



S. Bailly
R. Pirracchio
J. F. Timsit

What's new in the quantification of causal effects from longitudinal cohort studies: a brief introduction to marginal structural models for intensivists

Received: 7 May 2015
Accepted: 9 June 2015
Published online: 24 June 2015
© Springer-Verlag Berlin Heidelberg and ESICM 2015

Electronic supplementary material The online version of this article (doi:[10.1007/s00134-015-3919-6](https://doi.org/10.1007/s00134-015-3919-6)) contains supplementary material, which is available to authorized users.

S. Bailly
U823, Grenoble Alpes University, Rond-point de la Chantourne,
38700 La Tronche, France

S. Bailly (✉) · J. F. Timsit
UMR 1137, IAME Team 5, DeSCID: Decision Sciences in
Infectious Diseases, Control and Care Inserm/Paris Diderot,
Sorbonne Paris Cité University, 16 rue Henri Huchard,
75018 Paris, France
e-mail: sbailly@chu-grenoble.fr
Tel.: +33 140 25 77 02

R. Pirracchio
Department of Anesthesia and Perioperative Care, San Francisco
General Hospital and Trauma Center, University of California San
Francisco, San Francisco, CA, USA

R. Pirracchio
Division of Biostatistics, School of Public Health, University
of California at Berkeley, Berkeley, CA, USA

J. F. Timsit
Medical and Infectious Diseases ICU-Paris Diderot University/
Bichat Hospital, 46 rue Henri Huchard, 75018 Paris, France

When randomized controlled trials are not possible, observational longitudinal data may be the only option to quantify the impact of a treatment on an outcome [1]. As an illustration, we assessed the impact of empirical systemic antifungal therapy (SAT) on mortality in critically ill patients [2]. A regression may be used to model the

relationship between SAT and mortality, with adjustment on potential confounders. In many situations, the regression coefficient cannot be interpreted causally, particularly when SAT administration is a time-dependent variable [3–5]. Indeed, besides a direct causal effect, there may be several paths linking the treatment to the outcome through various confounders, i.e., variables associated with treatment allocation and with the outcome which may confound the association of interest. In this situation, the association measure provided by standard regression may differ from what clinicians often seek, which is a quantification of the direct causal effect between an exposure and an outcome. This situation may be illustrated by considering severe sepsis as a single confounder. Indeed, severe sepsis may trigger SAT while it also impacts mortality. Hence, SAT and mortality share a common cause, i.e., an indirect path linking SAT to mortality (Fig. E1 in the Electronic Supplementary Material). The standard approach based on logistic regression, for instance, may lead to a biased estimation of the causal effect of SAT on mortality because some time-dependent variables (i.e., septic shock) which may be affected by previous treatment history can in turn affect both further treatments and the outcome [3]. To overcome the limitation of standard regression approaches, new specific statistical methods have been developed to handle this type of bias and are often referred to as causal inference methods. They were first introduced in intensive care unit (ICU) literature by Bekaert et al. [6].

Let us consider ten ICU patients, three with and seven without severe sepsis (SS) at baseline, and suppose that SAT is administered to two patients with SS (67 %) and two patients without SS (29 %). To conclude about a causal relationship between treatment and mortality, the distribution of the confounders between the groups should be balanced. The propensity score (PS) may be used for this balance. In other words, with the use of the PS, a

difference in outcome between the treatment groups may be considered as causally related to the exposure [7]. PS matching estimators are usually used to estimate the average treatment effect in the treated (ATT) and answers to the question: “How would the outcomes of the treated individuals have differed had they received the control?” Nonetheless, this may not be the question of interest. When estimating the impact of SAT, the research question is rather: “What would be the outcome if all patients at risk were treated or if all remained untreated?” [8]. This is usually referred to as the average treatment effect (ATE). The PS may be also used to estimate the ATE through the inverse probability of treatment weight (IPTW) estimators [8]. The IPTW general concept is to weight each individual contribution by the inverse of his/her probability of receiving his/her treatment. The weights are calculated as $\frac{1}{PS}$ in treated individuals and $\frac{1}{1-PS}$ in the untreated individuals and are used to create a pseudo-population in which the exposure is independent of the measured confounders as illustrated in Fig. 1 [9, 10].

In the simple situation of a binary point treatment with non-time-varying confounders, and under a certain set of assumptions, one just has to compare the weighted outcome in the treated and the untreated to get an estimation of the causal effect of the exposure on the outcome in this pseudo-population. Hence, the first benefit of these new statistical approaches is that, under a set of assumptions, they may offer the possibility to estimate such causal quantities. However, the experimental situation is often more complex and may involve multiple time-point treatments and time-varying confounders. Marginal structural models (MSM) are a new class of statistical models developed by Robins [3] to handle this particular situation. Practically, MSMs refer to the regression of the outcome on the exposure in the pseudo-population [9]. In MSMs, IPTW estimators are used to estimate the parameter of interest (e.g., the causal risk difference, relative risk, or odds ratio). Specifically, the causal risk difference (i.e., the ATE) equals the slope of a weighted linear regression of the outcome on the exposure, using the weights as defined earlier. The added value of MSMs,

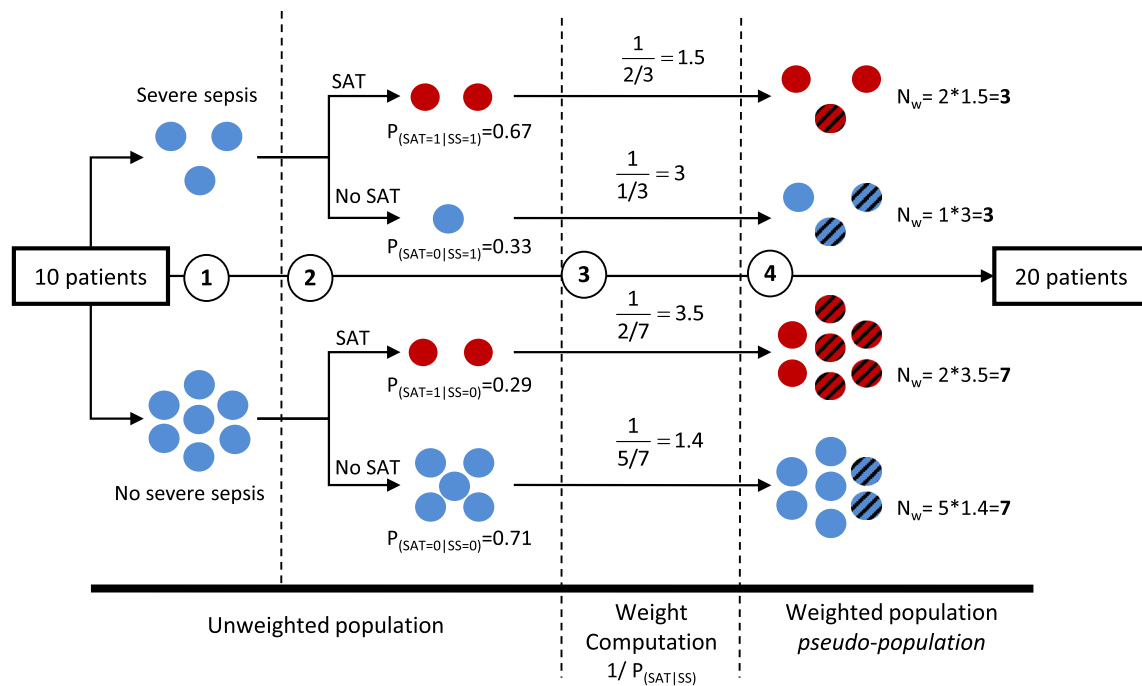


Fig. 1 Example of estimating the inverse probability of treatment weight. This example uses a virtual sample of ten patients. (1) The population may be divided into two groups according to the presence/absence of a severe sepsis (three versus seven). (2) In each subgroup, some patients are receiving SAT. For each individual in each subgroup, the probability of receiving his/her actual treatment ($P_{SAT|SS}$; i.e., probability of treatment given the presence/absence of SS) may be estimated from empirical proportions. (3) The inverse probability of treatment weight (IPTW) is computed from this probability and is equal to $1/P_{SAT|SS}$. The weight equals $1/PS$ when $SAT = 1$ and $1/(1 - PS)$ when $SAT = 0$. (4) These weights are used to build the pseudo-

population, where an individual with a low probability of SAT will be up-weighted and, conversely, an individual with a high probability of SAT will be down-weighted. The pseudo-population encompasses both factual and counterfactual observations. In this pseudo-population, the treated and untreated individuals are exchangeable and it is possible to compute directly the difference in mortality. If one death is observed in the treated group (without severe sepsis) and two deaths are observed in the untreated group (one with severe sepsis and one without severe sepsis), the estimated number of death in the pseudo-population is 3.5 in the treated group and 4.4 in the untreated group ($3 + 1.4$). The relative risk for mortality is $(3.5/10)/(4.4/10) = 0.795$

as compared to the previously described PS-based methods, is that they can handle time-dependent confounders when classical PS methods account only for baseline confounders bias.

In case of longitudinal data, the treatment probability has to be updated at each time point [11, 12]. This means that the treatment probability at time t is estimated from the variables measured up to time t , including the exposure history. The probability of being treated at time t is in turn defined as the product of all the treatment probabilities up to time t . Thus, the weights derived at each time point are combined into single weights to estimate the impact of the entire treatment regimen.

Causal interpretation of the MSM parameters relies on some assumptions [3, 13]. First, for each combination of covariates, there must be treated and untreated individuals. When, for certain characteristics, there are only treated or untreated, the so-called positivity assumption is violated [3, 9]. This means that, in this stratum, the causal effect is not identifiable. In the present issue, Muriel et al. used an MSM with IPTW to estimate the causal effect of analgesic and/or sedative drugs on the failure of non-invasive positive-pressure ventilation [14]. Although this approach seems well suited for the topic, the results should be interpreted with particular caution because of the positivity assumption. The relatively small number of individuals in each treatment group defies the positivity assumption. This may explain the great variability in the final estimates (the wide confidence intervals). A second

causal assumption is known as the ignorability assumption. It refers to the absence of significant unmeasured confounder. Specifically, in the context of longitudinal studies, at each measurement time one must have available the history of all risk factors of the exposure that are also associated with the outcome (time-dependent confounding factors). While the first assumption is verifiable from the data, the second is often non-testable. Some basic SAS codes adapted to our example as well as details about the positivity assumption are provided in the Electronic Supplementary Material. Finally, to obtain unbiased causal estimates, the model for the conditional probability of exposure has to be correctly specified. Although MSMs are increasingly used in the medical literature and offer a new appealing alternative to standard regression methods, they need a very complex analytic strategy, especially because IPTW estimators can be very unstable and need strong assumptions to be adequately interpreted. Thus, the use of MSM for observational and longitudinal ICU data analysis often requires extensive statistical background. To overcome the risk of PS model misspecification, recent advances have been proposed. Double-robust estimators (e.g., augmented IPTW or targeted maximum likelihood estimators) [15] may represent the future of causal inference.

Compliance with ethical standards

Conflicts of interest Authors have no conflict of interest to declare.

References

1. Benson K, Hartz AJ (2000) A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 342:1878–1886
2. Bailly S, Bouadma L, Azoulay E, Orgeas MG, Adrie C, Souweine B, Schwebel C, Maubon D, Hamidfar-Roy R, Darmon M, Wolff M, Cornet M, Timsit JF (2015) Failure of empirical systemic antifungal therapy in mechanically ventilated critically ill patients. *Am J Respir Crit Care Med* 191:1139–1146
3. Robins JM, Hernan MA, Brumback B (2000) Marginal structural models and causal inference in epidemiology. *Epidemiology* 11:550–560
4. Resche-Rigon M, Pirracchio R, Robin M, De Latour RP, Sibon D, Ades L, Ribaud P, Ferman JP, Thieblemont C, Socie G, Chevret S (2012) Estimating the treatment effect from non-randomized studies: the example of reduced intensity conditioning allogeneic stem cell transplantation in hematological diseases. *BMC Blood Disord* 12:10
5. Gerhard T, Delaney JA, Cooper-Dehoff RM, Shuster J, Brumback BA, Johnson JA, Pepine CJ, Winterstein AG (2012) Comparing marginal structural models to standard methods for estimating treatment effects of antihypertensive combination therapy. *BMC Med Res Methodol* 12:119
6. Bekaert M, Timsit JF, Vansteelandt S, Depuydt P, Vesin A, Garrouste-Orgeas M, Decruyenaere J, Clec'h C, Azoulay E, Benoit D (2011) Attributable mortality of ventilator-associated pneumonia: a reappraisal using causal analysis. *Am J Respir Crit Care Med* 184:1133–1139
7. Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70:41–55
8. Pirracchio R, Carone M, Rigon MR, Caruana E, Mebazaa A, Chevret S (2013) Propensity score estimators for the average treatment effect and the average treatment effect on the treated may yield very different estimates. *Stat Methods Med Res*. doi: [10.1177/0962280213507034](https://doi.org/10.1177/0962280213507034)
9. Cole SR, Hernan MA (2008) Constructing inverse probability weights for marginal structural models. *Am J Epidemiol* 168:656–664
10. Hernan MA, Robins JM (2006) Estimating causal effects from epidemiological data. *J Epidemiol Community Health* 60:578–586
11. Hernan MA, Brumback B, Robins JM (2000) Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* 11:561–570

-
12. Do DP, Wang L, Elliott MR (2013) Investigating the relationship between neighborhood poverty and mortality risk: a marginal structural modeling approach. *Soc Sci Med* 91:58–66
 13. Yang S, Eaton CB, Lu J, Lapane KL (2014) Application of marginal structural models in pharmacoepidemiologic studies: a systematic review. *Pharmacoepidemiol Drug Saf* 23:560–571
 14. Muriel A, Penuelas O, Frutos-Vivar F, Arroliga A, Abaira V, Thille A, Brochard L, Nin N, Davies A, Amin P, Du B, Raymondos K, Rios F, Violi D, Maggiore S, Soares MA, Gonzalez M, Abroug F, Bulow H, Hurtado J, Kuiper M, Moreno R, Zeggwagh AA, Villagomez A, Jibaja M, Soto L, D'Empaire G, Matamis D, Koh Y, Anzueto A, Ferguson N, Esteban A (2015) Impact of sedation and analgesia during non-invasive positive pressure ventilation on outcome. A marginal structural model causal analysis. *Intensive Care Med*. doi: [10.1007/s00134-015-3854-6](https://doi.org/10.1007/s00134-015-3854-6)
 15. Munoz ID, van der Laan M (2012) Population intervention causal effects based on stochastic interventions. *Biometrics* 68:541–549