Social Science & Medicine 75 (2012) 23–31

Contents lists available at SciVerse ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Commentary

The disappearing health effects of rapid privatisation: A case of statistical obscurantism?

David Stuckler [a,b,*], Lawrence King [a], Martin McKee [b]

[a] University of Cambridge, Department of Sociology, London WC1E 7HT, United Kingdom
[b] London School of Hygiene & Tropical Medicine, Department of Public Health & Policy, United Kingdom

ARTICLE INFO

Article history:
Available online 21 March 2012

Keywords:
Statistics
Privatisation
Mortality
Post-communism

Introduction

Gerry (2012) represents a further challenge to our original paper on mass privatization and the Post-Communist Mortality Crisis (Stuckler, King, & McKee, 2009). These challenges do not seek to explain the phenomenon in question, for example Gerry, Mickiewicz, and Nikoloski (2009) state that their goal " is not to establish per se what does cause mortality.. . but to demonstrate that there is no evidence in the data used by Stuckler et al that mass privatisation resulted in increased mortality"(p. 2). In this paper we refute each of the arguments presented by Gerry (2012) in turn. First, we review the background to the post-communist mortality crisis, contrasting our hypothesis with the biological implausibility of the arguments presented by Gerry; second, we revisit the methodological issues we addressed in our data analysis, showing how the methods used by Gerry distort the analysis and fail to adhere to statistical conventions; third, we show how once the models used by Gerry are estimated appropriately, none of our results is changed.

Key facts about the post-communist mortality crisis

The former Soviet countries experienced a devastating mortality crisis in the early 1990s. The following key facts are relevant to the arguments in this commentary.

1. *The increases in mortality rates during the 1990s were a radical break with past trends*

Cornia and Paniccia (2000) notes how, "contrary to widespread opinion, the mortality changes of the 1990s are not a continuation of past trends" (p. 4, 2000). Rosefielde noted that "Adult excess deaths...are not statistical phantoms, and constitute radical departures from both past Russian patterns and Western norms" (Rosefielde, 2001, p 1162). Further Cornia and Paniccia (2000) demonstrate that levels of mortality in the 1990s were not predictive of subsequent rises and were a radical break from past trends, a view substantiated our replication of their unit-root tests (Stuckler et al., 2009). This was not, as some have argued, simply a delayed effect of Gorbachev's anti-alcohol campaign, with those spared at the time dying subsequently, most obviously because the excess deaths in the early 1990s would be among people around 5 years older than those spared, which was not the case (Leon et al., 1997). We have further shown statistically that regions in which mortality fell more greatly were not simply those which had the greater transition era increases (Stuckler, 2009; Stuckler, King, & McKee, 2010).

2. *These increases were concentrated among those most exposed to labour market conditions*

The bulk of deaths in the post-Soviet mortality crisis occurred at working-age (WHO, 2011). Fig. 1 shows how mortality rates rose by over 90% among those aged 25–39, 70% in ages 40–59, 45% in ages 15–24, 30% in people over 60, and dropped or remained steady among those under 15. Those in lower educational and

* Corresponding author. University of Cambridge, Department of Sociology, London WC1E 7HT, United Kingdom. Tel.: +447726422212.
E-mail address: ds450@cam.ac.uk (D. Stuckler).

0277-9536/$ – see front matter
doi:10.1016/j.socscimed.2011.12.054

Fig. 1. Trends in mortality among Russian men, by age group, 1989–1996.

occupational groups, who were least able to adapt to changing labour markets, were worst affected (Bessudnov et al., in press; Chenet, McKee, Leon, Shkolnikov, & Vassin, 1998; Shkolnikov et al., 2006; Shkolnikov, Leon, Adamets, Andreev, & Deev, 1998).

3. *Despite a shared history, countries followed very different trajectories*

Despite similarities in their historical, cultural, and social legacies, mortality in post-communist countries became increasing diverse over the course of the transition. Adult age-standardised mortality rates were nearly identical in Russia and Poland in 1991, at slightly over 1400 deaths per 100,000 population (WHO, 2011). Three years later, Russian mortality rates had increased by 35% while in Poland they dropped by about 10%. Had Russia followed Poland's path, an estimated 1.6 million fewer lives would have been lost (D. Stuckler, 2009). Overall, outcomes ranged from a fall of 7.2% in the best-five performing countries to an increase of 24.4% in the worst-five performing countries.

4. *The main immediate causes of death were cardiovascular disease, poisoning, suicides and homicides*

We agree with Gerry (2012) that the increased dangerous use of alcohol (and surrogate alcohol) and the stress associated with socio-economic upheavals across the region were closely associated with these fluctuations (Leon et al., 2007; Walberg et al., 1989). Where we differ is that we wish to understand what underlies these factors. Crucially, the interval between changes in social and economic conditions, hazardous drinking, and death is short (Saburova, Keenan, Leon, & Elbourne, 2011), providing a biological explanation for the observed break with earlier conditions.

5. *The post-communist transition had social, political, and economic dimensions*

The age and socio-demographic profile of mortality makes it difficult to avoid the view that it may, in some way, be connected to labour market changes. This is consistent with an extensive body of evidence that the pace of economic reform contributed to rises in mortality (Cornia & Paniccia, 2000; Walberg, McKee, Shkolnikov, Chenet, & Leon, 1998), with a number of scholars specifically invoking rapid privatization using quantiative (Brainerd, 1998) and qualitative (Hertz, 2005) methods as well as research using related measures of economic liberalisation (Brainerd, 2001). So as Brainerd and Cutler (2005) argue "best guess is that [the mortality crisis] is related to increased fear of very low income brought about by economic dislocation and the absence of a social safety net, though the exact factors behind this effect still require some work" (p. 128, E. Brainerd, Cutler, DM., 2005).

Gerry (2012) neglects the extensive body of literature, to argue that the mortality changes in this period were "dynamic", meaning that present changes were a function of past trends, a view shared by others, who suggest that we "erroneously attribute the post-1991 rise in mortality to privatization rather than to the continuation of the previous trend"(p. 1, Åslund & Sachs, submitted for publication).

This dynamic hypothesis is biologically and empirically implausible. It is inconsistent with the changes in causes of deaths driving the fluctuations we seek to explain. Its advocates propose no plausible mechanism whereby these causes of death, which exhibit very short lag periods between exposure and outcome, would affect mortality rates in a subsequent period. Gerry et al. (2009) invoked "disease stemming from some past exposure to pollution"(p. 379) as explaining the trends although the putative mechanism linking this to an episode of heavy and fatal drinking has not been presented.

Having shown the absence of a plausible alternative hypothesis, we turn to the methods used to challenge our findings.

Methodological challenges

Cross-national longitudinal analysis faces multiple challenges (Greene, 2002; Wooldridge, 2002); here we note two main econometric challenges:

1. Confounding – there were many contextual changes occurring at differing places and times in post-communist countries that may have correlated with both privatization and mortality. To address this endogeneity problem we introduced several time-varying control variables, based on the literature, including GDP, war, education rates, urbanization, dependency ratios, an index of democratization, and other economic policies including price and trade liberalization. To adjust for pre-existing differences between countries we used "fixed effects" models using 'within-country' variation and adjusting for country-specific intercepts (which additionally controls for falls in mortality between 1985–1987 during Gorbachev's antialcohol campaign). This also enabled us to evaluate before-and after-effects of rapid mass privatization. This approach is appropriate where a long-run average or trend is interrupted by shocks and is recommended for assessing effects of policy interventions (Jones, 2000). We then added further controls for social, economic, and political factors that, crucially, were theory-derived, and we employed a range of diagnostics and robustness checks, including time-dummy variables and linear time-trend controls (both within the 1990s and longer-term, estimated from the 1970s to 2000s), instrumental variable approaches, and Granger-causality tests (King & Stuckler, 2006; Stuckler et al., 2009, 2010)
2. Serial correlation – if the present level of mortality correlates with itself over time it would violate the uncorrelated error terms assumption of Ordinary Least Squares (OLS). This does not bias the coefficient estimates, just the standard errors. To ensure that statistical significance tests are unbiased, the recommended approach is to estimate clustered standard errors that are robust in the presence of any serial correlation (Bertrand, Duflo, & Mullainathan, 2004; Wooldridge, 2002).

Gerry (2012) argues that serial correlation reflects the dynamic nature of the data and should be directly modelled. Variables that should be modelled dynamically are those where the outcome has short-term "memory", visible in recent lags. There are often plausible theoretical mechanisms making dynamic models sensible. For example, the election of a party in one year may affect their chances

of election in a subsequent year. In models of decision-making, especially Bayesian updating frameworks, all information relevant to a present decision may be available from past choices and any updates to that information that occurred in the contemporary period. However, no such explanation is plausible here, except possibly mimetic suicides, although this phenomenon is mainly seen in very small communities where many of those involved know each other personally (Gladwell, 2000).

Beyond the absence of a theoretical justification, there is a practical issue. If serial correlation is inaccurately estimated, it can substantially bias the coefficient estimates. Methodological approaches vary, but usually focus on either estimating the serial correlation in the residuals or using a lagged dependent variable to capture the relationship between the present and past year. Both assume these dynamic relationships are identical over time and across countries and among explanatory variables, and transform the data to remove this correlation. However, as Plumper, Troeger, and Manow (2005) notes, "this assumption is not very convincing and almost certainly wrong" (p. 335). When estimating a common serial correlation factor, using either a lagged dependent variable or a serial correlation transformation, autocorrelation corrections can have a severe downward bias (Achen, 2001; Plumper et al., 2005), and involve "dire consequences for the reliability of inference based on such models" (p. 3, McGuirk & Spanos, 2002). Statisticians offer "a simple message to autocorrelation correctors: Don't" (p. 1, Mizon, 1995).

Removal of serial correlation from the mortality data fundamentally changes the remaining components to be explained and thus the implied hypothesis and appropriate coding methods (for a graphical depiction, see Stuckler et al., 2009). In the present example this becomes highly problematic because the task is to explain why some countries had greater short-term rises in mortality, deviating from long-term trends, than other countries in a short period of time. Such an adjustment removes a significant component of these short-term fluctuations we sought to explain; as has been documented, such transformations involve a downward bias in the estimates on the coefficients of interest (Achen, 2001; Keele & Kelly, 2005). As Achen describes it: "In many time series applications in the social sciences, lagged dependent variables have no obvious causal interpretation, and researchers omit them. When they are left out, the other coefficients take on sensible values. However, when an autoregressive term is put in "as a control", it often acquires a large, statistically significant coefficient and improves the fit dramatically, while many or all of the remaining substantive coefficients collapse to implausibly small and insignificant values. Occasionally, the substantive variables even take on the wrong sign" (p. 0, 2001). The paper helpfully provides several examples of how introducing nonsensical lagged dependent variables render obvious findings statistically insignificant. Here, imposing a dynamic structure is thus a case of fitting the wrong model to the data.

Dynamic Panel estimation

Suppose that there was a plausible mechanism linking last year's mortality to the present outcomes (a weakly dynamic process). How should one estimate such a relationship? First, we can eliminate potential bias by introducing controls for country-specific fixed effects:

$$y_{i,t} = \alpha + \beta x_{i,t} + \mu_i + \varepsilon_{i,t} \quad (1)$$

Here i is country and t is year. y is mortality rate; x is a set of covariates; μ is the country fixed effect; and ε is the error term.

Correcting for fixed effects is equivalent to subtracting the mean level from each observation:

$$y_{i,t} - \overline{y}_i = \alpha + \beta x_{i,t} - \overline{x}_i + \varepsilon_{i,t} - \overline{\varepsilon}_i \quad (2)$$

Then, to make the model dynamic, we add a lagged dependent variable:

$$y_{i,t} - \overline{y}_i = \alpha + \varphi y_{i,t-1} - \overline{y}_{i,t-1} + \beta x_{i,t} - \overline{x}_i + \varepsilon_{i,t} - \overline{\varepsilon}_i \quad (3)$$

But as noted by Nickell (1981), OLS estimation of equation (3) is biased because the mean error term, $\overline{\varepsilon}_i$, correlates with the lagged dependent variable, y_{it-1}. This violates the assumption of OLS that the error is uncorrelated with the regressors. In panels with a short period, the bias is particularly high (up to 50% when there are only 2 time periods).

A large methodological literature has developed to cope with this 'Nickell-bias'. These strategies have focused on generalized-method-of-moments, rather than OLS, because of its greater flexibility. Anderson and Hsiao (1982) made an early proposal for datasets with a large number of units but few observations. The first step is to first difference the equation so to remove the fixed effects:

$$\Delta y_{i,t} = \alpha + \varphi \Delta y_{i,t-\mathbf{1}} + \beta \Delta x_{i,t} + \Delta \varepsilon_{i,t} \quad (4)$$

This transformation alters the theory, as it looks at year-to-year changes whereas fixed effects modelling looks at the levels over the time frame of the dataset. Firstly, differencing our main variable of interest, a dummy for the implementation of mass privatization, changes the coding from '0-before and 1-after' to a 1 *only* in the first year of implementation. This restriction modifies our hypothesis to specify: 'mass privatization had a negative effect on working-age mortality only in the first year it was implemented'. Secondly, the Δy_{it-1} is itself correlated with Δx_{it-1}, so that only contemporaneous effects of the changes in the covariates are observed and the Δy_{it-1} absorbs all longer-term effects of controls such as GDP, democracy, and war as well as the hypothesised relationship between privatization and mortality.

These two transformations would make sense if the impacts of privatization on mortality were immediate and disappeared within a year. Indeed, as we noted previously (Stuckler et al., 2010), privatization could have very rapid effects, even during anticipation of the economic shock (Perlman & Bobak, 2009). However, most countries actually implemented rapid mass privatization of a substantial share (>25%) of their state-owned assets over a two-year period (Lieberman & Kopf, 2007). The effects are likely to be felt for several years. In the case where privatization was implemented but never undone, the differenced model identifies a change in only the first year. Rather than interpreting the return of mortality rates to long-run trends as the effect wearing off, the first difference model interprets this as a random shock that is as large as the initial shock (a pattern which can manifest as serial correlation and be removed from the data by adjustments). Thus, it enters the model as noise. But this obscures the actual pattern: the recovery is anything but random, as it includes adaptation to the initial shock, survival bias, policy responses, etc. Such patterns can be seen visually in the figures presented in the following section. In other words, the first differencing of the covariates of interest, while seemingly innocuous, constrains time patterns from the data to the contemporary period and no longer appropriately tests our hypothesis about the effects of privatization, or any of the other covariates on mortality (However, this issue does not arise when evaluating the second measure of rapid privatization we used, privatization indices from the European Bank for Reconstruction and Development. But, without justification, the authors of the accompanying paper did not investigate this measure.)

Although these transformations are therefore flawed, let us continue with the derivation of methods in Gerry (2012). In the

first-differenced framework of equation (4), Δy_{it-1} remains correlated with $\Delta \varepsilon_{it}$. To correct for this bias, Anderson and Hsiao use the lagged second year difference of the dependent variable, Δy_{it-2}, to predict the first year difference, Δy_{it-1}. This removes the bias because the second lagged difference is not correlated with the error but is correlated with the first year difference (a key property of instrumental variables). However, the ability to predict the first year difference from the second year difference may be weak so the predicted value will have very little variation, which can give rise to biased estimates of the effect of past on the present. Arellano and Bond (1991) use a greater number of predictors, drawing on additional lags of differenced variables (so-called 'difference GMM'). Blundell and Bond (1998) observed that these difference estimators can also be biased in datasets with short time dimensions, especially when lag first differences poorly predict contemporaneous levels (weak instruments). To strengthen the predictive power of the instruments, they additionally instrument the levels equation with first differences, building a system of two equations: the original equation in levels and the transformed in differences (so-called 'system GMM').

As with any statistical model, difference and system GMM contain important assumptions. Both assume the correct model is dynamic. Difference GMM assumes that the lagged differences can predict contemporaneous levels. Higher lagged differences tend to predict higher levels (since a past increase will increase the current level), although these lags may be weak predictors. System GMM contains a further crucial assumption that the changes are uncorrelated with the country mean levels (the fixed effects). This assumption only holds if the fixed effect and dynamic process offset each other across the whole panel. It requires that "throughout the study period, individuals [here countries] sampled are not too far from steady-states, in the sense that deviations from long-run means are not systematically related to fixed effects."(p. 44, Roodman, 2006) As Roodman notes, "One leading estimator, Difference GMM, often suffers from weak instrumentation. The favoured alternative, System GMM, works only under arguably special circumstances" (p. 156, 2009). In this study period, countries are not converging to long-run means at an equal rate, especially in early years of transition, making the lagged differences in mortality rates correlate with the errors. Additionally, the use of internal instrumental models require large sample sizes to yield unbiased results: "a weakness of IV and GMM estimators is that their properties hold when N is large, so they can be severely biased and imprecise in panel data with a small number of cross-sectional units. This is often the case in most macro panels" (p. 2, Bruno, 2005) (To cope with this bias, Bruno recommends applying a bias correction based on Monte Carlo simulations and using a bootstrap covariance matrix for calculating the standard errors. We replicated our findings using these methods, finding results were unchanged.). Thus, as described above, at least two crucial assumptions of the dynamic estimators are violated.

Lastly it is important to note that executing system GMM involves many statistical decisions. For this reason, as described in 'How to do xtabond2' – the module for system GMM used by Gerry (2012), it is essential to report all specification choices. As the manual states, "Using these estimators involves many choices, and researchers should report the ones they make – difference or System GMM; first differences or orthogonal deviations; one- or two-step estimation; non-robust, cluster-robust, or Windmeijer-corrected cluster-robust errors; and the choice of instrumenting variables and lags used" (p. 44, Roodman, 2006; see also 2009). Although Gerry report that they used system GMM, they fail to note any of the following choices in the main text, leaving us to identify them from their statistical routines: their use of orthogonal deviations; one-step estimation; cluster-robust errors; t-statistics instead of z-statistics; lags of four years depth; the number of instruments used; and that they only instrumented the level equation. As we will show below, notwithstanding their distortion of our hypothesis, this constellation of choices attenuates our findings to non-significance by estimating a model with a very high coefficient on the lagged dependent variable.

To recap, the methodological adjustments introduced in the accompanying paper distort our original analysis. The data manipulations reduce statistical power, introduce downward bias of coefficients, violate main assumptions of the models they choose, and hence fail to replicate a test of our main hypothesis.

Results: how to make a mortality crisis disappear in five steps

We are grateful that Gerry (2012) provided his statistical code so that we are able to replicate his analysis.

Case study: Russia and Belarus

Before turning to their results, we can take a simplified example from our dataset, Russia and Belarus. These two countries had parallel mortality trajectories since the 1960s. Russia announced plans for rapid mass privatization in 1991 and began implementation in 1992; Belarus did not implement a rapid mass privatization programme. At this time the trends in mortality rates begin to diverge. As shown in Fig. 2, Russia's rates recover to the trajectory they were on prior to the early 1990s reform period, but then, in 1998 coinciding with the Russian rouble crisis, once again began to diverge from Belarus. Notably, these comparisons were based on comparative methods of matching similar, neighboring, and core post-Soviet countries which differed mainly on the implementation of mass privatization. One alternative possibility was Georgia, although health system privatization led to fees for registering deaths, known to lead to data anomalies. Another non-mass privatizing countries was Estonia, which was exceptional in that it did not mass privatize to achieve a rapid privatization but experienced similar risks by rapidly privatizing directly to foreign owners (see Stuckler et al., 2010 for further details about case selection).

Fig. 2 show the implications of transforming the original adult male working-age age-standardised mortality rate data, covering the period 1987 to 2005 in Russia and Belarus. Note that Panels B (lagged dependent variable) and C (first difference) are almost identical, as B is based on an estimated effect of the previous year's mortality on the current year of $\gamma = 0.92$.

Recall that our hypothesis is that the implementation of rapid mass privatization causes a short-run increase in mortality among working-age men, largely through stress- and alcohol-related pathways, but also through fiscal shocks to the state (undermining social welfare).

Suppose that mass privatization did have a causal effect in the first few years of implementation, depicted in the gap that emerged between Russia and Belarus which endured through about 1997. Would this difference be identifiable with the transformations of the data? First, we regressed the original data in Panel A on the dummy for mass privatization and used robust-clustered standard errors by country, finding mass privatization is associated with an increase of 330.0 ($p < 0.001$). Visually, this is approximately the magnitude of the actual increase. Then, adjusting for the lag of mortality rates, shown in Panel B, yields a much lower estimate of 86.7, $p < 0.001$. When the data are first-differenced, including a transformation of the binary mass privatization policy variable (so that the code now *only* includes a 1 for the first year of implementation), the association in Panel C drops to 95.5 ($p = 0.036$) and in Panel D the coefficient drops further to 90.9, $p = 0.045$.

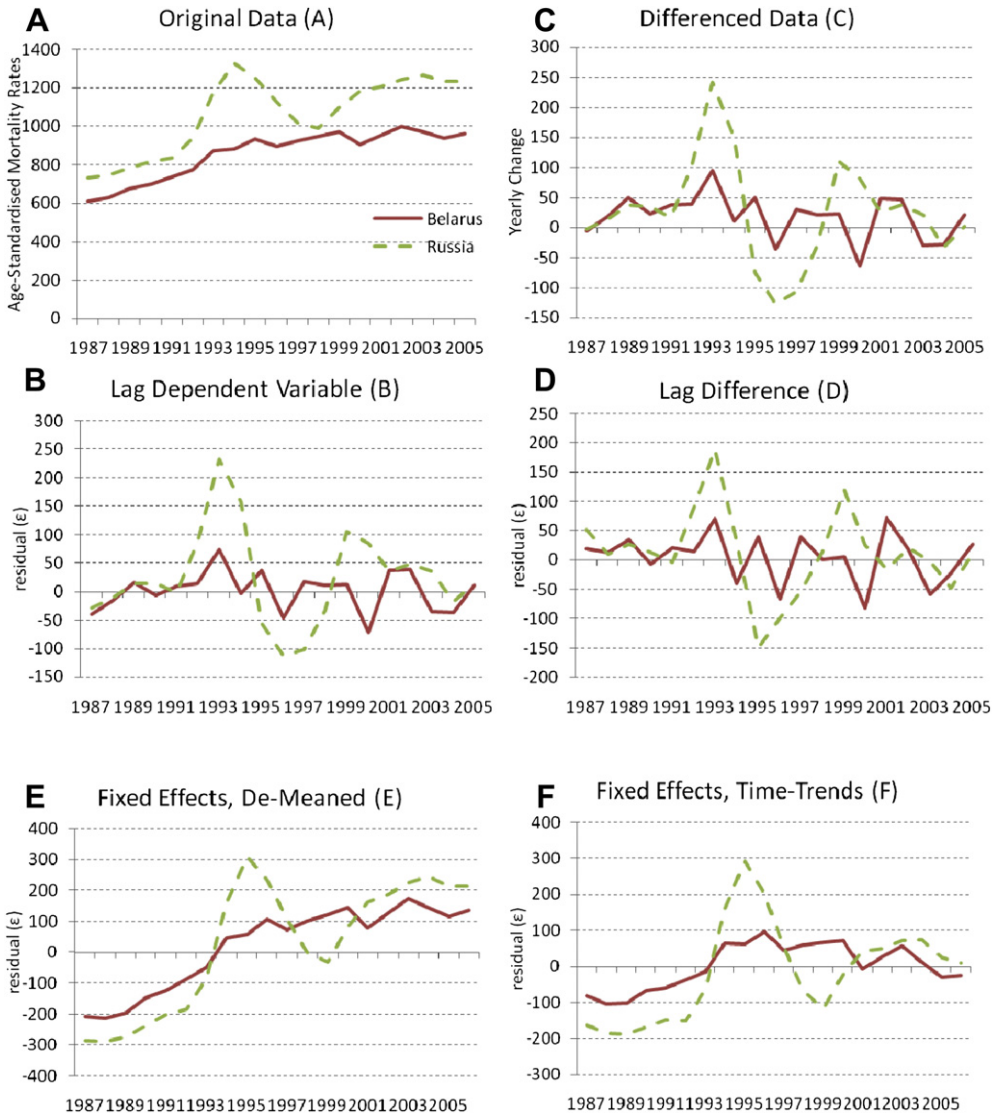

Fig. 2. Alternative transformations to age-standardised working-age male mortality rate data in Russia and Belarus, 1987–2005.

As discussed above, these estimates containing the lagged dependent variable are known to be biased (Nickell, 1981). Thus, we can introduce alternative instrumental variable approaches: the Difference GMM estimator gives 198.5, $p = 0.009$; the System GMM estimator yields 93.2, $p = 0.043$. By way of comparison, applying the estimator using the arbitrary choices of estimation options applied in Gerry (2012) – orthogonal deviations; t-statistics instead of z-statistics; instruments lagged from the second year; collapsed instruments; and instrumenting only for the levels equation of system GMM – gives the most conservative estimate of 66.3, $p = 0.313$.

This attenuation largely arises from using a model with a very high estimated effect of past mortality on the present (the estimated coefficient on the lagged dependent variable, γ is 0.96, so that prior mortality change almost perfectly predicts the present evolution of mortality). Fig. 3 depicts the contrast between the remaining data to explain where the coefficient, γ, of the lagged dependent variable is 1 (essentially first differences) and 0 (levels). As can be seen in the figure, estimating a series of regression models with a constant term while forcing different effect sizes on the lagged dependent variable, the mortality crisis progressively diminishes at higher levels of the lagged dependent variable. While the peak in mortality occurred in 1994, for $\gamma \geq 0.70$ the highest point occurs in 1993 (note that the differences trend cross zero when the levels trend changes direction, a pattern of a first-derivative). In the figure we also report the estimated effect of

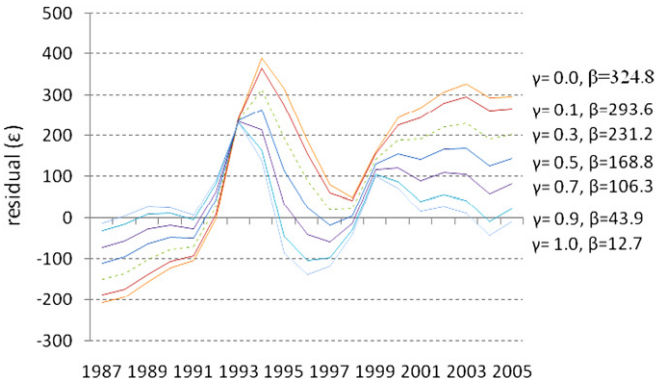

Fig. 3. Russian data transformations using alternative lagged dependent variable coefficients (γ).

mass privatization on mortality rates for each estimated effect size of the lagged dependent variable. As shown, there is clearly an inverse correlation of the estimated effect sizes of the lagged dependent variable and of rapid mass privatization. At the 0.96 effect size level estimated in the accompanying paper, there is no statistically significant effect in this hypothetical example, even though there should be; effects only turn significant when the mortality peak occurs in 1994 (i.e., when the coefficient is <0.70).

Thus, the data manipulations have largely eliminated the variation that could potentially be explained with our measure of mass privatization. The results presented in the accompanying paper have nothing to do with past mortality causing present trends but are an artefact of modelling assumptions.

To further illustrate this point, we note that even if we introduced a dummy variable, simply designed to capture the overall difference that emerged between Russia and Belarus between 1991 and 1994, it too would fail to register a statistically significant difference (167.4, $p = 0.107$) when using the method in Gerry (2012). A similar issue arises in the papers' covariates (Box 1).

Thus, our preferred models followed standard econometric practices to address serial correlation by clustering standard errors by country, rather than remove it from the data (Wooldridge, 2002) (although at behest of a reviewer we did include specifications with direct estimation of serial correlation in Stuckler et al., 2009, reporting our preferred results clustering standard errors). Our hypothesis was tested by comparing country-specific outcomes before and after the implementation of mass privatization. Here the unadjusted fixed effects models estimate gives rise to 316.1, $p < 0.001$ (Fig. 2, Panel E) and including a time-trend for each country from the 1980–2005 gives rise to a similar estimate of 325.8, $p < 0.001$ (Fig. 2, Panel F). Nonetheless, we presented models with and without adjustments for serial correlation, showing that the results were not qualitatively changed (see the web appendix to (Stuckler et al., 2010)).

Box 1. Alternative estimation of the effect of war on mortality in Tajikistan and Uzbekistan.

An example of the inappropriate adjustment for serial autocorrelation can be seen with the devastating ethnic conflict that began in Tajikistan in 1992, resulting a fall of male life expectancy by 11 years. This contrasts with Uzbekistan, which had otherwise followed similar trends. Our critics, after finding no statistically significant effect of war in their models, implausibly suggest conflicts may not have significantly affected for health in this region when widely accepted estimates are of between 36,000 and 150,000 excess deaths in Tajikistan (Kolsto, 2000; Kutz & Turpin, 1999). Continuing with the example above, the model in the accompanying paper would fail to identify any effect of a dummy variable for civil war (either constrained to 1992 or the period of the war between 1992 and 1997) in Tajikistan on health ($p = 0.473$).

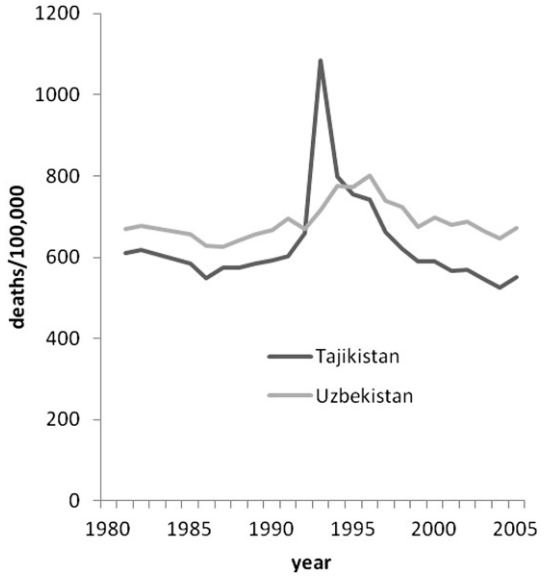

Fig. Age-standardised death rate/100,000 Tajikistan and Uzbekistan

Data torture: Replicating system GMM results and coding choices

Turning finally to the latest results presented in Gerry (2012), we first briefly summarise our original results (Stuckler et al., 2010). We showed that the association of both the indicator of mass privatization implementation and higher levels of the EBRD average privatization index were associated with higher male working-age mortality rates only in the former Soviet countries. In the paper we modelled log mortality rates to look at percentage changes, although as we noted this calculation did not affect the patterns we observed. To be clear, based on our preceding discussion we believe the dynamic model is a case of fitting the wrong model to the data and is inappropriate for testing our hypothesis; in this section, we seek not to identify a plausible dynamic specification but only to replicate Gerry's findings and show their sensitivity to several unspecified assumptions and best-practice guidelines.

In Table 1, we report estimates from a dynamic panel data model with a lagged dependent variable in which the equation has been first-differenced and report cluster-robust standard errors (Difference GMM). All the covariates were included as instruments for the differences equation. We begin with a lag depth of 2, the minimum distance needed to purge Nickell-bias, while avoiding issues of instrument proliferation. As shown, mass privatization is associated with a 13.2% increase in mortality rates ($p = 0.020$). This estimate is similar in magnitude to the original reported in the Lancet of 13.5%, $p < 0.001$. The model also passes important diagnostics: the Hansen test of overidentification is non-significant ($p = 0.645$), as is the Arellano-Bond test of second-order serial correlation ($p = 0.258$). The model's controls, such as war and income, take on plausible associations, significantly negative and positive, respectively, as in our original models. In the second column we report estimates using System GMM, using the standard method of instrumenting both the differences and levels equation. Similar patterns of covariates were observed. The resulting estimate is 9.6%, ($p = 0.050$). Again, the model passes the diagnostic tests. Taken together, this evidence directly refutes the assertions of Gerry (2012) that "there is no plausible dynamic specification in which mass privatization is significant."

Next we sequentially introduce the manipulations used by Gerry (2012). First, it reports the entire sample and does not constrain the lag depth; the estimate halves in magnitude, $p = 0.069$. Next, using t-statistics instead of the usual z-statistics, reduces the p-value to 0.094. Although Gerry suggests they instrumented both levels and differences, the model was actually constrained to use only instruments for the levels equation. Here the coefficient falls to 0.03, $p = 0.189$. A further transformation is introduced of forward orthogonal deviations, which instead of the usual first differences subtracting the past from the present value, subtracts the average of all future values from the present value. The p-value drops to 0.197. As a final step, the instrument count is reduced by collapsing the instruments, so that one set of lags

Table 1
Sequential replication of manipulations in the accompanying paper.

Covariate	(1) Former Soviet countries, difference GMM	(2) Former Soviet countries, system GMM	(3) All countries, instrumenting with lag depth 2 or greater	(4) All countries, *p*-value based on *t*-statistics; instrumenting with lag depth 2 or greater	(5) All countries, instruments for levels equation only; *p*-value based on *t*-statistics; instrumenting with lag depth 2 or greater	(6) All countries, orthogonal deviations; instruments for levels equation only; *p*-value based on *t*-statistics; instrumenting with lag depth 2 or greater	(7) All countries, collapsed instruments; orthogonal deviations; instruments for levels equation only; *p*-value based on *t*-statistics; instrumenting with lag depth 2 or greater
Mass privatization	0.132	0.096	0.048	0.048	0.033	0.032	0.024
	(0.056)	(0.049)	(0.026)	(0.028)	(0.024)	(0.024)	(0.033)
	$p = 0.020$	$p = 0.050$	$p = 0.069$	$p = 0.097$	$p = 0.189$	$p = 0.197$	$p = 0.473$
Lagged mortality rate	0.275	0.785	0.922	0.922	0.924	0.931	0.958
	(0.189)	(0.134)	(0.053)	(0.055)	(0.053)	(0.050)	(0.100)
	$p = 0.147$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
Number of instruments	21	33	100	100	100	100	23
Arellano-Bond test for AR(1) in first differences	$z = -1.46$ $p = 0.144$	$z = -1.33$ $p = 0.184$	$z = -1.27$ $p = 0.205$	$z = -1.27$ $p = 0.205$	$z = -1.27$ $p = 0.203$	$z = -1.27$ $p = 0.204$	$z = -1.34$ $p = 0.179$
Arellano-Bond test for AR(2) in first differences	$z = -1.13$ $p = 0.258$	$z = -0.37$ $p = 0.712$	$z = -0.57$ $p = 0.566$	$z = -0.57$ $p = 0.566$	$z = -0.618$ $p = 0.537$	$z = -0.610$ $p = 0.542$	$z = -0.603$ $p = 0.546$
Sargan test of overidentification	$\chi^2(11) = 36.99$ $p < 0.001$	$\chi^2(22) = 82.26$ $p < 0.001$	$\chi^2(89) = 133.7$ $p = 0.002$	$\chi^2(89) = 133.7$ $p = 0.002$	$\chi^2(89) = 138.2$ $p < 0.001$	$\chi^2(89) = 140.8$ $p < 0.001$	$\chi^2(12) = 30.82$ $p = 0.002$
Hansen test of overidentification	$\chi^2(11) = 8.75$ $p = 0.645$	$\chi^2(22) = 8.30$ $p = 0.996$	$\chi^2(89) = 20.90$ $p = 1.00$	$\chi^2(89) = 20.90$ $p = 1.00$	$\chi^2(89) = 18.16$ $p = 1.00$	$\chi^2(89) = 19.75$ $p = 1.00$	$\chi^2(12) = 18.71$ $p = 0.096$
Difference-in-Hansen tests of exogeneity of instrument subsets							
GMM instruments for levels							
Hansen test excluding groups	n/a	$\chi^2(11) = 8.29$ $p = 0.687$	$\chi^2(77) = 20.90$ $p = 1.00$	$\chi^2(77) = 20.90$ $p = 1.00$	$\chi^2(77) = 18.16$ $p = 1.00$	$\chi^2(77) = 19.75$ $p = 1.00$	$\chi^2(11) = 16.54$ $p = 0.122$
Difference (null H = exogenous)	n/a	$\chi^2(11) = 0.00$ $p = 1.000$	$\chi^2(12) = 0.00$ $p = 1.00$	$\chi^2(12) = 0.00$ $p = 1.00$	$\chi^2(12) = 0.00$ $p = 1.00$	$\chi^2(12) = 0.00$ $p = 1.00$	$\chi^2(1) = 2.17$ $p = 0.141$
Instrumental variables							
Hansen test excluding groups	$\chi^2(2) = 1.52$ $p = 0.468$	$\chi^2(13) = 2.49$ $p = 0.999$	$\chi^2(80) = 16.03$ $p = 1.00$	$\chi^2(80) = 16.03$ $p = 1.00$	$\chi^2(80) = 13.48$ $p = 1.00$	$\chi^2(80) = 18.75$ $p = 1.00$	$\chi^2(3) = 1.81$ $p = 0.612$
Difference (null H = exogenous)	$\chi^2(9) = 7.22$ $p = 0.614$	$\chi^2(9) = 5.81$ $p = 0.759$	$\chi^2(9) = 4.87$ $p = 0.846$	$\chi^2(9) = 4.87$ $p = 0.846$	$\chi^2(9) = 4.68$ $p = 0.861$	$\chi^2(9) = 1.00$ $p = 0.999$	$\chi^2(9) = 16.90$ $p = 0.050$

Notes: Former Soviet sample includes 15 countries and 161 country-years; Full country sample includes 24 countries and 263 country-years.

predicts each period's lagged dependent variable value. Thus the final column of Table 1 shows the Gerry (2012) model, which reported only the greatest estimate of the lagged dependent variable for the many possible model options and also the lowest statistical significance estimate of our hypothesis ($p = 0.473$). As shown, these also yield implausible results, such as no effect of war or GDP at conventional significance levels.

This lack of evidence strongly suggests an error in the models. The pattern observed here is precisely that in the preceding case study of Belarus and Russia. Each manipulation, increases the estimated effect of the lagged dependent variable, so that the hypothesis actually being tested increasingly fails to test the original hypothesis. Higher estimated effects of past mortality come to 'dominate the regression' explaining current mortality trends and

Table 2
Re-estimates of system GMM accounting for lagged effects and original models with time-trend controls, 26 countries 1989–2002.

Covariate	System GMM (first differences with lags)	Original model (fixed effects with time-trend controls)
Lag of Log of age-standardised working-age male mortality rates	0.849*** (4.98e-13)	–
Implementation of mass privatisation[a]	0.133** (0.00281)	0.139** (0.0442)
One year lag implementation of mass privatisation[a]	−0.0461 (0.155)	–
Two-year lag of implementation of mass privatisation[a]	−0.0521 (0.0812)	–
Log GDP per capita in current USD (2000)	0.0122 (0.510)	−0.134*** (0.0292)
EBRD index of price liberalization	−0.0340 (0.140)	0.0220 (0.0226)
EBRD index of foreign exchange and trade liberalisation	−0.0242* (0.0321)	−0.00464 (0.0157)
Freedom house democratisation index	−0.00205 (0.626)	−0.0126* (0.00586)
War	−0.0311 (0.279)	0.234*** (0.0347)
Population dependency ratio	0.000854 (0.705)	0.0126* (0.00506)
Tertiary education levels	0.00287 (0.157)	−0.00166 (0.00125)
Urbanization rate	0.000830 (0.670)	0.0327** (0.0100)
Linear time-trend	–	0.0033 (0.0086)
Period dummies	No	Yes

Notes: Robust standard errors in parentheses clustered by country to reflect non-independence of sampling.
*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.
[a] Mass privatization covariates in system GMM model jointly significant $\chi^2(3) = 14.08$, $p = 0.0028$.

'squash' the coefficients on mass privatization and all other covariates, generating implausible results.

This pattern of spurious findings is documented in the instruction manuals for system GMM, 'How to do xtabond' and a 'note on too many instruments'. It cautions: "As implemented in popular software packages, the estimators carry a great and underappreciated risk: the capacity *by default* to generate results that are invalid and appear valid." [italics in original]" (p. 156, Roodman, 2009; see also 2006)and lists recommended actions to avoid this risk. Unfortunately, Gerry fails to follow these recommendations, presenting only a selection of the recommended diagnostics (In some cases, diagnostic tests are misreported – although admittedly it is difficult to discern as they fail to report in the text their precise model. For example, they appear to misreport their Hansen test statistic, which was 0.096 not 0.13, a level, which according to the guidance manual on xtabond2, further indicates the invalidity of their model.).

As a final analysis, we revisit a dynamic model specification that is adjusted to test our hypothesis as well as our preferred fixed effect modelling framework with adjustments for long-term trends. The first adjustment is to relax the assumption of Gerry that constrains the effects of mass privatization to the current year to incorporate lagged effects of the first- and second year. As shown in Table 2, when we do so, our results are slightly larger overall than those of our main paper: mass privatization is associated with a 13.3% rise in the initial year, dampening to 8.7% and finally 3.5% by the third year. However, the other covariates like GDP remain statistically insignificant as any of their effects beyond the current period are captured by the lagged dependent variable; these covariates too would need adjustment to test appropriately hypothesis that an economic downturn increases risks of mortality. In contrast, our preferred fixed effects model incorporating a linear year time-trend and year dummy variables estimates a before/after association of mass privatization with mortality of 13.9%. War, GDP, and democratization have expected signs, an indication of validity.

Conclusion

King (1990) suggested that statistical methods should be evaluated by asking the question: "What did it do to the data?" (p. 11). We have documented how a series of biologically implausible manipulations was introduced to the data; how there was a failure to report the number of comparisons made; and ultimately how the mortality crisis that we sought to explain was made to disappear. These manipulations fulfil established criteria of 'data torture' (Mills, 1993).

Our analysis has implications for social science research and policy analysis. Researchers should not unthinkingly introduce an ever-more complicated set of statistical corrections for perceived statistical problems in the absence of a theoretical justification. The principles of Ockham's razor – with its stress on the importance of parsimony – apply to statistical inference as to any area of social science.

As Achen (2001) has shown, when researchers substitute *p*-values and good statistical fit for theory and common sense, they can talk themselves into accepting inappropriate regressions. A lack of coherence with existing knowledge, e.g., war often involves people being killed, should alert researchers that something is wrong.

It is possible to 'blow up' any true statistical finding by distorting fundamentally the hypothesis or fitting the wrong model. Here, differencing the data modifies the hypothesis to constrain the effect of mass privatization and all covariates to the contemporary period. Imposing a dynamic specification is a case of fitting the wrong model to the data; without theoretical justification it artificially inflates the effect of the past in driving the present and consequently biases the substantive coefficients towards negligible and nonsensical values.

Applying sophisticated modelling techniques does not necessarily enhance understanding. Researchers sometimes fall into the trap of trying to introduce a large body of control methods and alternative specifications to make sure their estimates are 'robust' to modelling choices. As Roodman (2009) notes "there is also a larger reminder here about the dangers of automated sophistication. It is all too easy to employ complicated estimators without fully appreciating their risks—indeed sometimes it takes years for their disadvantages to come to light" (p. 156). Researchers need to understand the consequences of correcting potential biases. Introducing a correction can create new sources of bias.

As we noted at the outset, our original paper has been the subject of sustained criticism. We welcome studies seeking to understand the underlying causes of the post-communist mortality crisis. However, some of our harshest critics include who argued for and supported radical restructuring in Eastern Europe (Sachs, 2009; The Economist, 2009; Wedel, 2001). Some of the efforts to discredit our work can best be understood as "statistical obscurantism": seeking to the muddy the waters, rather than to understand the ultimate causes of post-communist mortality in this region. This can be considered as one aspect of the wider phenomenon of "denialism" (Diethelm & Mckee, 2009). Data torture is notoriously difficult to document (Mills, 1993). Huff (1954) identifies a number of tactics followed in the 1940s and 1950s; today, the options are much more sophisticated. Examples include the 'confounder studies' by tobacco companies and manipulations of data by climate change sceptics. With growing sophistication of statistical programmes, a basic understanding of what has been done to the data and hypothesis may be obscured.

Acknowledgements

The authors are grateful to Sanjay Basu and Jacob Bor for constructive comments on various drafts of this manuscript.

References

Achen, C. (2001). *Why lagged dependent variables can suppress the explanatory power of other independent variables*. Working Paper. Department of Political Science, University of Michigan.

Anderson, T., & Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics, 18*, 47–82.

Arellano, M., & Bond, S. R. (1991). Some tests of specification of panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies, 58*, 277–297.

Åslund, A., Sachs, J. D. Mass privatization did not cause the post-communist mortality crisis. *The Lancet*, submitted for publication.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics, 119*, 249–275.

Bessudnov, A., McKee, M, Stuckler, D. Inequalities in male mortality by occupational class, perceived social status, and education in Russia, 1994–2006. *European Journal of Public Health*, in press.

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics, 87*, 11–143.

Brainerd, E. (1998). Market reform and mortality in transition economies. *World Development, 26*, 2013–2027.

Brainerd, E. (2001). Life and death in Eastern Europe: economic reform and mortality in the former Soviet Union: a study of the suicide epidemic in the 1990s. *European Economic Review, 45*, 1007–1019.

Brainerd, E., & Cutler, D. M. (2005). Autopsy on an empire: understanding mortality in Russia and the former Soviet Union. *Journal of Economic Perspectives, 19*, 107–130.

Bruno, G. S. F. (2005). *Estimation and inference in dynamic unbalanced panel data models with a small number of individuals*. Working Paper n. 165. Milan: Bocconi University.

Chenet, L., McKee, M., Leon, D., Shkolnikov, V., & Vassin, S. (1998). Alcohol and cardiovascular mortality in Moscow, new evidence of a causal association. *Journal of Epidemiology and Community Health, 52*, 772–774.

Cornia, G. A., & Paniccia, R. (2000). *The mortality crisis in transitional economies*. Oxford University Press.

Diethelm, P., & McKee, M. (2009). Denialism: what is it and how should scientists respond? *European Journal of Public Health, 19*, 2–4.
Gerry, C. (2012). The journals are full of great studies, but can we believe the statistics? Revisiting the mass privatisation – Mortality debate. *Social Science & Medicine, 75*, 14–22.
Gerry, C., Mickiewicz, T., & Nikoloski, Z. (2009). Did mass privatization really increase post-communist mortality? *The Lancet, 375*, 371.
Gladwell, M. (2000). *The tipping point*. New York: Little Brown.
Greene, W. (2002). *Econometric analysis*. Upper Saddle River, NJ: Prentice Hall.
Hertz, N. (2005). *BBC World service interview*. London.
Huff, D. (1954). *How to lie with statistics*. Meredith Mincey Readings.
Jones, A. (2000). Health econometrics. In A. Cuyler, & J. P. Newhouse (Eds.), *Handbook of health economics* (pp. 265–344). Amsterdam; New York; and Oxford: Elsevier Science.
Keele, L., & Kelly, N. J. (2005). Dynamic models for dynamic theories: the ins and outs of lagged dependent variables. *Political Analysis, 14*, 186–205.
King, G. (1990). On political methodology. *Political Analysis, 2*, 1–29.
King, L., & Stuckler, D. (2006). *Mass privatization and the post-communist mortality crisis*. Amherst: UMass Political Economy Research Institute.
Kolsto, P. (2000). *Political construction sites: nation-building in Russia and the post-Soviet states*. Westview Press.
Kutz, L., & Turpin, J. E. (1999). *Encyclopedia of violence, peace, & conflict*. Academic Press.
Leon, D., Chenet, L., Shkolnikov, V. M., Zahkharov, S., Shapiro, J., Rakhmanova, G., et al. (1997). Huge variation in Russian mortality rates 1984–94: artefact, alcohol, or what? *Lancet, 350*, 383–388.
Leon, D. A., Saburova, L., Tomkins, S., Andreev, E., Kiryanov, N., McKee, M., et al. (2007). Hazardous alcohol drinking and premature mortality in Russia: a population based case-control study. *Lancet, 369*, 2001–2009.
Lieberman, I., & Kopf, D. J. (2007). *Privatization in transition economies: The ongoing story*. Emerald Group Publishing Limited.
McGuirk, A., & Spanos, A. (2002). The linear regression model with autocorrelated errors: just say no to error autocorrelation. In *Paper presented at the American agricultural economics association*.
Mills, J. (1993). Data torturing. *New England Journal of Medicine, 329*, 1196–1199.
Mizon, G. (1995). A simple message for autocorrelation correctors: don't. *Journal of Econometrics, 69*.
Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica, 49*, 1417–1426.
Perlman, F., & Bobak, M. (2009). Assessing the contribution of unstable employment to mortality in posttransition Russia: prospective individual-level analyses from the Russian longitudinal monitoring survey. *American Journal of Public Health, 99*, 1818–1825.
Plumper, T., Troeger, V. E., & Manow, P. (2005). Panel data analysis in comparative politics: linking method to theory. *European Journal of Political Research, 44*, 327–354.
Roodman, D. (2006). *How to do xtabond: An introduction to "difference" and "system" GMM in STATA*. Washington D.C.: USA: Center for Global Development.
Roodman, D. (2009). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics, 71*(1), 135–158.
Rosefielde, S. (2001). Premature deaths: Russia's radical economic transition in Soviet perspective. *Europe–Asia Studies, 53*(8), 1159–1176.
Saburova, L., Keenan, K., Leon, D. A., & Elbourne, D. (2011). Alcohol and fatal life trajectories in Russia: understanding narrative accounts of premature male death in the family. *BMC Public Health, 11*, 481.
Sachs, J. (2009). *'Shock therapy' had no adverse effect on life expectancy in Eastern Europe*. London: Financial Times.
Shkolnikov, V., Andreev, E. M., Jasilionis, D., Leinsalu, M., Antonova, O. I., & McKee, M. (2006). The changing relation between education and life expectancy in central and Eastern Europe in the 1990s. *Journal of Epidemiology and Community Health, 60*, 875–881.
Shkolnikov, V., Leon, D., Adamets, S., Andreev, E., & Deev, A. (1998). Educational level and adult mortality in Russia: an analysis of routine data 1979 to 1994. *Social Science & Medicine, 47*, 357–369.
Stuckler, D. (2009). *Social causes of post-communist mortality: A macro-sociological analysis*. Cambridge, UK: University of Cambridge.
Stuckler, D., King, L., & McKee, M. (2009). Mass privatisation and the post-communist mortality crisis. *Lancet, 373*, 399–407.
Stuckler, D., King, L., & McKee, M. (2010). Reply: mass privatisation and mortality. *The Lancet, 373*, 1247–1248.
The Economist. (2009). Mass murder and the market: economic reform in Russia was accompanied by millions of early deaths. But it was not the cause. *The Economist*.
Walberg, P., McKee, M., Shkolnikov, V., Chenet, L., & Leon, D. A. (1998). Economic change, crime and mortality crisis in Russia: regional analysis. *British Medical Journal, 317*, 312–318.
Wedel, J. (2001). *Collision and Collusion: The strange case of Western aid to Eastern Europe*. New York: St. Martin's.
WHO. (2011). *WHO European health for all database*. January 2011 update. World Health Organization.
Wooldridge, J. (2002). *Econometric analysis of cross section and panel data*. Cambridge, MA.