SMART Designs in Observational Studies of Opioid Therapy Duration

John W. Jackson, ScD and Joshua J. Gagne, PharmD, ScD

Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.

J Gen Intern Med 29(3):429–31

DOI: 10.1007/s11606-013-2725-5

© Society of General Internal Medicine 2014

Chronic pain affects around 100 million US adults and is associated with \$635 billion in annual direct and lost productivity costs.¹ With the growing acceptance of palliative treatment for pain, prescription opioid use has risen dramatically over the past decade.² Hydrocodone plus acetaminophen is the most commonly prescribed drug in the US.³ Though effective in relieving pain, long-term opioid therapy has been linked to many safety concerns, including substance abuse, fractures, premature labor and neonatal abstinence syndrome, and deaths from accidental overdose.²

In this issue of JGIM, Scherrer and colleagues add to the litany of potential adverse events associated with long-term opioid use. Using propensity-score methods and administrative billing records, including medical encounter and prescription dispensing data from the Veterans Administration health care system (VA), the authors investigated whether duration of prescription opioid use was associated with increased risk of incident depression over up to 7-year follow-up.⁴ Among nearly 50,000 individuals who filled a first prescription for an opioid, 91 % discontinued use by 90 days. However, the four percent who continued use between 90 and 180 days had 25 % higher risk of receiving a diagnostic code for major depression, and those with even longer use had 50 % higher risk. If confirmed, these results could have important implications for managing both treatment of chronic pain and mental health in patients undergoing such treatment.

While randomized trials are the gold standard for evaluating drug outcomes, observational studies are often the best option for assessing drug safety when adequately powered randomized trials cannot ethically be performed. To address the question posed by Scherrer et al., it would be unethical to randomize patients with chronic pain to discontinue opioids. However, putting ethical, resource, and other logistical constraints aside, envisioning the ideal randomized trial for a particular study question can aid the design and analysis of a similar observational study.⁵

To determine whether longer opioid use causes depression, we can imagine a hypothetical discontinuation sequential multiple assignment randomized trial (SMART)⁶ that enrolls patients with chronic pain who screen negative for history of depression and substance abuse using a validated semi-structured clinical interview.⁷ After an initial phase of opioid therapy for 90 days, patients enter a second phase and are randomized to either placebo for the remainder of the study (group 1) or to another 90 days of opioid therapy (group 2). When the second phase is completed, patients in group 2 enter a third phase and are randomized to either placebo for the remainder of the study (group 2A) or to continue therapy (group 2B). Only patients who survive and respond to treatment undergo the sequential randomization phases. Incident depression would be prospectively assessed for all patients using a validated psychometric rating scale or semi-structured clinical interview.⁷

In this hypothetical randomized trial, both treatment decisions are made randomly, but in an observational analog, the decision to continue opioids at each point would be driven by conditions that may be risk factors for depression, such as persistent severe pain. Moreover, these depression risk factors may be influenced by prior opioid therapy. This dynamic temporal relationship among risk factors and treatment decisions is known as time-dependent confounding, and conventional methods for confounder adjustment will generally fail in its presence.⁸ To overcome this problem, Scherrer et al. sought to implement a marginal structural model,⁹ but the validity of the results hinges on how well their design and modeling choices for observational data maps to the hypothetical SMART.

Marginal structural models use propensity scores to address time-dependent confounding. Propensity scores were first developed to mimic randomized trials in observational studies by modeling and accounting for the initial treatment decision mechanism. Patients' propensity scores represent their probabilities of receiving a treatment given pre-existing characteristics, and they are usually estimated with a regression model, which should include all risk factors for the outcome (in this case, depression). Using the propensity score to address baseline confounding creates treatment groups that, on average, share the same distribution of measured risk

factors (i.e., covariate balance), as would be observed in an ideal randomized trial—though actual randomization would also balance unmeasured risk factors. Marginal structural models extend this covariate balancing property of the propensity score to groups of patients who follow different treatment regimes over time.

For marginal structural models to address time-dependent confounding, separate propensity scores are needed for each treatment decision that patients encounter during the study. In the hypothetical randomized trial, decisions to continue therapy occur at 90 and 180 days. In the observational version, the propensity score model for the 90-day treatment decision would incorporate covariates assessed before day 90, and the model for the 180-day treatment decision would include treatment status at day 90 and covariates assessed before day 180; bias could ensue if the model for a given treatment decision includes variables measured after that decision. The predicted values from these models are the probabilities of continuing treatment at that point in time, and their complements are the probabilities of discontinuing treatment. For each patient, the probability of following his or her observed treatment decision at days 90 and 180 is multiplied. When the data are weighted by the inverse of this product—the inverse probability weight—a crude analysis yields valid effect estimates provided that the models are correctly specified, that all risk factors for depression are included in the propensity score models, and that the so-called positivity condition is satisfied (i.e., that each covariate pattern in the data contains patients within each course of treatment).

In the hypothetical randomized trial, the comparisons of interest would be between opioid-treated and placebo-treated patients in the latter two phases (i.e., group 2 versus group 1 patients in phase 2, and group 2B versus group 2A patients in phase 3). If the corresponding analysis of observational data accounted for the inherent selection bias in the SMART design (i.e., prior response to therapy is included in both propensity score models and separate models and weights for censoring are used¹⁰), one could directly compare patients corresponding to group 1 patients (i.e., those with <90 days of use) and 2A (i.e., those with 90-180 days of use) to those corresponding to group 2B patients (i.e., those with ≥180 days of use). This comparison, as implemented by Scherrer et al., requires strong assumptions about positivity and the nature of selection bias because response to therapy (i.e., pain severity) and informative censoring were not accounted for in their analysis. In addition, the model used to estimate the propensity scores included variables measured before and after each treatment decision (i.e., up to the point of recorded depression), so the estimated propensity scores do not accurately reflect the treatment decisions. It is not clear

how these issues impacted the study results, but they could be empirically examined in subsequent analyses.

Even when marginal structural models closely match their corresponding ideal randomized trials, their use requires care because the results can be very sensitive to certain assumptions. For example, when there are clinical reasons for patients with a particular covariate profile to always or never follow a certain regimen—a positivity violation and strong form of confounding by indication—the results will be biased. This might happen if severe pain is a risk factor for depression and patients with chronic severe pain always use opioids for more than 180 days, then there are no comparable patients in the other treatment groups. Another related concern is that for some covariate patterns, the probability of following a given treatment regimen will be extremely low, so patients with these covariate patterns will dominate the analysis in the weighted data. These issues could be diagnosed by examining the distributions of the propensity scores and basic descriptive statistics for the weights;¹¹ without these diagnostics, it is difficult to know whether these issues impacted the analysis by Scherrer and colleagues.

For an observational study to further approximate the validity of a randomized trial, necessary study variables (i.e., exposures, outcomes, and confounders) must be accurately measured and available in the data. While Scherrer and colleagues report impressive performance characteristics of the depression algorithms that they used to determine cohort eligibility and outcomes, a recent systematic review suggests that they do not perform as well as these estimates suggest.¹² Using these algorithms to define cohort entry and track outcomes could result in recurrent depression being misclassified as incident depression, which could introduce bias since depression itself can lead to long-term opioid use.¹³ Moreover, the two most important determinants of chronic opioid use—pain severity and degree of opioid dependence—are not well captured in claims data. To the extent that these affect depression, there will be residual confounding.

Despite these limitations, we applaud Scherrer and colleagues for applying advanced analytic methods to observational data to address an exquisitely complex study question. Integrated healthcare systems such as the VA offer great promise for rigorous observational studies.¹⁴ Incorporating routine dimensional measures of psychiatric symptoms and pain severity at regular visits for patients with chronic pain⁷ could not only help mitigate the potential association of opioids on depression that Scherrer and colleagues have reported, but their assimilation into electronic databases would permit the capture of important variables for future studies of long-term opioid use. In the meantime, the findings by Scherrer and colleagues should be confirmed as soon as possible. While a randomized trial on the topic will not be performed, an observational study

designed and analyzed like the proposed SMART trial would be a valuable next step.

Corresponding Author: Joshua J. Gagne, PharmD, ScD; Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 1620 Tremont Street, Suite 3030, Boston, MA 02120, USA (e-mail: jgagne1@partners.org).

REFERENCES

1. Institute of Medicine (US). Committee on Advancing Pain Research Care and Education. *Relieving pain in America : a blueprint for transforming prevention, care, education, and research*. Washington, DC: National Academies Press; 2011. xvii, 364 p. p.
2. **Sullivan MD, Howe CG.** Opioid therapy for chronic pain in the United States: promises and perils. *Pain*. 2013. doi:[10.1016/j.pain.2013.09.009](https://doi.org/10.1016/j.pain.2013.09.009). PubMed PMID: 24036286.
3. IMS Institute for Healthcare Informatics. *The Use of Medicines in the United States: Review of 2011*. Parsippany, NJ: IMS Institute for Healthcare Informatics; 2012.
4. **Scherrer JF, Svrakic DM, Freedland KE, Chrusciel T, Balasubramanian S, Bucholz KK, et al.** Prescription opioid analgesics increase the risk of depression. *J Gen Intern Med*. 2013. doi:[10.1007/s11606-013-2648-1](https://doi.org/10.1007/s11606-013-2648-1). PubMed PMID: 24165926.
5. **Rosenbaum P.** *Delimnas and Craftsmanship. Design of Observational Studies*. New York, NY: Springer; 2010:3–13.
6. **Almirall D, Compton SN, Rynn MA, Walkup JT, Murphy SA.** SMARTer discontinuation trial designs for developing an adaptive treatment strategy. *J Child Adolesc Psychopharmacol*. 2012;22(5):364–74.
7. **Rush AJ, First MB, Blacker D.** American Psychiatric Association. Task Force for the Handbook of Psychiatric Measures. *Handbook of psychiatric measures*, vol. xxxvi. 2nd ed. Washington, DC: American Psychiatric Pub; 2008. 828 p. p.
8. **Robins JM, Hernan MA.** Estimation of the causal effects of time-varying exposures. In: Fitzmaurice GM, editor. *Longitudinal data analysis*. Chapman & Hall/CRC handbooks of modern statistical methods. Boca Raton: CRC Press; 2009:p. xiv, 618 p.
9. **Robins JM, Hernan MA, Brumback B.** Marginal structural models and causal inference in epidemiology. *Epidemiology (Cambridge, Mass)*. 2000;11(5):550–60.
10. **Joffe MM.** Administrative and artificial censoring in censored regression models. *Stat Med*. 2001;20(15):2287–304.
11. **Cole SR, Hernan MA.** Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*. 2008;168(6):656–64.
12. **Townsend L, Walkup JT, Crystal S, Olfson M.** A systematic review of validated methods for identifying depression using administrative data. *Pharmacoepidemiol Drug Saf*. 2012;21(Suppl 1):163–73.
13. **Braden JB, Sullivan MD, Ray GT, Saunders K, Merrill J, Silverberg MJ, et al.** Trends in long-term opioid therapy for noncancer pain among persons with a history of depression. *Gen Hosp Psychiatry*. 2009;31(6):564–70.
14. **Hennessy S, Freeman CP, Cunningham F.** *US Government Claims Databases. Pharmacoepidemiology*. 5th ed. Chichester, West Sussex, UK: Wiley-Blackwell; 2012:p. p.