Original Article

How to find nothing

David Hemenway

Health Policy and Management, Harvard School of Public Health,
677 Huntington Avenue, Boston, MA 02115, USA.

Abstract   Hypothesis testing can be misused and misinterpreted in various ways. Limitations in the research design, for example, can make it almost impossible to reject the null hypothesis that a policy has no effect. This article discusses two examples of such experimental designs and analyses, in which, unfortunately, the researchers touted their null results as strong evidence of no effect.
Journal of Public Health Policy (2009) **30**, 260–268. doi:10.1057/jphp.2009.26

> A common axiom in social science research is that a test that
> fails to reject the null hypothesis should not automatically lead
> the researchers to accept the null hypothesis.[1]

Introduction

Hypothesis testing can be misused and misinterpreted in various ways. It is, for example, possible to reach an incorrect conclusion, such as finding an effect that is not actually there (for example, incorrectly rejecting the null hypothesis) because of problems of design, measurement, modeling, and so on. This paper discusses the opposite problem – not finding an actual effect, and worse, then claiming that not being able to reject the null hypothesis shows there is no effect. In the two examples that follow, the researchers had a null finding (that is, no significant effect), but the research had serious limitations. These limitations meant that it was unlikely to find anything, even if there had been something to be found.

  The two examples deal with one issue – the role of firearms in public health – and one type of analysis – time series. It must be emphasized that similar problems are too often found in many types

of analyses throughout the medical and public health literatures. See Note A for an illustration.[1]

In the first example, the researchers so limited their search that they could expect to find very little effect. In the second example, the researchers designed the counterfactual (control) so that it was almost impossible to find an effect. Unfortunately, in these cases, the researchers touted their findings to the media, proclaiming that their research had shown that there was no effect – there was nothing to be found. In reality, from their analyses we learn very little, except that there may be a tendency among researchers to overstate their results.

Example 1

'Gun Shows Do Not Increase Homicides, Suicides' shouted the headlines.[2,3] The headlines accurately summarized the press release of an unpublished study examining the impact of gun shows on gun-related deaths. The authors analyzed data from Texas and California and claimed that 'this analysis makes an important contribution to understanding the influence of gun shows'.[4]

Unfortunately, the study was effectively designed to find no effect. The study mentioned two crucial caveats: the authors only examined the possible effect of gun shows on gun deaths occurring within 4 weeks and within 25 miles of the gun show. These limitations mean that the study probably missed 98 per cent of the possible effect of the gun shows on homicides and suicides.

In 2007, for example, fewer than 6 per cent of criminal guns recovered in Texas, and fewer than 3 per cent in California, had a time from initial sale to recovery of 3 months or less. For Dallas and Los Angeles in 2000 (the most recent year for which city-specific data are available), only half of traced guns were recovered within 25 miles of their point of initial sale.[5] In the Columbine school shootings, the killers obtained guns from two separate individuals who purchased the guns for them months before, at gun shows. In the Duggan study, this would be evidence of the lack of influence of gun shows on violent death.

Even for suicides, most gun suicides occur years after the gun purchase. In one study, the median time between gun purchase and

suicide was 11 years; only 13 per cent of the firearm suicides occurred in the first year after purchase.[6]

In addition, the public health and safety concern is not so much about gun shows *per se*, but the lack of background checks and regulatory oversight of sales made by private sellers at gun shows. Because in most states, unregulated private gun sales are permitted through many other avenues (for example, flea markets, classified ads, or over the internet), eliminating gun shows is not expected to have nearly as large an effect on gun sales to criminals as would requirements (already enacted in some states) that every gun transfer be made through licensed dealers, with appropriate background checks and oversight.

So in another sense, the study was designed not to find an effect. It would be like being interested in whether or not, in a sinking ship with several large holes, plugging only one of them would save the vessel.

Example 2

'Buyback Has No Effect on Murder Rate' and 'Australia No Safer With Gun Buyback: Study' proclaimed the news headlines.[7,8] These headlines accurately reflected the authors' claims: 'The findings were clear, she said: the policy has made no difference. There was a trend of declining deaths which has continued'.[8]

The study in question,[9] published in the *British Journal of Criminology*, and written by two Australians from the pro-gun lobby (the Sporting Shooters Association of Australia and the Australia and International Coalition for Women in Shooting and Hunting), analyzed the effect of the 1996 National Firearms Agreement (NFA). The NFA, passed in response to the 28 April 1996 Port Arthur, Tasmania massacre of 35 people, effectively banned assault weapons, bought back over 12 months more than 650000 of these weapons from existing owners, and tightened requirements for licensing, registration, and safe storage of firearms. The buyback is estimated to have reduced the number of guns in private hands by 20 per cent.

At first blush, the NFA seems to have been incredibly successful. Although 11 gun massacres occurred in Australia in the decade before the NFA, resulting in more than 100 deaths, in the decade following (and up to the present), there were no gun massacres.

It was also hoped that the NFA might reduce firearm homicide and firearm suicide. And again, the results seem to indicate a resounding success. For example, in the 7 years before the NFA (1989–1995), the average annual firearm suicide death rate per 100 000 was 2.6 (with an yearly range of 2.2–2.9); in the 7 years after the buyback was fully implemented (1998–2004), the average annual firearm suicide rate was 1.1 (yearly range: 0.8–1.4). In the 7 years before the NFA, the average annual firearm homicide rate per 100 000 was 0.43 (range: 0.27–0.60), whereas for the 7 years after NFA, the average annual firearm homicide rate was 0.25 (range: 0.16–0.33).

In any time series analysis of an intervention, researchers will compare what actually happened to what would have happened without the law (the counterfactual). Unfortunately, the counterfactual is never known. In this instance, the researchers made the assumption that the historical trend would have continued unabated. They made no effort to explain why the historical trend had been what it was, nor why they expected it to continue. The trend was downward.

The researchers chose 1979 as the beginning year for the trend analysis. They gave no explanation for this choice, and data were available for each year back to 1915. The Australian firearm suicide and the homicide rates in 1979 were the highest and the third highest, respectively, for any year 1932–1996. Identical analyses using data from 1915 to 2004 found that both firearm suicide and firearm homicide declined significantly after the NFA.[1]

The researchers' assumed counterfactual was that a linear trend of the actual death rate from 1979 to 1996 would continue forever. In other words, the assumed counterfactual was that if the historical rate fell from 3/100 000 to 2/100 000 in the initial period, it would fall to 1/100 000 in next period, then to 0/100 000, and then to −1/100 000. This assumption meant that the counterfactual predicted an ever-increasing percentage fall in death; indeed, the model predicted that without the NFA, the number of firearm homicides in Australia would be negative by 2015. Critics labeled this a 'Resurrection Problem'.[1] It would be very difficult for an intervention to be an improvement on that counterfactual; indeed, if in 2004, the Australia firearm homicide rate had been zero (and remained there), that rate would not have been low enough to reject the null hypothesis that the NFA had no effect.

The log of the death rate (with the analysis focusing on rates of change of fatalities rather than absolute levels of change) is commonly used to eliminate the absurdity of a negative death rate. Using such an approach (and even examining the 1979–2003 period), researchers found support for a statistically significant effect of the NFA on total firearm deaths (Note B).[10]

Although the gun lobby authors did not acknowledge the study limitations, they were more than willing to state to the press that 'In 1996 we were told that buying back those civilian firearms off licensed firearm owners would make society safer and would reduce firearm deaths. The evidence isn't there to support that' (Note C).[7]

Conclusion

When (for whatever reason) a problem has been decreasing, using a time trend as the counterfactual (with no other explanatory variables) makes it difficult to find a significant effect of a policy intervention. One could equally well make the claim that no policy has influenced infectious disease deaths in the United States. These have been declining since 1900 (Figure 1), and nothing (except perhaps the 1918 flu epidemic) seems to have significantly impacted that trend. Indeed, the Salk vaccine appears to have made the non-logged trend worse. In reality, most scientists believe that improvements in sanitation, hygiene, antibiotics, and immunization have been key in reducing infectious disease mortality. But, it would be incredibly easy to use national time trend analysis and be unable to reject the hypothesis that chlorine in municipal water supplies, penicillin, or vaccines had no effect whatsoever. (In 1988 in Victoria, home to some 20 per cent of Australians, gun control policies are reported to have significantly reduced firearm deaths in that state from 1988 to 1995.[11] If true, the null hypothesis for the NFA was partly to determine whether it was a statistically better intervention than that previous one.)

Similarly, using national time trend analysis on US national data on Gross Domestic Product, one could easily be unable to reject the null hypothesis that new technologies, such as steam power, electrification, or computers, had no effect on national output levels. Our economic capacity has been growing more or less steadily for hundreds of years; if the counterfactual is continued economic
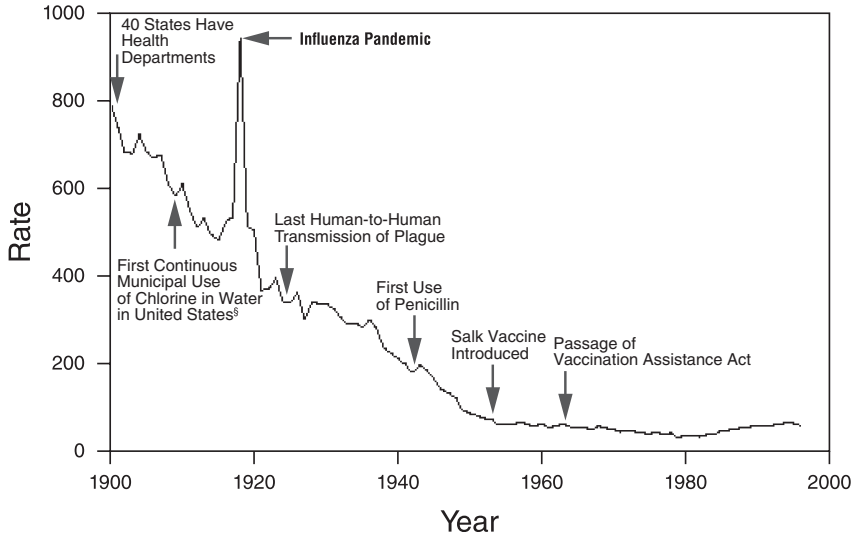
Figure 1: Crude death rate for infection diseases – the United States, 1990–1996.
Source: Achievements in public health 1900–1999: Control of infectious diseases. *Morbidity and Mortality Weekly Report* (1999), 48(**29**):621–629.

growth, few new technologies can be shown to have had any significant effect. But, indeed, it was new and improved technologies that actually caused the continued growth.

Deciding on the method for determining the counterfactual is one of the most difficult and important decisions for a social scientist engaged in policy evaluation. In time series analysis, the assumption that past trends will continue into the future is only an assumption, and it is not always right. Ideally, one would want to understand the reasons for the trend, and whether these causes would have continued into the future.

In statistics, as in life, it is always possible not to find what one is supposedly looking for, or in other words, to find nothing. There are various reasons, including not having the right tools (for example, enough data), not looking in the right place, and not being particularly interested in finding something. Another reason is that there is nothing to be found. However, when one does not undertake a good search, finding nothing does not mean very much.

Unfortunately, researchers who should know better often report their inability to reject the null hypothesis as proving that the null

hypothesis is true. Certainly, it is less compelling, and less newsworthy, to report: 'we can't be sure (at the 95 per cent confidence level) that the there was an effect (that any observed differences were not due to chance)' compared to 'we found no effect', which we interpret to mean that 'THERE WAS NO EFFECT.' This latter conclusion is typically far too strong. The inability of even a well-designed and well-executed single study to reject the null hypothesis is almost never enough to accept the null hypothesis.

Notes

(A) Headlines, such as 'X-ray Evidence Shows Popular Supplements Fail to Slow Knee Osteoarthritis'[12] and 'Supplements No Better Than Placebo in Slowing Cartilage Loss in Knees of Osteoarthritis Patients'[13] accurately summarized the written conclusion of a '24-month double-blinded, placebo-controlled study', which was 'undertaken to evaluate the effect of glucosamine … on progressive loss of joint space width in patients with knee osteoarthritis'.[14]

Taking glucosamine and other supplements has been promoted as a way to reduce cartilage loss. The scientific study was able to test this claim objectively, in part because loss of cartilage in osteoarthritis can be assessed radiographically as interbone distance.

So, can we largely discount the claims of supplement benefits as 'no statistically significant difference in mean joint space width loss was observed in any treatment group compared with the placebo group'?[14] The mean joint space loss for the placebo group was 166 micrometers. By contrast, the mean bone loss for the glucosamine group was 13 micrometers. Glucosamine appears to cut bone density loss by over 90 per cent! Then, why the dismissive headlines?

The authors explain, in this paper, that 'the power of the study was diminished by the limited sample size, variance in joint space width measurement and a smaller than expected loss in joint space width.' In other words, the study did not have the statistical power to find that a supplement that performed nine times better than placebo. The study sample size was too small; the study was (inadvertently) designed to find nothing.

(B) Although their analysis made it very difficult to find an effect of the NFA, the gun lobby researchers still found that firearm suicides fell significantly after the NFA. They legitimately wanted to

determine not only whether the NFA was associated with a fall in firearm suicide, but whether (a) the NFA led to method substitution (for example, hanging suicide replacing gun suicide) and (b) whether something other than the NFA may have affected suicide post-1996. They used non-firearm suicides as evidence for both these concerns. They set up the discussion so that if non-firearm suicides increased after the gun buyback, they could claim this was due to method substitution (that is, the NFA may have reduced firearm suicide, but there was substitution, causing non-firearm suicides to rise, so the NFA really did not have much effect on overall suicides). And if non-firearm suicides decreased, they could claim this showed that some factor other than the buyback was the real cause of the decrease in firearm suicides. When non-firearm suicides briefly fell after the NFA, they attributed this to method substitution, and then when non-firearm suicide began to rise, the authors concluded that 'society changes' (for example, suicide prevention programs) could have been the cause of the observed reduction in firearm suicides.

(C) Other researchers have used sophisticated analyses to search for a single year structural time series break date as a means of identifying the impact of the NFA. They could not find any such break, and concluded that 'the result of these test suggest that the NFA did not have any large effects on reducing firearm homicide or suicide rates'.[15] However, when policies have even modest lags, the structural break test can easily miss the effect. It can also miss the effect of the policy that occurs over several years. The massive Australian gun buyback occurred over two calendar years, 1996–1997. Firearm homicide and firearm suicide dropped substantially in both years, for a cumulative 2-year drop in firearm homicide of 46 per cent and in firearm suicide of 43 per cent. Never in any 2-year period, from 1915–2004 had firearm suicide dropped so precipitousljphpy.

About the Author

David Hemenway, PhD, is Director of the Harvard Injury Control Research Center and the Youth Violence Prevention Center and a professor of Health Policy and Management, Harvard University School of Public Health.

References

1. Neill, C. and Leigh, A. (2007) Weak tests and strong conclusions: A re-analysis of gun deaths and the Australian firearms buyback. Australian National University, Centre for Economic Policy Research. Occasional Paper no. 555.
2. *Targeted News Service*. (2008) Gun shows do not increase homicide, suicide. 1 October.
3. *States News Service*. (2008) Gun shows do not increase homicide, suicide. 1 October.
4. Duggan, M., Hjalmarsson, R. and Jacob, B.A. (2008) The effect of gun shows on gun-related deaths: Evidence from California and Texas. National Bureau of Economic Research, Working Paper no. 14371.
5. Wintemute, G., Hemenway, D., Webster, D., Pierce, G. and Braga, A.A. (2008) Critique of the Dugan *et al* 2008. NBER Working Paper no. 14371, http://economix.blogs.nytimes .com/2008/12/01/the-gun-show-loophole-revisited/, accessed 23 January 2009.
6. Cummings, P. *et al* (1997) Association between purchase of a handgun and homicide or suicide. *American Journal of Public Health* 87: 974–978.
7. ABC Online. (2006) Australia no safer with gun buyback: Study. 24 October.
8. *Sydney Morning Herald*. (2006) Buyback has no effect on murder. 23 October.
9. Baker, J. and McPhedran, S. (2006) Gun laws and sudden death: Did the Australian firearm legislation of 1996 make a difference? *British Journal of Criminology* 47: 455–469, (published online on 18 October 2006).
10. 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: 365–372.
11. Ozanne-Smith, J., Ashby, K., Newstead, S., Statakis, V.Z. and Clapperton, A. (2004) Firearm related deaths: The impact of regulatory reform. *Injury Prevention* 10: 280–286.
12. *Med Page Today*. (2008) X-ray evidence shows popular supplements fail to slow knee osteoarthritis. 30 September.
13. *Science Daily*. (2008) Supplements no better than placebo in slowing cartilage loss in knees of osteoarthritis patients. 1 October.
14. Sawitzke, A.D., Shi, H. and Finco, M.F. *et al* (2008) The effect of glucosamine and/or chondroitin sulfate on the progression of knee osteoarthritis. *Arthritis and Rheumatism* 58: 3183–3191.
15. Lee, W.S. and Suardi, S. (2008) The Australian firearms buyback and its effect on gun deaths. Melbourne Institute Working Paper no. 17/08. Victoria, Australia: Melbourne Institute of Applied Economic and Social Research, The University of Melbourne.