A Simulation-Based Evaluation of Methods to Estimate the Impact of an Adverse Event on Hospital Length of Stay

Matthew H. Samore, MD,*†‡ Shuying Shen, MS,*†‡ Tom Greene, PhD,*†‡ Greg Stoddard, MPH,*†‡
Brian Sauer, PhD,*†‡ Judith Shinogle, PhD,¶ Jonathan Nebeker, MD, MS,*†‡
and Stephan Harbarth, MD, MS§

Introduction: We used agent-based simulation to examine the problem of time-varying confounding when estimating the effect of an adverse event on hospital length of stay. Conventional analytic methods were compared with inverse probability weighting (IPW).

Methods: A cohort of hospitalized patients, at risk for experiencing an adverse event, was simulated. Synthetic individuals were assigned a severity of illness score on admission. The score varied during hospitalization according to an autoregressive equation. A linear relationship between severity of illness and the logarithm of the discharge rate was assumed. Depending on the model conditions, adverse event status was influenced by prior severity of illness and, in turn, influenced subsequent severity. Conditions were varied to represent different levels of confounding and categories of effect. The simulation output was analyzed by Cox proportional hazards regression and by a weighted regression analysis, using the method of IPW. The magnitude of bias was calculated for each method of analysis.

Results: Estimates of the population causal hazard ratio based on IPW were consistently unbiased across a range of conditions. In contrast, hazard ratio estimates generated by Cox proportional hazards regression demonstrated substantial bias when severity of illness was both a time-varying confounder and intermediate variable. The direction and magnitude of bias depended on how severity of illness was incorporated into the Cox regression model.

Conclusions: In this simulation study, IPW exhibited less bias than conventional regression methods when used to analyze the impact of adverse event status on hospital length of stay.

A central challenge for epidemiology and health services research is to address confounding as effectively as possible.[1–3] Whenever the causal question of interest concerns the effect of a time-varying exposure or treatment, the potential for time-varying confounding arises. Time-varying confounding is produced when outcome and exposure are influenced by the updated values of a third variable that changes over time.[4] A common setting where time-varying confounding is likely is when disease severity influences the decision to initiate drug therapy.[5]

Robins and colleagues[5–8] developed a method for estimating the causal effect of a time-varying exposure in the context of an underlying marginal structural model of the counterfactual outcomes. They applied a regression technique, called inverse probability weighting (IPW), to the problem of time-varying confounding. Another term for this method is inverse probability of treatment weighted estimation. The situation where time-varying confounders are also intermediate variables produces a dilemma.[6,9–11] Failure to adjust for a time-varying confounder leaves residual confounding and may exaggerate an effect estimate. If the time-varying confounder is also an intermediate variable at the same time, its inclusion in the analysis as a covariate can produce bias toward the null.

Experience with IPW in epidemiology and health services research has been relatively limited to date, particularly in comparison to propensity scores, another method for removing confounding due to measured variables.[12–17] Few studies have used simulation to systematically compare IPW to conventional regression methods when time varying confounding is present.[9,11,18] The application of IPW to outcomes research in hospital settings has not been explored. In this study, we examine the use of IPW in the context of a significant problem in patient safety and infection control, the estimation of the impact of an adverse event on cost, and hospital length of stay.

Many studies have addressed outcomes and costs associated with adverse events in hospitalized patients, applying a

variety of study designs and analytic methods.[19–29] Fluctuating severity of illness is recognized as a likely time-varying confounder in this situation.[30,31] Because severity of illness may also mediate the effect of the adverse event on subsequent length of stay, it is also likely to be an intermediate variable. To characterize the bias of conventional regression methods in comparison to IPW when applied to this problem, we developed a simulation model of the occurrence of adverse events among hospitalized patients. We sought to define the conditions under which conventional regression methods may yield biased results and to illustrate the use of IPW in removing this bias.[32–34]

METHODS

Agent-Based Simulation

We used a form of simulation called agent-based modeling, a modeling technique in which individual entities are assigned internal states. Rules (or behaviors) are encoded to govern the transitions between states over time. Thus, individuals are represented directly and populations are established as groups of individuals. The advantages of this technique for this study are its flexibility and its ability to create completely synthetic data sets suitable for statistical analysis.

The individual agent model was constructed in Anylogic 5.4 (XJ Technologies, Saint-Petersburg, Russia) using Java-based code to create replicated objects. A cohort of patients admitted to a 500-bed hospital during a 1-month period was simulated. During the simulated hospital stay, individuals transitioned between 3 states: state 1—hospitalized, with no adverse event experienced; state 2—hospitalized, with an adverse event experienced; state 3—discharged. All patients were in state 1 at the beginning of observation (time of admission). During the hospital stay, individuals in state 1 were eligible to transition to state 2 but not vice versa. That is, once the adverse event was experienced, the adverse event status variable was set to the value of 1 for the remainder of the hospitalization. Simulated patients moved from either of the hospitalized states to the discharged state. Death was combined into the discharge state in the simulations presented here to avoid problems related to competing risks. Individuals were removed from observation after entering the discharge state.

Each run of the agent-based model generated a synthetic data set. At intervals of 0.25 time units, each simulated individual's updated values for adverse event status, severity of illness, and discharge status, were appended to a table. The simulation was calibrated for 1 time unit to indicate 1 day. Thus, 0.25 time units represented 6 hours.

The discharge rate was modeled as a hazard function, using the following equation:

$$\lambda(t) = \lambda_0(t) \exp\{\beta_1 V(t) + \beta_2 A(t)\}$$

where $\lambda_0(t)$ represented the baseline hazard rate of discharge as a cubic spline function of time since admission, $A(t)$ was adverse event status at time t, $V(t)$ was the severity of illness score at time t, β_1 represented the individual β coefficient for the effect of severity of illness on log hazard rate of dis-

TABLE 1. The Parameter Values Used in the Simulation Are Listed (See Methods for Details)

	Values	Model Conditions
Discharge rate equation		
Coefficient for the effect of a unit increase in severity of illness on logarithm of the discharge hazard rate, β_1	−1	All
Coefficient for direct effect of adverse event on logarithm of the discharge hazard rate, β_2	0	No effect
	0	Indirect effect only
	−0.693	Dual effect
Severity of illness equation		
Standard deviation of baseline severity, s	1	All
First order autoregressive coefficient, c	0.5	All
Ratio of second to first order autoregressive coefficient, f	0.67	All
Standard deviation of random error term, z	1	All
Coefficient for effect of adverse event status on severity, β_3	0	No effect
	0.693	Indirect effect only
	0.693	Dual effect
Adverse event equation		
Intercept, α	−4.5	All
Factor that influences adverse event, $V'(t)$*	Independent score	No confounding
	Baseline severity	Baseline confounding
	Lagged severity	Time-varying confounding
Coefficient for effect of $V'(t)$ on logarithm of the odds of occurrence of adverse event, β_4	1	All

*See text for details.

charge, and β_2 equaled the beta coefficient for the direct effect of adverse event status on log hazard rate of discharge. The values for $V(t)$ and $A(t)$ were updated at intervals of 0.25 days. Modeling the baseline discharge hazard rate with a cubic spline function allowed more flexibility than alternative parametric forms. The coefficient β_1 was assigned a value of −1.0 (Table 1). Therefore, for each unit increase in severity of illness score, an individual's discharge hazard rate decreased by 63%. When β_2 was set to 0, adverse event status did not have a direct effect on the discharge hazard rate.

Synthetic patients were assumed to vary with respect to severity of illness, the consequence of which was heterogeneity across individuals in the hazard rates of discharge. The value of the severity of illness score depended on baseline severity of illness, recent prior severity of illness, and adverse event status. On admission, each individual was assigned a baseline severity of illness score, drawn from a normal distribution, with mean 0 and standard deviation s. The severity of illness $V(t)$ was updated at intervals of 0.25 time units using a

second order autoregressive equation,[35] as a compromise between complexity and over-simplification:

$$V(t) = V(0) + c(V(t - 0.25) - \mu) + cf(V(t - 0.5) - \mu)$$
$$+ \beta_3 A(t) + \epsilon,$$

where $V(0)$ equaled the individual's baseline severity of illness, $V(t - 0.25)$ equaled the first-order lagged severity, $V(t - 0.5)$ equaled the second-order lagged severity, c equaled the first-order autoregressive coefficient; f equaled the ratio of the second-order autoregressive coefficient to the first-order autoregressive coefficient; ϵ was an error term, drawn from a normal distribution with mean 0 and standard deviation z; β_3 represented the indirect effect of adverse event status on discharge hazard rate, mediated through severity. For the first-order lag term, μ equaled $V(0) + \beta_3 A(t - 0.25)$ and for the second-order lag term, μ equaled $V(0) + \beta_3 A(t - 0.50)$, where $A(t - 0.25)$ and $A(t - 0.50)$ were the first- and second-order lagged values for adverse event status.

   This equation was structured to produce a well-defined relationship between adverse event status and subsequent severity. Before occurrence of the adverse event, the mean value of $V(t)$ equaled $V(0)$. After occurrence of the adverse event, the mean value of $V(t)$ equaled $V(0) + \beta_3$. For synthetic individuals who experienced the adverse event, this effect was assumed to be constant for the remainder of hospitalization. When β_3 was set to 0, the indirect effect of the adverse event on discharge hazard rate was removed. Severity of illness was assumed to be constant between intervals of 0.25 days. In practice, severity of illness is not likely to be measured at intervals more frequently than 4 times per day.

   Adverse event status was a time-varying exposure. Confounding was produced by allowing severity of illness to influence the risk of experiencing an adverse event. The probability (risk) of experiencing an adverse event during the time interval between t and $t + 0.25$ was modeled as a logistic equation:

$$R(V(t)) = \exp(\alpha + \beta_4 V'(t))/[1 + \exp(\alpha + \beta_4(V'(t))],$$

where $V'(t)$ represented 1 of 3 alternative forms of severity of illness and β_4 equaled the β coefficient for its effect on the logarithm of the odds of experiencing an adverse event during the next time interval. When $V'(t)$ was specified to be the lagged severity of illness score ($V(t - 0.25)$), time varying confounding was produced. To engender baseline confounding alone, $V'(t)$ was specified as the value of severity of illness at the time of admission ($V(0)$). To abolish confounding, $V'(t)$ was set equal to an independent score with the same distributional characteristics as severity of illness. The independent score did not influence the discharge hazard rate and was uncorrelated with severity of illness.

Model Conditions

   The simulations were varied across 2 dimensions. One dimension was the impact of the adverse event on the hazard rate of discharge. Either the adverse event had no impact on discharge or the effect was indirect only or the effect was
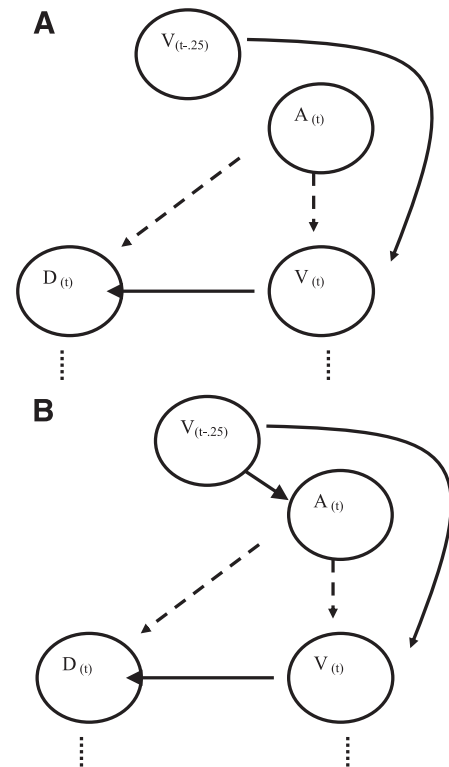
FIGURE 1. A, A directed acyclic graph that corresponds to the unconfounded simulations is depicted. The lack of influence of lagged severity of illness [$V(t - 0.25)$] on adverse event status at time t [$A(t)$] is reflected by the absence of an arrow between these 2 variables. The dashed arrows correspond to the indirect and direct causal pathways from the adverse event to the discharge hazard rate [$D(t)$]. B, This directed acyclic graph corresponds to simulations with time-varying confounding. Lagged severity of illness influences the probability of occurrence of an adverse event during the next time interval.

dual, indirect plus direct. Within each of these groups, the simulations were varied with respect to confounding (none, baseline only, time-varying). Thus, simulations were executed under 9 different conditions. We constructed directed acyclic graphs (DAGs)[36,37] corresponding to these conditions, which are depicted in Figures 1A and B.

   Parameter values used in the simulations are displayed in Table 1. The discharge hazard rate was set to produce an average length of stay of approximately 5.5 days when the adverse event did not have an effect and confounding was absent. Adverse event risks were calibrated such that approximately 30% of synthetic individuals experienced an adverse event. In all simulations, each unit increase of severity of illness decreased the hazard of discharge by 63%.

   A total of 250 replications were executed for each of the 9 conditions, using a random seed that varied randomly from run to run. Each replication produced a data set of synthetic patients admitted to a hospital during a 1-month period, run until all subjects were discharged. The standard errors for the mean hazard ratio (HR) (for each set of

conditions) were calculated by dividing the standard deviation by the square root of 250.

Cox Proportional Hazards Regression

Cox proportional hazards regression was applied to each synthetic data set to estimate the HRs for the exposure of interest, adverse event status. The outcome was discharge. For each run, 4 Cox models were fit that differed according to the included covariates: (1) adverse event status alone; (2) adverse event status plus baseline severity; (3) adverse event status plus lagged severity of illness as a time varying covariate; and (4) adverse event status plus lagged severity frozen during the postadverse event follow-up at its level before occurrence of the adverse event.

IPW

We performed a weighted regression analysis based on IPW. This method is related to the inverse weighting techniques that are used in survey statistics to account for unequal sampling probabilities. Weighting by the inverse of the probability of being sampled allows each observation to statistically represent the nonsampled members of a population. Robins et al[5,6,9] extended IPW to address the problem of time-varying confounding, in the context of an underlying marginal structural model. The model posits the existence of counterfactual variables to represent what would have happened to subjects under alternative levels of exposure or treatment. The observed portion of the data constitutes the exposure or treatment actually experienced; exposure or treatment not experienced is unobserved. Exposure status is guaranteed to be independent of history in the full (counterfactual) data set, because every subject experiences all levels of exposure. In the full data set, the association between exposure and outcome is unconfounded. In contrast, when exposure status is influenced by covariates, the observed data are a nonrandom sample of the full data set. The purpose of IPW is to quantify the number of unrealized observations in the full data set represented by each realized observation, thereby restoring the independence of exposure status and covariate history.

To perform IPW, a logistic regression model was fitted to the data to estimate the probability of experiencing an adverse event given lagged severity, baseline severity, and time from admission. This value was, in turn, used to estimate the probability of each individual's observed adverse event status at each time interval. For instance, among individuals who never experienced an adverse event, or had not yet experienced an adverse event, the probability of the observed adverse event status was equal to 1 minus the probability of experiencing an adverse event. The inverse of the probability of the observed adverse event status history up to the current time, given baseline severity, lagged severity, and time from admission represented the unstabilized weight. The weight was stabilized by using as the numerator the estimated probability of the individual's observed adverse event status history, given only baseline severity and time from admission.[38,39] The denominators of the stabilized and unstabilized weights are the same. The amount of variation of stabilized weights is an indication of the degree to which confounding is potentially time-varying.[39]

Pooled logistic regression was fitted incorporating the calculated stabilized weights to approximate estimates of the marginal structural Cox model. A natural cubic spline was used to model the time intercepts (each quarter of a day is 1 intercept for the logistic regression) with knots at 5th, 27.5th, 50th, 72.5th, and 95th percentile. Odds ratios derived from pooled logistic regression approximate HRs when the risk at each interval is less than 10%.[40,41] In this simulation, the interval-specific risks were less than 10%.

Estimation of Bias

The beta coefficients (β_1, β_2, and β_3) specified in the simulation represented individual-level effects on discharge hazard rates (individual causal HRs). However, Cox proportional hazards regression estimates population-level (marginal) associations. Parameter estimates generated by IPW are interpretable as population-level causal effects, linked to an underlying marginal structural model. Therefore, we considered the target parameter to be the population causal HR for the effect of the adverse event on discharge, rather than the individual causal HR.

Our goal was to estimate the true population causal HR to estimate bias. To guarantee exchangeability of populations, a synthetic population of 200,000 individuals was randomly assigned to experience the adverse event at the beginning of hospitalization (time 0) or to never experience the adverse event. A Cox proportional hazards regression model, which included terms for the adverse event and baseline severity of illness, was then fit to the synthetic data. This procedure was performed for each category of effect of the adverse event: (1) no effect; (2) indirect effect only; (3) dual indirect and direct effect.

The magnitude of bias for each analytic method for each of the 9 conditions was calculated by dividing the mean HR (derived from the 250 runs of the simulation) by the estimated population causal HR for the corresponding category of effect. Values less than 1 indicated bias away from the null, values of 1 indicated an absence of bias, and values greater than 1 indicated bias toward the null.[42] In this simulation, confounding produced unadjusted HR estimates that were deviated away from the null.

It should be noted that the population causal HR does not equal the individual causal HR when there is uncontrolled heterogeneity of the hazard rate. The phenomenon of uncontrolled heterogeneity is called the frailty effect in survival analysis. In this simulation, the variation in severity of illness represented uncontrolled heterogeneity. Another consequence of uncontrolled heterogeneity is that the population causal HR is not proportional. Therefore, the estimated population HR should be interpreted to be an approximation of the average population causal HR during follow-up.

Incorporation of individual random effects into a survival analysis potentially allows estimation of the individual HR under the assumption that the model is correctly specified. We did not explore these models in detail

because the frailty analysis targets a different parameter than IPW.[43,44] To feasibly fit an individual frailty model, it also would have been necessary to assume a parametric form for the hazard function. The practical value of individual frailty models has been questioned because of problems of identifiabilty.[45] Most frailty models do not retain the proportional hazards assumption.[44]

RESULTS

Descriptive Analyses

The results of each simulation run varied stochastically. Depending on the conditions, the average number of synthetic individuals per run ranged from 1793 to 2590 and the overall average length of stay ranged from 4.05 to 11.18 (Table 2). The fraction of synthetic individuals who experienced an adverse event ranged from 24% to 35%. Even when confounding was absent and the adverse event itself had no effect, length of stay was longer in individuals who experienced an adverse event (10.51 days), compared with individuals who did not experience an adverse event (3.62 days). The reason for this difference was that individuals who remained in the hospital had longer period of risks for occurrence of an adverse event.

The relationship between prior severity of illness and occurrence of the adverse event depended on the simulation conditions. When confounding was absent, adverse event status was not associated with prior severity of illness (odds ratio = 1). Similarly, longitudinal models with adverse event status as the predictor and severity of illness as the outcome yielded results that depended on whether the adverse event had an effect on severity of illness. When the adverse event did not have an effect on severity and confounding was absent, adverse event status was not associated with severity.

Impact of the Adverse Event on Discharge Hazard Rate

Cox Proportional Hazards Regression

For each of the 250 runs for each set of simulation conditions, 4 different Cox proportional hazards regression models were fit to estimate the effect of adverse event status on discharge (Table 3). Aggregating results across the 250 runs produced small standard errors around the estimate of the HR, less than 0.005 for all models. Therefore, for each set of conditions, the 95% confidence intervals of the HR were narrow, encompassing a range within 0.015 above and below the mean HR.

As explained in the Methods, the population causal HR and the individual ratio of hazards were discordant because of individual heterogeneity (frailty). The results of Cox proportional hazards regression for unconfounded simulations reflected this discrepancy. In the "unconfounded-indirect effect" simulation, the crude HR estimated by Cox was 0.6, whereas the ratio of individual hazards was 0.5. In the unconfounded-dual effect simulation, the crude HR was 0.40 whereas the ratio of individual hazards was 0.25. Including baseline severity as a covariate brought the Cox-derived HR estimate closer to the individual ratio of hazards because it removed a portion of the heterogeneity in background hazards.

When only baseline confounding was present, the crude (unadjusted) HR was 0.61 when the population causal HR was 1.0; 0.37 when the population causal HR was 0.58; and 0.23 when the population causal HR was 0.36. Thus, the bias which was calculated as the estimated HR divided by true HR, ranged from 0.61 to 0.64. Adding baseline severity as a covariate to the Cox proportional hazards regression model effectively removed the bias (range, 1.00–1.03).

When time-varying confounding was present, the crude (unadjusted) HR was 0.45 when the population causal HR

TABLE 2. Summary Statistics Derived from 250 Replicate Runs for Each Set of Conditions

Simulation Conditions (Average of 250 Simulations Each)	Adverse Event Experienced		Adverse Event not Experienced		Total	
	N	Mean Length of Stay	N	Mean Length of Stay	N	Mean Length of Stay
Adverse event effect						
Confounding by severity						
None						
None	758	10.51	1832	3.62	2590	5.63
Baseline only	632	12.76	1954	3.34	2586	5.64
Time-varying	912	11.45	1668	2.52	2580	4.05
Indirect only						
None	663	15.57	1594	3.60	2257	7.12
Baseline only	549	19.71	1692	3.35	2241	7.36
Time-varying	760	17.22	1385	2.52	2145	7.73
Indirect and direct						
None	576	23.82	1382	3.64	1958	9.58
Baseline only	484	30.19	1480	3.35	1964	9.97
Time-varying	634	27.00	1158	2.53	1793	11.18

TABLE 3. Hazard Ratio (HR) Estimates According to Simulation Conditions and Method of Analysis

| | | | HR Estimated from Cox Proportional Hazards Regression | | | | |
Model	Ratio of Individual Discharge Hazard Rates	Population Causal HR	Crude	Baseline Severity	Lagged Severity	Frozen Severity	Weighted Regression (IPW)
Adverse event effect							
Confounding by severity							
None	1	1.00					
None			1.00	1.00	1.00	1.19	1.00
Baseline only			0.61	1.03	0.91	1.06	1.04
Time-varying			0.45	0.64	0.90	1.59	1.02
Indirect only	0.5	0.58					
None			0.63	0.58	0.90	0.66	0.57
Baseline only			0.37	0.59	0.81	0.83	0.59
Time-varying			0.27	0.38	0.82	0.89	0.57
Indirect and direct	0.25	0.36					
None			0.43	0.36	0.51	0.41	0.35
Baseline only			0.23	0.36	0.45	0.34	0.35
Time-varying			0.17	0.23	0.45	0.54	0.35

Crude indicates unadjusted analysis; Baseline severity, adjusted for baseline severity; Lagged severity, adjusted for lagged severity; Frozen severity, similar to lagged severity except that severity frozen at value just before adverse event occurrence; IPW, time-varying severity adjusted.

was 1.0; 0.27 when the population causal HR was 0.58; and 0.17 when the population causal HR was 0.36. These crude HR estimates exhibited bias that ranged from 0.45 to 0.47. In this setting, all of the other HR estimates generated by Cox proportional hazards regression exhibited substantial bias, either toward or away from the null, depending on how severity of illness was included in the Cox regression model. When baseline severity was included as a covariate, the manifested problem was residual confounding (magnitude of bias: 0.64–0.65). Including lagged severity or frozen severity as a covariate produced bias in the opposite direction, toward the null (range of bias, 1.25–1.59).

IPW

For simulations, which lacked confounding or were constructed to possess only baseline confounding, the stabilized weights generated by the IPW method were clustered around 1 (0.95–1.05). For simulations in which time varying confounding was present, the stabilized weights ranged from 0.018 to 40.19. The range of stabilized weights reflected the high degree to which time-varying severity of illness was predictive of the adverse event.

Regardless of the presence or absence of confounding and whether the confounding was baseline or time-varying, IPW generated estimates of the HR for adverse event status that were unbiased or minimally biased (range of bias, 0.97–1.04; Table 3). Thus, when only baseline confounding was present, IPW performed comparably to the Cox proportional hazards regression models, which included baseline severity as a covariate; the differences in mean HR were less than 0.015. When time-varying confounding was present, the HR estimates generated by IPW were substantially less biased than all of the Cox proportional hazards regression estimates.

DISCUSSION

We used agent-based simulation to illustrate the application of IPW to the clinically relevant scenario of hospitalized patients who are at risk for experiencing an adverse event such as nosocomial infection or adverse drug reaction or other medical injury. Our goal was to identify methods that yield unbiased estimates of the effect of the adverse event on discharge rates. An estimation of effect on length of stay is a necessary part of the evaluation of impact of adverse events on costs. In the many studies that have examined this issue,[46–48] severity of illness has been recognized as a potential confounder. However, the implications of time-varying confounding have not been adequately addressed.

The simulation demonstrated that fitting Cox proportional hazards regression models which included severity of illness as a time-varying covariate, yielded estimates of effect that were biased toward the null when time varying confounding was present. Cox models fit with baseline variables only were also biased, but in the opposite direction. In contrast, IPW provided an unbiased estimate of the population effect of the adverse event on discharge rate across a range of simulation conditions.

Regardless of whether IPW or conventional regression methods are used, an assumption of no unmeasured confounding is necessary. With real data, the magnitude of bias associated with a particular method of analysis cannot be known with certainty. In any given study, Cox proportional hazards regression may or may not produce biased effect estimates. Because measures of severity are approximations of "true" severity, biases due to residual confounding and to blocking of an intermediate variable may cancel each other out. An advantage of IPW is that lack of bias does not depend on this type of chance occurrence.

A common practice in survival analysis is to evaluate for confounding on the basis of a comparison of crude and adjusted HR. However, this procedure may be misleading. An adjusted HR that is closer to the null than the crude HR may reflect blockage of an intermediate variable or removal of positive confounding. The process of performing IPW yields more information about the potential magnitude of time-varying confounding than a simple comparison of crude and adjusted HR. As a method for confounding control, IPW illustrates the key principle that making exposure independent of the measured covariate history removes confounding due to measured variables. In our study, adverse event status was independent of lagged severity of illness when a logistic model was fit using the IPW derived stabilized weights. When examining events that occur in hospitalized patients, it may be advantageous to approach confounding by focusing on factors that influence exposure rather than from the perspective of predicting outcome. This framework avoids pitfalls associated with intermediate variables.

Our simulation highlights the discrepancy between population and individual effects in the presence of uncontrolled heterogeneity, which in this simulation was due to severity of illness. Heterogeneity in the individual hazards also typically induces nonproportionality of the population HR. These manifestations of individual heterogeneity may represent intrinsic features of effect estimation when time-varying confounders are also intermediate variables. As demonstrated in this study, removing heterogeneity by adding severity of illness into the model of the outcome produces bias toward the null when severity of illness is an intermediate variable.

In the simulation described here, an adverse event influenced hospital length of stay through its effect on discharge rate. We in turn used analytic methods that were parameterized with respect to discharge HRs. In contrast, many of the published studies that examined outcomes of adverse events directly estimated effects in terms of length of stay or costs, because they used matched cohort study designs or applied linear regression. However, the matched cohort design does not solve the problems reported here. Further, it is associated with other types of biases.[49,50]

Our simulation incorporated several simplifying assumptions to highlight key underlying principles and to increase the ease of interpretation. One simplification was that the impact of the adverse event on discharge rate remained constant. In reality, the effects of an adverse event likely abate with time, except when a cascading series of new problems are triggered. The simulation reported here also ignored the consequence of the competing risk of death. Application of IPW within a framework of inverse probability of censoring makes it possible to address censoring and competing risks.[51] The flexibility of agent-based simulation supports a range of options for increasing complexity. A discussion of the analysis of synthetic data generated by these alternative simulation structures is beyond the scope of this article.

A simulation study can provide useful insights but it does not replace collection of real data. The parameter estimates used in this simulation were selected to be plausible. However, the simulation was not calibrated to a specific data set, and therefore was not designed to directly estimate the magnitude of bias therein. Measured indicators of severity of illness do demonstrate substantial day-to-day variability, comparable to our simulated severity of illness score.[52] In outcome studies of adverse events in hospitalized patients, evidence of time-varying confounding has been found.[31]

In summary, when an exposure occurs during the period of observation, the use of IPW can provide a more robust approach to analysis of observational data than conventional Cox proportional hazards regression. Unlike propensity scores, IPW can be readily adapted to account for time-varying confounders that are also intermediate variables. We have demonstrated that this is a relevant problem when estimating the effect of an adverse event on hospital length of stay. Developing accurate measures of outcomes associated with different types of adverse events is a critical first step toward understanding the cost-effectiveness of interventions to prevent or mitigate their occurrence.

REFERENCES

1. Robins JM, Greenland S. The role of model selection in causal inference from nonexperimental data. *Am J Epidemiol*. 1986;123:392–402.
2. Greenland S, Morgenstern H. Confounding in health research. *Annu Rev Public Health*. 2001;22:189–212.
3. Greenland S, Brumback B. An overview of relations among causal modelling methods. *Int J Epidemiol*. 2002;31:1030–1037.
4. Robins J. The control of confounding by intermediate variables. *Stat Med*. 1989;8:679–701.
5. Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*. 2000;11:561–570.
6. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11:550–560.
7. Cole SR, Hernan MA, Robins JM, et al. Effect of highly active antiretroviral therapy on time to acquired immunodeficiency syndrome or death using marginal structural models. *Am J Epidemiol*. 2003;158:687–694.
8. Hernan MA, Brumback BA, Robins JM. Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures. *Stat Med*. 2002;21:1689–1709.
9. Hernan MA, Lanoy E, Costagliola D, et al. Comparison of dynamic treatment regimes via inverse probability weighting. *Basic Clin Pharmacol Toxicol*. 2006;98:237–242.
10. Sterne JA, Hernan MA, Ledergerber B, et al. Long-term effectiveness of potent antiretroviral therapy in preventing AIDS and death: a prospective cohort study. *Lancet*. 2005;366:378–384.
11. Hernan MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health*. 2006;60:578–586.
12. Rubin D. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med*. 1997;127:757–763.
13. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med*. 1998;17:2265–2281.
14. Braitman L, Rosenbaum P. Rare outcomes, common treatments: analytic strategies using propensity scores. *Ann Intern Med*. 2002;137:693–695.
15. Cepeda MS, Boston R, Farrar JT, et al. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *Am J Epidemiol*. 2003;158:280–287.
16. Brookhart MA, Schneeweiss S, Rothman KJ, et al. Variable selection for propensity score models. *Am J Epidemiol*. 2006;163:1149–1156.
17. Sturmer T, Schneeweiss S, Brookhart MA, et al. Analytic strategies to adjust confounding using exposure propensity scores and disease risk scores: nonsteroidal antiinflammatory drugs and short-term mortality in the elderly. *Am J Epidemiol*. 2005;161:891–898.

18. Haight T, Neugebauer R, Tager IB, et al. Comparison of the Inverse Probability of Treatment Weighted (IPTW) Estimator With a Naıve Estimator in the Analysis of Longitudinal Data With Time-Dependent Confounding: a Simulation Study U.C. Berkeley Division of Biostatistics Working Paper Series; 2003.

19. Cosgrove SE, Kaye KS, Eliopoulous GM, et al. Health and economic outcomes of the emergence of third-generation cephalosporin resistance in Enterobacter species. *Arch Intern Med*. 2002;162:185–190.

20. Bates DW, Spell N, Cullen DJ, et al. The costs of adverse drug events in hospitalized patients. Adverse Drug Events Prevention Study Group. *JAMA*. 1997;277:307–311.

21. Bates DW, Yu DT, Black E, et al. Resource utilization among patients with sepsis syndrome. *Infect Control Hosp Epidemiol*. 2003;24:62–70.

22. Classen DC, Pestotnik SL, Evans RS, et al. Adverse drug events in hospitalized patients. Excess length of stay, extra costs, and attributable mortality. *JAMA*. 1997;277:301–306.

23. Engemann JJ, Carmeli Y, Cosgrove SE, et al. Adverse clinical and economic outcomes attributable to methicillin resistance among patients with *Staphylococcus aureus* surgical site infection. *Clin Infect Dis*. 2003;36:592–598.

24. Carmeli Y, Eliopoulos G, Mozaffari E, et al. Health and economic outcomes of vancomycin-resistant enterococci. *Arch Intern Med*. 2002;162:2223–2228.

25. Graves N, Weinhold D, Roberts JA. Correcting for bias when estimating the cost of hospital-acquired infection: an analysis of lower respiratory tract infections in non-surgical patients. *Health Econ*. 2005;14:755–761.

26. Asensio A, Torres J. Quantifying excess length of postoperative stay attributable to infections: a comparison of methods. *J Clin Epidemiol*. 1999;52:1249–1256.

27. Wey SB, Mori M, Pfaller MA, et al. Hospital-acquired candidemia. The attributable mortality and excess length of stay. *Arch Intern Med*. 1988;148:2642–2645.

28. Pittet D, Tarara D, Wenzel RP. Nosocomial bloodstream infection in critically ill patients: excess length of stay, extra costs, and attributable mortality. *JAMA*. 1994;271:1598–1601.

29. Zhan C, Miller MR. Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. *JAMA*. 2003;290:1868–1874.

30. Harbarth S, Burke JP, Lloyd JF, et al. Clinical and economic outcomes of conventional amphotericin B-associated nephrotoxicity. *Clin Infect Dis*. 2002;35:E120–E127.

31. Soufir L, Timsit JF, Mahe C, et al. Attributable morbidity and mortality of catheter-related septicemia in critically ill patients: a matched, risk-adjusted, cohort study. *Infect Control Hosp Epidemiol*. 1999;20:396–401.

32. Grimm V, Revilla E, Berger U, et al. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *Science*. 2005;310:987–991.

33. Bankes S, Lempert R. Robust reasoning with agent-based modeling. *Nonlinear Dynamics Psychol Life Sci*. 2004;8:259–278.

34. Lempert RJ. A new decision sciences for complex systems. *Proc Natl Acad Sci USA*. 2002;99(Suppl 3):7309–7313.

35. Donnelly JH, Shannon RE. Minimum mean-squared-error estimators for simulation experiments. *Commun ACM* 1981;24:253–259.

36. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10:37–48.

37. Robins J. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *J Chronic Dis*. 1987;40(Suppl 2):139S–161S.

38. Bodnar LM, Davidian M, Siega-Riz AM, et al. Marginal structural models for analyzing causal effects of time-dependent treatments: an application in perinatal epidemiology. *Am J Epidemiol*. 2004;159:926–934.

39. Fewell Z, Hernan MA, Wolfe F, et al. Controlling for time-dependent confounding using marginal structural models. *Stata J*. 2004;4:402–420.

40. Thompson WA Jr. On the treatment of grouped observations in life studies. *Biometrics*. 1977;33:463–470.

41. D'Agostino RB, Lee ML, Belanger AJ, et al. Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study. *Stat Med*. 1990;9:1501–1515.

42. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*. 2003;14:300–306.

43. Gutierrez RG. Parametric frailty and shared frailty survival models. *Stata Journal*. 2002;2:22–44.

44. Oakes D. Bivariate survival models induced by frailties. *J Am Stat Assoc*. 1989;84:487–493.

45. O'Quigley J, Stare J. Proportional hazards models with frailties and random effects. *Stat Med*. 2002;21:3219–3233.

46. Evans RS, Classen DC, Stevens LE, et al. Using a hospital information system to assess the effects of adverse drug events. *Proc Annu Symp Comput Appl Med Care*. 1993;161–165.

47. Hugonnet S, Eggimann P, Borst F, et al. Impact of ventilator-associated pneumonia on resource utilization and patient outcome. *Infect Control Hosp Epidemiol*. 2004;25:1090–1096.

48. Rosenthal VD, Guzman S, Migone O, et al. The attributable cost and length of hospital stay because of nosocomial pneumonia in intensive care units in 3 hospitals in Argentina: a prospective, matched analysis. *Am J Infect Control*. 2005;33:157–161.

49. Schulgen G, Kropec A, Kappstein I, et al. Estimation of extra hospital stay attributable to nosocomial infections: heterogeneity and timing of events. *J Clin Epidemiol*. 2000;53:409–417.

50. Samore MH, Harbarth SJ. In: Mayhall CG, ed. *A Methodologically Focused Review of the Literature in Hospital Epidemiology and Infection Control Hospital Epidemiology and Infection Control*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2004.

51. Beyersmann J, Gastmeier P, Grundmann H, et al. Use of multistate models to assess prolongation of intensive care unit stay due to nosocomial infection. *Infect Control Hosp Epidemiol*. 2006;27:493–499.

52. Harbarth S, Pestotnik SL, Lloyd JF, et al. The epidemiology of nephrotoxicity associated with conventional amphotericin B therapy. *Am J Med*. 2001;111:528–534.