Journal of Quantitative Criminology, Vol. 19, No. 2, June 2003 (© 2003)

Measurement and Other Errors in County-Level UCR Data: A Reply to Lott and Whitley

Michael D. Maltz[1,2] and Joseph Targonski[1]

Lott and Whitley note that our analyses of the errors in the county-level UCR data used in *More Guns, Less Crime* (J. R. Lott, University of Chicago Press, Chicago, 1998, 2000) ignore the fact that all data have measurement error, that the largest errors were in counties with low populations, and that population-weighted regressions were used. We agree that this mitigates some of the effects of the errors, but does not take them fully into account. We also note that this is but one of the problems associated with the analysis. We therefore find no reason to alter our original conclusion, "that in their current condition, county-level UCR crime statistics cannot be used for evaluating the effects of changes in policy."

KEY WORDS: Uniform Crime Reports; imputation; concealed weapons; guns; county-level crime data; measurement error.

1. INTRODUCTION

In their critique of our paper, which described the effect of missing data on the analyses in *More Guns, Less Crime* (Lott, 1998, 2000) that use county-level UCR data, Lott and Whitley (2003) make a number of statements. They include:

- we ignored the effect of population-weighted regression;
- all data sets have measurement error, and survey data have much more missing data than are found in Lott's (1998, 2000) analyses;
- we used the wrong states in our analyses;
- state-level analyses are similarly affected by missing data;
- state- and city-level analyses show the same pattern as the county-level analysis.

[1]Department of Criminal Justice, University of Illinois at Chicago, 1007 W. Harrison Street, Chicago, Illinois 60607-7140.
[2]To whom correspondence should be addressed. E-mail: mikem@uic.edu

0748-4518/03/0600-0199/0

Our original paper dealt primarily with the problems inherent in the UCR, not with the specifics of the analyses in *More Guns, Less Crime* (*MGLC*). This is the primary reason that we did not "directly test if it affects the results" (Lott and Whitley, 2003).

In this paper, however, we respond to the Lott and Whitley with respect to each of the above points. In Section 2 we discuss the effect of weighting the observations by population. In Section 3 we examine the effect of errors, including missing data. Section 4 discusses the choice of states we used. In Section 5 we discuss state-level analyses. Section 6 contains our conclusion.

2. POPULATION-WEIGHTED REGRESSION

Lott and Whitley are, of course, correct in their statement that a regression weighted by population would diminish the effect of the small counties that (generally) have less consistent reporting histories. As we stated in our paper, Lott (2000, pp. 143, 155) criticized Black and Nagin's (1998) analysis for ignoring the smaller counties and focusing their analysis on counties with 100,000 or more population, so we included all counties in our paper. Their reanalysis of our data shows that when weighted by population the problem is reduced. For example, we note from their Fig. 3 that, although there are 21 states that have over 10% of their *observations* with extensive missing data (defined by 30% or more coverage gaps), only 15 states have over 10% of their *population* affected by this magnitude of coverage gaps. This is still a substantial amount of error.

In both editions of *MGLC* the data were taken at face value, assumed to be error-free. Now that the nature of the errors is more thoroughly understood, Lott and Whitley suggest that they do not affect the results substantially. They reanalyzed the *MGLC* data removing the 16 most error-laden states; however they do not state how (or whether) they dealt with the errors that remained in the data of the remaining 44 states. If they did not adjust for the errors in those states, this step is inadequate.

3. ERRORS IN THE COUNTY-LEVEL DATA SET

The errors are not insignificant. Moreover, there are a number of different types of error that affect analyses, not just the ones to which Lott and Whitley refer. They are correct in their assertion that "[v]irtually all data have measurement error," in referring to the problems with UCR missing data that we documented (Maltz and Targonski, 2002). Actually, this is only one type of error that affects their findings; instrumentation error and error due to stochastic variation also need to be considered. We discuss these three types of error below.

3.1. Measurement Error Due to Missing Data

Lott and Whitley note that surveys often have much more missing data than do the UCR. But there are many different kinds of "missingness" in data, and they must be handled in different ways. Nonresponse from agencies may be due to computer problems, lack of personnel, insufficient training of staff, or lack of priority to record keeping (Maltz, 1999). Nonresponse from survey respondents may be due to refusal, language problems, or not being able to contact the respondent. Specific items may be missing, or whole observations may be missing. One needs to understand (and model) the process that generates the "missingness" to compensate for it.

Most programs that impute missing data make certain assumptions about their characteristics. For example, sophisticated techniques such as multiple imputation developed by Little and Rubin (1987) and Rubin (1987) provide consistent and efficient results. However, most multiple imputation procedures assume that the data are *missing at random*. When UCR data are missing specific items (e.g., no rape data when the other data are reported), they are not missing at random.[3] Nor are agencies that omit reporting their data omitting reports at random.

In addition, standard practice is to use methods that assume that the errors are independent, normal, and with zero mean. None of these hold for the errors in county-level UCR data: they are non-independent, non-normal, and virtually all negative; i.e., crime counts for specific months or for specific agencies within a county have been omitted, not overstated. What this does is generate a negative bias in the estimated crime rate for those agencies at those times, because the numerator (number of crimes) is reduced while the denominator (county population) is not.

3.2. Instrumentation Error

The instrument performing the measurement may also be in error. When we step on the bathroom scale in the morning, we note that when we jiggle the scale, successive measurements are different. We then may take the mean (or more optimistically, the lowest) value as correct. However, when the bathroom scale is broken, we don't accept the measurement at all. In

[3]The current method of imputation used by the FBI is cross-sectional and assumes that "similar" jurisdictions will have crime trends similar to those of the agency with missing data. An alternative method is a longitudinal approach, using an agency's past reporting behavior to impute for the missing values (Maltz, 1999).

many of the cases we uncovered in our analysis of the UCR crime data, the scale is broken. This was our concern in our original paper.

Lott and Whitley attempt to account for this by eliminating observations, known as *listwise deletion* in the imputation literature; however, we are concerned with the procedure they used. They did not delete only those *counties* that had many observations with high levels of missing data. Rather, even though it is a county-level analysis, they deleted entire *states* that had many counties with high levels of missing data.

It is unclear how this particular method of handling missing data affected the result. What is absent is a rationale (and a model) for this means of compensation for the missing data. And rationales do exist; for example, Dugan (2002) describes how one can use graphical methods to determine the sensitivity of the results of an analysis to specific time periods or observations.

Another instrumentation error results from the change in instruments. When the National Archive of Criminal Justice Data (NACJD) changed its method of imputation, it specifically stated, "*data from earlier year files should not be compared with data from 1994 and subsequent years because* [of] *changes in procedures used to adjust for incomplete reporting*" (data set 6669). Yet the county-level analysis in the second edition of *MGLC* did make this comparison. Although they do not mention it, we assume that Lott and Whitley do not take issue with our contention that the county-level analysis in the second edition is fundamentally flawed, and that their concern is with the county-level analysis in the first edition (Lott, 1998) and the state- and city-level analyses in the second edition.

3.3. Stochastic Variation

Implicit in the analysis of homicide (or any set of) data is the assumption that there is a "natural" rate that can be calculated, that the rate is affected by various factors, and that the factors included in the analysis both are sufficient to characterize the rate and are included in the analysis in the appropriate form. This natural rate is not fixed, but is instantiated by the specific homicide rate for the specific county and year and set of factors that are relevant for that county-year.

This natural rate also has a natural variation. To understand this, suppose a jurisdiction's homicide count fell from 10 in one year to 9 in the next year—we would not expect to see headlines trumpeting the 10% decrease in homicide. If the count fell from 100 to 90, we would feel more comfortable about this being a real decrease; and if it fell from 1000 to 900, we would feel even more comfortable. That is, even when the figures are themselves error-free there is an uncertainty associated with them.

Barnett (1981a, b; see also Maltz, 1994, 1996) estimated this uncertainty; he showed that a state's annual homicide count is approximately normal and the variance associated with it is slightly larger than the homicide count itself. This translates into an uncertainty, as measured by the standard error, approximately equal to the square root of the count.

If we adopt the customary two standard errors as a measure of uncertainty, such that any year-to-year variation that falls within two standard errors is considered to be indistinguishable from natural fluctuation, then if one year's homicide count was 10, it would not be inconceivable that the next year's count were to fall between 4 and 16 (or $\pm 60\%$) absent any change in legislation or any other factors that might affect homicide; if the "before" count were 100, the next year's count might easily be between 80 and 120 (or $\pm 20\%$); and if it were 1000, the next year's count could fall between 969 and 1031 (or $\pm 6\%$) and not be attributable to a change.

In other words, the uncertainty associated with each datum constitutes a high hurdle that a finding of deterrence must overcome. It is difficult enough to overcome this hurdle with state-level data, as Barnett showed; it is even more difficult to overcome this hurdle with county-level data.

4. CATEGORIZING STATES

Lott originally assigned states to three categories: those that had right-to-carry (RTC) laws in existence before 1977, those that did not have RTC laws during the time period 1977–1992, and those that changed their gun laws to include RTC during the time period 1985–1991. Lott's assignments were questioned by Marvell (2000) in an email sent to, among others, Lott and Maltz. Moreover, Vernick and Hepburn (forthcoming) analyzed the gun laws of the 50 states, and also have a different assignment of states to categories.

The assignment of states to categories is far from exact, and depends on interpreting state statutes. Moreover, the practices (what occurs on the ground, rather than what is written in the statutes) do not necessarily have much to do with the statutes: Texas passed a RTC law in 1996 while Connecticut has had one on the books for decades, but few would consider residents of Connecticut to be more inculcated in the gun culture than Texans (e.g., see Lundsgaarde, 1977). We stand ready to redo our analyses once all can agree on standards for categorizing states in terms of their RTC laws and practices.

5. STATE- AND CITY-LEVEL ANALYSES

It is not the case, as asserted by Lott and Whitley, that state-level analyses are affected to the same extent as county-level analyses by missing data. There are many instances where one of the largest agencies in a *county* has missing (and unimputed) data, thus having a major effect on that county's crime rate, as shown in Maltz and Targonski (2002). In contrast, rare is the case in which one of the largest agencies in a *state* has missing (and unimputed) data; this would be tantamount to not taking Chicago's data into consideration when the Illinois crime rate is being calculated.

The state-level files provided by the FBI (and BJS) take missing data into account by imputing *all* missing agency data. This contrasts with the county-level files provided by NACJD, which impute missing data only if an agency provides at least six months of data; otherwise, the agency is dropped completely (see Maltz, 1999).

We did not examine the city-level data set used by Lott, so we are unable to comment on it.

6. CONCLUSION

The changes in concealed carry laws did not take place in a vacuum. They were probably prompted to some extent by the dramatic increase in homicide that occurred during the late 1980s and early 1990s. The subsequent drop should have been anticipated, if only due to regression to the mean; in fact, there is some evidence that this was the case for the plummeting homicide rate in New York City (Maltz, 1998).

Success has a hundred fathers and failure is an orphan. Was the decrease in homicides due, as Lott suggests, to this legislation, or to the Federal program to increase police staffing, or to the advent of community policing, or to the waning of the crack epidemic, or to the efforts of the New York City Police Commissioner, or to the increased use of prisons over the past few decades, or to economic and demographic trends? All have been proposed as possible causes (Blumstein and Wallman, 2000), and all may have a limited claim to having some effect. It is obvious, however, that there is a natural "regression to the mean" in that when the times become more dangerous more (and more varied) efforts are put into play to bring the crime rate down.

One of our concerns is related to this issue. That is, to what extent is the finding affected by the particular epoch during which the analysis was conducted? Specifically, we have analyzed state-level data, using categories based on Vernick and Hepburn's (2003) analysis of gun laws, and found that there is indeed an association between the imposition of RTC laws and a slight reduction in homicide. However, this effect ends in 1994 and even this slight trend reverses itself from 1994–2001. It should also be noted that neither of two of the states that recorded the greatest declines in homicides, New York and Texas, changed their handgun laws during the period included in the analyses.

There is also the issue of stochastic variation. Ignoring this effect is all too common in social science studies, and is ignored equally in studies that show deterrence and negate deterrence. We have yet to see an analysis on *either side* of the gun control controversy that specifically addresses (and takes into account) stochastic variation. Until such analyses account for this effect, we are reluctant to place much credibility in these studies, let alone consider them as the basis for policy.

Even if the net effect of shall-issue laws appears to be (we say "appears to be" because at best there is an association, not causation) a reduction (or an increase) in homicide, the goal of firearms research should be to look beyond any net effect and determine if there is a difference between jurisdictions that appear to benefit from the laws and those that appear to be harmed by the laws. One size does not fit all, either in the shoe store or in the legislature.

ACKNOWLEDGMENTS

This research was supported in part by Grant No. 2001-IJ-CX-0006 from the National Institute of Justice. We thank Tom Marvell and Jon Vernick for furnishing us with information relating to states' gun laws. The opinions expressed herein are our own and do not reflect the opinions or policies of the U.S. Department of Justice.

NOTE ADDED IN PROOF

Those interested in the general issue of gun control should read Jacobs' book, *Can Gun Control Work?* and those interested in additional methodological problems with *More Guns, Less Crime* should read the forthcoming article by Ayres and Donohue, *Shooting Down the More Guns, Less Crime Hypothesis*.

REFERENCES

Ayres, I. and Donohue, J. (2003). *Shooting Down the More Guns, Less Crime Hypothesis*, Stanford Law Review, forthcoming.

Barnett, A. (1981a). Further standards of accountability for deterrence research. In Fox, J. A. (ed.), *Methods in Quantitative Criminology*, Academic Press, New York.

Barnett, A. (1981b). The deterrent effect of capital punishment: a test of some recent studies. *Operations Research* 29: 346–370.

Black, D., and Nagin, D. (1998). Do "Right-to-Carry" laws deter violent crime? *J. Legal Studies* 27: 209–219.

Blumstein, A., and Wallman, J. (2000). *The Crime Drop in America*, Cambridge University Press, New York.

Dugan, L. (2002). Identifying unit-dependency and time specificity in longitudinal analysis: a graphical methodology. *J. Quant. Criminol.* 18: 213–237.

Jacobs, J. (2002). *Can Gun Control Work?* Oxford University Press, New York.

Little, R. J. A., and Rubin, D. B. (1987). *Statistical Analysis of Missing Data*, John Wiley & Sons, New York.

Lott, J. R., Jr. (1998). *More Guns, Less Crime*, University of Chicago Press, Chicago.

Lott, J. R., Jr. (2000). *More Guns, Less Crime* (2nd Ed.). University of Chicago Press, Chicago.

Lott, J. R., Jr., and Whitley, J. (2003). Measurement error in county-level data. *J. Quant. Criminol.* 19**:** 185–198.

Lundsgaarde, H. P. (1977). *Murder in Space City: A Cultural Analysis of Houston Homicide Patterns*, Oxford University Press, New York.

Maltz, M. D. (1994). Operations research in studying crime and justice: its history and accomplishments. In Pollock, S. M., Barnett, A., and Rothkopf, M. H. (eds.), *Operations Research and the Public Sector* (Vol. 6, Chap. 7), Amsterdam, Elsevier North-Holland, pp. 201–262.

Maltz, M. D. (1996). From Poisson to the present: applying operations research to problems of crime and justice. *J. Quant. Criminol.* 12: 3–61.

Maltz, M. D. (1998). Which homicides decreased? Why? *J. Crim. Law Criminol.* 88: 1479–1486.

Maltz, M. D. (1999). *Bridging Gaps in Police Crime Data*. Report No. NCJ-1176365, Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice, Washington, DC, September, 1999. http://www.ojp.usdoj.gov/bjs/pub/pdf/bgpcd.pdf.

Maltz, M. D., and Targonski, J. (2002). A note on the use of county-level crime data. *J. Quant. Criminol.* 18: 297–318.

Marvell, T. (2000). Dates of SI laws, cont. E-mail sent on 10 December, 2000.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*, John Wiley & Sons, New York.

Vernick, J. S., and Hepburn, L. M. (2003). State and federal gun laws: trends for 1970–99. In Ludwig, J. and Cook, P. J. (eds.), *Evaluating Gun Policy*, Brookings Institution Press, Washington, DC, pp. 345–402.