_{bj}$, $i = a, b, j = 1, 2$.

Then after *LOCF* imputation, the observed/imputed values within each group have means

$$\begin{aligned} E(\tilde{X}_{a2}) &= \tilde{\mu}_{a2}=\tau\mu_{a1}+(1-\tau)\mu_{a2} \\ E(\tilde{X}_{b2}) &= \tilde{\mu}_{b2}=\tau\mu_{b1}+(1-\tau)\mu_{b2}. \end{aligned} \quad (4)$$

Denote the sample mean difference between the two groups in the observed/imputed values as $\tilde{D}_2$ with expectation

$$\tilde{\Delta}_2=E(\tilde{D}_2)=\tilde{\mu}_{a2}-\tilde{\mu}_{b2}=\tau\Delta_1+(1-\tau)\Delta_2 \quad (5)$$

s a function of the mean differences (Δ_1, Δ_2) between the groups at 1 and 2-years, respectively. The Appendix also presents the expression for the variance of the mean difference, say $V(\tilde{D}_2)$. A test of the null hypothesis of no difference between groups H_0: Δ_2 = 0 is then provided by the Z-test with value z equal to the sample mean difference $\tilde{D}_2$ relative to it SE. The probability of a significant result is then provided by

$$P(\tilde{z} \geq Z_{1-\alpha/2})=1-\Phi\left[Z_{1-\alpha/2}-\frac{\tilde{\Delta}_2}{V(\tilde{D}_2)^{1/2}}\right] \quad (6)$$

For illustration, assume a common variance ($\sigma_{ij}^2=\sigma^2$) among the values at 1 and 2-years within the two groups, so that the expected difference between groups at 2-years, and the probability of a significant result, depend principally on the expected differences among the observed values at years 1 and 2, the Δ_1 and Δ_2. Sample computations are presented in Table 2 with similar results to those shown in Table 1.

Figure 4 presents the probability of a statistically significant test ($\tilde{z}$) based on the values for Δ_1 and Δ_2 assuming that n = 200 in each group and σ^2 = 20. As was the case for a single sample, under the null hypothesis Δ_2 = 0, then the analysis of the observed/imputed values is unbiased if Δ_1 = 0 as well. Under the alternative hypothesis of a non-zero mean difference at 2-years ($\Delta_2 = \delta \neq 0$), then the analysis of the observed/imputed values is unbiased when the mean difference at 1-year equals that at 2-years ($\Delta_1 = \delta \neq 0$). Otherwise, the analysis based on the observed/imputed values at 2-years is positively biased $E(\tilde{D}_2) > \Delta_2$ when $\Delta_1 > \delta$ and negatively biased when $\Delta_1 < \delta$. Again, since the true mean differences Δ_1 and Δ_2 are unknown, all analyses using *LOCF* are suspect.

Note that when the difference between the two groups equals the change within a single group ($\mu_2 = \Delta_2$), and the variance is the same, then the probability of significance with $2n$ per group is the same as that for a single group of size n. Thus, the power to detect a difference of 1 in this case also equals 0.723.

Longitudinal Analyses

Now consider that a more general longitudinal mixed model is used to compare the overall group difference in the 1 and 2-year values combined (e.g. LSMEANS), possibly adjusted for a time effect (1 versus 2-years) and the baseline value, taking into account the intercorrelation among the repeated measures. The covariance structure then has three

components, the variance of the values at 1 and 2-years (σ_1^2, σ_2^2) and the covariance σ_{12}. If some of the X_2 measures are missing and are imputed using 1-year values, then the covariance will be inflated.

The reason is simple. For $\tau = 0.3$, the 70% of subjects with observed 2-year values have correlation ρ_{12} between the 1 and 2-year values, but the 30% with imputed values have correlation 1.0 since each imputed value at 2-years has exactly the same value as that subject's 1-year value. Thus, the correlation between the observed/imputed 2-year values and the observed 1-year values is inflated. See the Appendix.

In a repeated measures analysis, the higher the intercorrelaton among the longitudinal measures, the less the total information provided. Thus the inflation in the observed/imputed correlation reduces the information in the data and reduces power.

It should also be noted that *LOCF* is a form of "single-point imputation" in which the "estimate" of the expected value of the missing value is imputed, but without consideration of the conditional variance. Even if the conditional estimate is unbiased, such methods generally underestimate the overall variance and lead to inflation in the type I error probability.

Discussion

The origins of last observation carried forward are obscure. It was perhaps first used in analyses submitted to the FDA, and the method quickly became pervasive. However, there has not been, to this author's knowledge, a single refereed statistical publication demonstrating that the method is in general valid (unbiased). Rather, numerous papers have shown, mostly by statistical simulation, that the method can be biased.

Herein, I use a simple mixture model to show that the analysis of the means of a mixture of observed and imputed missing data using *LOCF* is biased. The only condition where *LOCF* is unbiased is when the missing data occurs completely by chance and the data used as the basis for the *LOCF* imputation has exactly the same distribution as does the unknown missing data. Since it can never be proven that these distributions are exactly the same, all *LOCF* analyses are suspect and should be dismissed.

In some applications *LOCF* is described as conservative or a "worst case" analysis, such as when patients tend to improve over time so that the 1-year values, and the imputed 2-year values, tend to be worse than the true (but missing) 2-year values. However, the bias introduced by *LOCF* is a function of the *difference* between two groups at 1-year, not just the individual group mean values at 1-year. Even though the values at 1-year tend to be worse, the mean difference at 1-year will likely be different from the true difference at 2-years, leading to a biased estimate of the treatment effect at 2-years.

LOCF is commonly applied in studies of new drugs to treat type 2 diabetes. Per FDA guidance,[10] the sponsor must show that the new drug can lower blood glucose levels, as measured by HbA1c %, over a period of 6 to 12 months versus placebo. Typically, the placebo group HbA1c rises slowly whereas the treated group HbA1c decreases substantially

(improves) at 3 months but then worsens (increases) over time.[11,12] Many studies also include a provision to terminate therapy if the HbA1c rises. While this may lead to missing data in both groups, the potential bias is clearly greater in the active group since *LOCF* replaces the missing data with earlier values that tend to be lower (better) than the true missing values, thus overstating the benefits of the treatment.

Unfortunately, many FDA guidances explicitly recommend that submissions use *LOCF* to address missing data. Further, numerous articles are published every year that claim an effect based on *LOCF* analyses, all of which are likely biased to some degree. Regulatory agencies and journal editors (and reviewers) should be critical of any study with a substantial fraction of missing data, and should be highly skeptical of the veracity of any results and pursuant claims based on *LOCF* analyses.

The NRC report,[2] commissioned by the FDA, describes other approaches to deal with missing data that can provide a less biased assessment of outcomes. However, all such methods require assumptions about the properties of the missing data that are inherently untestable. Accordingly, the NRC report stresses the importance of avoiding missing data to the extent possible. To this end, Lachin[13] describes the "intention to treat design" in which the complete follow-up of all subjects is encouraged.

In summary, the well-known statistical properties of the mixture of two distributions are employed to demonstrate that *LOCF* analyses can introduce a positive or negative bias that can grossly inflate or deflate, respectively, the probability of a statistically significant test result under either the null or alternative hypothesis. Accordingly, without exception, all analyses using *LOCF* are suspect and should be dismissed. Statistically, last observation carried forward is specious (def: appearing to be true but actually false).

Acknowledgments

References

1. Little, RJ.; Rubin, DB. Statistical analysis with missing data. 2. New York: Wiley; 2002.

2. National Research Council. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press; 2010. The prevention and treatment of missing data in clinical trials.

3. Little RJ, D'Agostino R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials. N Engl J Med. 2012; 367:1355–60. [PubMed: 23034025]

4. O'Neill RT, Temple R. The prevention and treatment of missing data in clinical trials: An FDA perspective on the importance of dealing with it. Clinical Pharmacology & Therapeutics. 2012; 91:550–554. [PubMed: 22318615]

5. Ware JH, Harrington D, Hunter DJ, D'Agostino RB Sr. Missing data. N Engl J Med. 2012; 367:1353–4.

6. Kenward MG, Molenberghs G. Last observation carried forward: A crystal ball? Journal of Biopharmaceutical Statistics. 2009; 9:872–888. [PubMed: 20183449]

7. Siddiqui O, Ali MW. A comparison of the random-effects pattern mixture model with last-observation-carried-forward (*LOCF*) analysis in longitudinal clinical trials with dropouts. J Biopharm Stat. 1998; 8:545–563. [PubMed: 9855033]

8. Veberke, G.; Molenberghs, G.; Bijnens, L.; Shaw, D. Linear mixed models in practice. New York: Springer; 1997.

9. Cohen AC. Estimation in mixtures of two normal distributions. Technometrics. 1967; 9:15–28.

10. Food and Drug Administration (Center for Drug Evaluation and Research). [accessed January 4, 2015] Guidance for industry: Diabetes mellitus: developing drugs and therapeutic biologics for treatment and prevention. 2008. http://www.fda.gov/downloads/Drugs/Guidances/ucm071624.pdf

11. UK Prospective Diabetes Study (UKPDS) Group. U.K. Prospective Diabetes Study 16. Overview of 6 years' therapy of type II diabetes: a progressive disease. Diabetes. 1995; 44:1249–1258. [PubMed: 7589820]

12. Kahn SE, Haffner SM, Heise MA, Herman WH, Holman RR, Jones NP, Kravitz BG, Lachin JM, O'Neill MC, Zinman B, Viberti G. for the ADOPT Study Group. Glycemic durability of rosiglitazone, metformin, or glyburide monotherapy. N Engl J Med. 2006; 355:2427–2443. [PubMed: 17145742]

13. Lachin JM. Statistical considerations in the intent-to-treat principle. Controlled Clinical Trials. 2000; 21:167–189. [PubMed: 10822117]

Appendix

A. Variance of Observed/Imputed Values Under *LOCF*

For a single sample, the variance of the combination of observed plus imputed (*LOCF*) values at time 2 is provided by

$$V(\tilde{X}_2) = \tilde{\sigma}_2^2 = \tau \left[(\mu_1 - \tilde{\mu}_2)^2 + \sigma_1^2 \right] + (1 - \tau) \left[(\mu_2 - \tilde{\mu}_2)^2 + \sigma_2^2 \right] \quad \text{(A.1)}$$

and the variance of the mean of the observed/imputed values is provided by

$V\left(\tilde{\overline{X}}_2\right) = \tilde{\sigma}_2^2 / n$. Note that this variance does not depend on the correlation of the X_1 and missing X_2 values because under the missing completely at random assumption, the *LOCF*-imputed values are obtained by sampling from the marginal distribution of X_1. Thus the resulting $\tilde{X}_2$ values are a mixture of values from two distributions, rather than a linear combination.

For the case of two independent groups of observed/imputed values, the variances of the observed/imputed values within each group are

$$V(\tilde{X}_{a2}) = \tilde{\sigma}_{a2}^2 = \tau \left[(\mu_{a1} - \tilde{\mu}_{a2})^2 + \sigma_{a1}^2 \right] + (1 - \tau) \left[(\mu_{a2} - \tilde{\mu}_{a2})^2 + \sigma_{a2}^2 \right]$$
$$V(\tilde{X}_{b2}) = \tilde{\sigma}_{b2}^2 = \tau \left[(\mu_{b1} - \tilde{\mu}_{b2})^2 + \sigma_{b1}^2 \right] + (1 - \tau) \left[(\mu_{b2} - \tilde{\mu}_{b2})^2 + \sigma_{b2}^2 \right].$$

and the variance of the mean difference with sample size *n* per group is

$$V(\tilde{D}_2) = \frac{V(\tilde{X}_{a2})}{n} + \frac{V(\tilde{X}_{b2})}{n}. \quad \text{(A.3)}$$

The above model, and the computations presented herein, assume that the fraction missing (τ) is pre-specified or fixed. In practice, however, this fraction could be the expectation of a Bernoulli random variable Y to denote missing versus not, again assuming missing completely at random, so that Y is jointly independent of X_1 and X_2. In this case it follows that $E(\tilde{X_2})$ remains as presented in (2). Decomposing $V(\tilde{X_2}) = E\,[V(X_2|Y)] + V\,[E\,(\tilde{X_2}|Y)]$ yields the same expression as (A.1).

B. Correlation of Measures Under *LOCF*

Under *LOCF* the covariance of the observed values at year 1 and the observed/imputed values at year 2 is

$$Cov(X_1, \tilde{X}_2) = \tilde{\sigma}_{12} = (1-\tau)\,Cov(X_1, X_2) + \tau\,Var(X_1)$$
$$= (1-\tau)\sigma_{12} + \tau\sigma_1^2 \qquad \text{(A.4)}$$

Since $\sigma_{12} \leq \sigma_1\sigma_2 \cong \sigma_1^2$ then

$$\tilde{\sigma}_{12} \geq (1-\tau)\sigma_1\sigma_2 + \tau\sigma_1^2 \cong \sigma_1^2 \geq \sigma_{12}. \quad \text{(A.5)}$$

Then the correlation is

$$\tilde{\rho}_{12} = \frac{\tilde{\sigma}_{12}}{\sigma_1\tilde{\sigma}_2} \geq \frac{\sigma_{12}}{\sigma_1\tilde{\sigma}_2}. \quad \text{(A.6)}$$

Since $\tilde{\sigma}_2^2 \cong \sigma_2^2$ when $\sigma_1^2 \cong \sigma_2^2$, then except for highly irregular situations, $\tilde{\rho}_{12} \quad \rho_{12}$.
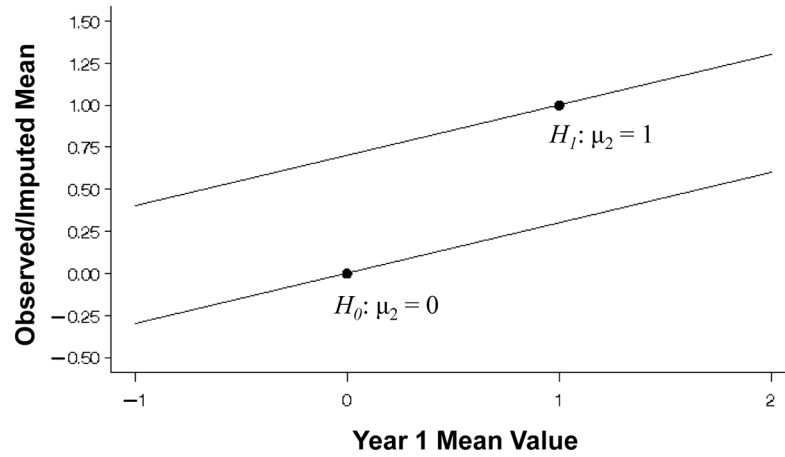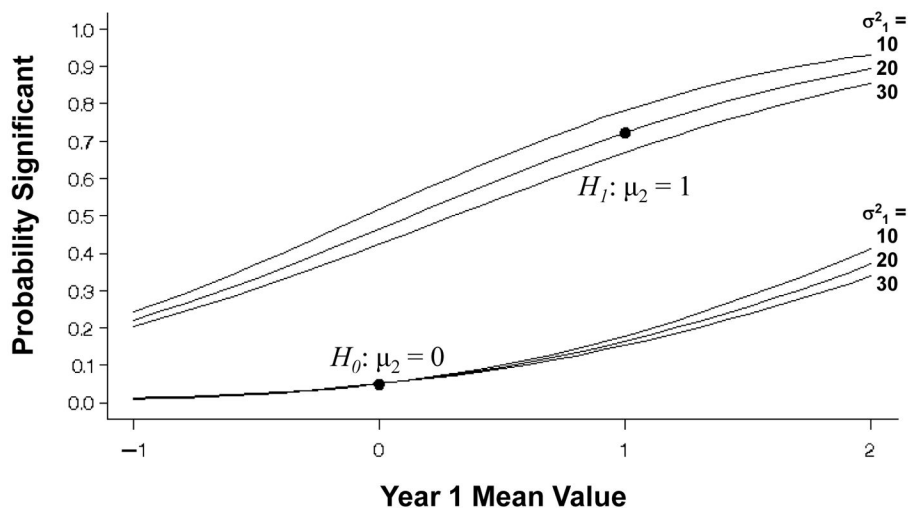
Figure 1.
Expected mean change from baseline $(\tilde{\mu}_2)$ in a one-sample analysis at 2-years from equation (1) under the null and the alternative hypotheses ($\mu_2 = 0$ or 1, respectively), for a single sample of $n = 100$ with 30% missing that are replaced by values at 1-year with mean μ_1 ranging from -1 to 2. The correct values with no missing data are designated by the dots.
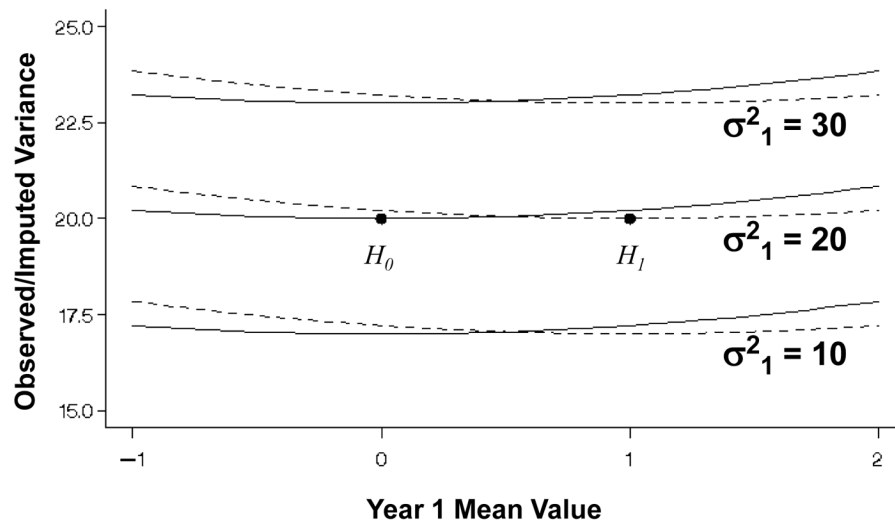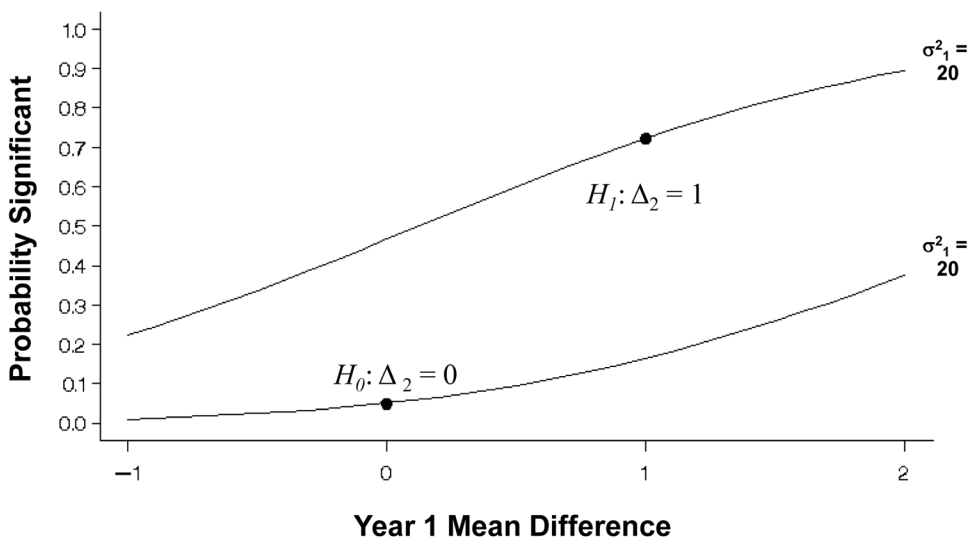
Figure 2.

Expected variance of the change from baseline ($\tilde{\sigma}_2^2$) in a one-sample analysis at 2-years from equation (A.1) under the null and the alternative hypotheses ($\mu_2 = 0$ or 1 designated by solid and dashed lines, respectively), for a single sample of $n = 100$ with 30% missing that are replaced by values at 1-year with mean μ_1 ranging from -1 to 2. Computations assume that the true variance at 2-years is $\sigma_2^2 = 20$ and the variance of the values at 1-year is $\sigma_1^2 = 10$, 20, or 30. The correct variances with no missing data are designated by the dots.

Figure 3.
One-sided probabilities of significance for a one-sample test of the change from baseline at 2-years under the null and the alternative hypotheses ($\mu_2 = 0$ or 1, respectively), with variance $\sigma_2^2 = 20$, for a single sample of $n = 100$ with 30% missing that are replaced by values at 1-year with mean μ_1 ranging from -1 to 2 and variance $\sigma_1^2 = 10$, 20, 30. The correct probabilities with no missing data are designated by the dots.

Figure 4.
One-sided probabilities of significance for a test of the difference between two groups *A* and *B* at 2-years with mean $\Delta_2 = \mu_{a2} - \mu_{b2}$ under the null and the alternative hypotheses ($\Delta_2 = 0$ or 1, respectively), for *n* = 200 per group with 30% missing that are replaced by values at 1-year with mean difference $\Delta_1 = \mu_{a1} - \mu_{b1}$ ranging from −1 to 2 and variance $\sigma^2 = 20$ within each group at both times. The correct probabilities with no missing data are designated by the dots.

Table 1

Mean μ_2 and variance $\tilde{\sigma}_2^2$ of the observed/imputed values at year 2, and the probability that a one-sample one-sided test of H_0: $\mu_2 = 0$ would be nominally statistically significant at the 0.05 level, for given mean and variance of the 1 and 2-year values $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2)$, with a sample size of $n = 100$, 30% of which are missing completely at random at 2-years and are replaced by the values at 1-year. Cases 1 and 5 also provide the type I error probability and power with no missing data.

Case	μ_2	σ_2^2	μ_1	σ_1^2	$\tilde{\mu}_2$	$\tilde{\sigma}_2^2$	Pr(significant)
1	0	20	0	20	0	20	0.05
2	0	20	−1	20	−0.3	20.21	0.010
3	0	20	1	20	0.3	20.21	0.164
4	0	20	1	30	0.3	23.21	0.153
5	1	20	1	20	1	20	0.723
6	1	20	−1	20	0.4	20.84	0.221
7	1	20	0	20	0.7	20.21	0.465
8	1	20	2	20	1.3	20.21	0.894
9	1	20	2	30	1.3	23.21	0.854

Table 2

The expected mean difference ($\tilde{\Delta}_2$) between two independent groups A and B of the observed/imputed values at 2-years, the expected variance of the group difference ($V(\tilde{D}_2)$), and the probability that a two-group test of H_0: $\Delta_2 = 0$ would be nominally statistically significant at the 0.05 level one-sided, for given means of the 1 and 2-year values, assuming $\sigma^2 = 20$ at each time in both groups, with a sample size of $n = 200$ per group, 30% of which are missing completely at random at 2-years and are replaced by the values at 1-year. Cases 2 and 7 also provide the type I error probability and power with no missing data.

Case	μ_{a2}	mu_{b2}	μ_2	μ_{a1}	μ_{b1}	μ_1	$\tilde{\Delta}_2$	$V(\tilde{D}_2)$	Pr(*significant*)
1	1	1	0	1	2	−1	−0.3	0.20105	0.01034
2	1	1	0	1	1	0	0.0	0.20000	0.050
3	1	1	0	1	0	1	0.3	0.20105	0.165
4	1	1	0	1	−1	2	0.6	0.20420	0.376
5	1	0	1	1	2	−1	0.4	0.20420	0.224
6	1	0	1	1	1	0	0.7	0.20105	0.467
7	1	0	1	1	0	1	1.0	0.20000	0.723
8	1	0	1	1	−1	2	1.3	0.20105	0.896