

**Neill & Leigh's "Weak Tests and Strong Conclusions: A
Reanalysis of Gun Deaths and the Australian Firearms
Buyback" Revisited:
The Value of a Conservative Conclusion**

Prepared by Dr Jeanine Baker¹ & Dr Samara McPhedran²

April 2007

1: Director, Research and Policy Unit, Sporting Shooters Association of Australia (SA)

2: Chairwoman, International Coalition for Women in Shooting and Hunting (WiSH)

Summary

Neill and Leigh argue that Baker and McPhedran (2006; henceforth, B&Mc) do not provide evidence to support the conclusion that the 1996 gun laws had no significant impact on firearm homicide and suicide in Australia. However, they fail to sustain this argument.

B&Mc's analysis showed that there was no significant impact of the 1996 National Firearms Agreement (NFA) on firearm homicides. In addition, B&Mc found a significant acceleration in the downwards trend in firearm suicides. They note this clearly, and observe that caution is necessary given suicides using other methods also declined. They suggest that social and other changes may have had overall effects, and highlight this as a future research question requiring evaluation.

It follows from these results that there is insufficient evidence to support the simple premise that the removal of large number of firearms from legal owners had an impact on the pre-existing downwards trend for firearm-related homicide and suicide.

Neill and Leigh (henceforth N&L) put forward a number of incorrect assertions and inaccurate statements. These are dealt with below.

Shortcomings

1) It is claimed Baker and McPhedran “...do not provide any justification for using less than one third of the available data”.

This refers to the fact that homicide and suicide data from the early decades of the 20th century onwards are held by the Australian Bureau of Statistics. B&Mc use these data to give an overview of long-term trends in Australia. However, data from 1979 onwards are used in formal analyses.

N&L intimate that the use of this particular time series is a feature peculiar to B&Mc, and is thus “dubious”. They fail to acknowledge that two other Australian studies into firearm-related deaths examine precisely this time period, and that those studies, like B&Mc, found no significant acceleration in the rate of decline in firearm homicide post-1996 (Chapman, Alpers, Agho, & Jones, 2006; Ozanne-Smith, Ashby, Newstead, Stathakis, & Clapperton, 2004).

Implying that the time period used by B&Mc was selected to bias the results, as opposed to a consistent feature between a number of studies, cannot be explained by a lack of knowledge of the other studies. N&L cite those very studies in the course of other arguments (see point 8 below).

2) N&L allege that, due to historically high levels of firearm homicide and suicide in the late 1970’s and early 1980’s, analysis using the period 1979-1996 vs. 1997 onwards does not reflect longer term trends across years where the rate of firearm homicide and suicide was lower than observed in the late 1970’s and early 1980’s.

N&L use this observation to contend that B&Mc do not control for the effects of social change.

This objection is, however, entirely consistent with arguments by B&Mc that it is premature to conclude the gun laws were responsible for declines in firearm related

deaths post-NFA, precisely due to the likely influence of social and other factors. N&L fail to recognise that such factors could also have contributed to the high levels of firearm related death (particularly homicide) in the late 1970's-early 1980's.

3) N&L claim that when a longer time series is investigated, a significant decline in firearm homicides post-NFA is found. They argue that using the time series from 1979 'biases' the findings against such a result.

This is a misleading argument, and entails N&L undermining their own arguments about social factors and high levels of firearm related homicide in the 1970's and 1980's.

Firearm-related homicides (and homicides by other methods) fluctuated in the decades before the 1970's.

Figure 1, below, shows this.

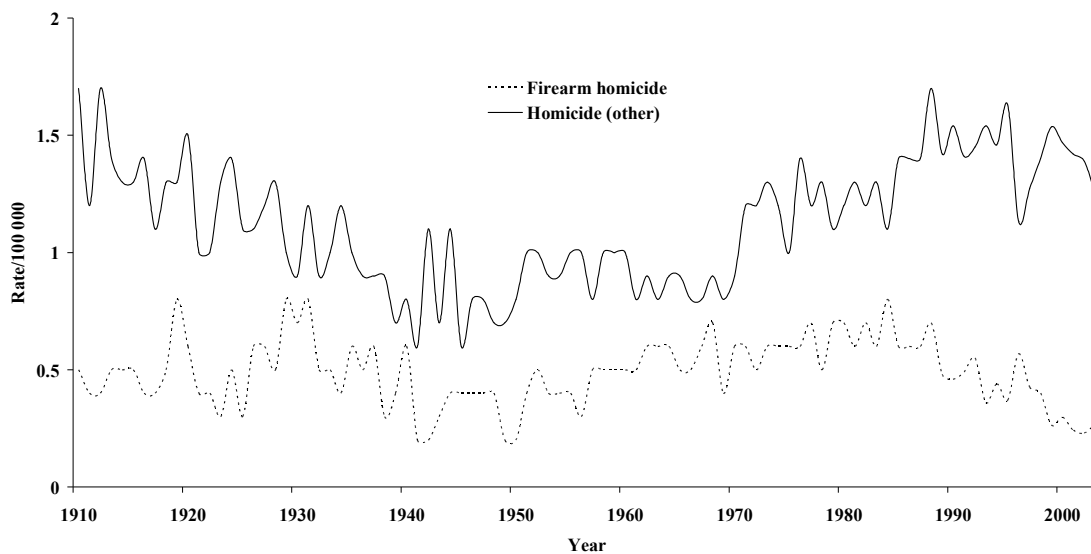


Figure 1: Historical trends in homicide in Australia, 1910-2003.

N&L compare a very long-term trend (the period 1915-1996, i.e., over 80 years) with an 8 year period post-NFA (1997-2004), and claim that firearm-related homicides fell more quickly after the legislative reforms. Thus, they argue that the laws were responsible for lowering firearm homicides. This is a flawed conclusion.

Firstly, the theoretical possibility that the decline in firearm homicides beginning after the peak in the early- to mid-1980's may represent levels of firearm-related homicide dropping back to 'historically average' levels, rather than any effect of legislation, must be acknowledged.

Secondly, the fluctuations in the data from 1915-1996 (or, indeed, from 1969 onwards) have a predictable impact on ARIMA modelling. When a trend in the 1915-1996 data is plotted, it is almost flat. This is because although the data contains numerous fluctuations, these 'balance' each other out over time.

Simply put, in the ARIMA procedure the 'shape' of the data over the specified time series (whether it contains a steady trend or whether the values remain, on average, steady over time) makes an important contribution to the predictions of what will happen (the forecast), based on what has previously happened (the predictive model). In simple terms, ARIMA uses a combination of past values, which can include a consistent trend in the data, to forecast future values. What has happened in the past will therefore influence what is predicted to happen in the future.

When the degree of fluctuation over the long-term time period is of the order observed for firearm homicide, this in turn will 'dampen' the influence of the downwards trend that emerged in the 1980's on the forecast of what is expected to happen in the future, based on the past. Being based on such a long time series that contains shifts in both directions means the changes that began in the 1980's may not be picked up as quickly (there will be a 'lag') or be as influential on the predictive model as they would in a shorter time series.

The forecast from 1997 onwards will still be based on a combination of the previously modelled value, and its error component, but will include a negligible 'change' component (which is why the values forecast by N&L barely move from 1997 on). When incorporated into a predictive model the almost steady long-term trend in the 1915-1996 data is all but guaranteed to produce a forecast for 1997-2004 that will, unless the downwards trend emergent from the 1980's virtually ceased after 1996, differ from what actually happened. In short, the forecast does not allow for continuation of the trend that began in the 1980's, except at a radically slower rate.

As a consequence, rather than B&Mc's predictions being 'relatively low', N&L's forecasts based on 1915-1996 data are unrepresentatively high. They cannot adequately respond to the real-world changes that began in the 1980's. Therefore, the claim that firearm homicides fell more quickly post-NFA in comparison with predictions based on historical trends from 1915 onwards is the result of statistical sleight of hand.

It is noteworthy that, despite their emphasis upon time periods, N&L do not acknowledge the substantial limitations inherent in using an 80 year time period to forecast an eight year time period, or the dangers of drawing conclusions from such techniques and comparisons.

Also, the use of such a long time period is necessarily confounded by the 'underlying factors' (e.g., social change) that N&L argue must be controlled for. Therefore, for N&L to use this time period, without acknowledging the contribution of periods of major social change in Australia or the overall trends in homicide over that time period, negates their own claims regarding the importance of controlling for changes in underlying factors.

4) N&L claim that the ARIMA (Auto-Regressive Integrated Moving Average, or "Box-Jenkins") model is 'dubious' for use on firearm homicides.

No justification is given for this, other than the statement that there are 'well known' difficulties associated with using ARIMA on such a 'short' time series. However, ARIMA is a widely used, flexible model employed across a range of social science and related disciplines.

In addition, it is difficult to sustain the claim that a period of 18 years (1979-1996) is 'short'. The statement is not supported, and appears to be included simply to augment claims that a longer time period should be examined (which, as shown above, carries a related set of difficulties when applied to homicides, and generates results that are foregone rather than meaningful).

Furthermore, questions raised about the suitability of ARIMA are better applied to the long-term data N&L use to support their contention that there was a significant change in firearm homicide rates post-1996, than to the 1979–1996 period.

Although ARIMA is able to deal with steady trends, and assess the impacts of an ‘intervention event’, the data from 1915 onwards contain very definite cyclical patterns (see for example the peaks around the 1930’s and the dips around 1940-1950) which may be tied to definite external events (economic depression in the former case and WW2 in the latter). N&L have not introduced measures to attempt to control for or accommodate these trends.

5) It is claimed that the results vary depending on the statistical package used, and that ARIMA is an unstable model, thus N&L were unable to reproduce the results of B&Mc.

Small differences in calculation processes may occur between some software packages performing the same general tasks, and it is acknowledged that rounding error may also create small differences. However, this does not seem a sufficient explanation for any differences in B&Mc and N&L’s results.

N&L claim their failure to reproduce the results of B&Mc signifies inappropriate model choice and demonstrates the sensitivity of ARIMA (thus implying that the results of B&Mc should not be considered valid, while maintaining in direct contradiction that the results of N&L should be accepted). However, they utterly overlook the fact that different results within ARIMA come largely from different user-controlled model defaults and specifications including, for example, how error terms are treated during estimation and forecasting, how constants within the data are treated, and how the forecasting period is initialised.

The general similarity of N&L to B&Mc’s predictions (a difference in the order of 0.02, on average) suggests the use of different algorithms for the treatment of variability and forecasting, rather than a major difference in results or inapplicability of the general model.

6) It is argued that ARIMA is unsuitable because “...statistical tests consistently reject the hypothesis that homicide rates follow a non-stationary process. BM present no statistical tests to show whether these models are appropriate.”

Footnote 1 in B&Mc discusses precisely this issue, and notes that the flexibility of the ARIMA model makes it suitable for application in a range of situations including those of autocorrelation (where the data for one year are related to, and influence, the data for subsequent years). Thus, the points raised by N&L are in fact considerations taken into account by the use of ARIMA, including the matter of a stationary versus non-stationary process.

It is not correct to state that B&Mc do not present tests to assess the appropriateness of the model used. B&Mc provide, for each application of ARIMA, a statistical ‘goodness of fit’ test (R^2 values) to indicate how well the model performed in predicting death rates in the period 1979-1996. The better the fit, the more ‘likely’ the forecasts are to be accurate post-1996. This is a similar method to that used by Ozanne-Smith *et al.* (2004) to test the fit of their models.

In cases where the model did not provide a high level of prediction, B&Mc are careful to note this (see, for example, comments concerning non-firearm homicide, p.7, or discussion of the difference in model fit between firearm suicide and firearm homicide, p.7).

7) It is argued that B&Mc predict a ‘negative probability’ (i.e., the model predicts that deaths fall below zero) for firearm homicide.

This is not correct.

B&Mc found, for accidental firearm deaths, an extremely wide confidence interval. Confidence intervals are a measure of how much ‘variation’ or ‘uncertainty’ there is in the data. In general, as the sample size decreases, variability increases. As B&Mc note, there are very few unintentional firearm-related deaths in Australia.

The lower limit of the confidence interval indeed for unintentional firearm-related deaths falls below zero. The predicted or forecast rate of death, however, does not.

It is not unusual for the lower limit of a confidence interval to fall below zero, and this event does not negate the validity of the predictive methods used. Furthermore, B&Mc clearly acknowledge that the degree of variation in the data for accidental deaths in particular entails that conclusions must be approached with due caution (p.9).

8) It is argued that Poisson or log models are used to ensure that death rates are always positive, and that N&L are “...unable to find another study examining firearm-related deaths that simply uses the death rate...”.

This is a disingenuous argument.

There exist relatively few studies examining firearm-related deaths in Australia. Two papers (Chapman *et al.*, 2006; Ozanne-Smith *et al.*, 2004) use Poisson regression or negative binomial regression, which involves a transformation of the data.

However, N&L fail to realise that both ARIMA and simple death rates rather than log-transformed death rates are commonly used in wider literature (see for example Morrell *et al.*, 2006, who use ARIMA and simple death rates to assess trends in young male suicide in Australia).

This raises two important points that N&L tellingly overlook.

Firstly, in instances where data show autocorrelation, Poisson models and their derivatives are not necessarily the best option for controlling for underlying trends, which is why alternative methods such as ARIMA are useful.

Secondly, and crucially, N&L do not acknowledge that the results of B&Mc fit with the results of both Ozanne-Smith *et al.* (2004), who studied one Australian state (Victoria) and Chapman *et al.* (2006), who examined Australia-wide data. All three studies found a significant acceleration in the rate of decline in firearm suicide post-

1996, compared with 1979-1996 trends, and all three studies found that there was no significant acceleration in the rate of decline in firearm homicides post-1996.

Therefore, if N&L's re-analysis entails rejection of B&Mc's results, it must by definition also necessitate rejection of the other two studies. Why B&Mc have been selected for specific attention is puzzling, given the consistency between the data series and results of the three studies.

The major difference between the three studies is that B&Mc found no evidence of an impact of the gun laws on firearm homicide, and concluded that on the basis of available evidence it cannot be said that the laws had an impact. Chapman *et al.* (2006) and Ozanne-Smith *et al.* (2004) found no evidence to demonstrate an impact of the gun laws on firearm homicide, but nonetheless concluded - in the absence of, or contrary to, statistical outcomes - that the gun laws had an impact.

It is concerning that Chapman *et al.* (2006) in particular concluded that the legislative reforms had an impact on firearm-related homicides, despite their own results showing otherwise. It is equally troubling that N&L do not appear bothered by the fact that the conclusions of that study were not supported by the results.

Mass shootings

N&L do not discuss the fact that the long-term data from 1915 onwards demonstrates the fallacy of Chapman *et al.*'s (2006) argument, made in the absence of any analysis, that Australia's gun laws have prevented mass shootings.

It is correct that there were 10 mass shootings (four or more fatalities) in the decade 1986-1996 (four of those shootings occurred in one year, 1987). However, this was a feature peculiar to that decade. When longer time series are evaluated on the basis of available, verifiable data concerning mass shootings, it emerges that in the decades prior to 1986, there were lengthy spans of time when no mass shootings took place. Thus, the absence of mass shootings in earlier years, well before the legislative reforms, suggests that the laws cannot be attributed with preventative properties.

In conjunction with their statement that data must be carefully evaluated over long time periods before conclusions can be drawn, the observation that many years pre-1986 did not contain mass shootings utterly undermines Chapman *et al.*'s (2006) bold assertion that the gun laws prevented mass shootings in Australia. Instead, the data used by N&L suggest that the declines post-96 may represent firearm related deaths (mass shootings or otherwise) dropping back to historically consistent levels.

Why N&L ignore this important implication of their data is unclear. This is a notable oversight given the purported goal of N&L is to assess congruence between results and conclusions in firearm-related death research. This omission, in conjunction with the lack of acknowledgement that B&Mc's results for firearm homicide were independently replicated by Chapman *et al.* (2006), raises serious questions about the objectivity and intent of the authors.

9) N&L argue that the inclusion of non-firearm homicides and suicides does not represent an adequate control group.

N&L again provide a misleading argument, by failing to acknowledge that the inclusion of deaths by 'other' methods is a common and widely used control group when examining deaths by a 'particular' method. See, for example, Chapman *et al.* (2006).

N&L do not apply logical consistency to their definition of a control group; they claim that the potential for method substitution negates the use of non-firearm deaths as a control, but simultaneously do not allow that the *absence* of method substitution (showing, by their own definition, that the group did not appear to be affected by the policy change) would subsequently allow that group to be used as an adequate control.

10) It is claimed that:

“...even with a finding of a statistically significant decrease in firearm homicides or suicides post-NFA, B&M could draw virtually any conclusion they wish. In the event that non-firearm deaths increased after the buyback, they could claim that this is due to method substitution, showing the NFA too have [sic] been ineffective. In the event that non-firearm deaths fell after the buyback, they could claim that non-firearm deaths are a control for firearm deaths and therefore that some factor other than the buyback must have caused the reduction in firearm deaths.”

And:

“The fact that overall violent deaths have fallen since 1996...strongly suggest there has not been any substantial method substitution.”

N&L are unable to provide any justification for why the theories outlined above should not be accepted as valid hypotheses, other than the meaningless statement that “B&Mc could draw virtually any conclusion they wish.”

If non-firearm homicides and suicides had increased post-NFA, this would provide reasonable support for speculation that method substitution *may* have occurred. N&L are unable to clarify why it would not be reasonable to propose this theory.

N&L argue “even with a finding of a statistically significant decrease in firearm homicides or suicides...B&Mc could draw virtually any conclusion they wish”. With reference to suicides, B&Mc *did* find a significant acceleration in the rate of decline, and concluded that firearm suicide may have been affected by the NFA. Therefore, contrary to the assertion that B&Mc could have drawn “virtually any conclusion”, the conclusion B&Mc in fact drew was that the laws may have had an impact on firearm suicides.

What also occurred was that B&M found no evidence for method substitution, given the observation – supported by N&L – that non-firearm deaths fell post-1996. Thus, N&L’s criticism is in fact entirely consistent with, and upholds, the conclusions of B&Mc.

N&L's statement on the apparent lack of method substitution highlights the importance of taking other factors into account. Here, N&L contradict their own claims that B&Mc do not consider underlying factors contributing to general trends.

On one hand, N&L erroneously argue that B&Mc do not control for external factors, and that this undermines B&Mc because those factors may contribute to declines in all types of deaths. On the other hand, they argue any suggestions that external factors may have contributed to observed declines in both firearm and non-firearm deaths should not be accepted if they come from B&Mc.

N&L state in endnote 3 that:

“We do not discuss our findings on non-firearm death rates in detail in this paper, because they have little bearing on our key arguments.”

This is an evasive position, given those findings have every bearing on the key argument that social and other changes and their influences over time must be considered, as well as on the argument that method substitution did not occur.

It is puzzling why N&L imply that B&Mc argue for the occurrence of method substitution, and then proceed to ‘demolish’ this argument by stating that there were declines in non-firearm homicide and suicide. B&Mc made no such claims about method substitution.

B&Mc point out that there were declines in non-firearm homicide and suicide, which suggests method substitution did not appear to have occurred, and that other factors may have influenced overall trends – a suggestion N&L claim as their own, while alleging that B&Mc did not acknowledge this, and simultaneously arguing that if B&Mc were to acknowledge this, it would be a questionable assertion. Thus, N&L again fully support B&Mc's conclusions, while undermining their own arguments through logical incoherence.

11) Neill & Leigh conclude by stating that:

“A common axiom in social science research is that a test that fails to reject the null hypothesis should not automatically lead the researcher to accept the null hypothesis.”

Noting that the null hypothesis is generally conceived of as the ‘hypothesis of no effect’, there are two scenarios associated with hypothesis testing.

- a) The researcher does not find evidence that enables them to reject the null hypothesis (there is no evidence that ‘anything happened’).
- b) The researcher finds evidence that allows them to reject the null hypothesis (there is evidence that ‘something happened’).

There are two ‘errors’ associated with hypothesis testing.

- a) The researcher incorrectly accepts the null hypothesis, and concludes nothing happened, when something actually did happen – ‘Type II error’.
- b) The researcher incorrectly rejects the null hypothesis, and concludes something happened, when nothing actually happened – ‘Type I error’.

N&L argue that although there was no evidence to enable *rejection* of the null (we cannot reject the hypothesis that “the gun laws had no effect”), it should not be *accepted* that the gun laws had no effect. They believe that even though there was no evidence that the gun laws had an effect, it should not be said that the gun laws had no effect.

Essentially, N&L object to B&Mc’s conclusion that the NFA did not have an impact on firearm homicide, despite the fact that the evidence shows no significant impact of the NFA on firearm homicide.

They claim that even though the evidence did not show an impact, it should not automatically be assumed there was not an impact. This equates to the proposal that

B&Mc's conclusions may run the risk of a Type II error (incorrectly concluding that nothing happened, when something may have happened).

However, this criticism is undermined by N&L's own interpretive framework. They attempt to find evidence, particularly through the selection and analysis of a time period that will necessarily produce a foregone outcome, that:

“...to the extent that the time series evidence points anywhere, it is towards the conclusion that the NFA reduced gun deaths.”

What N&L fail to realise is that rejecting the null ('the gun laws had no impact') does not mean that the alternative ('the gun laws had an impact') is true. As McCloskey and Ziliak (1996) correctly note: “...there may be other alternatives...which would cause rejection of the null” (for example, the social factors the N&L switch between invoking and rejecting, depending on their particular argument).

Instead, N&L:

- Reject three independently conducted studies pointing in the same direction.
- Reject their own statements about taking into account underlying social and other factors.
- Encourage the conclusion that the laws directly caused a decline in firearm homicides.

This not only overlooks that assumptions about 'causality' must be approached with due caution, but also lends itself to a Type I error. Within scientific research, a Type I error is regarded as the more serious type of error, and it is the error for which statistical analyses should attempt to control against.

In short, it is considered preferable to conclude there was no effect when there may have been an effect, than to conclude there was a proven effect when in reality there was not.

Conclusions

Far from representing the ‘strong conclusion’ N&L claim, acceptance of the null hypothesis – that the gun laws had no impact on firearm homicide – represents a conservative conclusion that takes into account the variable nature of the data and the potential for committing a Type I error.

In contrast, to insist on the basis of tenuous and/or demonstrably manipulated analyses that the gun laws ‘caused’ a decrease in firearm homicides represents a very strong conclusion with a high likelihood of Type I error.

Such observations should be taken in conjunction with N&L’s lack of criticism of studies that found no evidence of an impact of the gun laws on firearm homicide but nonetheless definitively concluded there was a major impact. Cumulatively, these factors suggest N&L may prefer to conclude that the gun laws had an unquestionable effect than to admit the possibility that the laws may *not* have had an impact or that other factors may have played a role in influencing firearm-related deaths in Australia.

References:

Baker, J. & McPhedran, S. (2006). Gun Laws and Sudden Death: Did the Australian Firearms Legislation of 1996 Make a Difference? *British Journal of Criminology*. Advance access October 18, 2006. doi:10.1093/bjc/azl084.

Chapman, S., Alpers, P., Agho, K., and Jones, M. (2006). Australia’s 1996 gun law reforms: faster fall in firearm deaths, firearm suicides and a decade without mass shootings. *Injury Prevention*, 12(6): 365-372.

McCloskey, D.N. & Ziliak, S.T. (1996). The Standard Error of Regressions. *Journal of Economic Literature*, 34(1): 97-114.

Morrell, S., Page, A.N., & Taylor, R.J. (2006). The Decline in Australian Young Male Suicide. *Social Science and Medicine*. Advance access September 27, 2006. doi:10.1016/j.socscimed.2006.09.027.

Ozanne-Smith, J., Ashby, K., Newstead, S., Stathakis V.Z., & Clapperton, A. (2004). Firearm related deaths: the impact of regulatory reform. *Injury Prevention*, 10: 280-286.