An Empirical Method of Detecting Time-Dependent Confounding: An Observational Study of Next Day Delirium in a Medical ICU

T.E. Murphy[1,*], P.H. Van Ness[1], K.L.B. Araujo[1] and M.A. Pisani[2]

[1]*Yale Program on Aging, Yale School of Medicine, New Haven, CT, USA*

[2]*Section of Pulmonary, Critical Care and Sleep Medicine, Yale School of Medicine, New Haven, CT, USA*

Abstract: Longitudinal research on older persons in the medical intensive care unit (MICU) is often complicated by the time-dependent confounding of concurrently administered interventions such as medications and intubation. Such temporal confounding can bias the respective longitudinal associations between concurrently administered treatments and a longitudinal outcome such as delirium. Although marginal structural models address time-dependent confounding, their application is non-trivial and preferably justified by empirical evidence. Using data from a longitudinal study of older persons in the MICU, we constructed a plausibility score from 0 – 10 where higher values indicate higher plausibility of time-dependent confounding of the association between a time-varying explanatory variable and an outcome. Based on longitudinal plots, measures of correlation, and longitudinal regression, the plausibility scores were compared to the differences in estimates obtained with non-weighted and marginal structural models of next day delirium. The plausibility scores of the three possible pairings of daily doses of fentanyl, haloperidol, and intubation indicated the following: low plausibility for haloperidol and intubation, moderate plausibility for fentanyl and haloperidol, and high plausibility for fentanyl and intubation. Comparing multivariable models of next day delirium with and without adjustment for time-dependent confounding, only intubation's association changed substantively. In our observational study of older persons in the MICU, the plausibility scores were generally reflective of the observed differences between coefficients estimated from non-weighted and marginal structural models.

INTRODUCTION

Observational studies of medication use among older patients in a medical intensive care unit (MICU) are complicated by myriad clinical and statistical issues [1-7]. Critically ill patients often concurrently receive multiple treatments such as intubation and differing families of medications such as sedatives and antipsychotics [8-12]. This makes it hard to disentangle the impact of intubation from medication use on outcomes during critical illness. If these concurrently administered treatments exhibit time-dependent confounding, and assuming compliance with pertinent assumptions such as the absence of any unmeasured confounders, a marginal structural model can adjust for any bias contributed by time-dependent confounding [13-16]. Because the implementation of a marginal structural model can be complex, it would be useful to have a simple, empirical measure indicating the relative plausibility of such confounding. In this report we propose a simple procedure for this purpose.

METHODS

Definition of Simple Confounding

We refer to simple confounding as that which is not time dependent. It occurs when a covariate is associated with the primary explanatory variable as well as the outcome. Simple confounding is not within the causal pathway between the primary variable and the outcome. In order to get a more accurate estimate of the association between primary variables and outcomes, inclusion of important confounders and their potential interactions is standard practice in multivariable regression.

Definition of Time-Dependent Confounding (TDC)

When estimating the longitudinal association between a time-dependent explanatory variable, e.g.., a time varying treatment such as daily dose of haloperidol, and a longitudinal outcome such as next day diagnosis of delirium, a special type of confounding that can occur is time-dependent confounding (TDC). TDC can also be introduced if there are other concurrent time-varying treatments (or covariates) that may themselves be predictors of the outcome and/or influence subsequent levels of the treatment of interest. It is also notable that past treatments of interest may influence subsequent levels of the time-dependent covariates. For purposes of illustration and to follow the structure presented by Robins, Hernan, and Brumback [16], we will describe a common treatment and disease scenario for critically ill older persons in the MICU. Figure **1** depicts some measured covariates (Covars), some unmeasured confounders (Unmeasured), and a treatment (Intubated) being evaluated. At the far right

*Address correspondence to this author at the Yale Program on Aging, 300 George St, Suite 775, New Haven, CT 06511, USA; Tel: 203-737-2295; Fax: 203-785-4823; E-mail: Terrence.murphy@yale.edu
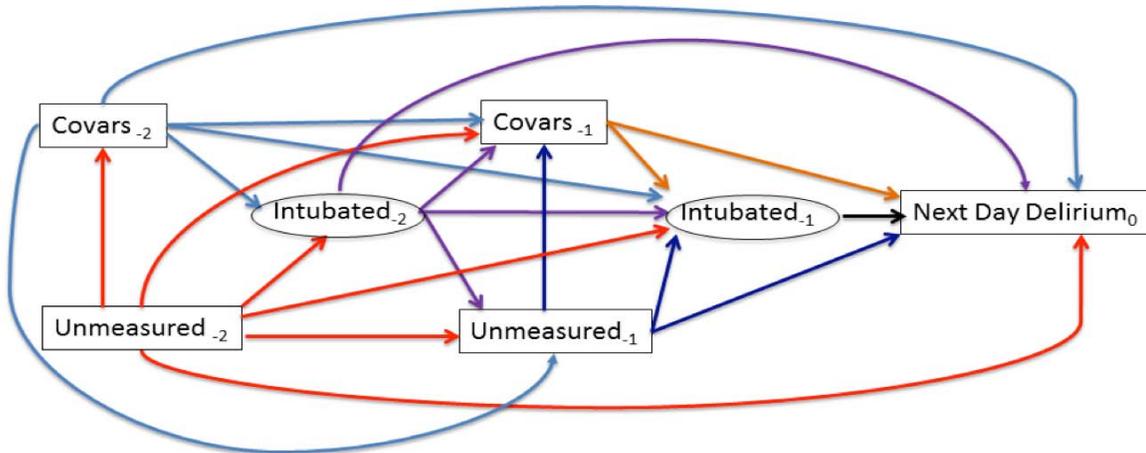
Figure 1: Illustration of Time-Dependent Confounding in the MICU.

of the figure is the outcome being modeled, i.e., a diagnosis of next day delirium. Intubation is performed on a high proportion of critically ill older patients, and the use of intubation is reasonably influenced by several measured covariates, e.g., severity of illness and use of sedating and or antipsychotic medications. Intubation on a given day is also influenced by unmeasured factors such as the latent respiratory condition, i.e., the individual's respiratory vulnerability manifesting as the acute condition.

In Figure **1** the measured covariates such as severity and medications are contained within the term 'Covars' and the latent respiratory factor is contained within the 'Unmeasured' term. The measured and unmeasured terms along with intubation are temporally indexed such that the subscripts -1 and -2 respectively represent one and two days before the measurement of the outcome. Note that previous values of the measured and unmeasured terms, i.e., $Covar_{-2}$ and $Unmeasured_{-2}$, influence subsequent use of intubation, which in turn influences the successive values of the measured and unmeasured terms. It is this temporal feedback among the explanatory variables that constitutes time dependent confounding. In the scenario represented in Figure **1**, the association of intubation with the outcome is confounded by the measured and unmeasured terms. In contrast with simple confounding, the arrows originating from time-dependent covariates and ending at the time-dependent treatment of primary interest (intubation) influence the estimation of any causal effect between that treatment and the outcome.

Figure **2** represents the same scenario after intubation has been adjusted for TDC and differs from Figure 1 in two ways. First, in accordance with the assumption of no unmeasured confounders, the unmeasured variables and all corresponding arrows have been removed. Second, those arrows originating from covariates and ending in intubation have also been removed. A marginal structural model is a method to remove the temporal confounding of the covariates on intubation's association with next day delirium. It does this by first calculating the probability of intubation as a function of the measured variables concurrent with or prior to intubation, and then uses the inverse of that probability to weight the observations used to model the outcome. Assuming that all measured terms that influence intubation are captured in the first stage model of probability, and that no unmeasured confounders exist, such a model yields an association for intubation that has been adjusted for TDC. Because the assumptions of capturing all covariates and the non-existence of unmeasured factors are very strong, we refrain from using the terms unbiased or causal in describing the resultant associations from this observational study.

Extension of Time-Dependent Confounding to Multiple Treatments

Figure **3** exemplifies the multiplicity of treatments experienced on a daily basis by many older persons in the MICU. Instead of the single treatment of intubation represented in Figures **1** and **2**, Figure **3** shows how the narcotic fentanyl, the antipsychotic haloperidol (Haldol), and intubation are routinely administered to older MICU patients on a daily basis. Because TDC is clinically plausible, all three treatments have been weighted by the inverse probability of their daily levels based on previous and concurrent covariates. For this reason the only arrows entering the treatments are from a previous treatment and all arrows originating
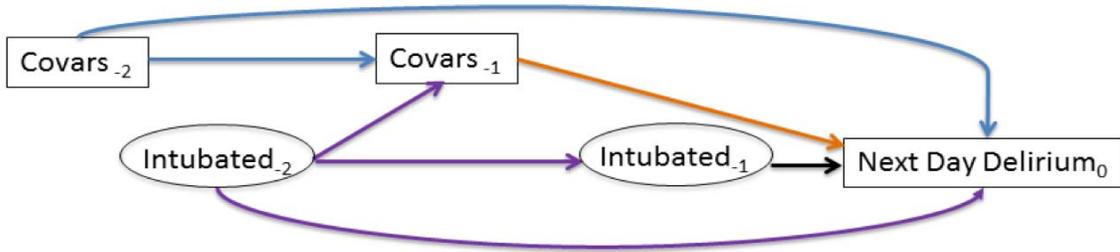
Figure 2: Removal of Time-Dependent Confounding of Association between Intubation and Next Day Delirium.
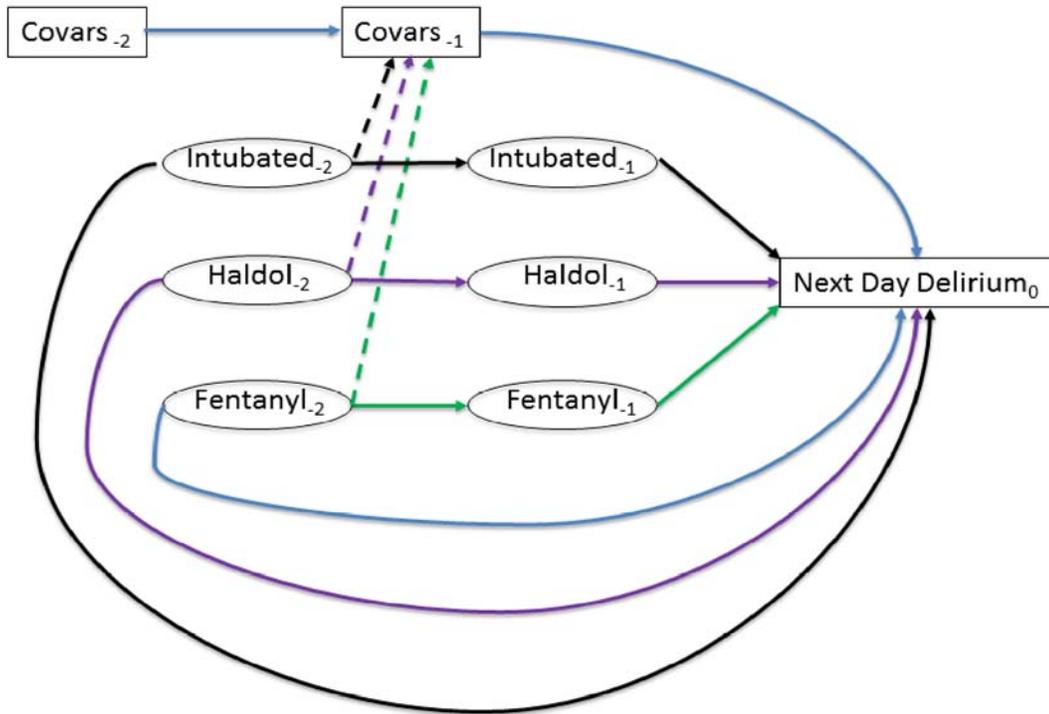

Figure 3: Removal of Time-Dependent Confounding between Three Concurrent Treatments and Next Day Delirium.

from the most recent treatments are directed toward the outcome, i.e., next day delirium.

Figure **4** shows the implementation of a marginal structural model intended to remove the TDC among these three treatments and clinically important covariates prior to evaluation of their individual associations with next day delirium. While our original clinical motivation was to evaluate the association between cumulative dose of haloperidol and next day delirium, we felt that a three-tiered marginal structural model was required to properly address the potential for TDC among these.

Description of Analytical Sample and Related Statistical Concerns

The original cohort of study participants consisted of 309 patients age 60 years and older who were

admitted to the MICU at Yale-New Haven Hospital from September of 2002 through September of 2004. As described previously [4, 11, 17, 18], proxy respondents served as the primary source of baseline information for critically ill patients. Hospital medical records were reviewed to obtain demographic information, admission diagnoses, laboratory data, and detailed, shift-based medication dosing. In a recent analysis the subgroup of 93 patients who received at least one dose of haloperidol during their MICU stay were followed through death or their first eight days. A marginal structural model evaluated the associations between three time-dependent variables, i.e., doses of fentanyl, haloperidol, and intubation, and the outcome of next day diagnosis of delirium [19]. This analysis showed that adjustment for time-dependent confounding among these three treatments resulted in a much larger association for intubation while those of fentanyl and
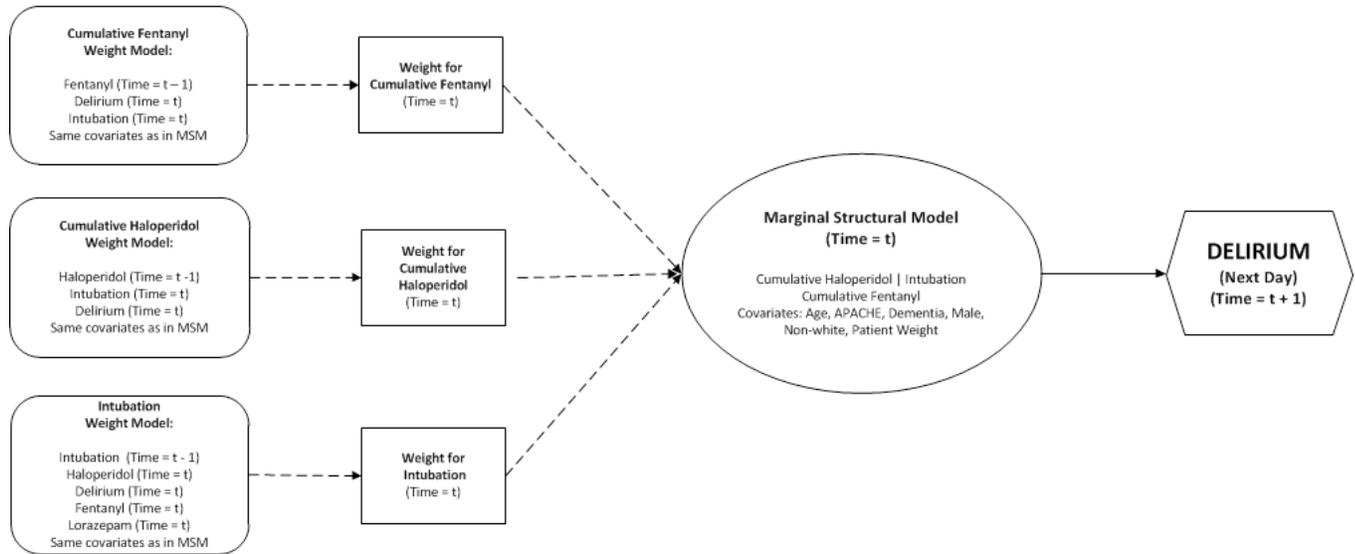
Figure 4: Marginal Structural Model (MSM) of Association between Three Concurrent Treatments in the MICU and Next Day Delirium.

haloperidol were unchanged. A simple, empirical technique providing evidence of time-dependent confounding would be a useful way to determine whether a marginal structural model is warranted. Although our concern for time-dependent confounding of the causal effect of intubation on a delirium outcome in this case was supported by the literature and clinical experience [20], in cases where such prior information is not readily available, such a tool might be helpful.

Plausibility of Time-Dependent Confounding Between Pairs of Explanatory Variables

Simple, exploratory techniques can be used to score the plausibility of time-dependent confounding between any given pair of explanatory variables. Using SAS software [21], we used longitudinal plots, a measure of correlation, a cross-correlation function, and bidirectional regression analyses as our root measures of time-dependent confounding. The three possible pairings among fentanyl, haloperidol, and intubation define the columns of Table **1**. The rows of Table **1** indicate whether the simple, descriptive metrics suggest the presence of TDC or not, and assign corresponding scores. The connection between each measure and TDC is delineated in the next section.

Combining the Exploratory Measures into an Overall Plausibility Score

The rows of Table **1** represent the five primary criteria that were measured and evaluated as evidence of time-dependent confounding, the total point score for a given pair of explanatory variables, and a qualitative

interpretation of the point scores. Because all indicators were calculated using SAS, the specific procedure is indicated in the text that follows in parentheses. The first were simple longitudinal plots (proc gplot) of the two explanatory variables where each graph lags one variable with respect to the other by one unit of time (day). If either of the plots showed trends that were roughly parallel and that did not cross, one point was assigned. The single point reflects the fact that this is very weak evidence. The second criterion assigned one point if there was correlation ≥ 40% at lag zero (proc corr), commensurate with low plausibility of TDC. The third criterion tested whether the cross-correlation function, which examines both variables across a range of positively and negatively lagged values, was significant in either temporal direction (proc timeseries). Cross-correlation merits two points because it indicates a substantive non-random linking of the two variables. The fourth and fifth criteria are each assigned three points, and respectively tested for a significant association when regressing one of the variables on the lagged values of the other. For instance in row four for the first column (Fentanyl and Haloperidol) of Table **1**, daily dose of fentanyl was regressed on the daily doses of haloperidol from the previous day. Row five in that column regressed daily doses of haloperidol on the daily doses of fentanyl from the previous day. Because these are statistical tests of significance, they were assigned three points each. Note that if there is a statistically significant association in both temporal directions, the score will be ≥ 6, automatically resulting in high plausibility for TDC between that pair of explanatory variables. The penultimate row is the total

Table 1: Exploratory Evidence for Detecting Time-Dependent Confounding Among Explanatory Variables

Criterion and Weighting (points assigned for criterion)	Pairs of MICU Treatments Being Examined for Time-Dependent Confounding		
	Fentanyl and Haloperidol	**Fentanyl and Intubation**	**Haloperidol and Intubation**
Similar Trends in Plots In Either Temporal Direction? (1 point)	Yes	Yes	No
Significant Correlation ≥ 40%? (1 point)	Yes	Yes	Yes
Cross-Correlation Function Significant in Either Temporal Direction? (2 points)	No	No	No
Significant Association in GEE Regression of First Variable on Lag of Second? (3 points)	Yes	Yes	No
Significant Association in GEE Regression of Second Variable on Lag of First? (3 points)	No	Yes	No
Total Point Score where higher indicates greater evidence of time-dependent confounding (0 to 10 points)	5 points	8 points	1 point
Qualitative Weight of Evidence for Time-Dependent Confounding	Moderate (4 – 5 points)	High (≥ 6 points)	Low (≤ 3 points)

MICU = medical intensive care unit.
GEE = generalized estimating equations.

point score and the final row is the interpretation of that score. Totals ≤ 3 are considered low, scores of 4 or 5 are considered moderate, and scores ≥ 6 are considered high.

RESULTS

Comparison of Plausibility Scores with Unweighted and Marginal Structural Model Results

Table **1** indicates moderate evidence of time-dependent confounding between fentanyl and haloperidol, strong evidence between fentanyl and intubation, and weak evidence between haloperidol and intubation. The major challenge in evaluating the utility of these scores is that any bias due to time-dependent confounding cannot be directly measured, but is often inferred from theoretical factors. The famous case presented by Hernan, Brumback, and Robins [22] showed that when the effect of antiretroviral medication on the survival of HIV positive patients was adjusted for its time-dependent confounding with red blood cell count *via* a marginal structural model, the association between use of these medications and survival went from negative to positive. We examined the changes in associations between our concurrent treatments, fentanyl, haloperidol, and intubation, and the outcome of next day diagnosis of delirium in un-weighted and weighted (marginal structural) models. The un-weighted and

weighted models each included all three concurrent treatments as depicted in Figure **4**. A comparison of the estimated associations from un-weighted and weighted models was used to assess whether the empirical scores were informative. The model results presented in Table **2** were previously published in a clinical study that concluded that cumulative dose of haloperidol was positively associated with higher odds of next day diagnosis of delirium among non-intubated patients who received it (Odds Ratio (Credible Interval) 1.05 (1.02 – 1.09)) [19].

The rows of Table **2** are explanatory variables in a longitudinal model of next day delirium and comprise common treatments given to older persons in the MICU. Because that model included a significant interaction between cumulative dose of haloperidol and intubation, the associations for haloperidol are presented separately for non-intubated and intubated patients. The columns are the estimated odds ratios and credible intervals estimated by un-weighted and marginal structural models, the latter denoted as the weighted model. The associations of neither fentanyl nor haloperidol among non-intubated patients change between un-weighted and weighted models. This suggests either of two possibilities. The first is that neither of the drugs exhibited time-dependent confounding and the second is that extant time-dependent confounding did not substantially bias their estimated associations with the outcome of next day

Table 2:   Multivariable Associations of Three Treatments with Next Day Delirium, N=93[a]

	Un-weighted Outcome (not adjusted for time dependent confounding)	Weighted Outcome[b] (adjusted for time dependent confounding)
Variables with Time Dependent Confounding [c]	Odds Ratio (95% CI) [d,e]	Odds Ratio (95% CI) [d,e]
Cumulative dose of haloperidol (mg) among non-intubated patients	1.06 (1.01 – 1.14)	1.05 (1. 02 – 1.09)
Cumulative dose of haloperidol (mg) among intubated patients	3.39 (1.61 – 8.01)	5.48 (2.44 – 12.50)
Intubation	3.38 (1.67 – 7.08)	5.66 (2.70 – 12.02)
Cumulative dose of fentanyl (mg)	1.03 (0.98 – 1.09)	1.02 (0.95 – 1.12)

Abbreviations: CI: credible interval (Bayesian equivalent of confidence interval), GCS: Glasgow Coma Scale, mg: milligrams, iqcodeA: informative questionnaire on cognitive decline in the elderly.
[a]The 93 participants contributed 598 patient-days of follow-up.
[b]Marginal structural model with weighting for cumulative fentanyl, cumulative haloperidol, and daily intubation. Model weight was product of individually standardized weights.
[c]Each variable with time dependent confounding measured on day preceding diagnosis of delirium.
[d]Significance defined as credible interval exclusive of 1.00.
[e]All Odds Ratios include adjustment for age, APACHE II score, cognitive impairment defined as IQcode score > 3.3, nonwhite race, and patient weight.

diagnosis of delirium. Note that intubation's association goes up in the weighted model for its main effect as well as in the subgroup of intubated patients taking haloperidol. This suggests that intubation did experience some bias from time-dependent confounding, and that when this was adjusted for, its association became stronger. The marginal structural model used in that analysis was quite complex in that it assigned daily weights, corresponding to the inverse probability of treatment, to the cumulative doses of fentanyl and haloperidol as well as for intubation.

So how does one decide whether the extra time and effort of fitting a marginal structural model is justified? We reconcile the evidence in Table **1** with the model results in Table **2** as follows. Apart from any content related reasons that justify a marginal structural model, we argue that if there is strong evidence of time-dependent confounding between any pair of explanatory variables, then a marginal structural model is justified. If there is some level of theoretical evidence and moderate or higher empirical evidence, then a marginal structural model is also justified. We believe the empirical evidence provided by the scores in Table **1** correctly flagged the need to use an MSM that adjusted for the time-dependent confounding between intubation and the other treatments. The shift in point estimates of intubation's associations with the outcome appear to corroborate that belief.

CONCLUSION

The clinical and statistical communities are increasingly aware of the risk of biased results from longitudinal analyses because of the time-dependent confounding between pairs of explanatory variables. Using a previously published longitudinal study of older persons in the MICU, we propose and demonstrate a simple plausibility score based on descriptive and exploratory statistics that may be used to justify the added complexity of fitting a marginal structural model.

REFERENCES

[1]   Ely EW, Gautam S, Margolin R, Francis J, May L, Speroff T, *et al*. The impact of delirium in the intensive care unit on hospital length of stay. Intensive Care Medicine 2001; 27(12): 1892-900.
http://dx.doi.org/10.1007/s00134-001-1132-2

[2]   Ely EW, Shintani A, Truman B, Speroff T, Gordon SM, Harrell FE, Jr., *et al*. Delirium as a predictor of mortality in mechanically ventilated patients in the intensive care unit. JAMA: the journal of the American Medical Association 2004; 291(14): 1753-62.
http://dx.doi.org/10.1001/jama.291.14.1753

[3]   Murphy TE, Van Ness PH, Araujo KL, Pisani MA. Bayesian time-series analysis of arepeated-measures Poisson outcome with excess zeroes. American Journal of Epidemiology 2011; 174(11): 1230-7.
http://dx.doi.org/10.1093/aje/kwr252

[4]   Pisani MA, Inouye SK, McNicoll L, Redlich CA. Screening for pre-existing cognitive impairment in older intensive care unit patients: use of proxy assessment. Journal of the American Geriatrics Society 2003; 51: 689-93.
http://dx.doi.org/10.1034/j.1600-0579.2003.00215.x

[5]   Pisani MA, Murphy TE, Van Ness PH, Araujo KLB, Inouye SK. Characteristics associated with delirium in older patients in a medical intensive care unit. Archives of Internal Medicine 2007; 167(15): 1629-34.
http://dx.doi.org/10.1001/archinte.167.15.1629

[6]   Milbrandt EB, Deppen S, Harrison PL, Shintani AK, Speroff T, Stiles RA, *et al.* Costs associated with delirium in mechanically ventilated patients. Critical Care Medicine 2004; 32(4): 955-62.
http://dx.doi.org/10.1097/01.CCM.0000119429.16055.92

[7]   Ouimet S, Kavanagh BP, Gottfried SB, Skrobik Y. Incidence, risk factors and consequences of ICU delirium. Intensive Care Medicine 2007; 33(1): 66-73.
http://dx.doi.org/10.1007/s00134-006-0399-8

[8]   Kalisvaart KJ, de Jonghe JF, Bogaards MJ, Vreeswijk R, Egberts TC, Burger BJ, *et al.* Haloperidol prophylaxis for elderly hip-surgery patients at risk for delirium: a randomized placebo-controlled study. Journal of the American Geriatrics Society 2005; 53(10): 1658-66.
http://dx.doi.org/10.1111/j.1532-5415.2005.53503.x

[9]   Pandharipande P, Shintani A, Peterson J, Pun BT, Wilkinson GR, Dittus RS, *et al.* Lorazepam is an independent risk factor for transitioning to delirium in intensive care unit patients. Anesthesiology 2006; 104(1): 21-6.
http://dx.doi.org/10.1097/00000542-200601000-00005

[10]   Patel SB, Poston JT, Pohlman A, Hall JB, Kress JP. Rapidly reversible, sedation-related delirium versus persistent delirium in the intensive care unit. American Journal of Respiratory and Critical Care Medicine 2014; 189(6): 658-65.
http://dx.doi.org/10.1164/rccm.201310-1815OC

[11]   Pisani MA, Kong SYJ, Kasl SV, Murphy TE, Araujo KLB, Van Ness PH. Days of delirium are associated with 1-year mortality in an older intensive care unit population. Am J Respir Crit Care Med 2009; 180: 1092-7.
http://dx.doi.org/10.1164/rccm.200904-0537OC

[12]   Pisani MA, Murphy TE, Araujo KLB, Slattum P, Van Ness PH, Inouye SK. Benzodiazepine and opioid use and the duration of intensive care unit delirium in an older population. Critical Care Medicine 2009; 37: 177-83.
http://dx.doi.org/10.1097/CCM.0b013e318192fcf9

[13]   Robins JM. Marginal structural models. Proceedings of the Section on Bayesian Statistical Science 1997: 1-10.
http://dx.doi.org/10.1002/(SICI)1097-0258(19980215)17:3<269::AID-SIM763>3.0.CO;2-J

[14]   Robins JM. Correction for non-compliance in equivalence trials. Statistics in Medicine 1998; 17: 269-302.

[15]   Robins JM. Marginal structural models versus nested models as tools for causal inference. . In: Halloran E, Berry D, editors. Statistical Models in Epidemiology: The Environment and Clinical Trials. New York: Springer-Verlag 1999; p. 95-134.

[16]   Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000; 11(5): 550-60.
http://dx.doi.org/10.1097/00001648-200009000-00011

[17]   Pisani MA, Redlich CA, McNicoll L, Ely EW, Inouye SK. Underrecognition of preexisting cognitive impairment by physicians in older icu patients. Chest 2003; 124: 2267-74.
http://dx.doi.org/10.1378/chest.124.6.2267

[18]   Pisani MA, Bramley K, Vest MT, Akgun KM, Araujo KL, Murphy TE. Patterns of opiate, benzodiazepine, and antipsychotic drug dosing in older patients in a medical intensive care unit. American journal of critical care: an official publication, American Association of Critical-Care Nurses 2013; 22(5): e62-9.

[19]   Pisani MA, Araujo KL, Murphy TE. Association of cumulative dose of haloperidol with next-day delirium in older medical ICU patients. Critical Care Medicine 2015; 43(5): 996-1002.
http://dx.doi.org/10.1097/CCM.0000000000000863

[20]   Baker JA, Lovell K, Harris N. A best-evidence synthesis review of the administration of psychotropic pro re nata (PRN) medication in in-patient mental health settings. Journal of Clinical Nursing 2008; 17: 1122-31.
http://dx.doi.org/10.1111/j.1365-2702.2007.02236.x

[21]   SAS. SAS User's Manual. 9.2 ed. Cary, NC: SAS Institute; 2009. p. Windows Based Software.

[22]   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(5): 561-70.
http://dx.doi.org/10.1097/00001648-200009000-00012