The Instrumental Variable Method to Study Self-Selection Mechanism: A Case of Influenza Vaccination

Byung-Kwang Yoo, MD, PhD, 1 Kevin D. Frick, PhD 2
1 Center for Health Policy, Stanford University, Stanford, CA, USA; 2 Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

ABSTRACT

Objective: To assess whether estimates of the effectiveness of influenza vaccination in reducing rates of hospitalizations and all-cause mortality derived from cross-sectional data could be improved by applying the instrumental variable (IV) method to data representing the community-dwelling elderly population in the United States in order to adjust for self-selection bias.

Methods: Secondary data analysis, using the 1996–97 Medicare Current Beneficiary Survey data. First, using single-equation probit regressions this study analyzed influenza-related hospitalization and death due to all causes predicted by vaccination status, which was measured by claims or survey data. Second, to adjust for potential self-selection of the vaccine receipt, for example, higher vaccination rates among high-risk individuals, bivariate probit (BVP) models and two-stage least squares (2SLS) models were employed. The IV was having either arthritis or gout.

Results: In single-equation probit models, vaccination appeared to be ineffective or even to increase the probability of adverse outcomes. Based on BVP and 2SLS models, vaccination was demonstrated to be effective in reducing influenza-related hospitalization by at least 31%. The BVP model results implied significant self-selection in the single-equation probit models.

Conclusions: Adjusting for self-selection, BVP analyses yielded vaccine effectiveness estimates for a nationally representative cross-sectional sample of the community-dwelling elderly population that are consistent with previous estimates based on randomized controlled trials, prospective cohort studies, and meta-analyses. This result suggests that analyses with 2SLS and BVP in particular may be useful for the analysis of observational data regarding prevention in which self-selection is an important potential source of bias.

Keywords: bivariate probit model, influenza vaccination, instrumental variable method, self-selection bias, vaccine effectiveness.

Introduction

Influenza and pneumonia ranked fifth among all causes of death for those aged 65 and older and ranked sixth among all age groups in the United States in 1997 [1]. Medicare reimbursement for excess hospitalization ranged from $750 million to $1 billion per epidemic from 1989 to 1991 [2]. At present, the main option for reducing the considerable impact of influenza in the United States is annual vaccination before the influenza season for people at high-risk for influenza and its complications and health-care workers [3]. People at high-risk include all those 50 years and older, residents of nursing homes and other chronic care facilities, and nonelderly individuals with specific chronic medical conditions, for example, chronic disorders of the pulmonary or cardiovascular systems, including asthma [3].

To the best of our knowledge, only two randomized clinical trials (RCTs) have been conducted to examine the efficacy of influenza vaccination in community elderly population in the Netherlands and Great Britain [4,5]. Over time, RCTs have become more difficult to conduct for some treatments because of ethical, financial, and administrative difficulties, particularly for studies that are large in scale, with frequent observations, and with a longer follow-up period. When observational studies are the only choice available, particularly for effectiveness rather than efficacy research, the analysis of observational data may yield a biased effectiveness estimate for interventions that can be chosen by individuals. The choice leads to a potential violation of an assumption for unbiased estimated, that is, that the residual term in a model is not correlated with an explanatory variable. The correlation that leads to violations of the assumption, an endogeneity problem, may stem from two problems: the omitted variable problem and the simultaneous causality problem, that is, two-way causality between an outcome and an intervention. The “omitted variable problem” can, in theory, be solved by an observational
data analysis including all variables possibly correlated with the intervention. This solution is still unable to solve the latter “simultaneous causality problem” with observational data.

The instrumental variable (IV) method, an econometric technique, can address both problems in evaluating intervention effectiveness using an observational data with a limited number of variables. Although this technique has been applied in evaluating various types of acute care such as acute myocardial infarction [6] and surgical treatments for breast cancer [7], this technique could make further contributions to the literature of preventive care effectiveness evaluation.

As McClellan and Newhouse stated, the IV method should be used as a complement rather than a substitute for a randomized controlled experiment [8]. Nevertheless, when RCT data are limited, IV approaches are useful but will help to control an endogeneity problem only if good IVs are identified in evaluating a medical intervention for which the choice is affected by unobservable factors. The IV method’s key advantage compared with other statistical methods used to analyze observational data is the ability to isolate exogenous effects of an intervention by excluding endogenous self-selection effects without having to directly measure such self-selection. Because of this advantage the IV method has the potential to be superior to a conventional single-equation model unless all variables possibly correlated to the focused intervention are available for a single-equation model, that is, no omitted variable problem.

Even if an intervention is truly effective in reducing the risk of adverse health outcomes, a single-equation model could mistakenly lead to the inference that the intervention is either ineffective or significantly “increases” the adverse outcome risks when high-risk individuals are more likely to receive the intervention. When the opposite type of self-selection occurs, intervention effectiveness could be overestimated to the extent that low-risk individuals are more likely to receive the intervention. Because a more intensive medical intervention tends to be offered to higher-risk individuals, its effectiveness is likely to be underestimated when using observational data due to the former type of self-selection as observed in the literature, for example, intensive treatment of acute myocardial infarction [6], and more and early prenatal care for women with a higher risk of bearing a low birth weight child [9].

The two-stage estimation method, a subtype of the IV method, is a general way to obtain consistent estimates by adjusting for a potential endogeneity problem, for example, self-selection [10]. Under the two-stage estimation method, the first-stage equation regresses a potentially self-selected variable on covariates. This stage aims at obtaining predicted values of the self-selected variable, that is, the flu shot (FS) receipt. The second-stage equation regresses a health outcome variable on covariates including the predicted values of the self-selected variable obtained from the first-stage regression in place of the self-selected variable [10].

Special caution is needed in choosing an appropriate subtype of the IV method depending on the types of outcome and intervention. In particular, the appropriate estimation method depends on whether variables are continuous or dichotomous. This is because the two-stage estimation method does not generally yield an unbiased effectiveness estimate, resulting from the second-stage probit regression maximizing a misspecified likelihood, when both an outcome variable and an endogenous intervention variable are dichotomous [10,11]. One solution is a bivariate probit (BVP) model. Another solution is to use a two-stage least squares (2SLS) model that treats these dichotomous dependent variables and an endogenous covariate as continuous variables and runs ordinary least squares (OLS) models, instead of probit models, for the first and the second-stages to obtain consistent estimates [11].

In general, BVP models have received relatively less attention in the clinical effectiveness evaluation literature, but have been used by studies focusing on the association between health condition and employment [12] and that between health insurance choice and health-care utilization [13].

The purpose of our article is to illustrate in detail how to evaluate medical intervention effectiveness with observational data, adjusting for self-selection of the intervention by the IV method when both an intervention and outcome measures are dichotomous. Among subtypes of the IV methods, we employed BVP and 2SLS analyses as our main models that can be easily implemented by statistical software such as STATA Version 7 or later [14]. As a potentially endogenous intervention example, influenza vaccination (flu shot, FS) among Medicare elderly population was used for two reasons. One is the substantial impact of influenza epidemics as illustrated above. The second is that findings from our IV method analyses with nationally representative data are expected to yield significant contributions to the influenza vaccination literature in terms of improving generalizability, compared with past RCT analyses. To the best of our knowledge, our study is the first to evaluate the effectiveness of any type of vaccination adjusting for self-selection by the IV method.

Concretely, we test two major hypotheses concerning influenza vaccination effectiveness estimated by BVP and 2SLS models:

**Hypothesis 1**

A single-equation probit model will underestimate the vaccine effectiveness in terms of magnitude and statis-
tical significance level compared with BVP and 2SLS models adjusting for self-selection of influenza vaccination. This is because individuals at higher risk are more likely to receive shots among the US Medicare elderly population. According to self-report in Medicare Current Beneficiary Survey (MCBS) data, the FS rate has steadily increased from 1991–92 season (49%) to 1999–2000 season (68%) [15]. FS rate for 1996–97 seasons for the entire Medicare elderly population was 62% in our data set.

**Hypothesis 2**
The qualitative differences in FS effectiveness estimates in hypothesis 1, depending on the adjustment for self-selection, are robust to sensitivity analyses in varying outcome periods, the scope of outcomes, and sources of data on FS receipt, that is, either claims data or survey data. In addition, the results of the sensitivity analyses in hypothesis 2 will be predictable. FS effectiveness estimates will be greater in magnitude when using a shorter outcome period and when using survey data. Flu epidemic was reported to occur for different periods and at different severity levels depending on a state, from October to March [16]. Because a larger number of states experienced flu epidemic during a shorter outcome period, FS effectiveness will be greater in magnitude using a shorter period than a longer period. Because some FSs received outside Medicare billing system were not included in claims data, FS effectiveness will be underestimated with claims data.

**Data and Methods**

**MCBS Data and Study Population**
The MCBS is a longitudinal survey conducted by the Centers for Medicare and Medicaid Services [17]. This study used MCBS Cost and Use data in calendar years 1996 and 1997, focusing on the influenza season starting in fall 1996 and ending in spring 1997. The study population was defined to be individuals aged 65 years or older whom were continuously covered by Medicare Part B from September 1, 1996 to March 31, 1997, including those who were alive on September 1, 1996 but died during the study time period. Medicare Health Maintenance Organization (HMO) enrollees were not included because Medicare claims with the exact dates of FSs were not available for Medicare HMO enrollees [17]. Furthermore, this study used only the community population following most of past literature, where FS effectiveness is explored in either the community population or the facility population because of the differences in individual characteristics, particularly health conditions [18,19]. To yield results generalizable for the entire Medicare population with the observed number of the study subjects (n = 4338), two sample structure variables were used: a cluster identifier variable and a sampling probability weight variable.

**Dependent Variables and Major Explanatory Variable**
Variables used in these analyses are defined and summarized in Table 1. Medicare inpatient claims data on primary diagnosis and admission diagnosis between October 1, 1996, and March 31, 1997, were used to identify hospitalizations resulting from influenza-related diseases, because Centers for Disease Control and Prevention (CDC) reported that the influenza season lasted from October 1996 to March 1997 in most states [17]. To account for the influenza season varying across geographic regions, we examined three additional outcome periods, starting November 1, December 1 and January 1.

On the basis of previous literature this study uses two definitions concerning the diagnoses of influenza and its complications [2,20–22]. The narrow definition included pneumonia and influenza (ICD-9-CM codes 480–487) only. The broader definition additionally included acute bronchitis (ICD-9-CM code 466).

<table>
<thead>
<tr>
<th>Variable definition</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>77.21 (7.10)</td>
</tr>
<tr>
<td>Ever smoked (d.v.)</td>
<td>0.560</td>
</tr>
<tr>
<td>Education: 13 years or more (d.v.)</td>
<td>0.264</td>
</tr>
<tr>
<td>Medigap supplemental insurance (d.v.)</td>
<td>0.752</td>
</tr>
<tr>
<td>Household member: 2 or more (d.v.)</td>
<td>0.646</td>
</tr>
<tr>
<td>Race: nonwhite (d.v.)</td>
<td>0.350</td>
</tr>
<tr>
<td>Chronic condition: high level (d.v.)</td>
<td>0.199</td>
</tr>
<tr>
<td>Chronic condition: low level (d.v.)</td>
<td>0.726</td>
</tr>
<tr>
<td>Ever received pneumococcal vaccination (d.v.)</td>
<td>0.475</td>
</tr>
<tr>
<td>Ever smoked (d.v.)</td>
<td>0.199</td>
</tr>
<tr>
<td>Death due to all causes</td>
<td>0.0106</td>
</tr>
<tr>
<td>Residence in metropolitan area (d.v.)</td>
<td>0.663</td>
</tr>
<tr>
<td>Arthritis (secondary diagnosis) or gout (d.v.)</td>
<td>0.0763</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospitalization due to pneumonia and influenza, or acute bronchitis</td>
<td>0.0106</td>
</tr>
<tr>
<td>Death due to all causes</td>
<td>0.0106</td>
</tr>
</tbody>
</table>

Table 1: Variable definitions and sample means (n = 4338)

**d.v., dichotomous variable.**
Effectiveness of Influenza Vaccination

MCBS survey data were used to identify cases of death from all causes that occurred between October 1, 1996 and March 31, 1997, because claims data do not include all death cases.

FS variables were created from either claims data or survey data. Based on the claims data, 50.2% of the study population received FSs, from September 1, 1996 to March 31, 1997 (Table 1). Survey data indicated that 68.8% received an FS during 1996–97 seasons.

Nichol and colleagues define three levels of influenza risk depending on chronic conditions: high-risk (those with heart or lung disease), intermediate-risk (those with diabetes, renal disease, rheumatologic disease or dementia and/or stroke but not underlying heart or lung disease), and low-risk (those without any of these diseases or conditions) [11]. We created indicator variables associated with these three levels. In addition to these indicators of chronic condition levels that were expected to reflect actual influenza risk more systematically than a list of specific diseases, ever smoking status and age were included in analyses and expected to be associated with a higher likelihood of adverse influenza-related outcomes.

Analytical Methods

As a preliminary analysis, a single-equation probit model was used to estimate associations between the influenza-related adverse health outcomes (e.g., influenza-related hospitalization and death due to all causes) and the receipt of an FS.

Our main analyses adopted BVP and 2SLS models. The fundamental goal shared by 2SLS and BVP is the same as using an additional exogenous “instrumental” variable to isolate a part of a treatment variable that is uncorrelated with the error term. A BVP model provides not only consistent but also efficient estimates, being dependent on the assumptions of the joint distribution of error terms for the dependent variable and the endogenous covariate [11]. In contrast, 2SLS models are robust to this assumption of bivariate normal distribution. One of our major interests, measuring the magnitude of the FS effectiveness, is expected to be more precisely estimated by BVP models than 2SLS models. The BVP model estimates Equation 1 and 2 simultaneously, treating both FS and outcome in these two equations as dichotomous [11,23]. IV1 represents an IV unique to Equation 1, and, in Equations 1 and 2, X1 and X2 are other exogenous controlling covariates. Our IV, having either arthritis as a secondary diagnosis or gout, was chronic diseases that require a regular clinic visit, at which an FS could be suggested and given, but are uncorrelated with influenza risk [21,24].

First-stage equation:

\[ y_1^* = \alpha_0 + \alpha_1 (IV_1) + \alpha_2 (X_1) + \varepsilon_1 \] (1)

Second-stage equation:

\[ y_2^* = \beta_0 + \beta_1 (FS) + \beta_2 (X_1) + \beta_3 (X_2) + \varepsilon_2 \] (2)

Following the expression by McClellan and colleagues [6], a BVP/2SLS estimator measures the ratio (hospitalization rate among those with arthritis – hospitalization rate among those without arthritis)/ (vaccination rate among those with arthritis – vaccination rate among those without arthritis) in “marginal” individuals. This ratio indicates the incremental or marginal effect of vaccination over the change in likelihood in vaccination across two groups, those with and without arthritis. In case of an RCT estimator, this ratio’s denominator is equal to one, that is, vaccination status between a treatment group and a control group. The vaccination effects based on BVP/2SLS estimators are applicable for “marginal” individuals for whom arthritis is an important factor in the vaccination decision.

Instrumental Variable Assumptions

Our IV passed the two applicable assumption tests for valid IVs as below. First, IVs should be correlated with an endogenous variable [10]. Concretely, in a first-stage Equation 1 where the FS receipt is an outcome variable, a standard t-test indicated that an IV’s coefficient, for example, \( \bar{\varepsilon}_1 \), was significant at the 5% level after controlling for other exogenous covariates when arthritis and gout were combined to one variable. That is, when these two conditions are included as two separate IVs, the assumption that these two IVs are jointly significant was rejected. Second, IVs should also have enough “explanatory power” for an endogenous variable after controlling for other exogenous covariates in the first-stage Equation 1. Explanatory power is tested by an F-statistic measuring the increase in \( R^2 \), when an IV is incrementally added as an explanatory variable, in case of a linear model. Because there is no comparable test in limited dependent variable models, we tested this assumption by mimicking the linear model by using pseudo-\( R^2 \). Standard F-test using pseudo-\( R^2 \)-squared indicated significant explanatory power at the 5% level when arthritis and gout were combined to one variable, not included as two separate IVs, in BVP analyses. For 2SLS analyses, the combined IV did not pass these relevancy tests. A general disadvantage of the IV method is the difficulty in finding a unique set of IVs.
for each endogenous variable. To use a less appropriate IV, for example, weakly correlated with an endogenous variable, could yield biased estimates [25]. Although applicable but not testable in our case, the third assumption for valid IVs is that IVs should be uncorrelated with the error term in the second-stage Equation 2 where a health outcome is a dependent variable. This exogeneity assumption cannot be directly tested when a model is exactly identified, that is, the number of IVs is equal to that of endogenous variables, as in our models.

All statistical analyses were performed using statistical software STATA Version 7 [14].

Results

The estimates of vaccine effectiveness were summarized in Table 2: those estimated by single-equation probit models (columns 1 and 2) and by BVP models (columns 3 and 4) where FS was measured by claims data (columns 1 and 3) and survey data (columns 2 and 4). These vaccine estimates were imputed based on the marginal effects of the probit models exploring associations between the influenza-related adverse health outcomes and the receipt of FS summarized in Table 3. In Table 3, column 1 presents the marginal effects estimated by single-equation probit models. Columns 2 and 3 indicate the marginal effects estimated by a BVP model and a correlation coefficient in a BVP model, respectively.

The imputation of FS effectiveness based on the FS marginal effects was exemplified as follows. The first row and second column in Table 3 shows that the FS’s marginal effect was 0.69% when outcome is hospitalization due to pneumonia and influenza and estimated by a BVP model. Because the observed probability of hospitalization among the unvaccinated individuals was 1.62%, FS was approximately 43% effective, marginal effect (0.69) divided by the hospitalization rate among the unvaccinated (1.62), in preventing hospitalization due to pneumonia and influenza between October and March.

All estimates controlled for three levels of chronic diseases detailed in the methods section, ever receiving the pneumococcal vaccination, ever smoking status, age, sex, race, educational attainment, presence of supplemental health insurance, the number in the household, metropolitan area, and nine census regions.

Hypothesis 1, the single-equation model underestimates FS effectiveness, was strongly supported by our empirical results. For instance, when an individual receives an FS, the individuals’ probability of hospitalization from December to March due to the narrow definition of influenza-related diagnoses decreased by 41% compared with those who missed an FS at a statistically significant level ($P < 0.01$) in a BVP model. In a single-equation probit model, FS effectiveness was 2%, which was statistically not different from zero (Table 2, row 3, columns 2 and 4).

The estimates of improved FS effectiveness, hypothesis 2, were partly supported by our empirical sensitivity analyses. Such improvement in estimates was robust to the changes in outcome periods, sources of data on FS receipt, and relatively robust to the scope of outcomes (Table 2). FS effectiveness tended to be greater in magnitude and statistical significance level when the outcome scope was defined more narrowly. Such obvious trend was not observed in sensitivity analyses.

### Table 2: Effectiveness of influenza vaccination adjusting for self-selection of vaccination (n = 4338)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Period</th>
<th>Single-equation probit model (%)</th>
<th>Bivariate probit model (adjusting for self-selection of vaccination) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Claims data*</td>
<td>Survey data*</td>
</tr>
<tr>
<td>Hospitalization</td>
<td>10/01/1996–3/31/1997</td>
<td>14.44</td>
<td>−17.56$^{13}$</td>
</tr>
<tr>
<td>due to pneumonia and influenza</td>
<td>11/01/1996–3/31/1997</td>
<td>17.09$^{5}$</td>
<td>−6.58$^{3}$</td>
</tr>
<tr>
<td>(P &amp; I)</td>
<td></td>
<td>12/01/1996–3/31/1997</td>
<td>16.41$^{7}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1/01/1997–3/31/1997</td>
<td>1.27</td>
</tr>
<tr>
<td>Hospitalization</td>
<td>10/01/1996–3/31/1997</td>
<td>9.60</td>
<td>−17.86$^{13}$</td>
</tr>
<tr>
<td>due to P &amp; I, or acute bronchitis</td>
<td>11/01/1996–3/31/1997</td>
<td>11.98</td>
<td>−7.96$^{1}$</td>
</tr>
<tr>
<td>12/01/1996–3/31/1997</td>
<td>12.84</td>
<td>0.30</td>
<td>−3.57$^{1}$</td>
</tr>
<tr>
<td>1/01/1997–3/31/1997</td>
<td>−6.35$^{1}$</td>
<td>−3.57$^{1}$</td>
<td>11.83</td>
</tr>
<tr>
<td>Death due to all causes</td>
<td>10/01/1996–3/31/1997</td>
<td>−19.56$^{11}$</td>
<td>−11.26$^{11}$</td>
</tr>
<tr>
<td>12/01/1996–3/31/1997</td>
<td>−19.56$^{11}$</td>
<td>−11.26$^{11}$</td>
<td>10.18</td>
</tr>
<tr>
<td>1/01/1997–3/31/1997</td>
<td>−19.56$^{11}$</td>
<td>−11.26$^{11}$</td>
<td>10.18</td>
</tr>
</tbody>
</table>

*Data source of vaccination status.
$^{1}$Significant at 5%.
$^{2}$Negative values indicate the increased risk of outcome.
$^{3}$Significant at 1% (based on t-statistic in Table 3).
$^{4}$Significant at 0.1%.

All models controlled for age, age-squared, sex, race, education, supplemental insurance, number of household, metropolitan residence, ever smoking, ever receiving pneumococcal vaccination, nine census regions and three levels of chronic conditions relating to flu risks (heart disease, lung disease, diabetes, renal disease, rheumatologic disease, dementia and stroke).
an analyses changing the sources of data on FS receipt and outcome periods.

Self-selection of FS due to unobservable factors suggesting high risk was implied by the estimated correlation coefficient (p) and the explanatory variables estimates in BVP models. That is, the estimated p was positive and statistically significant (p < 0.05) in most models presented in column 3 (Table 3). Both BVP models and 2SLS models (not presented in Tables) changed signs, magnitudes, and efficiency of some estimates, particularly FSs compared with single-equation probit models (Table 3). FS estimates in 2SLS models were statistically significant only when an outcome is hospitalization due to pneumonia and influenza from January to March, and implausibly large in magnitude, being greater than 200%.

Discussion

Although high-risk individuals could be either more or less likely to self-select a medically based intervention, only the former type of self-selection was apparent in all of our analyses and hence the FS marginal effect estimates were biased in the direction of less effectiveness. When the BVP and 2SLS models adjust for the potential endogeneity problem due to self-selected FSs, the marginal effects of FSs changed to the expected direction, being negative in all models (Table 3, columns 2 and 3). FS effectiveness, based on these negative marginal effects, was statistically significant in most BVP models with hospitalization due to pneumonia and influenza, and some 2SLS models with a narrowly defined adverse health outcome. These significant FS effectiveness estimates in BVP and 2SLS models were consistent with the CDC's report that delivered FSs contained a good antigen match throughout the 1996–97 influenza season nation-wide [16].

Possible unobservable factors motivating FS self-selection are preference for FS and unmeasured health status. These possible unobservable factors were implied by other empirical results. Namely, individuals with FSs were more likely to have a past season's FS and an influenza-related chronic condition than those who missed FSs at a statistically significant level (P < 0.001).

The major purpose of conducting sensitivity analyses in our study was to test the robustness of self-selection bias in terms of its direction and statistical significance, not to make a strict comparison with past studies by setting a common outcome. This was because it was difficult to make other factors comparable to past studies, such as the differences in study population characteristics, unmeasured heterogeneity between the vaccinated and the unvaccinated, degrees of vaccine-antigen match, and laboratory confirmation of influenza.

Regarding the magnitude of FS effectiveness among the community-dwelling elderly, previous RCTs reported that FS was effective in reducing clinical influenza by 47–58% and all-cause mortality by 14% [4,5]. In a meta-analysis, FS effectiveness was 33% for hospitalization due to pneumonia and influenza, and 50% for mortality due to all causes [26]. Large cohort
studies including community elderly in Europe reported that FS effectiveness was 21–44% in reducing hospitalization due to influenza and its related respiratory diseases [22,27] and 10–28% in reducing all-cause mortality [28]. None of these cohort studies employed IV methods, 2SLS models or BVP models, to adjust for potential self-selection of FSs. Our FS effectiveness estimates in preventing influenza-related hospitalization were comparable in magnitude and statistical significance level to previously published results. Our insignificant estimates in reducing all-cause mortality do not contradict an observational study using the US nationally representative data from 1968 to 2001, reporting fewer than 10% of all winter deaths were attributable to influenza in any flu season [29].

The FS effectiveness estimates in the literature may have changed if they had employed BVP models to adjust for potential self-selection of FS. For instance, because a higher FS rate among high-risk individuals was observed in the Swedish study like our study [27,30], BVP models’ estimates of FS effectiveness might have increased in the statistical significance level and magnitude in their study. In contrast, a low FS rate among high-risk elderly people was reported in the study in England and Wales [22] where FS effectiveness magnitude might have declined after adjusting self-selection by BVP models.

In our analyses, FS rates based on claims and survey data were 50.2% and 68.8%, respectively. It is hard to judge the extent of measurement error in each data source, in part because there was no literature directly examining the matching rates applicable to our study. Although 10% to 20% of Medicare beneficiaries are estimated to receive FSs outside the Medicare billing system [31], close validity examinations were conducted with local populations only [32,33]. Because claims data are unable to capture the benefit of FS in reducing adverse outcomes among those actually received FS but did not have a claims record, the effectiveness of FS based on claims data could be underestimated.

The adjustment of self-selection for non-HMO enrollees would be valid and generalizable for HMO enrollees as well, if high-risk individuals are more likely to receive FSs among HMO enrollees. Some studies indicated that Medicare HMO enrollees on average are healthier [34] and more likely to get FSs [35]. Thus, healthy individuals in HMOs are more likely to get FSs than healthy individuals in the general population. Our results find that we are underestimating the effectiveness of FSs in the non-HMO population. If we added HMO enrollees, we would have a more balanced mixed of individuals getting FSs based on health status. That is, the estimate including HMO enrollees would seem to be less biased toward ineffectiveness.

Although not presented in this article, most predictors of influenza-related adverse health outcomes conformed with the literature; an exacerbated chronic condition level, advancement in age, and smoking history were positively associated with adverse outcomes [21,36]. Such conformity reinforces the validity of this study’s empirical results regarding FS effectiveness.

Our study is expected to make two contributions to the literature. One is the validity of the IV method for adjusting substantial self-selection influence in evaluating FS effectiveness. BVP models and the example of IVs employed in our analyses are expected to be useful in evaluating other types of self-selected medical interventions where both an outcome and the intervention are treated as dichotomous variables. Furthermore, future studies are expected to control for additional self-selection problems of pneumococcal vaccination and smoking history through employing multivariate probit models.

The second contribution is improved generalizability in FS effectiveness based on the US nationally representative study population. For instance, FS was evaluated to be effective during the 1996–97 season after controlling for nine census regions, although these regions significantly differed (P < 0.01) in key individual factors such as FS rates, previous season’s FS rates, pneumococcal vaccination rates, chronic condition levels and subjective general health status levels, in addition to environmental factors like climate and influenza epidemic levels. Future studies using additional influenza seasons are expected to improve the generalizability of estimates of FS effectiveness even further. Also, effectiveness evaluation studies focusing on specific subpopulations at different levels of risks could contribute to the appropriate allocation of vaccines to maximize the health benefits of a vaccination program particularly when vaccine supply is limited or delayed as observed in the United States during 2004–05 and recent influenza seasons [37,38].

Conclusion

Adjusting for self-selection, BVP analyses yielded vaccine effectiveness estimates for a nationally representative cross-sectional sample of the community-dwelling elderly population that are consistent with previous estimates based on randomized controlled trials, prospective cohort studies, and meta-analyses. This result suggests that IV methods in general and analyses with 2SLS and BVP in particular may be useful for the analysis of observational data regarding prevention in which self-selection is an important potential source of bias.

Authors are grateful to Robert Moffitt, David Salkever, Karen Bandeen-Roche, David Bishai, Eric Slade, and Judy Kasper for their support as dissertation committee members.
Effectiveness of Influenza Vaccination

We also acknowledge Jay Bhattacharya, Marshall McBean, and Scott Grosse for their useful comments.

Source of financial support: This research was supported by a grant from the United States Department of Health and Human Services, Centers for Medicare and Medicaid (30-P-91295/3-01). Authors do not have any conflict of interest to declare.

References

30. Christenson B, Lundbergh P. Comparison between cohorts vaccinated and unvaccinated against influ-


