American Journal of Infection Control ■■ (2017) ■■-■■

Contents lists available at ScienceDirect

American Journal of Infection Control

journal homepage: www.ajicjournal.org

Major Article

Controlling for endogeneity in attributable costs of vancomycin-resistant enterococci from a Canadian hospital

Patrick Lloyd-Smith PhD *

Department of Resource Economics and Environmental Sociology, University of Alberta, Edmonton, AB, Canada

Key Words:
Vancomycin resistance
hospital-acquired infections
cost analysis
endogeneity
nonlinearity

Background: Decisions regarding the optimal provision of infection prevention and control resources depend on accurate estimates of the attributable costs of health care–associated infections. This is challenging given the skewed nature of health care cost data and the endogeneity of health care–associated infections. The objective of this study is to determine the hospital costs attributable to vancomycin-resistant enterococci (VRE) while accounting for endogeneity.

Methods: This study builds on an attributable cost model conducted by a retrospective cohort study including 1,292 patients admitted to an urban hospital in Vancouver, Canada. Attributable hospital costs were estimated with multivariate generalized linear models (GLMs). To account for endogeneity, a control function approach was used.

Results: The analysis sample included 217 patients with health care–associated VRE. In the standard GLM, the costs attributable to VRE are $17,949 (SEM, $2,993). However, accounting for endogeneity, the attributable costs were estimated to range from $14,706 (SEM, $7,612) to $42,101 (SEM, $15,533). Across all model specifications, attributable costs are 76% higher on average when controlling for endogeneity.

Conclusions: VRE was independently associated with increased hospital costs, and controlling for endogeneity lead to higher attributable cost estimates.

To properly evaluate infection prevention and control (IPAC) programs in hospitals, information on the attributable costs of health care–associated infections is needed. One approach to estimating attributable costs is through the use of statistical models. Recent statistical model applications have focused on the skewed distribution of health cost data, and consequently used nonlinear models in estimation.[1] However, if the infection variable is endogenous in the attributable cost analyses, this may lead to biased and inconsistent cost estimates.[2] Furthermore, recent evidence suggests that nonlinearity and endogeneity need to be addressed at the same time.[3,4]

An endogenous variable is an explanatory variable that is correlated with the model's error term.[3] In attributable cost of infection, endogeneity problems can arise if the infection variable is associated with any unobserved or uncontrolled factors that also affect costs or is measured with error. For example, the 2-way relationship between infections and length of stay (LOS) can lead to potential endogeneity problems.[1,2] Although an infection increases the LOS of a patient, therefore influencing total costs, an increased LOS is associated with an increased risk of becoming infected. Therefore, the estimated cost of infections is biased if LOS is not controlled for in the model. However, LOS cannot be included in the model because a primary channel in which infections increase patient costs is by increasing LOS. Furthermore, other relationships between infections and other unobserved variables, such as prior health problems that are not included in models, may bias results. These issues require novel approaches to correct for this bias.

The instrumental variable approach has been identified as a promising strategy for controlling for the endogeneity of infections.[1,5] Only 1 study, Graves et al,[2] has attempted to correct for this bias by applying the 2-stage least squares approach in a linear setting. However, as previously noted, health care cost data are more robustly analyzed using nonlinear models.[3,4] The main objective of the present study is to estimate the attributable costs of vancomycin-resistant enterococci (VRE) while controlling for endogeneity concerns in a nonlinear model. A second objective is to compare these cost estimates with a previous study that did not control for endogeneity to quantify the direction and magnitude of the endogeneity bias.

* Address correspondence to Patrick Lloyd-Smith, MA, Department of Resource Economics and Environmental Sociology, University of Alberta, 515 General Services Building, Edmonton, AB T6G 2H1, Canada.

E-mail address: lloydsmi@ualberta.ca.

Funding/support: Supported by Infection Prevention and Control, Providence Health Care.

Conflicts of interest: None to report.

https://doi.org/10.1016/j.ajic.2017.08.044

METHODS

Study design

The data used in this study were obtained from databases maintained by the Finance Department and the IPAC Department at Providence Health Care. The study population includes patients admitted to an urban hospital (St. Paul's Hospital) in Vancouver, Canada, from April 1, 2008-March 31, 2009. The sample included laboratory-confirmed cases with VRE colonization or infection (n = 219) and a random sample of 1,166 control patients that were admitted to St. Paul's Hospital. Patient characteristics are provided in Lloyd-Smith et al.[6] This study was approved by the University of British Columbia and Providence Health Care Ethics Boards.

Generalized linear model

The cost equation is modeled with the generalized linear model (GLM) and a log link using a gamma distribution based on the results of a modified Park test. This specification is a common choice in attributable cost analyses of health care data.[7-9] Variable selection was conducted using a backward selection process. Variables were excluded one by one based on their statistical significance and the Akaike information criterion. A Wald test of the joint significance of all the dropped variables revealed that it is appropriate to exclude these variables. To examine cost-savings, the average treatment effects on the treated were calculated using the model coefficients and the mean costs of patients with VRE.

Control function approach

The control function approach uses a 2 equation setup, treatment, and cost. First, the treatment equation specifies the probability of acquiring an infection as a function of observed covariates and instrumental variables.[10] To reflect the binary nature of the VRE variable, the first-stage analysis uses a GLM model with a binomial distribution and probit link. Second, the cost equation models hospital costs as a function of observed covariates, the infection variable, and residuals from the first equation estimation. A GLM with a log link and a gamma distribution is used for the cost equation.

Statistical theory only demonstrates that some function of the residuals is the appropriate control function, not the specific type of residual nor its functional form.[10] To assess the robustness of the results, I consider 4 different types of residuals (response, Pearson, Anscombe, and deviance).[11] To account for a flexible functional form, I estimate models including residuals and second-order polynomials of the residuals.[4] Nonparametric bootstrap replication is used to calculate estimates of empirical SEMs to account for constructed variables in the second equation.[10]

Instrumental variables

Specific hospital procedures performed on patients that are correlated with VRE and uncorrelated with LOS are potential instruments. The initial instrument is whether or not a patient had a nasogastric feeding tube inserted (NGF) and was used by Graves et al[2] in their study of costs of respiratory tract infections and is a commonly identified risk factor for VRE and similar drug-resistant organisms.[12-17] Two other potential instruments I consider are whether the patient received parenteral nutrition via the percutaneous infusion approach (PNA), and whether the patient received a vascular access device (VAD), which have both been found to be risk factors with VRE in the medical literature.[12,18] The prevalence of the 3 procedures among the study patients is 11% for NGF, 3% for PNA, and 19% for VAD. The study patients received the procedures during the specific hospital stay that is the focus of this analysis. For each combinations of instruments, I conduct 3 different instrumental variable diagnostic tests: the Durban-Wu-Hausman test (endogeneity), a first-stage F test (instrument strength), and an overidentification test (an indirect test of instrument validity).[19] Results of these formal instrument tests show that the instrumental variables are valid.

RESULTS

The mean cost per day ± SD was $13,069 ± $17,783 for the control patients and $46,924 ± $55,881 for the patients with VRE (all figures are in 2009 Canadian dollars, which are equal to 0.88 2009 U.S. dollars). Across the whole sample, the mean cost was $18,755 ± $30,755, with a median of $8,574. These differences between mean and median costs suggest the costs are severely skewed, which is a common finding in health care cost data. The attributable costs of VRE are calculated using the model's VRE coefficient and the mean cost of a patient with VRE as the reference group. Using a GLM specification, the attributable costs of VRE without addressing endogeneity concerns are estimated to be $17,949 (SEM, $2,993) (data not shown).

Table 1 summarizes the attributable costs of VRE using the GLM setup with corrections for endogeneity. Results are shown for the 2 instrument specifications that represent the lower and upper bound of the cost estimates. Using the NGF instrument, the attributable costs range from a low of $14,706 (SEM, $7,612) using first-degree Pearson residuals to a high of $39,435 (SEM, $23,199) using second-degree deviance residuals. The models using all 3 instruments (NGF + VAD + PNA) report attributable costs ranging from $17,626 (SEM, $7,217) to $42,101 (SEM, $15,533). The simple average attributable costs of VRE across all instrument and residual specifications for models that address endogeneity is estimated to be $31,764 (data not shown).

Table 1
Attributable costs of vancomycin-resistant enterococci using generalized linear models while controlling for endogeneity

	Instrument specification			
	NGF		NGF + VAD + PNA	
Control function specification	Mean	SEM*	Mean	SEM*
Response residuals				
First degree	$29,782	$15,752	$36,506	$12,023
Second degree	$29,902	$15,689	$36,391	$11,953
Anscombe residuals				
First degree	$24,243	$17,361	$33,099	$13,324
Second degree	$36,804	$21,351	$40,662	$14,447
Pearson residuals				
First degree	$14,706	$7,612	$17,626	$7,217
Second degree	$26,299	$13,758	$32,974	$11,495
Deviance residuals				
First degree	$32,635	$21,197	$39,129	$14,609
Second degree	$39,435	$23,199	$42,101	$15,533

NOTE. Values are in Canadian dollars. The instrument specifications are different procedures conducted on the patients. These results are calculated using the model's vancomycin-resistant enterococci coefficient and the average cost of a patient with vancomycin-resistant enterococci in the sample. The exact calculation is $(1 - \exp[-\beta_{VRE}]) \times \$46{,}924$ where β_{VRE} is the coefficient for the vancomycin-resistant enterococci variable. All variables included in table 1 of Lloyd-Smith et al[6] are included as explanatory variables in the attributable cost models except for death, length of stay, and male sex. The age variable is recoded as a dummy variable with all individuals >75 years of age classified as one.

NGF, nasogastric feeding tube inserted; *PNA*, parenteral nutrition via the percutaneous infusion approach; *VAD*, vascular access device.

*The SEMs for these models were calculated using the bootstrap method and 400 draws.