Estimated legacy effects from simulated post-trial data were less biased than from combined trial/post-trial data

Lin Zhu\textsuperscript{1*}, Katy J. L. Bell\textsuperscript{2}, Andrew Hayen\textsuperscript{1}

\textsuperscript{1}Australian Centre for Public and Population Health Research, Faculty of Health, University of Technology Sydney, Sydney, NSW, Australia

\textsuperscript{2}School of Public Health, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia

* Corresponding author:

Australian Centre for Public and Population Health Research, Faculty of Health, University of Technology Sydney, Ultimo, NSW, Australia

Tel.: +61 295145014;

E-mail address: Lin.Zhu@uts.edu.au
Abstract:

Objectives

“Legacy effects” describe the phenomena where treatment effects are apparent during the post-trial period that are not attributable to the direct effects observed within the trial. We investigate different approaches to analysis of trial and extended follow up data for the evaluation of legacy effects.

Study design and setting

We conducted a simulation to compare three approaches, which differed in terms of the time period and selection of trial participants included in the analysis.

Results

The most common approach used for estimating legacy effects in the literature, which combines initial trial and post-trial follow-up data, gave the most biased estimates. Approaches using post-RCT data had better performance in most scenarios. When the size of the legacy effect was set to differ according to whether or not drugs were taken post-trial, the stratified approach using post-trial data but only from participants taking the drug post-trial performed were less biased but often had lower power to detect a legacy effect.

Conclusion

When estimating legacy effects, approaches to analysis that are restricted to post-trial follow-up data are preferred. If data are available on participant drug use post-trial, then both stratified and un-stratified approaches to analysis of the post-trial data should be investigated.
Keywords: Legacy effects; randomized controlled trial; post-trial follow-up; cardiovascular disease

Declaration of interest

The author declares no conflict of interest.
1. Introduction:

The term “legacy effect” was first used in the context of cardiovascular disease prevention, in reports of the post-trial follow up after the United Kingdom Prospective Diabetes Study (UKPDS).[1] In that randomized controlled trial of intensive versus conventional glycaemic control, participants who were allocated to conventional treatment had a higher risk of microvascular complications than those on intensive treatment over the ten-year period of the active trial.[2] After the RCT ended, the trial investigators recommended that all participants aim for more intensive control, and the glycated haemoglobin levels of the two groups converged after one year. However, among participants undertaking follow-up after the trial, the statistically significant relative reduction in microvascular disease was found to have persisted, and additional statistically significant reductions in myocardial infarction and all-cause mortality also emerged for those originally randomised to the intensive-control group compared to those in the original control group. These findings were hypothesised to be a “legacy effect” of the earlier tighter glycaemic control for intervention group during the trial period that was only being realised years later.

Although the definition of legacy effects is not well specified, the term has generally been used to describe long term effects of a treatment that are observed after the trial has ended and that are not due to the direct (shorter term) effects of the treatment that were observed during the trial. These effects are thought to occur despite a similar proportion of individuals in the intervention and placebo group taking the active drug after the trial has ended, and attaining similar mean levels of the intermediate outcome (such as glycated haemoglobin, blood pressure or total cholesterol) in the post-trial period.[3–5]

Determining possible legacy effects may be of particular interest for interventions aimed at primary cardiovascular disease prevention. Here, the legacy effect concept has been used to support the case that early preventative treatment at a relatively young age may prevent cardiovascular disease at a much older age.[6] Many large-scale randomized controlled trials examining the effect of cardiovascular preventative treatment (drugs to control glucose in people with diabetes, and to lower blood pressure or cholesterol in people with or without diabetics) have reported the long-term health outcomes beyond the end of the trials.[4,7,8] The basic design of these studies is shown in Figure 1.
In most of these studies, the long-term effects of the drugs have been calculated using data from both the trial period and the post-trial follow-up period. These reports focus on whether there is a survival benefit to the group randomized to active treatment, which is still detectable at long-term follow-up.[9] Although such findings have been used to argue that legacy effects exist, it is possible that the observed effects might be due in part, or entirely, to the direct treatment effects observed during the within-trial period.[10] Analysis without disentangling the contribution of within-trial and post-trial effects will result in biased estimates. In this simulation study, we investigated how we might best analyse data from a matching RCT and post-trial follow-up study, to detect a legacy effect. Our objectives were to compare the performance of three different approaches to the choice of time period and trial participants to include in analysis in terms of ability to correctly detect when a legacy effect was, or was not, present.

2. Method:

2.1 Simulation Design

We formulated a setting that combined an RCT and an extended follow-up study. Independent datasets were generated with a known legacy effect or no legacy effect of the drug, in addition to a direct effect of the drug in all scenarios. We then evaluated three different approaches to analysis by applying each of them to the simulated data. The simulated data was designed to broadly reflect data that might be observed in a clinical trial for cardiovascular disease prevention, and the distributions of the simulation variables was based on the review of legacy effects of statin drugs.[10]

2.2 Clinical question of interest: do statins have legacy effects in preventing cardiovascular disease?

A current subject of clinical debate is the age at which drugs to prevent cardiovascular disease (such as statins, blood pressure lowering drugs and diabetic drugs for tighter glucose control) should be offered to people at risk.[11] Advocates of earlier intervention argue that some people who are currently displaying no symptoms or signs of disease, and who are not at high short-term absolute risk of cardiovascular disease, may benefit from starting preventative medication at an early age. [12] The hypothesis that the earlier
one starts these drugs, the lower one's risk of a cardiovascular event in the long term, has not been directly tested in an RCT, and because of feasibility issues it is unlikely to be. However, the hypothesis may be indirectly tested using data from post-trial follow-up after large controlled RCTs. A finding that randomisation to the active drug rather than comparator during the trial, has a “legacy” effect in protecting the person from cardiovascular disease after the trial, would suggest that earlier intervention may be worthwhile. An approach to data analysis that is able to reliably and accurately identify legacy effects is needed in order to use post trial follow up for this purpose.

2.3 Scenarios Investigated

Simulation settings were divided into two main scenarios, based on different assumptions on the size of the legacy effect for people who continued or discontinued using the drug post-trial. In the first scenario, “non-compounding legacy effect”, we assumed legacy effects were the same among the participants randomised to active treatment in the trial, irrespective of whether they continued to use the drug or not post-trial. This scenario simulates the situation where the protective effect from earlier treatment (with statins for example) occurs whether or not the person continues to take the drug post-trial. In the second scenario, “compounding legacy effect”, we assumed that there were no legacy effects if a participant stopped using the drug post-trial. This scenario simulates the situation where there is only a protective effect from earlier treatment if the person continues to take the drug post-trial, as otherwise the underlying natural disease progression catches up and the protection is undone.

We varied the size of legacy effects and proportion using the drug in the post-trial follow-up across simulations. The size of the legacy effect was defined as relative to the initial (direct) treatment effect and included 0 (no legacy effect), 50% and 100% of the direct
treatment effect. Proportions using the drug in the post-trial follow-up were assumed to be the same in the two groups and ranged from 20% to 100%. This resulted in a total of 30 sub-scenarios, and 10,000 simulations were run for each sub-scenario. Table 1 provides a summary of all variables considered in the simulations.

2.4 Data Generation

The starting point for simulation was to generate a cohort of patients with an underlying distribution of survival times. These survival times were generated from a Weibull distribution[13]. The shape parameter \( \nu \) was set at 0.5 which assumes the event rate is increasing over time, a situation often observed empirically in cardiovascular disease[14]. The scale parameter \( \lambda \) was chosen so that approximately 90% of participants who receive no treatment will survive after the 5-year RCT. The initial (direct) treatment effect was set as 0.8 (HR=0.8) and the sample size for each simulation was chosen at 8,000 people (4000 in each randomised group).[8,15]

For greater generalizability, the risk for each participant \( h_{\text{risk}} \) was simulated with a log-normal distribution. The mean and standard deviation on the log scale were set as -0.4 and 0.5 respectively, and so about 80% of patients were at low risk \( (h_{\text{risk}} < 1) \)[16,17].

Participants were assumed to have entered the randomized controlled trial (RCT) at \( t_0 \), and the initial trial ended at \( t_1 \). The surviving participants were followed until the end of the follow up period at \( t_2 \). The timepoints \( t_1, t_2 \) were set as 5 years and 15 years, which assumes the duration of the initial trial and post-trial follow-up were 5 and 10 years respectively.

If \( U(0,1) \) denotes a standard uniform distribution, then individual survival times can be generated using the formula:

\[
T = \begin{cases} 
\left( \frac{-\log (1-U)}{\lambda h_{\text{risk}} \exp (\beta X(t))} \right)^{1/\nu} & - \log (1-U) < \lambda h_{\text{risk}} \exp (\beta X(t))t_1^\nu \\
\left( t_1^\nu + \frac{-\log (1-U) - \lambda h_{\text{risk}} \exp (\beta X(t))t_1^\nu}{\lambda h_{\text{risk}} \exp (\beta X(t) + \beta_{LE})} \right)^{1/\nu} & - \log (1-U) \geq \lambda h_{\text{risk}} \exp (\beta X(t))t_1^\nu
\end{cases}
\]
In the equation above, \( X(t) \) is a time-varying variable that takes a value of 0 while a participant is assigned to placebo and 1 when assigned to active treatment\(^{[18]} \). \( \beta \) is the regression coefficient associated with the assigned treatment (representing direct treatment effects), while \( \beta_{LE} \) is the regression coefficient associated with legacy effects. If the algorithm generates a participant’s survival time beyond the end of post-trial follow-up, then the participant is considered censored and the survival time replaced by \( t_2 \). For simplicity, nonadherence with treatment and loss to follow-up were not considered in this study.

2.5 Comparison of approaches to analysis of data from long term follow up after RCT

Analysis for each scenario was performed according to the intention-to-treat principle, whereby participants are analysed in the groups to which they were randomly allocated at the start of the trial. Cox proportional hazard models were fitted to estimate the legacy effects, and individual risk was included as a covariate in the model. Three approaches were compared; these differed in terms of the choice of time period and trial participants to include in analysis:.

1. All trial participants. Data from the start of the RCT to end of post-trial follow-up was used (All data).

2. Participants surviving post-trial. Post-trial follow-up data was used. (Post-trial data)

3. Participants surviving post-trial and who took the drug during post-trial follow-up. Post-trial data used. (Post-trial data – drug strata)

2.6 Performance Indicators

Indicators used to assess the approaches to analysis were the bias, mean square error (MSE), coverage of 95% CIs and empirical power/size\(^{[19]} \). Bias was calculated as the difference between the average estimates over all simulations and the true value. Mean square error was calculated as the average squared difference between the estimated values and true value, according to the formula: \( \text{MSE} = (\hat{\beta} - \beta)^2 + (\text{SE}(\hat{\beta}))^2 \) where \( \text{SE}(\hat{\beta}) \) is the empirical standard error of the estimate of interest over all simulations. MSE is a useful measure of the overall accuracy which incorporates both measures of bias and variability. The coverage of 95% confidence interval was calculated as the proportion of times that the obtained confidence interval included the true specified parameter value.
Empirical size/power were used to indicate the probability of making a correct statistical
inference. Empirical size was calculated as the percentage of rejections of the null
hypothesis for each data scenario created under the null hypothesis, and empirical power
was calculated as the percentage of rejections of the null hypothesis for each data
scenario created under the alternative hypothesis[20].

3. Results:

The hazard ratios (95% CI) estimated by the different approaches to analysis in
simulations for each of the two scenarios are presented in Figure 2. In most simulated
scenarios, the estimates using the ‘All Data’ approach were considerably biased. The
hazard ratios were over-estimated (i.e. away from null) when there was no or a small
legacy effect, and under-estimated (i.e. towards null) when there was a large legacy effect.
In the “non-compounding legacy effect” scenario, where legacy effects for patients
allocated to active treatment in the RCT were the same whether or not they continued to
use the drug post-trial, both ‘Post-trial Data’ and ‘Post-trial data – drug strata’ approaches
gave unbiased estimates for hazard ratios, but the former had better performance in
terms of MSE (Figure 3). The ‘Post-trial data – drug strata’ approach, which use the least
amount of the available data, generally had estimates with a larger MSE than the other
approaches. In the “compounding legacy effect” scenario, where there were legacy effects
for patients allocated to active treatment in the RCT only if they continued to use the drug
post-trial, ‘Post-trial Data’ showed better performance than the ‘All Data’ approach, but
‘Post-trial data – drug strata’ approach was the least biased estimates. In addition, it
showed more robust performance in terms of MSE compared with other methods.

The 95% coverage of the three approaches across the different scenarios are displayed
Figure 4. The ‘Post-trial data – drug strata’ approach had consistently good coverage
among all scenarios investigated. In the “non-compounding legacy effect” scenario, the
‘Post-trial Data’ approach also had good coverage. The ‘All Data’ approach to analysis had
the worst coverage, with especially poor performance where there was no or small legacy
effects. In the “compounding legacy effect” scenario, lower coverages were observed at
different levels for ‘All Data’ and ‘Post-trial Data’ approaches.

Figure 5 shows the results of empirical power/size. The ‘All Data’ approach to analysis
always tended to accept the hypothesis of legacy effect. In addition, the power of the ‘Post-
trial data – drug strata’ approach increased with the amount of the available data used in analysis in most scenarios. In the “non-compounding legacy effect” scenario, power was mainly lower using the ‘Post-trial data – drug strata’ approach than the ‘Post-trial Data’ approach in, whereas in the “compounding legacy effect” scenario, the reverse was true.

4. Discussion

An ‘All Data’ approach to analysis is usually taken for studies that contain an RCT and post-trial follow-up, and is an indication of long term treatment efficacy.[21,22] However, this approach appears to be inappropriate for the evaluation of legacy effects, as it results in biased estimates of the true legacy effect in most situations. In addition, this approach often falsely concludes that there is a legacy effect when in truth there is not, (i.e. type I error). The results of our simulation confirm our earlier hypothesis, that in order to disentangle the direct effects of treatment during the RCT from legacy effects occurring post-trial, we need to restrict our analysis to post-trial period.[10]

Our study also shows that approaches to analysis that use only post-trial data may miss detecting a small size legacy effect, especially where sample sizes are small, which is to be expected.[23] The sample size calculated for the initial trial, while sufficient to detect the (usually larger) direct treatment effect within the trial, may be insufficient for detecting a legacy effect in post-trial period. Pooling data from several post-trial follow up studies in an individual participant data meta-analysis may be needed to overcome issues of insufficient power in the primary studies. In addition, for many pragmatic post-trial studies which use linkage to administrative data to track the participants’ health outcomes, some important individual information, such as medication status, is not available.[8,10] Therefore, only ‘Post trial data’ (not stratified) analysis can be conducted. Although the ‘Post-trial data – drug strata’ analysis performed better in the “compounding legacy effect” scenarios, we need information on the use of drugs in the post-trial period for this analysis. This requires the post-trial study to be pre-specified and funded.

Where data on post-trial drug use are available, analysis limited to people who took the drug post-trial may provide evidence about the benefits of starting treatment at a younger age, while analysis limited to people who didn’t take the drug post-trial may provide evidence about the safety of stopping drugs at an older age. We also found the ‘Post-trial data – drug strata’ approach to analysis to be more robust to variation in the
assumptions made about the size of the legacy effect for people who continue or discontinue using the drug in the post-trial period. But as the stratified approach uses relatively less amount of the available data, estimates tended to have a larger MSE.

Our study has some limitations. Although the risks for each individual were adjusted in the post-trial analysis in the simulation, we did not consider other confounders for simplicity. Potential sources of confounding include the differential use of medication, imbalanced levels of risk factors, differential loss to follow-up, and other differences between trial arms. These are likely to occur in post-trial follow up studies in real life, especially when the size of treatment effect is large.[24] We did not allow for competing events in our study, and these will become increasingly important the longer the duration of the post-trial follow-up, with the potential to bias the estimates of legacy effects.[25] In addition, our simulations were based on plausible scenarios of treatment using statins for primary prevention of cardiovascular disease, and these might not be generalisable to other types of interventions, such as surgical and behavioural interventions.[26] We were necessarily constrained in the number of possible scenarios that we were able to investigate in this study.

In conclusion, our study found that the approach most commonly used to estimate legacy effects is usually not appropriate as results may indicate the persistence of the direct effects of treatment rather than a legacy effect. When estimating the extent of legacy effects, approaches to analysis that are restricted to post-trial follow-up data are preferred. The selection of participants to include in the post-trial analysis is less clear cut, and we recommend using both the unstratified and stratified approaches to analysis if data on drug use post-trial are available. Importantly, more attention should be paid to the design of post-trial studies for evaluating possible legacy effects to ensure adequate sample sizes and study power to detect possible legacy effects.[27]

References:


**Figure 1.** The basic design of study for evaluating legacy effect (randomized controlled trials and post-trial follow up)

**Figure 2.** Hazard ratio estimated by different methods in simulated scenarios *

*The points with confidence intervals show estimated hazard ratio and corresponding 95% CI for each method in different scenarios and the dash lines indicate the truth values in each setting. Bias can be obtained by comparing the estimate and true value.

**Figure 3.** Mean square error of different methods in simulated scenarios

**Figure 4.** Coverage Probabilities of 95% CI of different methods in simulated scenarios *

*The lines for "Post-trial Data" and "Post-trial data – drug strata" are basically overlapped in "No legacy effect" scenario and "non-compounding legacy effect" scenario.

**Figure 5.** Empirical Power/Size of different methods in simulated scenarios *

*The lines for "Post-trial Data" and "Post-trial data – drug strata" are basically overlapped in "No legacy effect" scenario and "Legacy=100%" of "non-compounding legacy effect" scenario.

**Table 1.** Summary of variables used in the simulation

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weibull parameter</td>
<td>scale parameter $\lambda = 0.01$, shape parameter $\nu = 1.45$</td>
</tr>
<tr>
<td>Log-normal parameter</td>
<td>mean and standard deviation on log scale mean = -0.4, SD = 0.5</td>
</tr>
<tr>
<td>Sample size</td>
<td>8000</td>
</tr>
<tr>
<td>Length of RCT and PTFU</td>
<td>5 years and 10 years</td>
</tr>
<tr>
<td>Initial Treatment effect</td>
<td>0.8</td>
</tr>
<tr>
<td>Size of legacy effect (compared with treatment effect)</td>
<td>0, 50%, 100%</td>
</tr>
<tr>
<td>Proportion of treatment receiving in PTFU</td>
<td>20%, 40%, 60%, 80%, 100%</td>
</tr>
</tbody>
</table>
Randomized controlled trial (In-trial)

Participants

Randomized Allocation

Intervention Group

Control Group

Post-trial follow-up (PTFU)

Original Intervention Group

Original Control Group

Outcomes

Compare results for treatment effect

End of RCT

Survivors eligible for follow-up after trial

Compare results for legacy effect

Outcomes