# Evaluating Methods and Findings from a Study of State Gun Policies

*Terry L. Schell and Andrew R. Morral*

In January 2016, the RAND Corporation began work on Gun Policy in America, a research initiative that aims to create a resource where policymakers and the general public can access unbiased information that informs and enables the development of fair and effective firearm policies. Building on a long history of providing policymakers with objective, independent analyses of complex topics, RAND is developing policy analysis tools and research syntheses, grounded in science, with the goal of clarifying the effects of current and proposed firearm measures. We expect to release our initial research findings in the winter of 2016–2017.

The Gun Policy in America initiative employs a mixed-methods approach that incorporates statistical modeling, expert elicitation, a systematic review and synthesis of the research literature, and other techniques that draw on RAND's expertise in objectively analyzing complex policy challenges.

In March 2016, as RAND's research effort was getting under way, the British medical journal *The Lancet* published an article by Bindu Kalesan and colleagues titled "Firearm Legislation and Firearm Mortality in the USA: A Cross-Sectional, State-Level Study" (Kalesan et al., 2016a). The article examined the effects of 25 state firearm laws on gun deaths. Given the relevance of the findings to our ongoing research, we read the findings with great interest and care. In the course of that review, we identified a number of serious analytical errors that we suspected could undermine the article's conclusions.

It is important to note that many—perhaps most—scientific articles include some mistakes, and normally this would not merit further attention. In this case, however, we could see that the article's unusually strong and significant effects were probably wrong but would dominate our synthesis of findings from the research literature. Moreover, we could not ignore the article without violating the rules of our systematic review. The reported findings appeared likely to support bad gun policies and to hurt future research efforts that could treat the many highly significant findings reported in the article as valid. We contacted the article's authors with our concerns but ultimately determined that new analyses were needed to test the validity of the published findings. We reconstructed the article's data set using information in the article and reanalyzed the article's models using the procedures documented in this report.

This short technical report serves as a supplement to our comment published in *The Lancet* in response to the article and will be of greatest interest to readers who are familiar with statistical methods. Because our comment in *The Lancet* was limited to 250 words, we were not able to include detail on our reanalysis and additional statistical support for our conclusions. Therefore, this report provides the background on our assessment and a detailed account of the errors we detected. Data and code for the analyses described here accompany this report on RAND's website at www.rand.org/t/RR1642.

## OVERVIEW OF ANALYTICAL FLAWS IN THE ARTICLE ON STATE GUN POLICIES PUBLISHED IN *THE LANCET*

In their article, Kalesan and colleagues address the important question of whether firearm laws implemented in U.S. states are associated with reductions in firearm-related mortality (Kalesan et al., 2016a). Their surprising results suggested, among other things, that nationwide implementation of just three laws would reduce firearm homicides, suicides, and accidental deaths by 98 percent—and that laws concerning the use of firearm locks to prevent unauthorized use are associated with a nearly elevenfold increase in firearm homicides.

However, the central statistical model that served as the basis for these findings used 39 parameters (38 predictor degrees of freedom and an intercept) to explain just 50 observations. This suggests model overfitting, a phenomenon in which researchers attempt to estimate a model that is more complex than can be supported with the available data. As a result, the model's parameter estimates and significance tests may be inaccurate. We reanalyzed the article's data and found clear evidence that its substantive conclusions were invalid. Specifically, we found that the article's multivariate effects were *not* statistically significant, that the point estimates for multivariate incidence rate ratios (IRRs) reported in the article deviated substantially from estimates produced with more appropriate statistical methods, and that there were factual errors and inconsistencies in the text and tables of the article.

## STATISTICAL PROBLEMS: MODEL FIT, STANDARD ERRORS, AND ESTIMATES

The article's parameter estimates and their confidence intervals were incorrect, making them appear to support invalid conclusions. Two problems led to these errors: overdispersion in the data relative to the assumed likelihood function and inaccurate estimates of standard errors due to model overfitting.

### Errors in the Evaluation of Model Fit

The article posits that the residuals of the models used to generate IRR estimates are Poisson-distributed. The article acknowledges this assumption, stating, "Because the variance of our outcome was equal to the mean, we used Poisson regression with population as an offset to normalise population sizes, and

We reanalyzed the article's data and found clear evidence that its substantive conclusions were invalid.

robust standard errors" (Kalesan et al., 2016a, pp. 1849–1850). However, it later correctly reports the mean of the outcome to be 632 firearm deaths per 100,000 people, with a standard deviation of 629. Thus, the variance is *not* equal to the mean but approximately 600 times larger. Although this discrepancy in the mean and variance does not prove overdispersion, substantial overdispersion is evident in the models used to estimate the crude IRRs presented in the article. After the population offset was applied, the models used to estimate the 25 crude IRRs had residual variance approximately 15 times larger than expected under a Poisson distribution. The article then presents Huber-corrected standard errors (robust standard errors in Stata), which widened the confidence intervals around the crude IRRs, potentially mitigating this violation of model assumptions.

Although robust standard errors can correct for the effect of overdispersion on the standard errors of model parameters, this procedure does not correct the likelihood statistics that are used to evaluate model fit. The article's supplementary appendix describes how the final model was selected (Kalesan et al., 2016b). Using the Akaike Information Criterion (AIC), residual deviance, and McFadden's R-squared, according to the article, the final model had a better fit than models with fewer predictors. However, because overdispersion in the data was not addressed in the likelihood function (only in the post-estimation calculation of standard errors), all three of these fit criteria are biased in favor of excessively complex models (see Burnham and Anderson, 2002). As such, adding parameters to the models may appear to improve model quality, but a model that correctly accounts for overdispersion would reveal that the opposite is true. Moreover, these criteria are valid only asymptotically in large samples. When the number of predictors is close to the number of observations, a correction to the likeli-

hood, such as the corrected AIC, is required (Burnham and Anderson, 2002). Both of these problems (overdispersion and violation of asymptotic assumptions) resulted in a bias toward selecting a more complex model than the data could support.

By several metrics, the multivariate model estimated and interpreted in the article was far too complicated for the data. Although the article did not explicitly report that the multivariate model used 39 parameters to fit 50 data points, this can be computed from the AIC and deviance values presented in the article's appendix. Specifically, the model included an intercept, 25 laws, four categorical covariates with four levels each, and one continuous covariate. Such a high ratio of the number of predictors to the number of observations violates key assumptions of the article's statistical methods, which are asymptotically correct only in large samples. While the field does not have precise definitions for when large sample assumptions are violated, requiring at least ten cases per parameter is a common rule of thumb (see Draper and Smith, 1998; Peduzzi et al., 1996). However, others suggest no fewer than 15 cases per parameter, or that sample size requirements grow exponentially with increasing numbers of parameters (Good and Hardin, 2012; Cavanaugh, 1997). By all standard recommendations for the methods described in the article, anything above ten or even five parameters should be suspected as leading to an overfit model when modeling 50 data points (Harrell, Lee, and Mark, 1996).

By definition, overfit models yield residual deviances that are misleadingly low, meaning that the in-sample deviances provide inaccurate information about true prediction or generalization error. As a result, *p*-values and standard errors based on large-sample asymptotic distributions are smaller than they should be and yield false significance.

We demonstrated that the residual deviance for the article's multivariate model was too low by comparing it to a cross-validated estimate of the residual deviance. (Data and code for all analyses described here accompany this report on RAND's website at www.rand.org/t/RR1642, and we encourage others to investigate and replicate our methods.) We used "leave one out" cross-validation to estimate prediction error. With this method, we ran the model for 49 states and used it to predict the state that was not included in the estimation. We repeated this process for all 50 states and summed the total Poisson deviance of those 50 predictions.

Unfortunately, the article's model was not sufficiently identified to cross-validate the full model; dropping even a single state resulted in predictors that are perfectly correlated or have

no variance, violating the model's requirements for estimation. Indeed, the predictor "Child handgun restrictions" applies to only a single state (i.e., it is a variable indicating "New Jersey"). To overcome this problem, we cross-validated a simpler model that dropped the "Child handgun restrictions" and "Large magazine ban" predictors, neither of which was a significant predictor in the article's multivariate model.

This reduced model had a cross-validated residual deviance of 465.5, which is approximately 28 times larger than the in-sample deviance from the article's model, 16.7 (Kalesan et al., 2016b, Supplementary Table 1). In contrast, models that are not overfit provide very similar cross-validated and in-sample deviance. For example, a model for the crude effect of "Gun dealer license" had a cross-validated deviance that was larger than the in-sample deviance by a factor of just 1.12. In short, the article's multivariate model included so many predictors that it almost perfectly explains the data used to estimate it, yet it also has a high generalization error—meaning that it makes very inaccurate predictions about "new" data.

## Errors in the Calculation of Confidence Intervals and Significance

While this analysis demonstrates that the model is overfit (that is, it yields a deceptively small residual variance in the sample on which it is estimated and makes inaccurate predictions about new data), this does not, itself, indicate that the article's conclusions are wrong. It does, however, indicate that one needs to estimate confidence intervals and statistical significance in a manner that does not rely on large-sample distributional assumptions (Harrell, Lee, and Mark, 1996).

Resampling methods (e.g., bootstrapping, permutation, exact tests) are frequently used for hypothesis testing and for generating confidence intervals when the requirement for a large sample and distributional assumptions of maximum likelihood methods are not met (Good, 2005). However, none of the multivariate IRRs on which the article's conclusions depend are statistically significant when evaluated with such methods. The table on the following page compares confidence intervals for the article's statistically significant multivariate effects with corresponding bootstrapped confidence intervals (see Efron, 1979). The bootstrapped intervals are wider by orders of magnitude. The predictors with the three smallest IRRs in the article (bold in the table) were the primary focus of the article's conclusions. Our comparison revealed that these predictors have effectively unbounded bootstrapped confidence intervals.

**Comparing the Article's 95% Confidence Intervals to Bootstrapped Confidence Intervals for Multivariate IRRs Interpreted as Significant in the Article**

| Predictor in Multivariate Model | Kalesan et al. | | Bootstrap | |
|---|---|---|---|---|
| | Lower Limit | Upper Limit | Lower Limit | Upper Limit |
| Gun dealer license | 0.85 | 0.97 | 0.36 | 2.92 |
| Record keeping and retention | 0.74 | 0.85 | 0.06 | 4.58 |
| Mandatory theft reporting | 1.26 | 2.13 | 0.03 | 132.96 |
| Gun store security precaution | 0.76 | 0.92 | 0.07 | 5.31 |
| Police inspection | 1.10 | 1.30 | 0.20 | 12.10 |
| Bulk purchases limitation | 1.39 | 1.70 | 0.08 | 25.67 |
| **Firearm identification** | **0.09** | **0.29** | **0.00** | **>1,000** |
| Owner theft reporting | 0.40 | 0.74 | 0.01 | 25.58 |
| **Universal background checks** | **0.23** | **0.67** | **0.00** | **91.54** |
| Safety training | 0.45 | 0.73 | 0.02 | 16.54 |
| Extension of background-check limit | 1.09 | 1.63 | 0.05 | 37.04 |
| Permit law involvement | 0.61 | 0.80 | 0.03 | 8.26 |
| Closure of gun show loophole | 1.03 | 1.15 | 0.21 | 6.48 |
| **Ammunition background checks** | **0.09** | **0.36** | **0.00** | **>1,000** |
| Firearm locks | 2.12 | 7.15 | 0.00 | >1,000 |
| Assault weapon ban | 1.11 | 2.59 | 0.02 | 96.38 |
| Discretion allowed when issuing permits to carry a concealed weapon | 1.11 | 1.29 | 0.24 | 9.81 |
| Stand-your-ground | 1.03 | 1.12 | 0.57 | 1.72 |

NOTES: Bootstrap estimates are based on 20,000 replicated data sets using a standard case-resampling bootstrap procedure. Bold indicates the predictors with the three smallest IRRs in the article that were the primary focus of the article's conclusions.

The multivariate model used to generate our bootstrapped confidence intervals was slightly different from the model used in the article. Because of these differences, the bootstrapped intervals may underestimate the true confidence intervals. The multivariate model used in the bootstrap analysis dropped the seven predictors with IRRs near 1 reported as nonsignificant in the article.[1] We made this modification because many of those variables had little variance or were highly collinear, and, thus, the model was not estimable in most of the resampled data sets. For example, a model that included "Child handgun restrictions" could be estimated only in resampled data sets that include New Jersey because all other states have identical values. Essentially, the full model was overfit to such an extent that it could not be estimated within almost any random subset of the 50 states. Dropping those reportedly nonsignificant predictors resulted in a more parsimonious model as judged by cross-validation. In addition, all of the laws that were statistically significant in the article's multivariate model were also significant in the simplified version used here when significance is assessed using the same method as was used in the article.

[1] These predictors were "Report records to state," "Fingerprinting," "Ammunition purchaser records," "Child handgun restrictions," "Child access not permitted," "Juvenile handgun purchases," and "Large magazine ban."

Thus, the model used to produce our bootstrapped estimates is anticonservative because it underestimates the width of the true confidence intervals for the article's model. Despite this underestimation, none of the effects reported in the article approach statistical significance using bootstrapped standard errors.

The use of the bootstrap does require certain assumptions (see Good, 2005), and readers may wonder whether the case-resampling method we used is the most appropriate for estimating confidence intervals in this data set. We acknowledge that it is difficult to know whether our bootstrapped confidence intervals actually reflect 95-percent coverage of the true population value. To partially address this concern, it is helpful to note that the bootstrapped confidence intervals for the crude IRRs are almost identical to the Huber-corrected intervals reported in the article. These confidence intervals are presented as an appendix to this report. The discrepancy between the two methods occurs only in the more complicated multivariate model. More generally, the case-resampling bootstrapped confidence intervals very closely correspond to the Huber-corrected confidence intervals *until* the models get highly complex. When the number of parameters becomes a substantial fraction of the sample size (e.g., 25 degrees of freedom on 50 cases), the two methods dramatically diverge. The Huber-corrected intervals get progressively narrower as additional parameters are added and residual deviance necessarily approaches zero; in contrast, the bootstrapped intervals correctly get larger as the model becomes less well identified.

## Errors in the Estimated IRRs

These analyses demonstrate that the effects detected with the article's multivariate model are not statistically significant, but we also found that the reported IRR estimates are very likely incorrect and should not be interpreted. When the number of predictors is close to or greater than the number of data points, standard methods to estimate effects may return estimates that are dramatically different from those of the true data-generating model.

Unfortunately, there is no generally accepted solution to the problem of estimating interpretable coefficients when the number of parameters is close to (or exceeds) the amount of available data. However, there are some methods that are commonly used to address this model selection problem. Often, these methods use a model quality indicator (e.g., AIC, cross-validated error) to select a best subset of predictors (e.g., stepwise regression) that can be well estimated—or they

shrink model coefficients toward the null (e.g., IRR = 1) until the model predictions minimize cross-validated error. (See, for example, Hoerl and Kennard, 1970; Tibshirani, 1996). When we applied such methods, they produced multivariate IRR estimates that differed widely from those reported in the article. For example, the LASSO method (Tibshirani, 1996) is an alternate approach to estimating the multivariate Poisson model that uses coefficient shrinkage to avoid overfitting. LASSO estimation yields a solution in which all multivariate IRRs are between 0.95 and 1.05. The pattern of effects is also substantively different; the LASSO multivariate IRRs have a correlation of just 0.14 with those reported in the article. The LASSO model has a cross-validated residual deviance of 75, compared to 465 for the article's model, indicating that it is much more accurate in predicting the rate of firearm deaths for states that were not used in the model estimation.

It is possible that the LASSO method, which relies on shrinking the model coefficients, may have resulted in coefficients that are biased toward being too small (i.e., yielding IRRs too close to 1) when applied to such a small data set. However, even if one used conventional stepwise regression—which suffers from the opposite problem, producing estimates that are biased away from IRR = 1—the best multivariate model estimated IRRs that are completely different from those reported in the article. Specifically, we used "forward-backward" stepwise variable selection with the minimum cross-validated error as the stopping criterion. With this method, the model includes the strongest predictors in the sample, and attempts to add or subtract predictors until no predictor can be added or removed without increasing the cross-validated error. The best model using this method of variable selection included only five laws as predictors, all of which had IRRs that fell into a narrow range between 0.93 and 1.07. In contrast, the article reported multivariate IRRs for laws that varied from 0.16 to 3.90. In short, even a stepwise regression approach to variable selection—which is known to exaggerate effect sizes—found no evidence that any of these laws had large effects on firearm fatalities when conditioned on the available covariates (most notably the prior year's firearm fatalities, which entered the model as the best predictor). The large multivariate effect sizes reported in the article emerge only if one continues to add predictors to the model even when they degrade the cross-validated predictive accuracy.

Published statistical analyses often violate assumptions in ways that underestimate standard errors. In many cases, these violations are harmless because the estimate is essentially

sound, though the *p*-values may be modestly anticonservative. In the article, however, the multivariate IRRs bear little relationship to the estimates derived from the same data using more-appropriate models. Therefore, we concluded that the reported multivariate IRRs should not be used or interpreted.

## ERRORS IN THE DESCRIPTION OF STATE LAWS

We recreated the article's data set using (1) the state-level data on covariates and laws provided in the article and its supplementary appendix (Kalesan et al., 2016a and 2016b, respectively), and (2) the state-level firearm-related death counts and population sizes from the U.S. Centers for Disease Control and Prevention's Web-Based Injury Statistics Query and Reporting System (WISQARS). The total number of firearm deaths in 2010 reported in WISQARS exactly matched the article's numbers.

However, there were inconsistencies in the article and its appendix in the description of which laws applied to each state. Table 1 in the article describes each state's laws in a way that is sometimes inconsistent with the descriptions in the subsequent table (Kalesan et al., 2016a, Table 2) and in Supplementary Appendix Table 2 (Kalesan et al., 2016b). For example, California is listed as not having a law closing the gun show loophole in the article's Table 1, but it is shown as having such a law in Table 2. We identified 11 such inconsistencies, which are listed in the table below. Moreover, neither the data presented in Table 1 nor the data in Table 2 and Supplementary Appendix Table 2 exactly reproduced all IRRs presented in the article's crude associations (Kalesan et al., 2016a, Figure). That is to say, neither set of tables accurately represents the data actually used in the article's statistical analyses.

By trying various combinations of possibilities when the tables disagreed, we identified a state law listing that matched the "Crude IRR" estimates and confidence intervals shown in the figure in the article. There was one exception ("Gun store security precaution"), for which we matched the confidence interval listed in the figure, but not the point estimate. However, the article's crude IRR for this law was inconsistent with its own confidence intervals, indicating an error in the article's tabled value (see Kalesan et al., 2016a, Figure). Thus, we are confident that our reconstructed data set is functionally identical to the data used to create the crude IRR estimates in the published article. This reconstructed data set produced

multivariate IRR estimates that were slightly different from the original results; they have a correlation of 0.99 with the article's multivariate IRRs. Similarly, the AIC values were close for the article's full model, which was used for all study conclusions (485.7 versus 484.7 for the original and reconstructed data sets, respectively). The table on the following page explains how we resolved the inconsistencies across the article's tables.

## CONCLUSIONS

The IRRs on which the article's conclusions were based are not statistically significant, and the IRR estimates themselves are substantively different from estimates using more-appropriate methods.

Uncorrected, the article offers high-profile support for claims about gun policy for which there is no evidence. Some of the policies for which the article claims support could do harm. For instance, if policymakers accept the article's finding that laws requiring gun locks *increase* firearm deaths, this could lead to policies relaxing gun lock requirements, even though the article presents no credible evidence to support such a change. Our corrected analyses found no statistically significant evidence for any of the article's major conclusions.

Although we identified problems with this particular article, we hold past work by the research team in high regard. Our interest in reporting this reanalysis is to correct the scientific record and to justify our decision to exclude this article from RAND's ongoing systematic examination of the effects of gun policies. The article's unusually large—and erroneous—effects would otherwise dominate our synthesis of the gun policy research and those of other meta-analyses for decades to come. We have opted to publicly share this evaluation because effective gun policies depend on accurate, objective research, and the article's findings could damage the analytical support for these decisions and lead to flawed policies.

Finally, we hope that carefully documenting these problems—and demonstrating statistical methods that can identify and correct them—may serve as a guide to researchers in this field. The problems caused by an overfit model and a misspecified likelihood function are not unique to this particular article. The underlying problems are present in a substantial fraction of published studies that attempt to link state-level policies with mortality or crime data. We hope that this document can encourage the field to adopt better statistical methods, resulting in improved substantive conclusions.

**Resolving Inconsistencies for Use in the Reconstructed Data**

| Law | State | Value Used | Data Source |
|---|:---:|:---:|:---:|
| Bulk purchases limitation | Virginia | Y | Table 1 |
| Stand-your-ground | Louisiana | Y | Table 1 |
| Stand-your-ground | Arkansas | N | Sup. Table 2 |
| Stand-your-ground | Arizona | Y | Sup. Table 2 |
| Stand-your-ground | Nevada | N | Sup. Table 2 |
| Stand-your-ground | Oregon | N | Sup. Table 2 |
| Stand-your-ground | North Dakota | N | Table 1 |
| Gun store security precaution | Connecticut | Y | Sup. Table 2 |
| Closure of gun show loophole | California | Y | Table 2 |
| Assault weapons ban | North Dakota | N | Sup. Table 2 |
| Assault weapons ban | Maryland | Y | Sup. Table 2 |

NOTES: A "Y" value indicates that we concluded that the article's analysis assumed the state had such a law in place; an "N" value indicates that the article assumed no such law was in place. For example, our reconstructed data set indicated that Virginia had a bulk purchase limit as shown in Kalesan et al., 2016a, Table 1, even though the opposite claim was made in Kalesan et al., 2016b, Supplementary Table 2.

**Appendix: Comparing the Article's Confidence Intervals to Bootstrapped Confidence Intervals for Crude IRRs**

| Predictor | Kalesan et al. | | Bootstrap | |
|---|---|---|---|---|
| | **Lower Limit** | **Upper Limit** | **Lower Limit** | **Upper Limit** |
| Gun dealer license | 0.60 | 0.89 | 0.60 | 0.91 |
| Record keeping and retention | 0.65 | 0.92 | 0.65 | 0.92 |
| Report records to state | 0.57 | 0.85 | 0.56 | 0.88 |
| Mandatory theft reporting | 0.56 | 0.84 | 0.39 | 0.88 |
| Gun store security precaution | 0.63 | 0.95 | 0.62 | 0.99 |
| Police inspection | 0.65 | 0.95 | 0.65 | 0.97 |
| Bulk purchases limitation | 0.62 | 0.89 | 0.51 | 1.01 |
| Firearm identification | 0.52 | 0.82 | 0.46 | 0.87 |
| Owner theft reporting | 0.45 | 0.87 | 0.43 | 0.88 |
| Universal background checks | 0.60 | 0.86 | 0.47 | 0.88 |
| Fingerprinting | 0.47 | 0.76 | 0.41 | 0.75 |
| Safety training | 0.57 | 0.87 | 0.37 | 0.91 |
| Extension of background-check limit | 0.50 | 0.75 | 0.47 | 0.75 |
| Permit law involvement | 0.46 | 0.88 | 0.45 | 0.88 |
| Closure of gun show loophole | 0.55 | 0.79 | 0.53 | 0.80 |
| Ammunition purchaser records | 0.61 | 0.84 | 0.40 | 0.89 |
| Ammunition background checks | 0.42 | 0.85 | 0.37 | 0.80 |
| Firearm locks | 0.53 | 0.76 | 0.50 | 0.76 |
| Child handgun restrictions | 0.46 | 0.56 | 0.45 | 0.55 |
| Child access not permitted | 0.64 | 0.98 | 0.63 | 0.98 |
| Juvenile handgun purchases | 0.51 | 0.71 | 0.49 | 0.70 |
| Assault weapon ban | 0.45 | 0.72 | 0.41 | 0.68 |
| Large magazine ban | 0.43 | 0.73 | 0.35 | 0.68 |
| Discretion allowed when issuing permits to carry a concealed weapon | 0.52 | 0.75 | 0.50 | 0.77 |
| Stand-your-ground | 1.27 | 1.74 | 1.27 | 1.76 |

NOTE: Bootstrap estimates are based on 10,000 replicated data sets using a standard case-resampling bootstrap procedure.

## REFERENCES

Burnham, Kenneth P., and David R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed., New York: Springer, 2002.

Cavanaugh, Joseph E., "Unifying the Derivations for the Akaike and Corrected Akaike Information Criteria," *Statistics and Probability Letters*, Vol. 33, No. 2, April 1997, pp. 201–208.

Draper, Norman R., and Harry Smith, *Applied Regression Analysis*, 3rd ed., New York: John Wiley and Sons, 1998.

Efron, Bradley, "Bootstrap Methods: Another Look at the Jackknife," *Annals of Statistics*, Vol. 7, No. 1, 1979, pp. 1–26.

Good, Phillip I., *Permutation, Parametric, and Bootstrap Tests of Hypotheses*, 3rd ed., New York: Springer-Verlag, 2005.

Good, Phillip I., and James W. Hardin, *Common Errors in Statistics (and How to Avoid Them)*, 4th ed., New York: John Wiley and Sons, 2012.

Harrell, Frank E., Jr., Kerry L. Lee, and Daniel B. Mark, "Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors," *Statistics in Medicine*, Vol. 15, No. 4, February 29, 1996, pp. 361–387.

Hoerl, Arthur E., and Robert W. Kennard, "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics*, Vol. 12, No. 1, February 1970, pp. 55–67.

Kalesan, Bindu, Matthew E. Mobily, Olivia Keiser, Jeffrey A. Fagan, and Sandro Galea, "Firearm Legislation and Firearm Mortality in the USA: A Cross-Sectional, State-Level Study," *The Lancet*, Vol. 387, No. 10030, April 30, 2016a, pp. 1847–1855.

———, supplementary appendix to "Firearm Legislation and Firearm Mortality in the USA: A Cross-Sectional, State-Level Study," *The Lancet*, Vol. 387, No. 10030, April 30, 2016b. As of July 8, 2016:
http://thelancet.com/cms/attachment/2057150372/2061572438/mmc1.pdf

National Centers for Injury Prevention and Control, Centers for Disease Control and Prevention, Web-Based Injury Statistics Query and Reporting System (WISQARS), online database. As of July 8, 2016:
https://www.cdc.gov/injury/wisqars/

Peduzzi, Peter, John Concato, Elizabeth Kemper, Theodore R. Holford, and Alvin R. Feinstein, "A Simulation Study of the Number of Events per Variable in Logistic Regression Analysis," *Journal of Clinical Epidemiology*, Vol. 49, No. 12, December 1996, pp. 1373–1379.

Tibshirani, Robert, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society, Series B (Methodological)*, Vol. 58, No. 1, January 1996, pp. 267–288.

## About This Report

As part of RAND's Gun Policy in America research initiative, this critique of "Firearm Legislation and Firearm Mortality in the USA: A Cross-Sectional, State-Level Study," published by Kalesan et al. in the journal *The Lancet*, serves as a technical appendix to the RAND study team's comment on the article, also published in *The Lancet*. Because our comment was limited to 250 words, we were not able to include detail on our reanalysis and additional statistical support for our conclusions. Therefore, this report provides technical background on our assessment, including a detailed account of the errors detected. It will be of greatest interest to readers familiar with statistical methods. Data and code for the analyses described here are available for download at www.rand.org/t/RR1642.

Gun Policy in America aims to create a resource where policymakers and the general public can access unbiased information that informs and enables the development of fair and effective firearm policies. RAND's projects and analyses do not necessarily reflect the opinions of its donors or supporters. More information is available at www.rand.org/research/gun-policy.

## Limited Print and Electronic Distribution Rights

# www.rand.org