



Published in final edited form as:

*J R Stat Soc Ser C Appl Stat*. 2011 August 1; 60(4): 475–496. doi:10.1111/j.1467-9876.2010.00757.x.

## Estimating the Causal Effect of Low Tidal Volume Ventilation on Survival in Patients with Acute Lung Injury†

**Weiwei Wang,**

University of Texas Health Science Center at Houston, Houston, TX U.S.A.

**Daniel Scharfstein,**

The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD U.S.A.

**Chenguang Wang,**

Food and Drug Administration, Silver Spring, MD U.S.A.

**Michael Daniels,**

University of Florida, Gainesville, FL U.S.A.

**Dale Needham,**

The Johns Hopkins School of Medicine, Baltimore, MD U.S.A.

**Roy Brower,** and

The Johns Hopkins School of Medicine, Baltimore, MD U.S.A.

**the NHLBI ARDS Clinical Network**

### Summary

Acute lung injury (ALI) is a condition characterized by acute onset of severe hypoxemia and bilateral pulmonary infiltrates. ALI patients typically require mechanical ventilation in an intensive care unit. Low tidal volume ventilation (LTVV), a time-varying dynamic treatment regime, has been recommended as an effective ventilation strategy. This recommendation was based on the results of the ARMA study, a randomized clinical trial designed to compare low vs. high tidal volume strategies (The Acute Respiratory Distress Syndrome Network, 2000). After publication of the trial, some critics focused on the high non-adherence rates in the LTVV arm suggesting that non-adherence occurred because treating physicians felt that deviating from the prescribed regime would improve patient outcomes. In this paper, we seek to address this controversy by estimating the survival distribution in the counterfactual setting where all patients assigned to LTVV followed the regime. Inference is based on a fully Bayesian implementation of Robins' (1986) G-computation formula. In addition to re-analyzing data from the ARMA trial, we also apply our methodology to data from a subsequent trial (ALVEOLI), which implemented the LTVV regime in both of its study arms and also suffered from non-adherence.

### Keywords

Bayesian inference; Causal inference; Dynamic treatment regime; G-computation formula

### 1. Acute Lung Injury

Acute lung injury (ALI) is a condition characterized by acute onset of severe hypoxemia and bilateral pulmonary infiltrates. Acute respiratory distress syndrome (ARDS) is a sub-set of

†This research was sponsored by NIH contracts NO1-HR-46046-64, NO1-HR-16146-54, R01-CA-85295.

ALI with more severe hypoxemia. The 1994 American-European consensus criteria for ALI are (1) acute onset, (2) bilateral infiltrates on chest radiography, (3) pulmonary-artery wedge pressure less than or equal to 18mm Hg when measured or the absence of clinical evidence of left atrial hypertension, and (4) ratio of the partial pressure of arterial oxygen ( $PaO_2$ ) to the fraction of inspired oxygen ( $FiO_2$ ) less than or equal to 300. A patient with ALI is considered to have ARDS if the  $PaO_2 : FiO_2$  ratio is less than or equal to 200 (Bernard et al., 1994).

ALI can be caused by direct injury to the lung as in pneumonia and aspiration of gastric contents, or indirect lung injury as in sepsis from a source other than the lung and pancreatitis. The majority of deaths among ALI patients are due to sepsis or multi-organ dysfunction rather than primary respiratory causes (Ware and Matthay, 2000). A recent large prospective study of the incidence and mortality of ALI in the United States (Rubenfeld et al., 2005) reported an incidence of 86 per 100,000 person-years with an in-hospital mortality rate of 38.5 percent.

### 1.1. Mechanical ventilation for ALI

ALI patients frequently require mechanical ventilation in an intensive care unit (ICU). There are many different ways to provide mechanical ventilation support for ALI patients. In this paper, we focus on the assist-control (volume-control) mode of ventilation, during the most acute phase of respiratory failure. This mode is often used immediately after initiation of mechanical ventilation. In this mode, the physician selects a minimum respiratory rate and tidal volume. If a patient's respiratory rate falls below the prescribed rate, the ventilator will "control" the breathing by raising pressure in the airway to deliver breath at the prescribed tidal volume and at the prescribed minimum respiratory rate. If a patient's respiratory rate exceeds the prescribed minimum set rate, the ventilator assists the patient's inspiratory efforts by raising pressure in the airway until the prescribed tidal volume is achieved. The best ventilation strategy for ALI patients has been controversial. The resting tidal volume of healthy people is about 6 to 7 ml per kilogram of predicted body weight (PBW) (Tobin et al., 1983). The predicted body weight is calculated from sex and height, which are good predictors of lung volume. Traditionally, a generous tidal volume of 10-15 ml/kg PBW was recommended for ALI patients because they were useful for achieving clinical goals of gas exchange. However, in ALI patients, the volume of aerated lung available for ventilation is substantially reduced from the disease process. Consequently, generous tidal volumes may cause injury to the aerated lung from over-distention, causing further lung injury, known as barotrauma or volutrauma (Webb and Tierney, 1974; Dreyfuss and Saumon, 1998). This issue has led to several clinical trials of lung-protective methods of mechanical ventilation with low tidal volumes for patients with ALI (Fan et al., 2005).

### 1.2. The ARMA Trial

The National Heart, Blood and Lung Institute ARDS Clinical Network conducted a randomized trial comparing mechanical ventilation using traditional tidal volumes versus low tidal volume ventilation (The Acute Respiratory Distress Syndrome Network, 2000). This study was dubbed the ARMA trial. A total of 429 and 432 patients were randomized to traditional and low tidal volume ventilation, respectively. The assist-control ventilation mode was used until a patient was ready to be weaned from the ventilator. Figure 1 shows a flowchart of traditional and low tidal volume ventilation strategies. In the group treated with traditional tidal volumes, the initial tidal volume was set to 12 ml/kg PBW. This tidal volume was then adjusted downward if necessary to keep the airway pressure measured after a 0.5-second pause at the end of inspiration (plateau pressure) less than or equal to 50 cm  $H_2O$ . If the plateau pressure subsequently decreased below 45 cm  $H_2O$ , then tidal volumes were increased so that plateau pressure was greater than or equal to 45 cm  $H_2O$  or tidal

volume was 12 ml/kg PBW. In the group treated with low tidal volumes, the tidal volume was initially set to 6 ml/kg PBW and then decreased to a minimum of 4 ml/kg PBW if necessary to maintain the plateau pressure less than or equal to 30 cm  $H_2O$ . If the plateau pressure subsequently fell below 25 cm  $H_2O$ , then tidal volume was increased so that plateau pressure was greater than or equal to 25 cm  $H_2O$  or tidal volume was 6 ml/kg PBW. Tidal volumes greater than 6 ml/kg PBW were allowed if patients experienced severe dyspnea, provided the plateau pressure remained below 30 cm  $H_2O$ . In both arms, plateau pressures were allowed to go above the threshold levels (50 and 30 cm  $H_2O$  for the traditional and low tidal volume arms, respectively) if the tidal volume was 4 ml/kg PBW or arterial pH was less than 7.15. In both study groups, most other aspects of the mechanical ventilation strategies, including positive end-expiratory pressure (PEEP),  $FiO_2$ , respiratory rate set on the ventilator, and inspiratory:expiratory ratio (I:E) were controlled according to explicit protocol rules.

The study demonstrated lower mortality, more ventilator-free days, and more organ failure-free days in ALI patients who received mechanical ventilation with the LTVV versus traditional strategy. The ARDS Network concluded that the LTVV strategy was better than the traditional strategy and recommended that the LTVV strategy be used for clinical management of ALI.

### 1.3. Controversy

The ARMA trial, its results, and the ARDS Network recommendation were considered controversial by some clinicians and researchers. Some argued that the recommendation of a tidal volume goal of 6 ml/kg PBW is not optimal because the traditional tidal volume strategy used in the trial did not accurately represent usual care approaches that clinicians were using when the trial was conducted (Eichacker et al., 2002; Tobin, 2000). In the trial, the most common tidal volumes prescribed by clinicians before their patients were enrolled were in the range of 9 - 11 ml/kg PBW (Thompson et al., 2001). Another perspective that is consistent with this concern was expressed in the editorial that accompanied the ARMA trial publication (Tobin, 2000). The editorial argued that in 5 randomized trials of ALI mechanical ventilation strategies (including the ARMA trial), clinical outcomes were better in the study groups that received low tidal volumes only when the mean plateau pressures in the traditional study groups exceeded 32 cm  $H_2O$ . This was interpreted by some to suggest that it was not necessary to reduce tidal volumes to 6 ml/kg PBW in patients whose plateau pressures were below 32 cm  $H_2O$  while they were receiving tidal volumes that exceeded 6 ml/kg (Tobin, 2004). In another review of these 5 clinical trials, the authors suggested that clinical outcomes would be optimized in most ALI patients with intermediate tidal volumes between 6 and 12 ml/kg and plateau pressures of 28-32 cm  $H_2O$  (Eichacker et al., 2002).

In the ARMA trial, adherence to the LTVV protocol rules was not perfect. Since only the selected tidal volume and resulting plateau pressure can be recorded, not the selection procedure, a set of specifications has been carefully developed to determine whether the LTVV protocol was followed, using the values of tidal volume, plateau pressure, arterial pH, set respiratory rate, etc (see Table 1). Ventilation on a given day is considered adherent to LTVV if the settings are adherent to both the plateau pressure specification and the tidal volume specification. Plateau pressure adherence follows if lines 1 or 2 (Table 1) are satisfied; tidal volume adherence follows if any of lines 3-7 are satisfied. In the ARMA trial, clinicians sometimes deviated from these rules. Figure 2 (a) displays the observed plateau pressure/tidal volume combinations in the LTVV arm of the ARMA trial. Dots and crosses denote adherence and non-adherence with the LTVV protocol, respectively. Points falling inside region (1) are adherent to LTVV. Points falling inside region (2) are not adherent. Points falling inside regions (3), (4) and (5) must satisfy additional restrictions of arterial pH and set respiratory rate to be adherent to LTVV. Overall, 28% of the observed ventilation

days deviated from the LTVV protocol. It is important to note that the majority of the non-adherent points within the regions (3), (4) and (5) of Figure 2 were due to violations of the pH restriction, potentially due to clinicians' concern regarding immediate adverse effect of low pH. According to the LTVV protocol, each non-adherent (cross) point could have been adherent (dot) by adjustment of ventilator settings. Consider a non-adherent (cross) point in region (4), which does not satisfy the LTVV regime because pH greater than or equal to 7.2. By lowering tidal volume (with a consequential decrease in plateau pressure), pH will decrease, yielding adherence in the same region or movement to an alternative region (due to change in tidal volume and plateau pressure).

Some have argued that physicians deviated from the LTVV protocol when they thought specific aspects of the protocol were detrimental to their patients. For example, when patients appeared to be very dyspneic (short of breath) while receiving 6 ml/kg PBW, some physicians raised tidal volume above 6 ml/kg PBW in an attempt to alleviate the dyspnea. By making these deviations, clinical outcomes of some patients could have improved (Deans et al., 2005, 2006). Thus, there is continued controversy regarding the ARDS Network recommendation to use LTVV. To address this controversy, we seek to draw inference about the distribution of survival that would have resulted had all patients in the LTVV arm of the ARDSNet study strictly adhered to the LTVV protocol. We will compare this distribution to the observed LTVV survival curve that reflects imperfect adherence with the LTVV protocol. We will also evaluate the effect of a simpler regime, which is easier for clinicians to implement.

#### 1.4. ALVEOLI Study

After the ARMA trial, the LTVV protocol was recommended by the ARDS Network and was used in subsequent trials conducted by ARDS Network. For example, ALVEOLI was a subsequent randomized trial to test the hypothesis that higher end-expiratory airway pressure/lower  $FiO_2$  ventilation strategy would reduce mortality in patients with ALI when compared to a lower end-expiratory airway pressure/lower  $FiO_2$  ventilation strategy (The National Heart, Blood and Lung Institute ARDS Clinical Trials Network, 2004). This ALVEOLI trial was stopped due to lack of efficacy after a total of 549 patients were enrolled. In the ALVEOLI trial, the LTVV regime was used in both study arms and non-adherence was observed. Overall, 22% of the observed ventilation days were non-adherent with the LTVV protocol. As with the re-analysis of the ARMA trial, in this paper, we will estimate the distributions of survival under strict adherence to the complex LTVV protocols and under a simpler regime. We will then compare these results to the observed survival curve. Consistency of results across the ARMA and ALVEOLI studies will provide greater support for our conclusions.

#### 1.5. Statistical Challenges

There are two main challenges in the statistical analysis.

**1.5.1. Time-Dependent Dynamic Treatment Regime**—LTVV is a time-dependent dynamic treatment regime. Dynamic treatment regimes are individually tailored treatments (Murphy et al., 2001; Moodie et al., 2007). In a dynamic treatment regime, the treatment is provided to patients only when it is needed and the level of treatment is adjusted based on time-dependent measurements of the patient's need for treatment. About 40% of ALI patients die in ICU generally while receiving mechanical ventilation. Recovering patients who remain on mechanical ventilation are evaluated daily by a weaning procedure to assess for discontinuation of ventilation. According to his/her performance in the weaning procedure, the patient may begin unassisted breathing, change to other ventilation modes, or remain on assist-control mode. Only if the patient is alive and on assist-control mode, is he/

she eligible or ‘at risk’ for LTVV. In contrast, in a non-dynamic (static) treatment regime the treatment level and type are specified before treatment is started. That is, the treatment regime is not changed with a patient's response, for example, administering a fixed drug dose for a fixed duration to a patient with a specific disease.

In observational studies of time-dependent treatments (dynamic or non-dynamic), there often exist time-dependent covariates which are influenced by past treatments as well as predict future treatments and the outcome. Such covariates are called time-dependent confounders. They are sometimes called endogenous or internal variables in the survival analysis literature (Kalbfleisch and Prentice, 2002). For example, a patient's organ failure status may be worse for patients who have not received LTVV. Moreover, those with worse organ failure also may be less likely to receive future LTVV (due to patients being perceived as “too sick” for LTVV by their doctors) and are more likely to die in-hospital. In this situation, the challenge is to isolate the effect of LTVV from the confounding effect of organ failure and other time-dependent confounders. An analysis which estimates the survival curve among those who are observed to adhere to LTVV will likely over-estimate the counterfactual survival curve of interest. If LTVV is an effective regime, the counterfactual survival curve will likely be higher than the observed (i.e, unconditional) survival curve. We illustrate this phenomenon in Section 4.1.

There is a significant and growing body of literature focused on drawing inference about dynamic treatment regimes from observational data with time-dependent confounders. In this paper, we will utilize the G-computation algorithm, as developed by (Robins, 1986, 1987b,a, 1988b,a) and described nicely in Lok et al. (2004), to estimate the counterfactual survival curve of interest. Alternative approaches that could be used include marginal structural (Murphy et al., 2001; Robins et al., 2000; Bembom and van der Laan, 2008; Robins et al., 2008; Moodie, 2009b; Hernan et al., 2006; Robins, 1999a,b) and structural nested (Robins, 1997, 1999b; Lok et al., 2004) modelling. Recently, the literature has been focused on identifying optimal dynamic regimes (Robins et al., 2008; Moodie, 2009a; Rich et al., 2010; Moodie and Richardson, 2010; Murphy, 2003; Moodie et al., 2007; Chakraborty et al., 2010; Zhao et al., 2009; Robins, 2004; Orellana et al., 2010). Although a very interesting direction of research, we will not seek to find the best ventilation strategy in this paper.

**1.5.2. Missing Data**—In the ARMA and ALVEOLI studies, reference measurements of ventilator parameters and other variables were taken on days 1,2,3,4,7,14,21,28. The ventilator parameters on the other days were not recorded. In addition to the reference measurements, the care providers recorded ventilator parameters, arterial pH and plateau pressure daily at randomly selected times to assess the accuracy of the ventilator settings relative to the protocol requirements. Because these random ventilator check data were scheduled to be collected daily during the entire study, we used them for our data analysis. However, missing data is a major problem for the random ventilator check data (see Table 2) and for many observational studies. We will address it by specifying a fully parametric model for the distribution of the data that are scheduled to be collected and assuming that the data that are not recorded are missing at random. In specifying our model, we must recognize that some of the ventilation parameters have biologically bounded support.

## 1.6. Outline

In Section 2, we introduce the data structure and notation including the definition of potential survival time under a dynamic treatment regime. In Section 3, we introduce assumptions necessary for identification of the distribution of the potential survival time. Section 4 introduces the G-computation method for estimating the potential survival curve

of a dynamic treatment regime and our fully Bayesian implementation. In Section 5, we apply the methods in the previous sections to our data. Section 6 is devoted to a discussion.

## 2. Data structure and notation

The notation used in this article is listed in Table 3. With this notation, the scheduled observed data on day  $t$  ( $t = 1, \dots, 28$ ) are

$$L_t = (S_t, HSP_t, ICU_t, VNT_t, G_t, D_t, W_t, OF_t)$$

where  $S_t$  is the indicator of surviving day  $t$ ;  $HSP_t$  is the indicator of hospitalization on day  $t$ , defined only when  $S_t = 1$ ;  $ICU_t$  is the indicator of being in the ICU on day  $t$ , defined only when  $S_t = HSP_t = 1$ ;  $VNT_t$  is the indicator of being on the assist-control mode of the ventilator on day  $t$ , defined only when  $S_t = HSP_t = ICU_t = 1$ ;  $R_t = I(HSP_t = ICU_t = VNT_t = 1)$  is the indicator of being at-risk on day  $t$ , defined only when  $S_t = 1$ ;  $G_t = (PE_t, LC_t)$  includes PEEP and lung compliance on day  $t$ , defined only when  $S_t = R_t = 1$ ;  $D_t = (VT_t, SR_t)$  includes tidal volume and set rate on day  $t$ , defined only when  $S_t = R_t = 1$ ;  $W_t = (PH_t, PP_t, DS_t, SO_t)$  includes pH, plateau pressure, indicator of severe dyspnea, and oxygen saturation on day  $t$ , defined only when  $S_t = R_t = 1$ ; and  $OF_t$  is an indicator of organ failure on day  $t$ , defined when  $S_t = HSP_t = 1$ . When  $S_t = R_t = 1$ , the variables  $G_t$ ,  $D_t$  and  $W_t$  are recorded simultaneously at a random time during the day. From a causal inference perspective, we will consider the variables that make up  $L_t$  to be sequentially measured in time; for example, when  $S_t = R_t = 1$ ,  $G_t$  is measured before  $D_t$  which is measured before  $W_t$  and  $OF_t$ . This implies that when  $S_t = R_t = 1$ ,  $D_t$  can affect  $W_t$  and  $OF_t$ , but not  $G_t$ .

In our study,  $L_0 = (AP, OF_0)$ , where  $AP$  denotes the baseline APACHE III score. Let  $\bar{L}_t = (L_0, \dots, L_t)$ . In general, we use a bar over a variable to denote that variable and all the past values of that variable. The scheduled observed data are  $\bar{L}_{28}$ . Our identification assumptions in the next section will be based on a subset of the observed data. Towards this end, it is useful to define  $H_0 = AP$ ,  $U_0 = OF_0$ ,  $H_t = (HSP_t, ICU_t, VNT_t, G_t)$  and  $U_t = (D_t, W_t, OF_t)$ , for  $t = 1, \dots, 28$ . With this notation, note that  $L_t = (S_t, H_t, U_t)$ .

When an individual is on the assist-control mode of ventilation, the tidal volume and set rate are directly manipulable by the treating physician. If the treating physician had followed the LTVV protocol (defined in the flowchart in Figure 1), there would be a unique ventilator setting (i.e.,  $D_t$ ) for each patient. Since the dynamic process of setting the ventilator dials was not recorded, a physician was considered to have followed the protocol if the tidal volume and plateau pressure specifications in Table 1 are satisfied. When  $S_t = R_t = 1$ , the adherence indicator on day  $t$ ,  $A_t$ , is a deterministic function of  $D_t$  and  $W_t$ . That is,  $A_t = I((D_t, W_t) \in \mathcal{S})$ , where  $\mathcal{S}$  is the set of combinations of tidal volume, set rate, pH, plateau pressure, and severe dyspnea, which satisfy the tidal volume and plateau pressure specifications in Table 1. We let  $A_t = 1$  if  $S_t = 0$  or  $S_t = 1, R_t = 0$ . We denote the counterfactual random variables (potential outcomes) that would have been observed had an individual been sequentially treated according to the LTVV protocol with the superscript \* notation. The goal is to draw inference about the counterfactual survival curve, i.e.,  $P[S_m^* = 1]$  for  $m = 1, \dots, 28$ .

## 3. Assumptions

There are four assumptions that we impose to identify the counterfactual survival curve of interest.

First, we assume that the potential survival outcome under the LTVV protocol is uniquely determined. This is the stable unit treatment assumption (Rubin, 1986) or consistency assumption (Robins, 1986). Under this assumption,  $S_t^* = S_t$  when  $\bar{A}_{t-1} = 1$ . This implies that, for each  $t$  ( $t = 1, \dots, 28$ ),

$$P \left[ S_t^* = 1 | S_{t-1} = 1, \bar{A}_{t-1} = 1, \bar{H}_{t-1} = \bar{h}_{t-1}, \bar{U}_{t-1} = \bar{u}_{t-1} \right] = P \left[ S_t = 1 | S_{t-1} = 1, \bar{A}_{t-1} = 1, \bar{H}_{t-1} = \bar{h}_{t-1}, \bar{U}_{t-1} = \bar{u}_{t-1} \right]$$

The LTVV protocol, as depicted in Figure 1, is designed to yield unique patient-level ventilator settings. As administration of the protocol was not directly observed, we assume, based on clinical considerations, that the physician followed the protocol if the tidal volume and plateau pressure specifications in Table 1 are satisfied. Accordingly, the treatment for an individual is unique and therefore her potential survival outcome is unique.

Second, we assume that the LTVV protocol is evaluable. Specifically, we assume that when a group of patients have been treated with the LTVV protocol until day  $t$ , at least some of them will continue to be treated with the protocol on day  $t + 1$ . This has been called the positivity assumption (Hernan and Robins, 2006), or equivalently, the evaluable treatment regime assumption (Lok et al., 2004). Mathematically, we require that, for each  $t$  and each possible value of  $\bar{h}_t$  and  $\bar{u}_{t-1}$ ,

$$P \left[ A_t = 1 | S_t = 1, \bar{A}_{t-1} = 1, \bar{H}_t = \bar{h}_t, \bar{U}_{t-1} = \bar{u}_{t-1} \right] > 0$$

For the LTVV arm in the ARMA study and both arms in the ALVEOLI study, the positivity assumption is reasonable since physicians were instructed to follow the LTVV protocol. In the traditional tidal volume arm of the ARMA study, for example, the positivity assumptions would not be plausible.

Third, we assume there is no unmeasured confounding. In particular, we assume that, for individual on assist-control mode ventilation, adherence to the LTVV protocol is sequentially randomized. That is, we assume, for each  $t$  ( $t = 1, \dots, 28$ ),

$$A_t \perp (S_k^*: k=2, \dots, 28) \mid \bar{A}_{t-1}, S_t = 1, R_t = 1, \bar{H}_t, \bar{U}_{t-1} \quad (3.1)$$

This assumption will hold if  $\bar{H}_t$  and  $\bar{U}_{t-1}$  includes all risk factors, other than treatment history, used by the physician to determine the patient's ventilator settings on day  $t$ . In the ARMA and ALVEOLI studies, there are many possible reasons why physicians could have deviated from the LTVV protocol, including respiratory acidosis resulting from LTVV and poor lung compliance (a measure of stiffness of lungs) which might cause some clinicians to increase the tidal volume in order to reduce acidosis. Therefore, arterial pH (a measurement of acidosis) and lung compliance are considered two key confounders. Other confounders considered in the analysis were baseline APACHE III score (a measurement of illness severity), organ failure, PEEP, most recent 'at-risk' status (on assist-control mode or not) and most recent LTVV adherence status.

The final assumption, which was alluded to earlier, is that missing data mechanism is ignorable, the parameters that govern a model for missing data mechanism are variation independent of those that govern the model for distribution of the data that are scheduled to be collected, and that the data that are not recorded are missing at random in the sense of Rubin (1976). Under this assumption, specification of a fully parametric model for  $L_{28}$  will

be sufficient for us to achieve our inferential objectives; the parameters of this model are identifiable from the distribution of the observed data.

#### 4. G-computation

Given the distribution of  $\bar{L}_{28}$  and the first three assumptions above, the potential survival function at day  $m$  ( $m \leq 28$ ) is identified via the following G-computation formula (Robins, 1986). The identification formula is derived as follows. By conditioning on  $H_0$ ,  $U_0$  and  $S_1$  and  $H_1$ , we can write

$$P[S_m^*=1] = \int_{h_0} \int_{u_0} \int_{h_1} P[S_m^*=1|S_1=1, \bar{H}_1=\bar{h}_1, U_0=u_0] dF(h_1|S_1=1, U_0=u_0, H_0=h_0) P[S_1=1|U_0=u_0, H_0=h_0] dF(u_0|H_0=h_0) dF(h_0)$$

Under sequential randomization,

$$P[S_m^*=1] = \int_{h_0} \int_{u_0} \int_{h_1} \underbrace{P[S_m^*=1|S_1=1, A_1=1, \bar{H}_1=\bar{h}_1, U_0=u_0]}_{\text{sequential randomization}} dF(h_1|S_1=1, U_0=u_0, H_0=h_0) P[S_1=1|U_0=u_0, H_0=h_0] dF(u_0|H_0=h_0) dF(h_0)$$

By further conditioning on  $U_1$ ,  $S_2$  and  $H_2$ ,

$$P[S_m^*=1] = \int_{h_0} \int_{u_0} \int_{h_1} \int_{u_1} \int_{h_2} P[S_m^*=1|S_2=1, A_1=1, \bar{H}_2=\bar{h}_2, \bar{U}_1=\bar{u}_1] dF(h_2|S_2=1, A_1=1, \bar{H}_1=\bar{h}_1, \bar{U}_1=\bar{u}_1) P[S_2=1|S_1=1, A_1=1, \bar{H}_1=\bar{h}_1, \bar{U}_1=\bar{u}_1]$$

Using sequential randomization,

$$P[S_m^*=1] = \int_{h_0} \int_{u_0} \int_{h_1} \int_{u_1} \int_{h_2} \underbrace{P[S_m^*=1|S_2=1, \bar{A}_2=1, \bar{H}_2=\bar{h}_2, \bar{U}_1=\bar{u}_1]}_{\text{sequential randomization}} dF(h_2|S_2=1, A_1=1, \bar{H}_1=\bar{h}_1, \bar{U}_1=\bar{u}_1) P[S_2=1|S_1=1, A_1=1, \bar{H}_1=\bar{h}_1, \bar{U}_1=\bar{u}_1]$$

By continuing to condition sequentially in time and employing sequential randomization,

$$P[S_m^*=1] = \int_{h_0} \int_{u_0} \int_{h_1} \int_{u_1} \cdots \int_{h_{m-1}} \int_{u_{m-1}} P[S_m^*=1|S_{m-1}=1, \bar{A}_{m-1}=1, \bar{H}_{m-1}=\bar{h}_{m-1}, \bar{U}_{m-1}=\bar{u}_{m-1}] \prod_{t=1}^{m-1} \left\{ P[S_t=1|S_{t-1}=1, \bar{A}_{t-1}=1, \bar{H}_{t-1}=\bar{h}_{t-1}, \bar{U}_{t-1}=\bar{u}_{t-1}] \right\}$$

Finally, by employing stochastic consistency,

$$P[S_m^*=1] = \int_{h_0} \int_{u_0} \int_{h_1} \int_{u_1} \cdots \int_{h_{m-1}} \int_{u_{m-1}} \underbrace{P[S_m=1|S_{m-1}=1, \bar{A}_{m-1}=1, \bar{H}_{m-1}=\bar{h}_{m-1}, \bar{U}_{m-1}=\bar{u}_{m-1}]}_{\text{stochastic consistency}} \prod_{t=1}^{m-1} \left\{ P[S_t=1|S_{t-1}=1, \bar{A}_{t-1}=1, \bar{H}_{t-1}=\bar{h}_{t-1}, \bar{U}_{t-1}=\bar{u}_{t-1}] \right\}$$

This latter multivariate integral has been called the ‘‘G-computation’’ integral or formula. To illustrate how the formula works, consider the following example.

##### 4.1. Example

Consider the treatment of patients with ALI on the first two days. Suppose that (1) all patients survive day 1, (2) all patients are on assist-control or ‘at-risk’ on day one, (3) 70 percent of the patients on day 1 are treated with LTVV ( $A_1 = 1$ ) and the others with traditional ventilation ( $A_1 = 0$ ), (4) the assignment of treatment on day one is completely



random, (5) organ failure on day 1 is highly associated with treatment on day 1, with patients on LTVV at much lower risk (20%) for organ failure as compared to those who are not (80%), (6) the risk of death on day 2 is higher for those with organ failure on day 1 and is lower for those treated with LTVV on day 1, (7) all surviving patients on day 2 are on assist-control ventilation on day 2, (8) treatment with LTVV on day 2 is positively associated with treatment with LTVV on day 1 and negatively associated with organ failure on day 1, (9) organ failure on day 2 is positively associated with organ failure on day 1 and negatively associated with treatment with LTVV on day 1 and (10) the risk of death on day 3 is higher for those with organ failure on day 2 and lower for those treated with LTVV on day 2. The distribution of the observed data is displayed in Figure 3.

In this example, we are interested in comparing the probability of surviving past day 3 under the LTVV regime to (a) the overall probability of surviving past day 3 and (b) the probability of surviving past day 3 among patients who are observed to adhere to the LTVV regime. Using the G-computation formula and noting that  $S_1 = 1$ ,  $H_t = \emptyset$ ,  $U_0 = \emptyset$ ,  $U_t = OF_t$  ( $t = 1, 2$ ), we obtain

$$\begin{aligned} P[S_3^* = 1] &= \sum_{o_1=0}^1 \sum_{o_2=0}^1 \left\{ P[S_3 = 1 | S_2 = 1, \bar{O} F_2 = \bar{o}_2, \bar{A}_2 = 1] P[OF_2 = o_2 | S_2 = 1, \bar{A}_2 = 1, OF_1 = o_1] \right\} \times P[S_2 = 1 | S_1 = 1, A_1 = 1, OF_1 = o_1] P[OF_1 = o_1] \\ &= 0.7 \times 0.8 \times 0.7 \times 0.2 \times 1 + 0.9 \times 0.2 \times 0.7 \times 0.2 \times 1 + 0.7 \times 0.2 \times 0.9 \times 0.8 \times 1 + 0.9 \times 0.8 \times 0.9 \times 0.8 \times 1 = 0.7228 \end{aligned}$$

In contrast, the overall three-day survival probability is 0.60 and the survival probability among those observed to adhere to the LTVV protocol is 0.844.

In this example, organ failure serves as a time-varying confounder. That is, organ failure on day 1 is affected by treatment on day 1 and predicts subsequent treatment. Organ failure on day 1 is also negatively related to survival on day 3. The three-day survival probability for those who adhere to the LTVV regime is higher than the counterfactual survival probability because those who adhere to therapy tend to have less organ failure than those who do not. The overall three-day survival probability is even lower because non-adherence to therapy is associated with greater organ failure and worse survival, regardless of organ failure status.

#### 4.2. Additional Modeling Assumptions

The G-computation formula teaches us that the  $P[S_m^* = 1]$  is a functional of the distribution of  $\mathcal{L}_{28}$ . We can write the likelihood of  $\mathcal{L}_{28}$  as follows:

$$Lik = dF(H_0) dF(U_0 | H_0) \prod_{t=1}^{28} \left\{ dF(U_t | \bar{H}_t, \bar{U}_{t-1}, S_t = 1) dF(H_t | \bar{H}_{t-1}, \bar{U}_{t-1}, S_t = 1) \right\}^{S_t} \times dF(S_t | S_{t-1} = 1, \bar{H}_{t-1}, \bar{U}_{t-1})^{S_{t-1}}.$$

In our analysis, we assume that a patient's status on day  $t$  only depends on his most recent history and the baseline APACHE III score ( $H_0$ ). Thus, the likelihood can be written as:

$$Lik = dF(H_0) dF(U_0 | H_0) \prod_{t=1}^{28} \left\{ dF(U_t | H_t, U_{t-1}, H_0, S_t = 1) dF(H_t | H_{t-1}, U_{t-1}, H_0, S_t = 1) \right\}^{S_t} \times dF(S_t | S_{t-1} = 1, H_{t-1}, U_{t-1}, H_0)^{S_{t-1}}.$$

Further modeling restrictions result from structural features of our datasets as well as from the clinical judgement of our physician co-authors (DN and RB). Structurally, we know that all patients survive day one and whenever a patient is on ventilator on day  $t - 1$ , she either dies or remains in the hospital and in the ICU at time  $t$ . Also, once a patient leaves the hospital (ICU), she does not return the hospital (ICU). With these considerations, we write

$$\begin{aligned}
 & dF(S_t|S_{t-1}=1, H_{t-1}, U_{t-1}, H_0) \\
 & = dF(S_t|R_{t-1}=1, G_{t-1}, U_{t-1}, H_0)^{R_{t-1}} \\
 & \times dF(S_t|ICU_{t-1}=0, OF_{t-1}, H_0)^{(1-ICU_{t-1})} \\
 & \times dF(S_t|VNT_{t-1}=0, OF_{t-1}, H_0)^{(1-VNT_{t-1})} \quad \text{for } t=1, \dots, 28
 \end{aligned} \tag{4.1}$$

$$dF(H_t|S_t=1, H_{t-1}, U_{t-1}, H_0) = dF(HSP_t|S_t=1, ICU_{t-1}=0, OF_{t-1}, H_0)^{(1-ICU_{t-1})} \times \tag{4.2}$$

$$dF(ICU_t|HSP_t=1, VNT_{t-1}=0, OF_{t-1}, H_0)^{HSP_t(1-VNT_{t-1})} \times \tag{4.3}$$

$$dF(VNT_t|ICU_t=1, R_{t-1}=1, G_{t-1}, U_{t-1}, H_0)^{ICU_t R_{t-1}} \times \tag{4.4}$$

$$dF(VNT_t|ICU_t=1, R_{t-1}=0, OF_{t-1}, H_0)^{ICU_t(1-R_{t-1})} \times \tag{4.5}$$

$$dF(LC_t|R_t=1, R_{t-1}=1, G_{t-1}, U_{t-1}, H_0)^{R_t R_{t-1}} \times \tag{4.6}$$

$$dF(LC_t|R_t=1, R_{t-1}=0, OF_{t-1}, H_0)^{R_t(1-R_{t-1})} \times \tag{4.7}$$

$$dF(PE_t|R_t=1, R_{t-1}=1, LC_t, PE_{t-1}, PD_{t-1}, DS_{t-1}, OF_{t-1}, H_0)^{R_t R_{t-1}} \times \tag{4.8}$$

$$dF(PE_t|R_t=1, R_{t-1}=0, LC_t, OF_{t-1}, H_0)^{R_t(1-R_{t-1})} \quad \text{for } t=1, \dots, 28 \tag{4.9}$$

$$dF(U_t|S_t=1, H_t, U_{t-1}, H_0) = dF(PH_t|R_t=1, R_{t-1}=1, LC_t, PH_{t-1}, DS_{t-1}, OF_{t-1}, H_0)^{R_t R_{t-1}} \times \tag{4.10}$$

$$dF(PH_t|R_t=1, R_{t-1}=0, LC_t, OF_{t-1}, H_0)^{R_t(1-R_{t-1})} \times \tag{4.11}$$

$$dF(PD_t|R_t=1, R_{t-1}=1, LC_t, PE_t, PD_{t-1}, H_0)^{R_t R_{t-1}} \times \tag{4.12}$$

$$dF(PD_t|R_t=1, R_{t-1}=0, LC_t, PE_t, OF_{t-1}, H_0)^{R_t(1-R_{t-1})} \times \tag{4.13}$$

$$dF(SO_t|R_t=1, R_{t-1}=1, LC_t, PE_t, PH_t, PD_t, SO_{t-1}, H_0)^{R_t R_{t-1}} \times \quad (4.14)$$

$$dF(SO_t|R_t=1, R_{t-1}=0, LC_t, PE_t, PH_t, PD_t, OF_{t-1}, H_0)^{R_t(1-R_{t-1})} \times \quad (4.15)$$

$$dF(SR_t|R_t=1, R_{t-1}=1, LC_t, PE_t, PH_t, SR_{t-1}, DS_{t-1}, OF_{t-1}, H_0)^{R_t R_{t-1}} \times \quad (4.16)$$

$$dF(SR_t|R_t=1, R_{t-1}=0, LC_t, PE_t, PH_t, OF_{t-1}, H_0)^{R_t(1-R_{t-1})} \times \quad (4.17)$$

$$dF(DS_t|R_t=1, R_{t-1}=1, LC_t, PE_t, PH_t, SO_t, SR_t, DS_{t-1}, OF_{t-1}, H_0)^{R_t R_{t-1}} \times \quad (4.18)$$

$$dF(DS_t|R_t=1, R_{t-1}=0, LC_t, PE_t, PH_t, SO_t, SR_t, OF_{t-1}, H_0)^{R_t(1-R_{t-1})} \times \quad (4.19)$$

$$dF(OF_t|R_t=1, LC_t, PE_t, PH_t, SO_t, SR_t, DS_t, OF_{t-1}, H_0)^{R_t} \times \quad (4.20)$$

$$dF(OF_t|R_t=0, R_{t-1}=1, LC_{t-1}, PE_{t-1}, PH_{t-1}, SO_{t-1}, SR_{t-1}, DS_{t-1}, OF_{t-1}, H_0)^{(1-R_t)R_{t-1}} \times \quad (4.21)$$

$$dF(OF_t|VNT_t=0, VNT_{t-1}=0, OF_{t-1}, H_0)^{(1-VNT_t)(1-VNT_{t-1})} \times \quad (4.22)$$

$$dF(OF_t|VNT_t=0, ICU_{t-1}=0, OF_{t-1}, H_0)^{(1-VNT_t)(1-ICU_{t-1})} \quad \text{for } t=1, \dots, 28 \quad (4.23)$$

where  $PD_t = PP_t - PE_t$ , conditioning on  $R_t$ ,  $HSP_t$ ,  $ICU_t$  or  $VNT_t$  implies that  $S_t = 1$ , conditioning on  $ICU_t$  or  $VNT_t$  implies  $HSP_t = 1$  and conditioning on  $VNT_t$  implies  $ICU_t = 1$ . It is important to note that we model  $PD_t$  instead of  $PP_t$  since we are conditioning on  $PE_t$  and that  $VT_t$  does not appear in the likelihood because it is a deterministic function of  $LC_t$  and  $PD_t$  (i.e.,  $VT_t = LC_t PD_t$ ). The distributions on the right hand side of the above three displays are modeled with separate regression models. The binary outcomes,  $S_t$ ,  $HSP_t$ ,  $ICU_t$ ,  $VNT_t$ ,  $DS_t$  and  $OF_t$ , are modeled using logistic regression. The continuous variables,  $H_0$ ,  $LC_t$ ,  $PE_t$ ,  $PH_t$ ,  $PD_t$ ,  $SO_t$ , and  $SR_t$ , are linearly transformed to the range [0, 1] (using the bounds displayed in Table 4) and modeled using Beta regression models (Ferrari and Cribari-Neto, 2004). In these models, the logit of the mean ( $\mu$ ) is modeled as a function of covariates and the variance is equal to  $\mu(1-\mu)/(\eta+1)$ , where  $\eta$  is a scalar dispersion parameter. In all regression models, there are separate intercepts for days 1 to 13, and common intercepts for days 14 to 21 and days 22-28; there are no interactions between covariates; the regression coefficients for time-varying covariates are assumed to be time-invariant as the dispersion parameters (in the Beta regression models).

### 4.3. Inference

We adopt a Bayesian approach to inference. We assume independently normally distributed mean zero, variance 100 priors on all the regression parameters. For each model, we built up a shrinkage prior on the time-varying intercept parameters as follows: the intercept on day one was assumed to be normally distributed with mean zero with variance 5, the intercept on day  $t$  ( $t = 2, \dots, 13$ ) was assumed to be normally distributed with mean equal to the intercept on day  $t - 1$  and variance  $\sigma^2$ , the common intercept for days 14 to 21 (days 22 to 28) was assumed to be normally distributed with mean equal to the intercept on day 13 (day 21) and variance  $\sigma^2$ , and  $\sigma^2$  was assumed to be uniform on the interval  $[0, 5]$ . Finally, the dispersion parameters  $\eta$  in the Beta regression models were assumed to be uniform on the interval  $[0, 100]$ .

The posterior sampling for the model parameters was performed in WinBUGS. Twelve-thousand iterations of which six-thousand were used for burn-in were performed. Trace plots of multiple chains were examined to check for convergence. For each iteration, we evaluated the G-computation formula by Monte-Carlo integration. The Monte-Carlo scheme is complicated by the fact that the integration requires us to sample from  $dF(S_t|S_{t-1} = 1, A_{t-1}, H_{t-1}, U_{t-1}, H_0)$ ,  $dF(H_t|S_t = 1, A_{t-1} = 1, H_{t-1}, U_{t-1}, H_0, S_t = 1)$ ,  $dF(U_t|S_t = 1, A_t = 1, H_t, U_{t-1}, H_0)$  instead of  $dF(S_t|S_{t-1} = 1, H_{t-1}, U_{t-1}, H_0)$ ,  $dF(H_t|S_t = 1, H_{t-1}, U_{t-1}, H_0)$ ,  $dF(U_t|S_t = 1, H_t, U_{t-1}, H_0)$ , where we know that  $A_t$  is a deterministic function of  $U_t$ .

Our sampling scheme proceeds according to Figure 4. For the sake of completeness, we document the procedure in each of the letter-marked boxes. In Box (A), the model in (4.1) is used. The sampling in Box (B) proceeds as follows:

- If  $ICU_{t-1} = 0$ , sample  $HSP_t$  from (4.2), else set  $HSP_t = 1$ .
- If  $ICU_{t-1} = 1$  and  $VNT_{t-1} = 0$ , sample  $ICU_t$  from (4.3), else if  $R_{t-1} = 1$ , set  $ICU_t = 1$ , else set  $ICU_t = 0$ .
- If  $R_{t-1} = 1$ , sample  $VNT_t$  from (4.4) else sample  $VNT_t$  from (4.5);

the sampling in Box (C) proceeds as follows:

- If  $R_{t-1} = 1$ , sample  $LC_t$  from (4.6), else sample  $LC_t$  from (4.7);

the sampling in Box (D) proceeds as follows:

- If  $R_{t-1} = 1$ , sample  $PE_t$  from (4.8), sample  $PH_t$  from (4.10), sample  $PD_t$  from (4.12), sample  $SO_t$  from (4.14), sample  $SR_t$  from (4.16) and sample  $DS_t$  from (4.18); else sample  $PE_t$  from (4.9), sample  $PH_t$  from (4.11), sample  $PD_t$  from (4.13), sample  $SO_t$  from (4.15), sample  $SR_t$  from (4.17) and sample  $DS_t$  from (4.19);

and sampling in Box (E) proceeds as follows:

- If  $R_t = 1$ , sample  $OF_t$  from (4.20), else if  $R_t = 0$  and  $R_{t-1} = 1$ , sample  $OF_t$  from (4.21), else if  $R_t = 0$ ,  $R_{t-1} = 0$ ,  $ICU_{t-1} = 1$ , sample  $OF_t$  from (4.22), else sample  $OF_t$  from (4.23).

This scheme is repeated 10,000 times and the resulting survival curves are averaged. As this is performed for each posterior draw from the model parameters, we obtain 6000 posterior draws of the survival curve of interest.

To evaluate, at least in part, the goodness of fit of our model, we can use a modification of the above procedure to obtain posterior draws for the distribution of the “observed” survival curve and compare it to the empirically observed survival curve. We note that we can express the “observed” survival curve as the following integral:

$$P[S_m=1] = \int_{h_0} \int_{u_0} \int_{h_1} \int_{u_1} \cdots \int_{h_{m-1}} \int_{u_{m-1}} P[S_m=1 | S_{m-1}=1, H_{m-1}=h_{m-1}, U_{m-1}=u_{m-1}] \prod_{t=1}^{m-1} \{P[S_t=1 | S_{t-1}=1, H_{t-1}=h_{t-1}, U_{t-1}=u_{t-1}]\} dF(h_t | S)$$

This suggests that we can obtain posterior draws by using the same algorithm above, except that, in Figure 4, we omit the evaluation of whether  $A_t = 1$  or not.

## 5. Analysis of ARMA and ALVEOLI Studies

In this section, we apply the G-computation formula to the ARMA and ALVEOLI trials. Figure 5 and Table 5 present the empirical observed (gray lines), posterior mean observed (black dashed lines), and posterior mean potential (black lines) survival curves for the ARMA and ALVEOLI studies. The table includes 95% posterior credible bands for these three curves and the figure depicts the confidence band for the potential survival curve by the shaded region. For reference, the empirical survival curves for those who were *observed to adhere* to the LTVV algorithm are included in the figures (gray dashed lines); this represents 92 and 141 patients in the ARMA and ALVEOLI studies, respectively. By visual comparison of the empirical observed and posterior mean observed survival curves, empirical and model-based, we see that our model provides a reasonable fit to the observed data. The posterior mean potential survival curve is below the posterior mean observed survival curve in the ARMA study, but the curves are barely indistinguishable for the ALVEOLI study. In both studies, the empirical survival curve for the adherers drops steeply during the first 7 days and then tapers off. This reflects the fact that those who are observed to adhere are those who either die or recover quickly, so that their need for ventilation is limited; those with longer ventilation periods are less likely to be reflected in these curves as they have a greater chance of violating the LTVV protocol or have missing data that precludes evaluation of their adherence status.

In displays not shown, we evaluated the posterior mean difference between the observed and potential survival curves along with a 95% posterior credible band. In both studies, the 95% credible bands for the difference include zero for all time, which does not statistically support the theory that imperfect adherence to the protocol led to improved outcomes.

Despite these non-significant findings, it is interesting to speculate why the potential survival curve under LTVV adherence is noticeably lower than the observed survival curve in the ARMA study, but virtually identical in the ALVEOLI study. We suspect that, with the positive results in ARMA, many doctors were more comfortable using a strategy that targeted lower tidal volumes and plateau pressures. With these data, doctors were more willing to allow their patients to stay on lower tidal volumes. With more experience they learned that patients may experience some “agitation” (e.g., dysynchrony with the mechanical ventilator) early in their clinical course on lower tidal volumes, but this “agitation” improves within a couple of days. In ALVEOLI, the use of heavy sedation or neuromuscular blockade (which would address this dysynchrony with the mechanical ventilator) was more limited, primarily occurring in the first couple of days, whereas in ARMA many patients remained heavily sedated or neuromuscular-blocked for more extended periods of time. It is this difference in medication use which may explain the discrepancy between the results across trials.

## 6. Discussion

In this paper, we used the G-computation formula to estimate the potential survival distribution under the dynamic LTVV treatment regime for patients with acute lung injury.

Our results demonstrate that strict adherence to the LTVV algorithm does not lead to increased mortality, as has been suggested in the clinical literature.

Our methodology relies on the untestable assumption of no unmeasured confounding. That is, we have assumed that we have measured all the time-dependent confounders that, when conditioned upon, make compliance with LTVV on a given day, a random process, akin to flipping a possibly biased coin (Assumption 3.1). While this assumption cannot be empirically validated from the observed data, the clinical experts on our team (Dale Needham and Roy Brower) have come to feel comfortable with this assumption. The key question here is why some physicians deviate from the LTVV protocol. The concern of such physicians is that a low tidal volume may cause worse gas exchange, especially among those hard patients with stiff lungs (lungs that are more difficult to ventilate). Gas exchange can be measured by arterial pH and the stiffness of lungs can be measured by lung compliance. With these two variables and a few others included, we feel that we have captured the key factors that govern adherence to the LTVV protocol. Nonetheless, future work should focus on exploring the sensitivity of our conclusions to deviations from this assumption.

The G-computation has been criticized because it can suffer from the “null paradox.” When one uses the G-computation formula to estimate the means of a collection of counterfactual outcomes indexed by levels of treatment, the paradox states that one can falsely reject the null hypothesis of no variation in means across of levels of treatment because of the need (due to the curse of dimensionality) to parametrically specify the component distributions in the G-computation integral. Since we are not using the G-computation formula in this fashion, we do not think this paradox is a major concern.

There are alternatives to our G-computation analysis. Murphy *et al.* (2001) proposed an inverse-weighting procedure in which individuals who adhere to the dynamic treatment regime are inverse-weighted by their probability of being adherent through their course of follow-up. In our study, these weights are the inverse of the product of as many as 28 conditional probabilities. In implementing this procedure, we found that some of the weights became prohibitively large and the variances of the resulting estimates were too large to be of practical use. Truncation of weights was considered as a way of trading off bias with variance, but the results were highly sensitive to the choice of the truncation threshold. Weight stabilization is not an option here as our interest is in the marginal survival curve.

Another approach that has been advocated is structural nested failure time models (Lok et al., 2004). In these models, one specifies, for each day  $t$ , the conditional (on time-varying factors up to day  $t$ ) quantile-quantile mapping between the distribution of the counterfactual survival time had patients followed the dynamic regime through day  $t$  and not followed the regime after day  $t$  and the distribution of the counterfactual survival time had patients followed the regime through day  $t - 1$  and not followed the regime after day  $t$ . This is an interesting avenue to explore, which may ultimately involve less overall modeling than the G-computation algorithm and does not suffer from the “null paradox”.

We hope that this unique illustration of the complexities of analyzing a dynamic treatment regime will serve to make these methods more accessible to statisticians and their clinical collaborators.

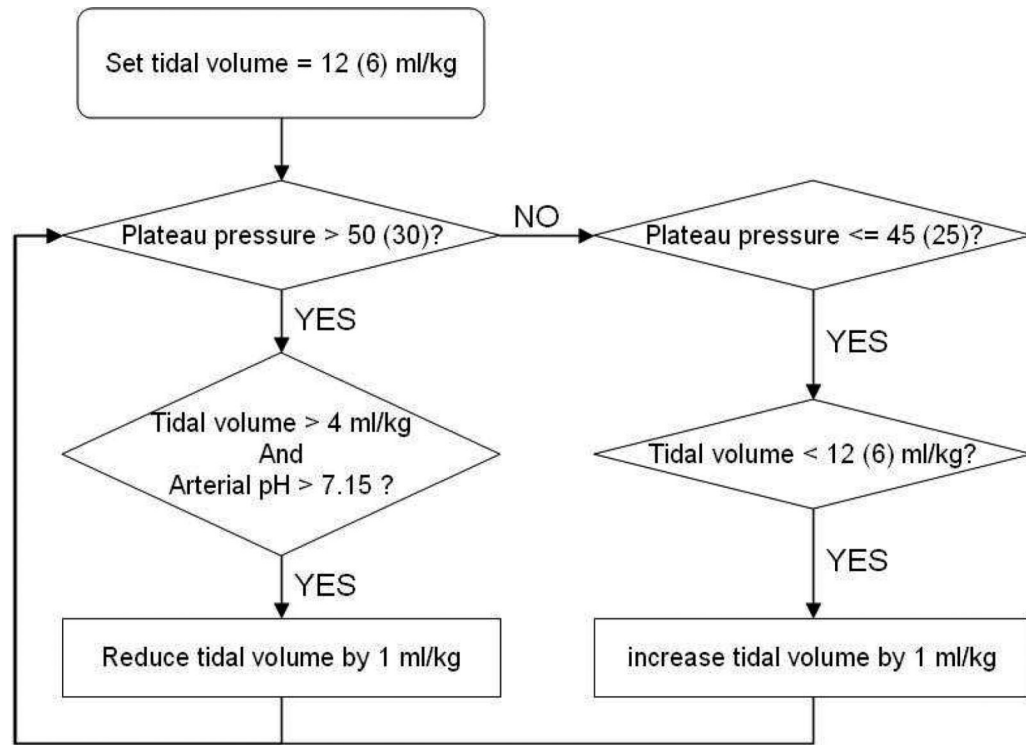
## References

- Bembom O, van der Laan MJ. Analyzing sequentially randomized trials based on causal effect models for realistic individualized treatment rules. *Statistics in Medicine*. 2008; 27:3689–3716. [PubMed: 18407580]

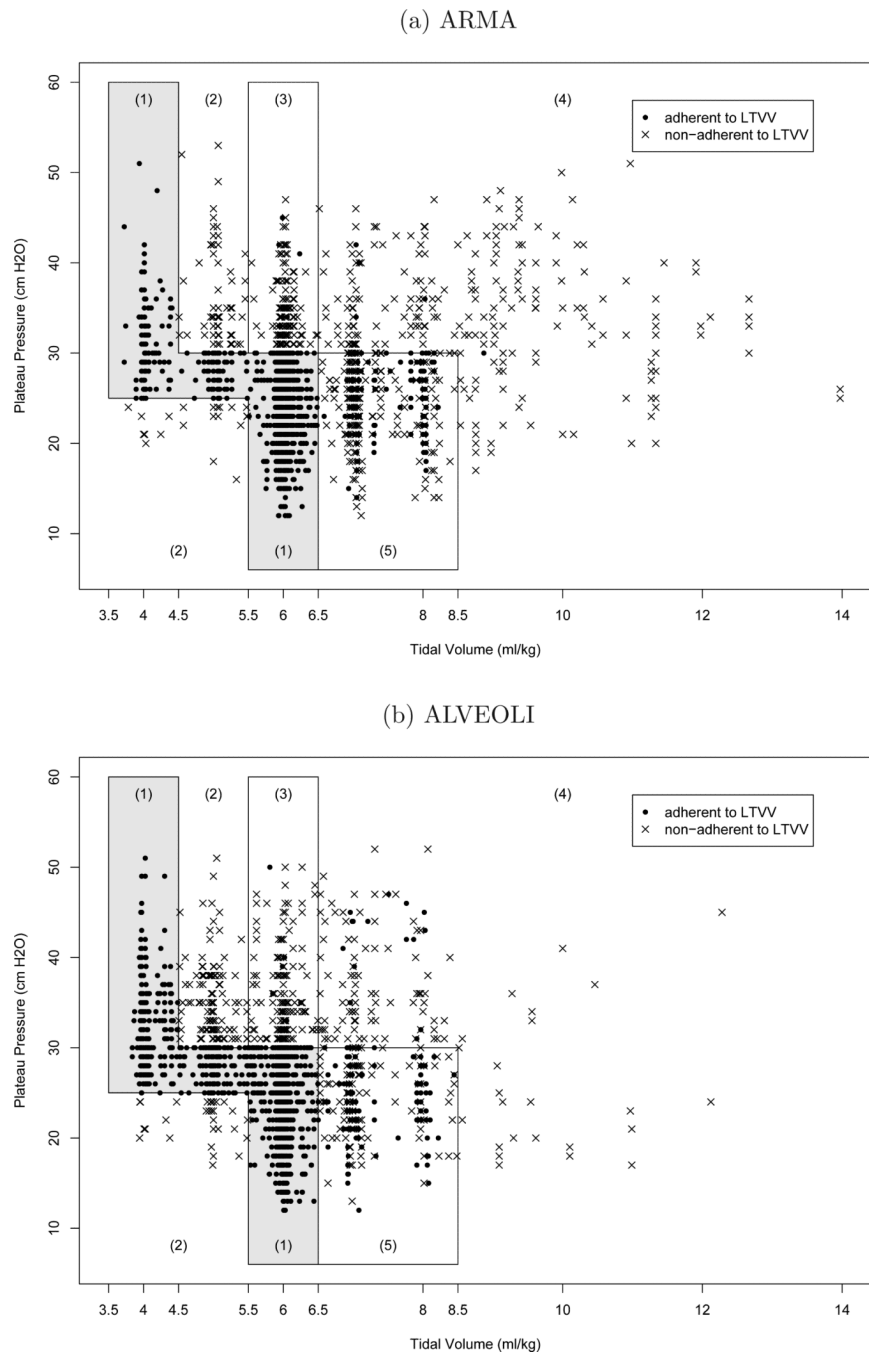
- Bernard GR, Artigas A, Brigham KL. The American-European Consensus Conference on ARDS. Definitions, mechanisms, relevant outcomes, and clinical trial coordination. *American Journal of Respiratory and Critical Care Medicine*. 1994; 149:818–24. e. a. [PubMed: 7509706]
- Chakraborty B, Murphy S, Strecher V. Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical Methods in Medical Research*. 2010; 19:317–343. [PubMed: 19608604]
- Deans KJ, Minneci PC, Cui X, Banks SM, Natan-Son C, Eichacker PQ. Tidal volumes in acute respiratory distress syndrome: One size does not fit all. *Critical Care Medicine*. 2006; 34:264–267.
- Deans KJ, Minneci PC, Cui X, Banks SM, Natanson C, Eichacker PQ. Mechanical ventilation in ards: One size does not fit all. *Critical Care Medicine*. 2005; 33:1141–1143. [PubMed: 15891350]
- Dreyfuss D, Saumon G. Ventilator-induced lung injury: Lessons from experimental studies. *American Journal of Respiratory and Critical Care Medicine*. 1998; 157:294–323. [PubMed: 9445314]
- Eichacker PQ, Gerstenberger EP, Banks SM, Cui X, Natanson C. A meta-analysis of ali and ards trials testing low tidal volumes. *American Journal of Respiratory and Critical Care Medicine*. 2002; 166:1510–1514. [PubMed: 12406836]
- Fan E, Needham D, Stewart T. Ventilatory management of acute lung injury and acute respiratory distress syndrome. *Journal of the American Medical Association*. 2005; 294(22):2889–2896. [PubMed: 16352797]
- Ferrari S, Cribari-Neto F. Beta regression for modelling rates and proportions. *Journal of Applied Statistics*. 2004; 31(7):799–815.
- Hernan M, Lanoy E, Costagliola D, Robins J. Comparison of dynamic treatment regimes via inverse probability weighting. *Basic and Clinical Pharmacology and Toxicology*. 2006; 98:237–242. [PubMed: 16611197]
- Hernan MA, Robins JM. Estimating causal effects from epidemiological data. *Journal of Epidemiology and Community Health*. 2006; 60:578–586. [PubMed: 16790829]
- Kalbfleisch, JD.; Prentice, RL. *The Statistical Analysis of Failure Time Data*. 2nd ed.. Wiley; 2002.
- Lok J, Gill R, van der Vaart A, Robins J. Estimating the causal effect of a time-varying treatment on time-to-event using structural nested failure time models. *Statistica Neerlandica*. 2004; 58(3):271–295.
- Moodie E, Richardson T, Stephens D. Demystifying optimal dynamic treatment regimes. *Biometrics*. 2007; 63:447–455. [PubMed: 17688497]
- Moodie EM. A note on the variance of doubly-robust g-estimators. *Biometrika*. 2009a; 96:998–1004.
- Moodie EM. Risk factor adjustment in marginal structural model estimation of optimal treatment regimes. *Biometrical Journal*. 2009b; 51:774–788. [PubMed: 19816876]
- Moodie EM, Richardson T. Estimating optimal dynamic regimes: Correcting bias under the null. *Scandinavian Journal of Statistics*. 2010; 37:126–146.
- Murphy S. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society - Series B*. 2003; 65:331–355.
- Murphy S, van der Laan M, Robins J. Marginal mean models for dynamic regimes. *Journal of the American Statistical Association*. 2001; 96:1410–1423. [PubMed: 20019887]
- Orellana L, Rotnitzky A, Robins J. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes. *The International Journal of Biostatistics*. 2010; 6
- Rich B, Moodie EM, Stephens D, Platt R. Model checking with residuals for g-estimation of optimal dynamic treatment regimes. *International Journal of Biostatistics*. 2010; 6
- Robins J. A new approach to causal inference in mortality studies with sustained exposure periods: Application to control of the healthy worker survivor effect. *Mathematical Modelling*. 1986; 7:1393–1512.
- Robins J. Addendum to "A new approach to causal inference in mortality studies with sustained exposure periods: Application to control of the healthy worker survivor effect". *Computers and Mathematics with Applications*. 1987a; 14:923–945.
- Robins J. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *Journal of Chronic Disease*. 1987b; 40:139S–161S.

- Robins J. The control of confounding by intermediate variables. *Statistics in Medicine*. 1988a; 8:679–701. [PubMed: 2749074]
- Robins, J. *Health Service Research Methodology: A Focus on AIDS*. U.S. Public Health Service, National Center for Health Services Research; Washington, D.C.: 1988b. Chapter The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies; p. 113-159.
- Robins, J. Structural nested failure time models.. In: Armitage, P.; Colton, T., editors. *The Encyclopedia of Biostatistics*. John Wiley and Sons; Chichester, UK: 1997.
- Robins, J. Optimal structural nested models for optimal sequential decisions.. In: Lin, D.; Heagerty, P., editors. *Proceedings of the Second Seattle Symposium on Biostatistics*. Springer; 2004.
- Robins J, Hernan M, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000; 11:550–560. [PubMed: 10955408]
- Robins J, Orellana L, Rotnitzky A. Estimation and extrapolation of optimal treatment and testing strategies. *Statistics in Medicine*. 2008; 27:4678–4721. [PubMed: 18646286]
- Robins JM. Association, causation, and marginal structural models. *Synthese*. 1999a; 121:151–179.
- Robins, JM. *Statistical Models in Epidemiology: The Environment and Clinical Trials*. Springer-Verlag; New York: 1999b. Chapter Marginal structural models versus structural nested models as tools for causal inference; p. 95-134.
- Rubinfeld G, Caldwell E, Peabody E, Weaver J, et al. Incidence and outcomes of acute lung injury. *The New England Journal of Medicine*. 2005; 353:1685–93. [PubMed: 16236739]
- Rubin DB. Comments on “Statistics and causal inference”. *Journal of the American Statistical Association*. 1986; 81:961–962.
- The Acute Respiratory Distress Syndrome Network. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. *The New England Journal of Medicine*. 2000; 342:1301–1308. [PubMed: 10793162]
- The National Heart, Blood and Lung Institute ARDS Clinical Trials Network. Higher versus lower positive end-expiratory pressures in patients with the acute respiratory distress syndrome. *The New England Journal of Medicine*. 2004; 351:327–336. [PubMed: 15269312]
- Thompson B, Hayden D, Matthay M, Brower R, Parsons P. Clinicians approaches to mechanical ventilation in acute lung injury and ards. *Chest*. 2001; 120:1622–1627. [PubMed: 11713144]
- Tobin M. Culmination of an era in research on the acute respiratory distress syndrome. *The New England Journal of Medicine*. 2000; 342:1360–1361. [PubMed: 10793169]
- Tobin M, Chadha TS, Jenouri G, Birch SJ, Gazeroglu HB, Sackner MA. Breathing patterns. *Chest*. 1983; 84:202–206. [PubMed: 6872603]
- Ware LB, Matthay MA. The acute respiratory distress syndrome. *New England Journal of Medicine*. 2000; 342:1334–49. [PubMed: 10793167]
- Webb HH, Tierney DF. Experimental pulmonary edema due to intermittent positive pressure ventilation with high pressures. *American Review of Respiratory Disease*. 1974; 110:556. [PubMed: 4611290]
- Zhao Y, Kosorok M, Zeng D. Reinforcement learning design for cancer clinical trials. *Statistics in Medicine*. 2009; 28:3294–3315. [PubMed: 19750510]

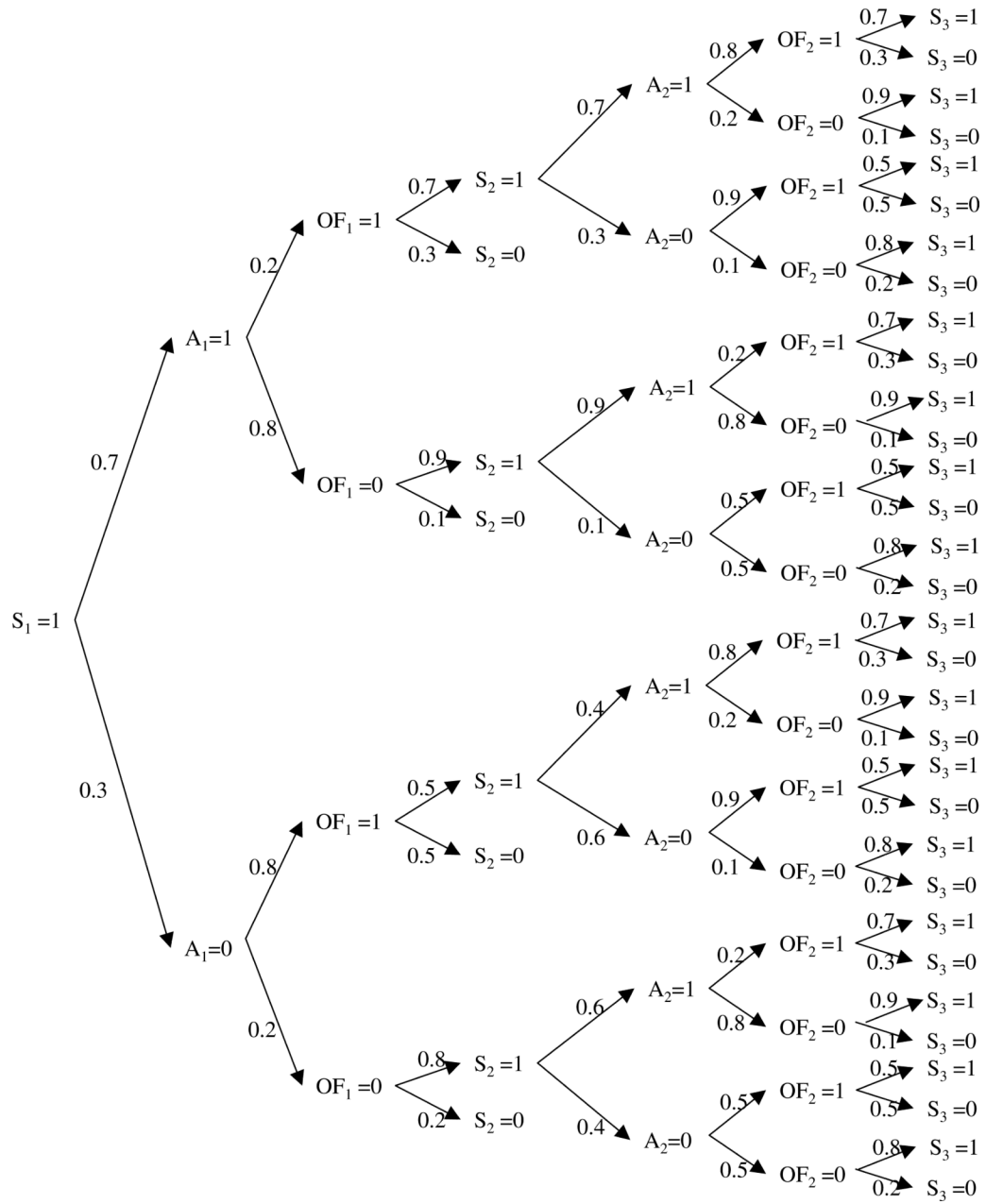




**Fig. 1.** Flowchart of the traditional and low tidal volume ventilation strategies in ARMA study. Numbers in parentheses are for low tidal volume ventilation.



**Fig. 2.** Observed plateau pressures and tidal volumes. Additional restrictions of LTVV adherence for points in region (1) none (3)  $pH < 7.2$  (4)  $pH < 7.2$  and set respiratory rate  $\geq 35$  (5) ( $pH < 7.2$  and set respiratory rate  $\geq 35$ ) or severe dyspnea. Points in region (2) are not adherent.



**Fig. 3.** Hypothetical example of G-Computation

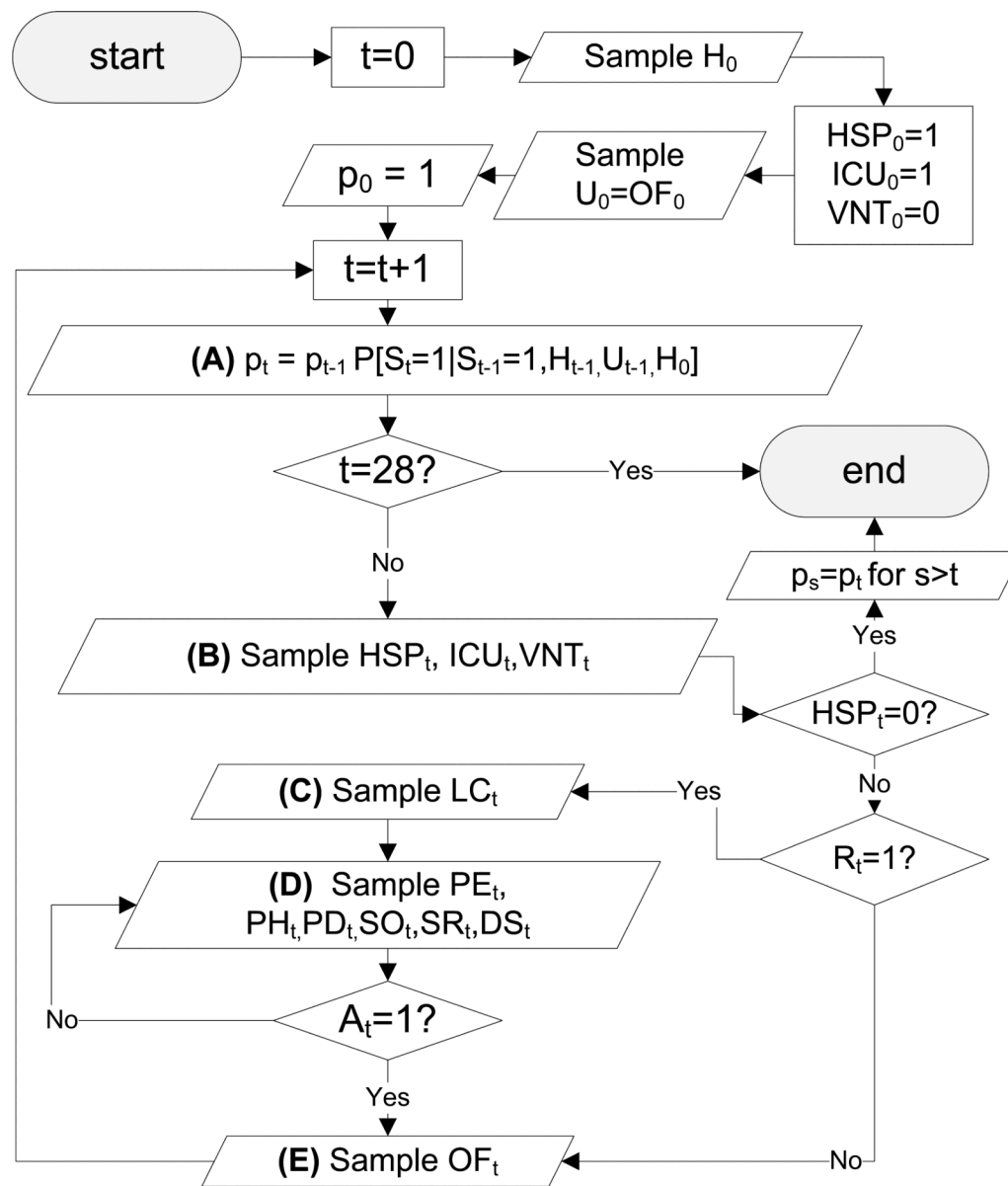
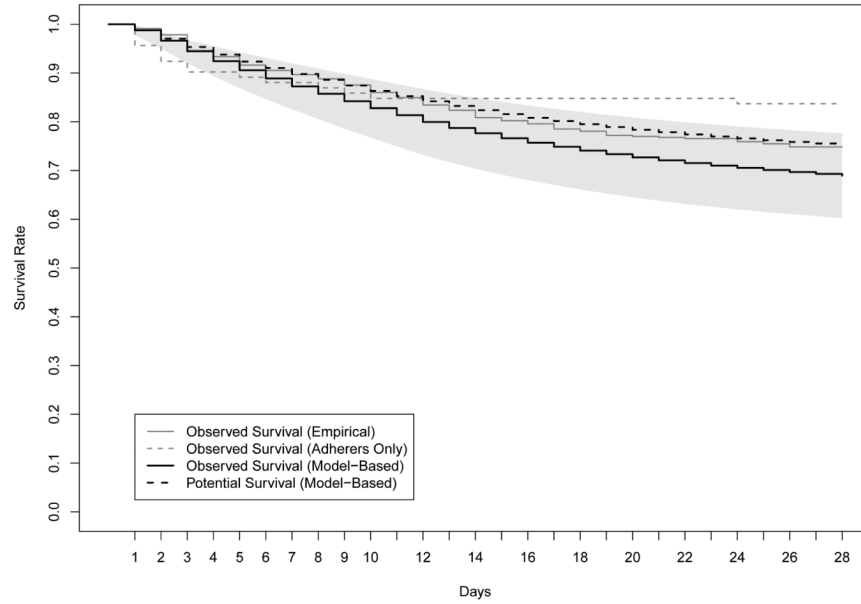
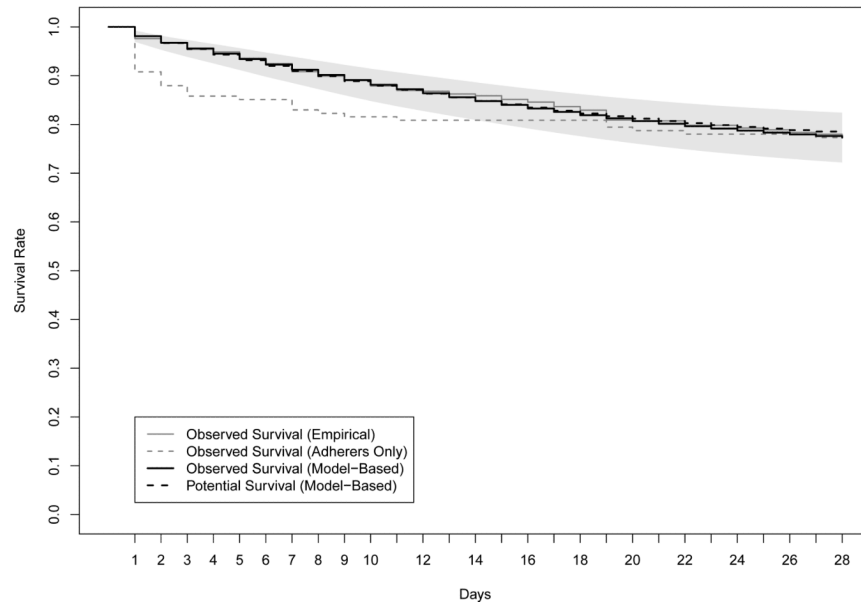


Fig. 4. Sampling scheme

(a) ARMA



(b) ALVEOLI



**Fig. 5.** Empirical observed survival curves (gray line). Empirical survival curve for adherers (gray dashed line). Posterior mean observed (black dashed line) and potential (black line) survival curves. Shaded area denote 95% posterior credible band for the potential survival curve.

**Table 1**

Specifications to Determine Adherence to LTVV

	<b>Plateau pressure specification</b>	<b>Additional specification</b>
1	$\leq 30$	
2	$> 30$	tidal volume $\leq 4.5$ or pH $< 7.2$
	<b>Tidal volume specification</b>	<b>Additional specification</b>
3	tidal volume $\geq 8.5$	pH $< 7.2$ and set respiratory rate $\leq 35$
4	$6.5 < \text{tidal volume} \leq 8.5$	(pH $< 7.2$ and set respiratory rate $\geq 35$ ) or (plateau pressure $\leq 30$ and severe dyspnea)
5	$5.5 \leq \text{tidal volume} \leq 6.5$	Plateau pressure $\leq 30$ or pH $< 7.2$
6	$4.5 < \text{tidal volume} < 5.5$	$25 \leq \text{Plateau pressure} \leq 30$
7	$3.5 \leq \text{tidal volume} \leq 4.5$	Plateau pressure $\geq 25$

**Table 2**

Missing values in the ARMA and ALVEOLI Studies

	ALI (n=2994)	ALVEOLI (n=2908)
Tidal Volume	2.8%	1.7%
Plateau Pressure	11.6%	13.2%
Set Rate	3.3%	1.4%
Arterial pH	28.7%	8.7%
Oxygen Saturation	16.7%	33.9%
PEEP	2.5%	1.4%

\*  $n$  denotes the number of at-risk person-days

Table 3

## Variable Definitions

Var	Definition and Description
$L_t$	$(S_t, H_t, U_t), t = 0, \dots, 28$
$S_t$	Indicator of surviving day $t$ , $S_0 = 1$
$HSP_t$	Indicator of hospitalization on day $t$ (defined only when $S_t = 1$ ) If $HSP_t = 0$ then $S_m = 1$ for all $m > t$ . $HSP_0 = 1$
$ICU_t$	Indicator of being in ICU on day $t$ (defined when $S_t = HSP_t = 1$ ) If $ICU_t = 0$ then $ICU_m = 0$ for $m > t$ . If $ICU_t = 1$ then $HSP_{t+1} = 1$ . $ICU_0 = 1$
$VNT_t$	Indicator of being on ventilator on day $t$ (defined when $S_t = HSP_t = ICU_t = 1$ ) If $VNT_t = 1$ then $HSP_{t+1} = ICU_{t+1} = 1$ . $VNT_0 = 0$
$R_t$	$I(HSP_t = ICU_t = VNT_t = 1)$ (defined when $S_t = 1$ )
$G_t$	$(PE_t, LC_t)$ (defined when $S_t = R_t = 1$ )
$D_t$	$(VT_t, SR_t)$ (defined when $S_t = R_t = 1$ )
$W_t$	$(PH_t, PPr, DS_t, SO_t)$ (defined when $S_t = R_t = 1$ )
$OF_t$	Indicator of organ failure (defined when $S_t = HSP_t = 1$ )
$H_t$	$(HSP_t, ICU_t, VNT_t, G_t)$ (defined when $S_t = 1$ ); $H_0 = AP$
$U_t$	$(D_t, W_t, OF_t)$ (defined when $S_t = 1$ ); $U_0 = OF_0$
$PE_t$	Positive end-expiratory pressure (defined when $S_t = R_t = 1$ )
$PP_t$	Plateau pressure (defined when $S_t = R_t = 1$ )
$PD_t$	$PP_t - PE_t$ (defined when $S_t = R_t = 1$ )
$VT_t$	Tidal volume (defined when $S_t = R_t = 1$ )
$LC_t$	Lung compliance, $VT_t/PD_t$ (defined when $S_t = R_t = 1$ )
$SR_t$	Respiratory set rate (defined when $S_t = R_t = 1$ )
$SO_t$	Oxygen saturation (defined when $S_t = R_t = 1$ )
$PH_t$	pH (defined when $S_t = R_t = 1$ )
$DS_t$	Indicator of severe dyspnea (defined when $S_t = R_t = 1$ )
$AP$	Baseline APACHE III score
$A_t$	Indicator of LTVV adherence, $I((D_t, W_t) \in S)$ , where $S$ is defined in Table 1 If $S_t = 0$ or if $S_t = 1$ and $R_t = 0$ , $A_t = 1$ .



**Table 4**

Bounds of selected variables

<b>Variable</b>	<b>Minimum</b>	<b>Maximum</b>
APACHE III Score ( $AP$ )	0	299
Lung Compliance ( $LC_t$ )	0.11	1.26
PEEP ( $PE_t$ )	0	28
pH ( $PH_t$ )	6.6	7.6
Plateau Pressure - PEEP ( $PD_t$ )	7	40
Oxygen Saturation ( $SO_t$ )	45	100
Set Rate ( $SR_t$ )	6	40

**Table 5**

## Survival Rates with 95% Confidence Band

Study ARMA			
Day	Observed (Empirical)	Observed (Model-Based)	Potential
1	0.991(0.964, 0.998)	0.988(0.978, 0.997)	0.988(0.978, 0.997)
2	0.978(0.947, 0.991)	0.971(0.957, 0.984)	0.966(0.951, 0.982)
3	0.946(0.906, 0.970)	0.953(0.935, 0.972)	0.944(0.921, 0.968)
4	0.933(0.890, 0.960)	0.938(0.916, 0.960)	0.924(0.894, 0.955)
5	0.916(0.870, 0.946)	0.923(0.898, 0.949)	0.906(0.869, 0.942)
6	0.905(0.857, 0.938)	0.910(0.882, 0.939)	0.889(0.847, 0.931)
7	0.897(0.847, 0.931)	0.898(0.867, 0.929)	0.873(0.826, 0.920)
14	0.809(0.749, 0.855)	0.824(0.783, 0.865)	0.776(0.704, 0.849)
21	0.768(0.705, 0.819)	0.779(0.731, 0.827)	0.721(0.638, 0.804)
28	0.748(0.684, 0.801)	0.752(0.700, 0.805)	0.690(0.603, 0.776)

Study ALVEOLI			
Day	Observed (Empirical)	Observed (Model-Based)	Potential
1	0.976(0.948, 0.989)	0.981(0.970, 0.992)	0.981(0.970, 0.992)
2	0.967(0.936, 0.983)	0.967(0.952, 0.981)	0.968(0.953, 0.982)
3	0.954(0.920, 0.974)	0.955(0.937, 0.972)	0.956(0.939, 0.973)
4	0.949(0.913, 0.970)	0.943(0.923, 0.963)	0.945(0.925, 0.965)
5	0.936(0.897, 0.960)	0.932(0.909, 0.955)	0.934(0.912, 0.956)
6	0.925(0.884, 0.951)	0.920(0.895, 0.945)	0.923(0.898, 0.947)
7	0.908(0.865, 0.938)	0.910(0.882, 0.937)	0.912(0.885, 0.939)
14	0.859(0.809, 0.896)	0.848(0.812, 0.884)	0.848(0.809, 0.886)
21	0.807(0.753, 0.851)	0.807(0.766, 0.849)	0.801(0.755, 0.848)
28	0.778(0.721, 0.825)	0.783(0.737, 0.828)	0.773(0.722, 0.824)