External Adjustment Sensitivity Analysis for Unmeasured Confounding: An Application to Coronary Stent Outcomes, Pennsylvania 2004–2008

Marco D. Huesch

Background. Assessing the real-world comparative effectiveness of common interventions is challenged by unmeasured confounding.

Objective. To determine whether the mortality benefit shown for drug-eluting stents (DES) over bare metal stents (BMS) in observational studies persists after controls for tests for confounding.

Data Sources/Study Setting. Retrospective observational study involving 38,019 patients, 65 years or older admitted for an index percutaneous coronary intervention receiving DES or BMS in Pennsylvania in 2004–2005 followed up for death through 3 years.

Study Design. Analysis was at the patient level. Mortality was analyzed with Cox proportional hazards models allowing for stratification by disease severity or DES use propensity, accounting for clustering of patients. Instrumental variables analysis used lagged physician stent usage to proxy for the focal stent type decision. A method originating in work by Cornfield and others in 1954 and popularized by Greenland in 1996 was used to assess robustness to confounding.

Principal Findings. DES was associated with a significantly lower adjusted risk of death at 3 years in Cox and in instrumented analyses. An implausibly strong hypothetical unobserved confounder would be required to fully explain these results.

Conclusions. Confounding by indication can bias observational studies. No strong evidence of such selection biases was found in the reduced risk of death among elderly patients receiving DES instead of BMS in a Pennsylvania state-wide population.

Key Words. Unmeasured confounding, confounding by indication, selection biases

Understanding the comparative effectiveness of treatments in the community requires valid approaches to control for selection biases such as confounding by indication. This can arise if characteristics of a patient that affect the

real-world choice of treatment also affect the outcomes of interest (Salas, Hotman, and Stricker 1999; Cox and Greenfield 2009).

Such biases are well known (Bhatt et al. 2004; Ko, Mamdani, and Alter 2004; Fox et al. 2007; Sy et al. 2009; Ko et al. 2010; Psaty and Siscovick 2010), although the issue of confounding in general is not always well addressed in published observational studies (Groenwold, Hoes, and Hak 2007).¹ Validated approaches that seek to control for such biases include propensity score adjustments (Austin 2008). However, propensity score methods may not be able to control for *unobserved* confounding if this is present (Rubin 1997; Rothman, Greenland, and Lash 2008).

For example, drug-eluting coronary stents (DES) were shown to be more effective compared with the older bare metal stents (BMS) in initial randomized clinical trials (Moses et al. 2003; Stone et al. 2004) that were underpowered to detect rare adverse effects (Melikian and Wijns 2008). After subsequent reports of rare late stent thrombosis emerged (Pfisterer et al. 2006; Shuchman 2006), several observational studies have examined real-world DES effectiveness (Malenka et al. 2008; Douglas et al. 2009; Ryan et al. 2009). In the largest patient-level study to date, elderly patients receiving DES had significantly lower rates of death both before and after propensity score adjustment (Douglas et al. 2009).

However, if patients who receive BMS are sicker in ways not captured by observed confounders, then propensity score approaches may fail. The resulting selection bias may distort the comparative effectiveness of DES. For example, physicians may assess the patient's ability to afford or to comply with subsequent antiplatelet therapy in ways that are unobserved by the analyst. BMS may then be used in patients more prone to adverse outcomes.

Given the widespread use of DES in the community, understanding the robustness of existing real-world studies to such potential biases is important. Recent studies controlling for such biases using instrumental variable (IV) approaches have unexpectedly found no mortality benefit for DES (Fedorpiel et al. 2012; Venkitachalam et al. 2011). However, IV studies are sensitive to choice of instrument, can have low power, may mask treatment heterogeneity, and inference beyond the marginal patients on which the IV estimates are identified can be limited (McClellan, McNeil, and Newhouse 1994; Harris and Remler 1998; Basu et al. 2007).

Address correspondence to Marco D. Huesch, M.B.B.S., Ph.D., USC Schaeffer Center for Health Policy and Economics, Gateway Unit A, 3335 S. Figueroa St, Los Angeles, CA 90089-7273; e-mail:huesch@usc.edu.

More broadly, observational studies of the real-world effectiveness of treatments that were shown to be efficacious in randomized clinical trials are likely to be increasingly used in comparative effectiveness research (Dreyer et al. 2010).

Accordingly, we gauged the sensitivity of observational long-term coronary stent outcomes results in Pennsylvania to unmeasured confounding. We used a technique originating in work by Cornfield and colleagues (1954) on the causal relationship between smoking and lung cancer, further described by Greenland (1996) and also conducted an IV analysis using observed physician preference as an instrument for stent type (Korn and Baumrind 1998; Brookhart and Schneeweiss 2007). We discuss limitations of the sensitivity analysis and threats to our choice of IV, highlighting persisting methodological problems in using observational data to quantify comparative effectiveness in the real world.

METHODS

Study Population

Discharge data from the Pennsylvania Health Care Cost Containment Council (PHC4) on all state-regulated hospital inpatient admissions for percutaneous coronary interventions (PCI) was linked to death certificate data from the Bureau of Health Statistics & Research at the Pennsylvania Dept of Health for in-state, all-cause deaths. All dates were interval censored to the calendar quarter.

We restricted subjects to be 65 years or older at time of index admission between January 1, 2004 and December 31, 2005 and in which the patient received DES (ICD-9-CM code 36.07), BMS (ICD-9-CM code 36.06), or both (coded as DES but excluded in one additional analysis). To allow feasible in-state follow-up, subjects were further restricted to an in-state address-of-record at the time of the index admission; mortality was followed through 2008. In subset analyses, we restricted the study population based on the presence or absence of myocardial infarction on admission, emergency admission, or Medicare fee for service insurance.

In additional analyses with greater homogeneity, and to allow greater comparability with prior work (Malenka et al. 2008; Douglas et al. 2009), we restricted inclusion to traditional Medicare fee-for-service beneficiaries. Medicare managed care insurance status is independently associated with survival benefits (Huesch 2011). In one subset analysis, we excluded patients

with an emergent index admission, myocardial infarction on admission, or those who did not survive at least 1 day in hospital (Malenka et al. 2008). In another, we excluded patients receiving both DES and BMS (Douglas et al. 2009).

Statistical Analysis

We report the event rate of all-cause death during the index admission, within the same quarter as the discharge quarter, and through 3 years after the quarter of discharge. Patients were at risk for death from the time of procedure and were considered censored after 3 years after discharge quarter. We coded comorbidity indicators to control for patient heterogeneity in coronary stent admissions (Block et al. 1998; Douglas et al. 2009). We controlled for diabetes, chronic obstructive pulmonary disease, congestive heart failure, peripheral vascular disease, cardiac dysrhythmia, liver disease, dementia, chronic renal disease, and nonacute stroke, using ICD-9 diagnosis codes and AHRQ's Clinical Classification codes. Our models also included the MediQual on-admission disease severity score, calculated by a Cardinal Health proprietary algorithm used by the data owner to predict in-hospital mortality from on-admission clinical and chart data.

We compared baseline characteristics by stent type using chi-squared tests. We compared the unadjusted event rates using Fisher's exact tests and plotted Kaplan–Meier survival graphs. Interval censoring yields graphs that plot time no finer than the quarter.

We used multivariate adjusted Cox proportional hazard models to adjust for baseline and clinical characteristics and to report hazard ratios in all Cox models. The Cox models were also refined using stratification and shared frailties. Shared frailty Cox models allow for clustering of patients by allowing a common term (a gamma distributed random effect for treating hospital or county of residence) to enter the specification multiplicatively.

Stratified Cox models relax the assumption of a single baseline hazard function, allowing unique baseline hazards in each decile of the on-admission MediQual disease severity score or the propensity for receiving DES. The propensity score was calculated from an unreported logit regression containing the same demographic and clinical characteristics and also indicators for hospital region and calendar quarter of discharge. The area under the receiver operating characteristic curve was acceptable (C-index 0.72; pseudo R^2 of 10.7 percent).²

Unmeasured Confounding: Mitigation

If confounding by indication is not correlated with observable characteristics, then the above approaches will not mitigate such selection biases. An IV analysis can mitigate for unobserved confounding (Landrum and Ayanian 2001; Stukel et al. 2007; Bhattacharya, Shaikh, and Vytlačil 2008). When physician preferences affect treatment type, an IV for treatment receipt can be derived from those observed preferences (Schneeweiss et al. 2005; Brookhart et al. 2006; Brookhart and Schneeweiss 2007).

Our IV was implemented as the proportion of a physician's admissions in the prior quarter in which a patient received at least one DES. This preference should (i) help to predict assignment of the patient to one treatment, but (ii) must not directly affect the outcome (Wang et al. 2005; Schneeweiss et al. 2008). The first condition can be conceptually justified by persistence in physician preferences over time. We found a Pearson correlation of 0.31 ($p < .001$) between the event of a patient receiving DES and his or her physician's prior quarter proportion of DES use. This condition was also tested using first-stage estimates (Staiger and Stock 1997).

The second, conditional exogeneity, condition can fail under several scenarios, which are further tested in the Results section below. A number of additional falsification tests are also reported. We used two-stage linear least squares to implement our IV approach, contrasting results with an analogous uninstrumented least squares regression. We report the marginal impact of DES use in these linear models as a risk difference in percent points.

Unmeasured Confounding: Sensitivity Analysis

To gauge the sensitivity of our results to unmeasured confounding, we simulated varying strengths of hypothetical associations between a single, binary unmeasured confounder and DES receipt as well as between such a confounder and death through 3 years. This technique is based on original work by Cornfield et al. (1954) on the relationship between smoking and cancer, and it is in increasing use (Rosenbaum 1987, 1991; Greenland 1996; Psaty et al. 1999; Wang et al. 2005; Schneeweiss 2006; Rothman, Greenland, and Lash 2008; Schneeweiss et al. 2008; Brookhart et al. 2010; Chiba 2012).

The intuition is straightforward. Suppose an association between a treatment type and an outcome is observed, with an odds ratio (OR) significantly different from 1. If the observed association is confounded, then external adjustment can "remove" such confounding and result in an externally adjusted OR of 1. If an implausibly strong hypothetical unmeasured

confounder is required to bring the observed OR “back” to 1, then this allows the analyst to be more confident that the observed association is more likely to be causal. If relatively weak confounding can fully explicate the observed results, then the observed association may not be causal in nature.

Unfortunately, this technique is subjective. To provide some objective comparisons and some perspective, we computed the strengths of association between *observed* confounders (i.e., patient-level risk adjustment variables) and both treatment type and the outcome (see Figure 2B). Another difficulty is that the simple approach used here fails in the case in which multiple, strongly positively correlated, unobserved confounders are each only weakly associated with outcome and treatment, but in aggregate represent a stronger association.

We implemented this analysis in an Excel spreadsheet (see Supplementary Table A1 and A5) following Greenland (1996) and Rothman, Greenland, and Lash (2008), pp. 348–351, although other approaches exist (Cabral and Luiz 2007; Lash, Fox, and Fin 2009; Luiz and Cabral 2010). Based on the observed associations, we used event rates of 16 and 20 percent for DES and BMS recipients, respectively, and assumed a single, binary hypothetical confounder with prevalence of 0.4 among BMS recipients.³

The initial, observed odds ratio between BMS stent type and death is then externally adjusted in an iterative process to vary strengths of association between the hypothetical confounder and DES receipt and to vary strengths of association between the hypothetical confounder and death at 3 years. A table (see Supplementary Table A2) summarizes columns representing the relative odds between confounder and outcome, and rows representing the relative odds between confounder and treatment type. In this table, the cells represent the external adjusted ORs. We graphically present results as an isobar representing combinations of strengths of association which would fully explain the observed association.

No adjustments were made for multiple comparisons (Rothman 1990). For all analyses, we used Stata, version 11.1 (Stata Corp, College Station, TX). This study was approved by Duke University’s Health System Institutional Review Board, and declared exempt from IRB review under 45CFR46.101(b)(4).

RESULTS

Baseline Characteristics

Table 1 shows that recipients of BMS were more likely to be sicker than recipients of DES based on observed characteristics.

Table 1: Characteristics of 38,019 Recipients of Drug Eluting or Bare Metal Stents (DES or BMS)

	Received DES (N = 31,934)	Received BMS (N = 6,085)
Age, median (IQR)	74 (69–79)	75 (70–80)*
Male	57.2	61.2*
White race	89.5	92.3*
Hispanic ethnicity	1.6	0.5*
Acuity		
Emergent	36.8	41.2*
Urgent	29.4	31.6
Elective	33.8	27.2
Presentation		
MI on admission	30.3	41.9*
Physician referral	44.1	38.6*
Emergency dept referral	29.6	34.8*
Weekend admission	10.6	14.2*
Afterhours presentation	33.8	34.3
MediQual score, median (IQR)	0.3 (0.1–2.0)	0.8 (0.2–3.5)*
Comorbidities		
Diabetes (II) w/o complications	25.4	23.6 [†]
Congestive heart failure	14.1	17.6*
Chronic pulmonary disease	9.4	11.5*
Peripheral vascular disease	5.2	5.4
Cardiac dysrhythmia	3.4	3.7
Stroke, not acute	3.3	3.2
Renal and ureteric disease	2.3	2.6
Dementia	1.4	1.7 [‡]
Liver disease	0.6	0.9 [‡]
Treatment in index admission		
Multi-vessel PCI	14.6	9.6*
CABG during index admission	0.3	1.2*
Days in hospital, median (IQR)	2 (1–4)	3 (1–5)
Discharge		
Home-health care	8.1	10.1*
Skilled nursing facility	3.5	5.1*
Short-term hospital	0.4	0.9*
Death in-hospital	1.3	3.2*
Alive, but died in qtr of discharge	2.4	4.9*
Died within quarters of discharge qtr		
4	7.1	12.3*
8	11.3	17.3*
12	15.4	22.5*

Note. *Difference significant at $p < .001$.

[†] $p < .01$.

[‡] $p < .05$.

IQR, interquartile range. Ethnicity and race are not mutually exclusive.

BMS users were more likely to be older, to be admitted on weekends through the emergency department, and to present urgently and emergently (all $p < .001$). A BMS user was significantly more likely to have had a coronary artery bypass graft in the same admission, although less likely to undergo multivessel repair (both $p < .001$). The MediQual predicted probability of in-hospital mortality, calculated by the data owner off on-admission clinical and chart data had a higher median among BMS patients (0.8 percent vs. 0.3 percent, $p < .001$).

The length of stay and discharge status and destination supports significantly longer or more complicated recovery. In-hospital death was more than twice as likely for BMS recipients (3.2 percent vs 1.3 percent, $p < .001$), as was discharge alive followed by death in the same quarter (4.9 percent vs. 2.4 percent, $p < .001$). Early mortality, especially within hospital, is unlikely to be due to stent type but to the severity of the underlying disease (see IV falsification test below).

Consistent with the apparent differences in disease severity, crude rates of death were higher among BMS recipients between the beginning of admission and through 1, 2, or 3 years (22.5 percent vs 15.4 percent, $p < .001$) after the discharge quarter (Figure 1).

Conventional Analyses Using Cox Models

In Table 2, results of traditional multivariate adjusted Cox proportional hazards models show that the crude survival benefit (unadjusted relative risk ratio 0.69; 95 percent CI: 0.65–0.73) persists after adjusting for observable patient risk factors, regional and calendar time variations (adjusted hazard ratio 0.79; 95 percent CI: 0.74–0.84).⁴ Confirmatory analyses using stratification, shared frailties, and subgroup analyses with greater homogeneity (Table 2, bottom rows) were similar.

Our conventional estimates are comparable with the largest recent real-world study, which estimated a similar benefit with hazard ratio of 0.75 (95 percent CI: 0.72–0.79) through 30 months (Douglas et al. 2009).

Instrumented Results

In linear least squares regressions ignoring possible treatment selection bias, the difference in risk of death at 3 years postdischarge was a decrease of 4.1 percent points (95 percent CI: 3.0–5.3 percent) among DES recipients (Table 3). In the two-stage least squares IV analysis, this difference was

Figure 1: Unadjusted Cumulative Deaths from Any Cause Following Index Admission

Note. Death in-hospital or after discharge in same quarter as discharge is coded as failing in quarter 0.

estimated as a much larger reduction of 7.5 percent points (95 percent CI: 1.2–13.9 percent) and the IV was strong (F test >10; see Supplementary Table A3).

Robustness Tests for Instrument Exogeneity

We assumed that, given the patient’s baseline and clinical characteristics, the likelihood of death was not correlated with the level of the instrument except through the use of DES. We considered five possible violations of this conditional exogeneity assumption (Staiger and Stock 1997). We also constructed four separate falsification tests for the IV.

First, physicians who tended to use DES more often as a proportion of their practice might also have been more aware of late stent thrombosis, the need for antiplatelet therapy, and the existence of high-risk subsets. A resulting selection bias could threaten our IV. However, over the horizon of this

Table 2: Relative Risk of Death within Three Years after Receipt of Drug-Eluting Stents (DES) versus Bare Metal Stents (BMS) in Cox Models

<i>Model</i>	<i>Hazard Ratio (95% CI)</i>
Unadjusted analysis	0.69 (0.65–0.72)*
Adjusted analysis	0.79 (0.74–0.84)*
Adjusted, stratified by deciles of	
On-admission severity	0.80 (0.75–0.85)*
Propensity to receive DES	0.78 (0.73–0.83)*
Adjusted, shared frailty at level of	
Treating hospital	0.78 (0.74–0.83)*
Patient county	0.78 (0.74–0.83)*
Adjusted analysis by patient subgroups	
With MI on admission	0.81 (0.74–0.88)*
Without MI on admission	0.76 (0.70–0.83)*
Emergency admission	0.80 (0.73–0.88)*
Not emergency admission	0.76 (0.70–0.83)*
With Medicare fee for service	0.80 (0.74–0.86)*
Without Medicare fee for service	0.75 (0.67–0.84)*
Sensitivity analyses	
Medicare fee for service, no MI or emergent admission, and survived at least 1 day in hospital	0.81 (0.72–0.92) [†]
Medicare fee for service, excludes patients receiving both DES and BMS	0.81 (0.75–0.87)*

Note. Except unadjusted analysis, all hazard ratios adjusted for age, gender, race, ethnicity, acuity, physician referral, ER referral, weekend admission, afterhours presentation, MediQual on-admission predicted probability of death, type II diabetes, congestive heart failure, chronic obstructive pulmonary disease, peripheral vascular disease, cardiac dysrhythmia, prior stroke, renal or ureteric disease, dementia, liver disease, index admission multivessel percutaneous coronary intervention, same stay coronary artery bypass graft, and indicators for one of nine regions in Pennsylvania and one of eight calendar quarters of time of discharge.

*Hazard ratio significantly different from 1 at $p < .001$.

[†] $p < .01$.

CI, confidence interval.

2004–2005 study understanding of the late stent thrombosis side effects only emerged slowly (Eisenberg 2004; Ong et al. 2005), with guidelines updated after study end.

Second, high DES use may be confounded by high stent volume: a volume-outcome effect could bias our IV estimates. However, although annual operator procedural volume is associated with improved in-hospital and short-term follow-up outcomes (Smith et al. 2006), it is not clear that long-term outcomes are also sensitive to the operator's experience. We entered each operator's observed total PCI cases over the 2 years as a covariate in our IV models. IV estimates did not change and volume was not a significant predictor of outcomes ($p = .11$).

Table 3: Difference in Risk of Death within 3 Years after Drug-Eluting Stents versus Bare Metal Stents in Linear Models

<i>Model</i>	<i>Risk Difference (95% CI)</i>
Adjusted analysis	
Ordinary least squares	-4.1 (-5.3 to -3.0)*
Instrumental variables least squares	-7.5 (-13.9 to -1.2)†
Instrument robustness checks	
MI on admission only	
Ordinary least squares	-4.9 (-6.8 to -3.0)*
Instrumental variables least squares	-13.0 (-23.7 to -2.4)†
Addition of facility size covariate	
Ordinary least squares	-4.1 (-5.3 to -3.0)*
Instrumental variables least squares	-7.6 (-14.0 to -1.3)†

Note. All risk differences adjusted for age, gender, race, ethnicity, acuity, physician referral, ER referral, weekend admission, afterhours presentation, MediQual on-admission predicted probability of death, type II diabetes, congestive heart failure, chronic obstructive pulmonary disease, peripheral vascular disease, cardiac dysrhythmia, prior stroke, renal or ureteric disease, dementia, liver disease, index admission multivessel percutaneous coronary intervention, same stay coronary artery bypass graft, and indicators for one of nine regions in Pennsylvania and one of eight calendar quarters of time of discharge.

*Risk difference significantly different from 0 at $p < .001$.

† $p < .05$.

Risk difference in % points. CI, confidence interval.

Third, our instrument could also fail if patients preferred physicians who used DES more often, and such patients were unobservably healthier. However, patients presenting with MI are less likely to be able to exercise such choices. In such subsets (Table 3, middle rows), the significant 3-year mortality risk difference in favor of DES persisted and strengthened. Fourth, a similar violation and the same rebuttal holds for a scenario in which physicians who used more DES as a proportion of their practice were able to choose unobservably healthier patients.

Fifth, it is possible that more skilled physicians might be sorted into larger academic medical centers in which they generally enjoyed earlier and better access to DES. Higher use of DES might then be spuriously correlated with unobserved qualities of the physician. However, we note that hospital DES adoption dates had also been conjectured as possibly correlated with contemporaneous rates of DES use (Malenka, Skinner, and Lucas 2009), and similar time-based instruments have been used more recently (Federspiel et al. 2012; Venkitachalam et al. 2011). The implicit assumption of these studies is that adoption timing is conditionally exogenous to outcomes. If this holds, it would downplay the threat to our IV.

However, to understand this potential threat to our IV validity better, we added a measure of hospital size, facility beds, to the linear regressions (Table 3, bottom row) to mitigate this bias. Estimates of DES impact on 3-year death rates were similar and remained significant, and the unreported coefficient on facility beds was insignificant ($p = .58$). We also note that adoption of DES was rapid between the first quarter of 2003 and the study start date, at which time the national rate of DES use had reached more than 50 percent of all percutaneous coronary interventions (Malenka et al. 2008). This makes it less likely that higher DES use was actually correlated with important unobserved physician or hospital quality factors as opposed to idiosyncratic physician preferences.

Although each of these five threats to the IV do not individually seem plausible or strong enough to threaten the robustness of our findings, it is nevertheless true that the combination of these threats could still render the IV invalid.⁵ To further explore the robustness of our IV, we conducted four additional falsification tests. We hasten to acknowledge that while our IV did not fail these falsification tests, this is necessary but not sufficient for ensuring the IV's validity.

First, we examined the rare (0.25 percent) event of self-discharge post stenting, against medical advice (AMA). This may be associated with unobserved good health, but it should not be caused by DES directly. AMA was positively associated with being younger, being a male, and negatively associated with being Caucasian (all $p < .001$). We found insignificant IV estimates near zero (+0.05 percent point risk difference; $p = .84$) in a poorly fitting model.

Second, we chose the subset of 5,884 patients who received only BMS and regressed the IV on mortality through 3 years postdischarge. In these BMS patients, who did not receive DES, our logic is that there should be no impact of the IV on outcomes unless the IV is capturing a beneficial selection effect. In a well-fitting OLS regression ($R^2 = 13.5$ percent), the IV had no significant effect on survival ($p = .52$). Third, we chose the subset of 1,163 patients who received both DES and BMS. Anatomical reasons (vessel size, deployment difficulties) typically lead to both stents being used (Abbott et al. 2007). In a well-fitting OLS regression ($R^2 = 12.9$ percent), no significant impact for the IV was found ($p = .73$).

Fourth, IV validity would be strengthened if earlier survival was not significantly impacted. We confirmed this in instrumented regressions on earlier mortality outcomes (Supplementary Table A4): death during the index admission (-0.021 ; 95 percent confidence interval [CI]: -0.043 to $+0.001$;

$p = .067$), and after live discharge but within the same quarter as the quarter of discharge (-0.022 ; 95 percent CI: -0.05 to $+0.007$; $p = .133$). However, this conclusion would be stronger if the early IV point estimates were closer to zero, rather than just including zero in their 90 or 95 percent confidence intervals. In summary, while our instrument appeared robust to considerations of instrument plausibility, weakness, or failure, standard concerns on instrument validity remain.

Treatment Heterogeneity and the Marginal Patient in Instrumented Analyses

It is not generally clear to what extent an IV estimate of survival benefit is applicable to the average patient (Brookhart and Schneeweiss 2007; Stukel et al. 2007), nor who the marginal patient is (Harris and Remler 1998). The marginal patient is a patient whose treatment status depends on the level of the instrument. Both patient factors and variation in the level of the instrument across physicians impact that likelihood.

Consider an appropriateness curve for receipt of stent type, ranging from those patients whose clinical characteristics easily meet the indications for DES at one end of this curve to those at the other end who are medically contraindicated for DES and thus get BMS. Our IV estimates do not pertain to patients who would have almost certainly received DES (or BMS). In these patients, the assumption of quasi-random assignment via the IV as randomizer is clearly violated. In our setting, the marginal patients are those patients in whom sufficient clinical uncertainty exists that would allow either DES or BMS to be used. Such patients likely resemble most those used in the falsification test above where both BMS and DES were actually received.

In addition to who the marginal patient most likely is, another important question is whether a particular patient type actually has different chances of encountering physicians with different preferences.⁶ Ideally, we would like patients to have similar chances of seeing a physician who inserts DES 95 percent of the time, as well as physicians who insert DES only 5 percent of the time. As Brookhart and Schneeweiss point out (2007), if patients instead face the chance of seeing physicians with less diverging preferences, then they are less marginal and “are therefore likely to be down weighted in an instrumental variable estimate.”

In our data, all hospitals have some physicians whose treatment preferences were for DES 100 percent of the time and some who were for BMS 100 percent of the time. A histogram of the mean within-hospital level of the IV

shows a broad mass in the 50–90 percent mean DES use range. This suggests that sufficient variation in physician preferences exists at the hospital level to ensure that marginal patients can be identified by the IV.

Finally, the IV estimates may still be biased if heterogeneous treatment effects exist. In this scenario, some patients will benefit more from the treatment based on unobservable factors. In settings like ours where treatment is more likely to be overused than underused in a population, and assuming the exclusion restriction actually holds, Brookhart and Schneeweiss (2007) show that the IV estimate bias is bounded above by zero. That is, the estimates will be more negative, showing an incorrectly large reduction in risk attributable to the treatment. Intuitively, the additional patients seen by physicians with high levels of the IV are less likely to need a DES, more likely to do well anyway, yet more likely to get a DES. In consequence, in a setting like ours where overuse of DES is more likely than underuse, the IV estimate of survival benefit may be overweighted on average and not applicable to the average patient.⁷

Sensitivity Analysis Using External Adjustment

Finally, we report a sensitivity analysis to judge the robustness of our results to a hypothetical confounder. The curve in Figure 2A denotes the strength of associations required to bring the odds ratio for the outcome among BMS recipients from 1.31 to 1 compared with DES (equivalently, the odds ratio for death among DES from 0.80 to 1), if solely due to unobserved confounding by a single binary unmeasured confounder.

Large OR of 3 or more, between such an unmeasured confounder and both the outcome as well as the use of a particular stent type, would be required to fully explain the observed association. To put these OR into perspective, we computed the odds ratios between individual observed confounders (i.e., medical comorbidities) in an analogous way in Figure 2B. It is readily apparent that none of the observed confounders came close to the strengths of association that a hypothetical unmeasured confounder would have to have to fully explicate the observed association between stent type and death.

Figure 2: (A) Strengths of Associations Required for Hypothetical Unmeasured Confounder to Fully Explain Observed Association between Drug-Eluting Stents (DES) and Death within Three Years. (B) Strengths of Associations between *Observed* Confounders and Stent Type and Death.

Note. Assumes hypothetical confounder is single, binary, and with prevalence of 0.4 among bare metal stents (BMS) recipients and varying lower prevalence among BMS patients (vertical axis) and varying strength of association between confounder and death (horizontal axis). Results robust to varying prevalence assumptions. Curved line segment in (B) represents magnified detail of curve in (A).

On one hand, although some comorbidities such as renal disease were strongly associated with longer term poor survival, they were not significantly related to stent type. On the other hand, while prior coronary artery bypass graft strongly predicted BMS stent type, it was only weakly associated with long-term survival. Better risk adjustment with richer clinical covariates may have found such confounders.²

Although this external adjustment sensitivity analysis (Cornfield et al. 1954; Greenland 1996) qualitatively supports the argument that unmeasured confounding may not be playing a large part in producing the observed association, it does not rule out such unmeasured confounding. For example, the combination of several, weaker unmeasured confounders may together fully explain our results. This challenge remains a weakness of external adjustment.

DISCUSSION

In this study, we addressed the robustness of observational studies of coronary stent comparative effectiveness to selection biases such as confounding by indication. We compared and contrasted the results of conventional analyses with analyses that either attempted to mitigate unmeasured confounding or sought to gauge the sensitivity of observed results to such confounding.

Controlling only for observable factors (traditional, adjusted, stratified, and shared frailty Cox models) our results are consistent in direction and magnitude with prior work in showing a substantial and significant survival benefit for DES use (Douglas et al. 2009).⁸

However, BMS recipients were observably sicker on admission, and had higher in-hospital mortality than DES recipients. Although our conventional analyses adjusted for observed confounders, unmeasured confounders may have guided a physician's decision to choose between BMS or DES.

Our instrumental variables (IV) approach attempted to control for such confounding by exploiting plausibly exogenous variation in physician preferences for stent type. The IV estimates results were in line with both our own uninstrumented results as well as past observational studies. However, the estimated risk difference in favor of DES of 7.5 percent points was nearly twice as high as the uninstrumented result of a 4 percent point risk difference.

Other recent studies find no benefit for DES using DES adoption time as an IV for stent type (Federspiel et al. 2012; Venkitachalam et al. 2011). This suggests the possibility that our instrumented results may not be valid, or that

our instrument is valid but identifies a different marginal patient than those identified by the other studies (McClellan, McNeil, and Newhouse 1994; Basu et al. 2007; Normand 2008).

In a complementary approach, we used an external adjustment sensitivity analysis to gauge the sensitivity of our observed results to unmeasured confounding (Cornfield et al. 1954; Greenland 1996). The results suggest that large odds ratios of 3 or more must be present between a single, binary, unmeasured confounder and both the choice of stent and the outcome of death. We note that no *observed* confounder had such strong associations with outcome and death. This analysis does not exclude the possibility that multiple, correlated, unmeasured confounders could in combination completely explicate the results.

Limitations

Our study has several important limitations. External validity is threatened by use of data from one small state, whereas internal validity is limited by the use of administrative data and possible cohort biases. Risk adjustment is poor in comparison with registry data or similar clinically audited and harvested data, although our data did include the MediQual Atlas predicted in-hospital mortality datum.

The coarsened time variables prevented more granular analyses such as landmark analyses for antiplatelet therapy (Malenka et al. 2008). A related limitation is our use of death certificate data, which prevented finer analyses of competing risks of deaths not due to ischemic heart disease and fails to capture out-of-state deaths.

A further limitation is unobserved postdischarge care differences (Leon et al. 2009). We did not observe whether longer term dual-acting antiplatelet therapy (in particular, clopidogrel) differed systematically between stent types (Eisenstein et al. 2007). It is thus possible that the entire benefits attributed to DES stent type in this study instead accrue to better and longer adherence to clopidogrel therapy. This alternative cannot be ruled out in this study.⁹

Conclusion

Our results strengthen the findings of prior observational work that DES use in real-world settings is associated with better long-term mortality rates than BMS use (Douglas et al. 2009). Controlling for unobserved and unmeasured

confounding by indication, our results do not change this conclusion. This study suggests that in some instances, the use of real-world comparative effectiveness data can survive the multiple challenges such data are exposed to by their construction in a nonrandomized setting. We advance the usefulness of external adjustment sensitivity analysis in helping to strengthen confidence in observational studies' results.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: The impetus for this study was a talk on observational data methods given by Dr. Sharon-Lise Normand at AcademyHealth's 2010 Annual Research Meeting in Boston. I am grateful to Dr. Normand for inspiring this paper. This study implements the method given by Sander Greenland in his 1996 external adjustment article. I thank, without implicating, Dr. Greenland for helping me understand several details of this study. I appreciate helpful comments from seminar participants at Memorial Sloan Kettering's Center for Health Policy & Outcomes and the Department of Epidemiology & Biostatistics in 2010, and from audience members and presenters in the CER Methods session chaired by David Melzer at AcademyHealth's 2011 Annual Research Meeting in Seattle.

Disclosures: I have no financial or other potential conflicts of interest to disclose.

Disclaimers: None.

NOTES

1. Groenwold, Hoes, and Hak (2007) searched all observational intervention studies between 1985 and 2005. As late as 2005, only 9 percent mentioned "confounding," "adjustment," or "bias." A sub-analysis in seven leading general medical journals showed that since the early 1990s, slightly more than 30 percent of non-randomized intervention studies mentioned any of the terms.
2. To put this model's explanatory power into context, the largest analogous observational study based off of initial registry data with rich chart and clinical data and more than 50 covariates yielded a risk model with a similar c-index of 0.71 (Douglas et al. 2009, p. 1633).
3. Results were not sensitive to changing the prevalence assumption between 0.4 and 0.8.

4. A log-log plot of survival by stent type showed reasonably parallel log-log curves, implying that the proportional hazards assumption was satisfied for the stent type variable.
5. We are grateful to an anonymous referee for pointing out this concern and acknowledge the referee's valuable suggestion of a falsification test centered on early mortality.
6. This is unrelated to the strength of the IV. If all physicians have identical very strong preferences for DES, this strong IV will clearly be highly correlated with treatment type. But there could be no marginal patients whose treatment is affected by the level of the IV by construction, since they are all very likely to get the same treatment.
7. This conclusion would also hold if patients who received a DES were less sick in an unobservable (to the analyst) manner and thus more likely to do well anyway. For example, suppose higher-volume interventional cardiologists favored DES stents, and they were more inclined to err on the margin of caution when deciding to insert a stent after the angiogram. Anecdotal reports of an "oculostenotic reflex" in some cardiologists, and scattered reports of billing fraud involving the insertion of stents in patients without angiographic evidence of their need support this concern.
8. Real-world effectiveness seems better than two original RCT results, which did not show survival differences attributable to DES (Moses et al. 2003; Stone et al. 2004). The RCTs restricted patients to "on-label" DES; off-label use was 69 percent of DES implantations in the Douglas and colleagues study (2009). Subset analyses better fitting the original RCT inclusion and exclusion criteria still showed similarly high DES effectiveness (Douglas et al. 2009).
9. Linking Medicare Part D data on Plavix[®] (clopidogrel) prescriptions to Part B data, or using commercial health plan integrated data, could ascertain after-care differences.

REFERENCES

- Abbott, J., V. Dawn, R. Matthew, M. Nakamura, H. A. Cohen, F. Selzer, Kip, K. E., Vlachos, H. A., Wilensky, R. L., and Williams, D. O. 2007. "Unrestricted Use of Drug-Eluting Stents Compared with Bare-Metal Stents in Routine Clinical Practice: Findings from the National Heart, Lung, and Blood Institute Dynamic Registry." *Journal of the American College of Cardiology* 50: 2029–36.
- Austin, P. C. 2008. "Primer on Statistical Interpretation or Methods Report Card on Propensity-Score Matching in the Cardiology Literature from 2004 to 2006: A Systematic Review." *Circulation Cardiovascular Quality and Outcomes* 1: 62–7.
- Basu, A., J. J. Heckman, S. Navarro-Lozano, and S. Urzua. 2007. "Use of Instrumental Variables in the Presence of Heterogeneity and Self-Selection: An Application to Treatments of Breast Cancer Patients." *Health Economics* 16: 1133–57.

- Bhatt, D. L., M. T. Roe, E. D. Peterson, Y. Li, A. Y. Chen, R. A. Harrington, et al. 2004. "Utilization of Early Invasive Management Strategies for High-Risk Patients with Non-ST-Segment Elevation Acute Coronary Syndromes: Results from the CRUSADE Quality Improvement Initiative." *Journal of the American Medical Association* 292: 2096–104.
- Bhattacharya, J., A. Shaikh, and E. Vytlačil. 2008. "Treatment Effect Bounds: An Application to Swan-Ganz Catheterization." *American Economic Review* 98: 351–6.
- Block, P. C., E. D. Peterson, R. Krone, K. Kesler, E. Hannan, G. T. O'Connor, and K. Detre. 1998. "Identification of Variables Needed to Risk Adjust Outcomes of Coronary Interventions: Evidence-Based Guidelines for Efficient Data Collection." *Journal of the American College of Cardiology* 32: 275–82.
- Brookhart, M. A., and S. Schneeweiss. 2007. "Preference-Based Instrumental Variable Methods for the Estimation of Treatment Effects: Assessing Validity and Interpreting Results." *International Journal of Biostatistics* [accessed on October 25, 2012]. Available at <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2719903/>
- Brookhart, M. A., P. S. Wang, D. H. Solomon, and S. Schneeweiss. 2006. "Evaluating Short-Term Drug Effects Using a Physician-Specific Prescribing Preference as an Instrumental Variable." *Epidemiology* 17 (3): 268–75.
- Brookhart, M. A., T. Stürmer, R. J. Glynn, J. Rassen, and S. Schneeweiss. 2010. "Confounding Control in Healthcare Database Research: Challenges and Potential Approaches." *Medical Care* 48 (6): S114–20.
- Cabral, M. D. B., and R. R. Luiz. 2007. "Sensitivity Analysis for Unmeasured Confounders Using an Electronic Spreadsheet." *Revista de Saude Publica* 41 (3): 446–52.
- Chiba, T. 2012. "Sensitivity Analysis for Unmeasured Confounding of Attributable Fraction." *Epidemiology* 23 (1): 175–6.
- Cornfield, J., W. Haenszel, E. C. Hammond, A. M. Lilienfeld, M. B. Shimkin, and E. L. Wynder. 1954. "Smoking and Lung Cancer: Recent Evidence and a Discussion of Some Questions." *Journal of the National Cancer Institute* 22:173–203. Reprinted. *International Journal of Epidemiology* 2009 (38): 1175–91.
- Cox, H. C., and S. Greenfield. 2009. "Comparative Effectiveness Research: A Report from the Institute of Medicine." *Annals of Internal Medicine* 151: 203–5.
- Douglas, P. S., J. M. Brennan, K. J. Anstrom, et al. 2009. "Clinical Effectiveness of Coronary Stents in Elderly Persons: Results from 262,700 Medicare Patients in the American College of Cardiology-National Cardiovascular Data Registry." *Journal of the American College of Cardiology* 53: 1629–41.
- Dreyer, N. A., S. R. Tunis, M. Berger, D. Ollendorf, P. Mattox, and R. E. Gliklich. 2010. "Why Observational Studies Should Be among the Tools Used in Comparative Effectiveness Research." *Health Affairs (Millwood)* 29 (10): 1818–25.
- Eisenberg, M. J. 2004. "Drug-Eluting Stents: Some Bare Facts." *Lancet* 364: 1466–7.
- Eisenstein, E. L., K. J. Anstrom, D. F. Kong, L. K. Shaw, R. H. Tuttle, D. B. Mark, et al. 2007. "Clopidogrel Use and Long-Term Clinical Outcomes after Drug-Eluting Stent Implantation." *Journal of the American Medical Association* 297 (2): 159–68.

- Federspiel, J. J., S. C. Stearns, B. C. Sheridan, J. J. Kuritzky, L. P. D'Arcy, D. J. Crespin, T. S. Carey, and J. S. Rossi. 2012. "Evaluating the Effectiveness of a Rapidly Adopted Cardiovascular Technology with Administrative Data: The Case of Drug-Eluting Stents for Acute Coronary Syndromes." *American Heart Journal* 164 (2): 207–14.
- Fox, K. A., F. A. Anderson, O. H. Dabbous, P. G. Steg, J. L. Lopez-Sendon, F. Van de Werf, et al. 2007. "Intervention in Acute Coronary Syndromes: Do Patients Undergo Intervention on the Basis of Their Risk Characteristics? The Global Registry of Acute Coronary Events (GRACE)." *Heart* 93: 177–82.
- Greenland, S. 1996. "Basic Methods for Sensitivity Analysis of Biases." *International Journal of Epidemiology* 25: 1107–16.
- Groenwold, R. H. H., A. W. Hoes, and E. Hak. 2007. "Confounding in Publications of Observational Intervention Studies." *European Journal of Epidemiology* 22 (7): 413–5.
- Harris, K. M., and D. K. Remler. 1998. "Who Is the Marginal Patient? Understanding Instrumental Variables Estimates of Treatment Effects." *Health Services Research* 33: 1337–60.
- Huesch, M. D. 2011. "Long-Term Follow-Up Outcomes after Medicare Managed Care of Coronary Disease." *Medical Care* 50 (1): 18–26.
- Ko, D. T., M. Mamdani, and D. A. Alter. 2004. "Lipid-Lowering Therapy with Statins in High-Risk Elderly Patients: The Treatment-Risk Paradox." *Journal of the American Medical Association* 291: 1864–70.
- Ko, D. T., J. S. Ross, Y. Wang, and H. M. Krumholz. 2010. "Determinants of Cardiac Catheterization Use in Older Medicare Patients with Acute Myocardial Infarction." *Circulation Cardiovascular Quality and Outcomes* 3: 54–62.
- Korn, M. A., and E. Baumrind. 1998. "Clinician Preferences and the Estimation of Causal Treatment Differences." *Statistical Science* 13 (3): 209–35.
- Landrum, M. B., and J. Z. Ayanian. 2001. "Causal Effect of Ambulatory Specialty Care on Mortality Following Myocardial Infarction: A Comparison of Propensity Score and Instrumental Variable Analyses." *Health Services Outcomes Research and Method* 2: 221–45.
- Lash, T. L., M. P. Fox, and A. K. Fin. 2009. *Applying Quantitative Bias Analysis to Epidemiologic Data*, 1st Edition. New York: Springer.
- Leon, M. B., D. J. Allocco, K. D. Dawkins, and D. S. Baim. 2009. "Late Clinical Events after Drug-Eluting Stents: The Interplay between Stent-Related and Natural History-Driven Events." *Journal of the American College of Cardiology Cardiovascular Interventions* 2: 504–12.
- Luiz, R. R., and M. D. B. Cabral. 2010. "Sensitivity Analysis for an Unmeasured Confounder: A Review of Two Independent Methods." *Revista Brasileira Epidemiologica* 13 (2): 188–98.
- Malenka, D. J., J. S. Skinner, and F. L. Lucas. 2009. "Outcomes in the Era of Bare-Metal Stents versus the Era of Drug-Eluting Stents—Reply." *Journal of the American Medical Association* 301: 34.
- Malenka, D. J., A. V. Kaplan, F. L. Lucas, S. M. Sharp, and J. S. Skinner. 2008. "Outcomes Following Coronary Stenting in the Era of Bare-Metal versus

- the Era of Drug-Eluting Stents." *Journal of the American Medical Association* 299: 2868–76.
- McClellan, M., B. J. McNeil, and J. P. Newhouse. 1994. "Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality? Analysis Using Instrumental Variables." *Journal of the American Medical Association* 272: 859–66.
- Melikian, N., and W. Wijns. 2008. "Drug-Eluting Stents: A Critique." *Heart* 94: 145–52.
- Moses, J. W., M. B. Leon, J. J. Popma, et al. 2003. "Sirolimus-Eluting Stents versus Standard Stents in Patients with Stenosis in a Native Coronary Artery." *New England Journal of Medicine* 349: 1315–23.
- Normand, S. L. T. 2008. "Some Old and Some New Statistical Tools for Outcomes Research." *Circulation* 118: 872–84.
- Ong, A. T. L., E. P. McFadden, E. Regar, P. P. T. de Jaegere, R. T. van Domburg, and P. W. Serruys. 2005. "Late Angiographic Stent Thrombosis Events with Drug-Eluting Stents." *Journal of the American College of Cardiology* 45: 2088–92.
- Pfisterer, M., H. P. Brunner-LaRocca, P. Buser, et al. 2006. "Late Clinical Events after Clopidogrel Discontinuation May Limit the Benefit of Drug-Eluting Stents." *Journal of the American College of Cardiology* 48: 2584–91.
- Psaty, B. M., and D. S. Siscovick. 2010. "Minimizing Bias Due to Confounding by Indication in Comparative Effectiveness Research: The Importance of Restriction." *Journal of the American Medical Association* 304: 897–8.
- Psaty, B. M., T. D. Koepsell, D. Lin, N. S. Weiss, et al. 1999. "Assessment and Control for Confounding by Indication in Observational Studies." *Journal of the American Geriatric Society* 47: 749–54.
- Rosenbaum, P. R. 1987. "Sensitivity Analysis for Certain Permutation Inferences in Matched Observational Studies." *Biometrika* 74 (1): 13–26.
- . 1991. "Discussing Hidden Bias in Observational Studies." *Annals of Internal Medicine* 115 (11): 901–5.
- Rothman, K. J. 1990. "No Adjustments Are Needed for Multiple Comparisons." *Epidemiology* 1: 43–6.
- Rothman, K. J., S. Greenland, and T. L. Lash. 2008. *Modern Epidemiology*, 3rd Edition. Philadelphia, PA: Lippincott, Williams & Wilkins.
- Rubin, D. B. 1997. "Estimating Effects from Large Data Sets Using Propensity Scores." *Annals of Internal Medicine* 127: 363–74.
- Ryan, J., W. Linde-Zwirble, L. Engelhart, L. Cooper, and D. J. Cohen. 2009. "Temporal Changes in Coronary Revascularization Procedures, Outcomes, and Costs in the Bare-Metal Stent and Drug-Eluting Stent Eras: Results from the US Medicare." *Circulation* 119: 952–61.
- Salas, M., A. Hotman, and B. H. Stricker. 1999. "Confounding by Indication: An Example of Variation in the Use of Epidemiologic Terminology." *American Journal of Epidemiology* 149: 981–3.
- Schneeweiss, S. 2006. "Sensitivity Analysis and External Adjustment for Unmeasured Confounders in Epidemiologic Database Studies of Therapeutics." *Pharmacoeconomics and Drug Safety* 15 (5): 291–303.

- Schneeweiss, S., R. J. Glynn, J. Avorn, and D. H. Solomon. 2005. "A Medicare Database Review Found That Physician Preferences Increasingly Outweighed Patient Characteristics as Determinants of First-Time Prescriptions for COX-2 Inhibitors." *Journal of Clinical Epidemiology* 58 (1): 98–102.
- Schneeweiss, S., J. D. Seeger, J. Landon, and A. M. Walker. 2008. "Aprotinin during Coronary-Artery Bypass Grafting and Risk of Death." *New England Journal of Medicine* 358: 771–83.
- Shuchman, M. 2006. "Trading Restenosis for Thrombosis? New Questions about Drug-Eluting Stents." *New England Journal of Medicine* 355: 1949–52.
- Smith Jr, S. C., T. E. Feldman, J. W. Jr, Hirshfeld, et al. 2006. "ACC/AHA/SCAI 2005 Guideline Update for Percutaneous Coronary Intervention: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (ACC/AHA/SCAI Writing Committee to Update the 2001 Guidelines for Percutaneous Coronary Intervention)." *Journal of the American College of Cardiology* 47: e1–121.
- Staiger, D. O., and J. H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65: 557–86.
- Stone, G. W., S. G. Ellis, D. A. Cox, et al. 2004. "A Polymer-Based, Paclitaxel-Eluting Stent in Patients with Coronary Artery Disease." *New England Journal of Medicine* 350: 221–31.
- Stukel, T. A., E. S. Fisher, D. E. Wennberg, D. A. Alter, D. J. Gottlieb, and M. J. Vermeulen. 2007. "Analysis of Observational Studies in the Presence of Treatment Selection Bias: Effects of Invasive Cardiac Management on AMI Survival Using Propensity Score and Instrumental Variable Methods." *Journal of the American Medical Association* 297: 278–85.
- Sy, R. W., P. G. Bannon, M. S. Bayfield, C. Brown, and L. Kritharides. 2009. "Survivor Treatment Selection Bias and Outcomes Research: A Case Study of Surgery in Infective Endocarditis." *Circulation Cardiovascular Quality and Outcomes* 2: 469–74.
- Venkitachalam, L., Y. Lei, J. M. Stolker, E. M. Mahoney, A. P. Amin, J. B. Lindsey, K. F. Kennedy, M. J. Pencina, J. J. Lopez, N. S. Kleiman, and D. J. Cohen. 2011. "Clinical and Economic Outcomes of Liberal Versus Selective Drug-Eluting Stent Use: Insights from Temporal Analysis of the Multicenter Evaluation of Drug Eluting Stents and Ischemic Events (EVENT) Registry." *Circulation*. Available at <http://circ.ahajournals.org/content/early/2011/08/13/CIRCULATIONAHA.110.978593.full.pdf+html> [accessed on October 25, 2012].
- Wang, P. S., S. Schneeweiss, J. Avorn, M. A. Fischer, H. Mogun, D. H. Solomon, and M. A. Brookhart. 2005. "Risk of Death in Elderly Users of Conventional vs. Atypical Antipsychotic Medications." *New England Journal of Medicine* 353 (22): 2335–41.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Table S1. Screenshot of Excel Spreadsheet Implementing Greenland's Sensitivity Analysis.

Table S2. Iterative Use of the External Adjustment Technique Populates a Table of Externally Adjusted Odds Ratios from Which Confounding at Varying Strengths Has Been Removed.

Table S3. First Stage 2SLS Regression on Receipt of DES.

Table S4. Comparison in OLS and IV Estimates of DES Impact at Different Times between Index Admission and Three Years Postdischarge.

Table S5. Editable Microsoft Excel Spreadsheet Implementing External Adjustment Method, Following Rothman, Greenland, and Lash (2008).