Government-Academic Partnerships in Randomized Evaluations: The Case of Inappropriate Prescribing

Adam Sacarny, David Yokum, and Shantanu Agrawal*

Abstract: There is growing evidence that inappropriate prescribing is harming patients and raising costs in the U.S. health care system. Through a partnership between the federal government and academics, we seek to develop evidence on reducing this prescribing. We conduct several randomized letter interventions targeting high-volume prescribers of drugs that can harm patients. We take a continuous improvement approach, rapidly evaluating each round and using the results to inform subsequent work. The first round of letters yielded no effects, and we responded with new interventions that are now under evaluation. We discuss lessons our work provides for future government-academic partnerships.

* Sacarny: Department of Health Policy and Management, Mailman School of Public Health, Columbia University; National Bureau of Economic Research; and Abdul Latif Jameel Poverty Action Lab (J-PAL). 722 W 168 St, 4th Floor, New York, NY 10032 (e-mail: ajs2102@columbia.edu). Yokum: Office of Evaluation Sciences, U.S. General Services Administration (previous); The Lab @ DC, Government of the District of Columbia (current). 1350 Pennsylvania Avenue, NW, Washington, DC 20004 (e-mail: david.yokum@dc.gov). Agrawal: Center for Program Integrity, Centers for Medicare and Medicaid Services, U.S. Department of Health and Human Services. 7500 Security Boulevard, Baltimore, MD 21244 (e-mail: Shantanu.Agrawal@cms.hhs.gov). We are grateful to Jason Doctor, Amy Finkelstein, Angela Kilby, Peter Meunnig, Maria Polyakova, Brad Shapiro, Melanie Wasserman, and many seminar participants for their comments on this work. We thank the staff of the CMS Center for Program Integrity and J-PAL North America for their assistance on these randomized evaluations. We gratefully acknowledge the support of the Robert Wood Johnson Foundation, J-PAL North America, and the Laura and John Arnold Foundation.

Introduction

There is growing evidence that inappropriate prescribing – prescribing that conflicts with clinical guidelines or violates fraud and abuse laws – is harming the health of patients and driving costly health care utilization. The near-quintupling of prescription opioid pain reliever overdose deaths since 1999 is a stark example of this phenomenon; health care utilization related to opioid use disorder has increased in lockstep, and the literature finds overprescribing and associated utilization in many other classes of drugs as well (see Paulozzi et al. 2011; Mack et al. 2015; Fleming-Dutra et al. 2016; and cites in Sacarny et al. 2016). On this topic, like many others, the interests of academic researchers, policymakers, and practitioners overlap. For researchers, there is uncertainty about the underlying drivers of overprescribing; in turn, for policy, the most effective approaches to stop overprescribing remain unclear. The opportunity and need for collaboration are significant.

This paper reviews a partnership that began in July 2014 between the federal government and academics that seeks to reduce inappropriate prescribing with randomized interventions. Our study group includes academic economists; the Centers for Medicare and Medicaid Services (CMS), which administers the Medicare program; and the Office of Evaluation Sciences (OES), a federal research team that applies behavioral science to improve government policies and programs. In this work, we are conducting a series of evaluations in Medicare Part D, the U.S. public insurance program that covers prescription drugs for many seniors and people with disabilities. The interventions involve high volume prescribers of drugs that can harm patients; in each round, prescribers are sent letters that aim to reduce overprescribing. We take a continuous improvement approach: we rapidly evaluate each round of the study, with preliminary results often ready just months after the letters are sent, then use the results to inform the subsequent round.

In this paper, we present a conceptual framework for harmful prescribing, describe our interventions and results, and discuss lessons that our work yields for future government-academic research partnerships.

Conceptual Framework

In our conceptual model of prescribing, based on Skinner (2011), health care providers maximize a convex combination of patient utility; financial gain, including the expected penalties from inappropriate prescribing (e.g. Becker 1968); and moral benefits, the utility providers receive from following norms (Levitt and List 2007).

Ordering patients from greatest to least benefit, allocative efficiency is achieved when the marginal benefit to the last treated patient equals the marginal cost of the treatment. Utilization past this point starts with the "flat of the curve," the segment where the marginal benefit is between cost and zero. This study focuses on prescribing in the next range: where the treatment causes harm. Prescribing in this segment conflicts with the aims of both the social planner and the patient.

We identify inappropriate prescribing by turning to violations of clinical practice norms, which include prescribing guidelines as well as fraud and abuse regulations. By this definition, a growing literature (cited in the introduction) finds signs of overprescribing in many classes of drugs.

Harmful prescribing could occur because providers place heavy weight on incentives that reward volume of care, since prescribing generates ongoing encounters with patients for which the provider can bill the patient's insurer. Providers can also accept illegal payments from patients in exchange for prescriptions. For prohibited activities like these, providers may perceive the probability of detection, or the penalties conditional on detection, as low. Providers may also value the treatment

above its social benefit because they fail to understand their own prescribing skill or they lack awareness of the benefits and harms of the drugs.

Letters to prescribers could influence behavior for several reasons. First, letters could change prescribers' perception of incentives: the probability of detection and penalties conditional on detection (a common approach in field experiments on taxes and fees; see Fellner, Sausgruber, and Traxler 2013; Castro and Scartascini 2015). A second channel for effects is through moral costs and benefits, the part of utility that reflects social or professional norms. Letters can trigger moral costs with messages that emphasize that one's behavior is unusual in comparison to his or her peers (Hallsworth et al. 2014). Lastly, letters could reduce overprescribing by educating prescribers about the benefits and harms of drugs. Messages that activate this channel can focus on evidence-based clinical guidelines, for example.

Interventions

Now, two and a half years after the partnership began, our work includes three rounds—one complete, one in evaluation, and one in design. In each round, high-volume prescribers in Medicare Part D are identified and randomized to a treatment or control arm. The treatment arm receives letters aiming to reduce overprescribing. We track the effects of the letters with Medicare administrative data available to CMS and its anti-fraud contractors. We register our work on public trial registries and pre-specify analysis plans prior to unblinding to study data.[1]

---

[1] See https://www.socialscienceregistry.org/trials/525 and https://www.socialscienceregistry.org/trials/729

First Intervention

For the initial intervention, CMS identified the top 1,525 prescribers of Schedule II controlled substances—the class of drugs defined by the Drug Enforcement Administration as that with the greatest risk of abuse and dependency for patients. This class includes most opioid pain relievers. Despite representing 0.2% of prescribers of these drugs, study prescribers accounted for about 10% of the Schedule II volume in Medicare Part D.

The study team then designed a letter to send to this group of prescribers building off a template from CMS. The letter contained a peer comparison message and graph stating that the provider was associated with much more controlled substance prescribing than his or her peers (other providers with the same medical specialization in the same state), a design to trigger the moral cost channel. We randomized half of the prescribers to a treatment group that received a letter and the remainder to a control group. Letters were sent in September 2014, and we then immediately turned to collecting and analyzing prescribing data.

The administrative data was updated with a very short lag, and our preliminary evaluation was ready in January 2015, 4 months after the intervention commenced. The preliminary and final analyses detected no effects on prescribing (Figure 1). At the 95% confidence level we ruled out that the letters drove more than a 1.4% reduction in prescribing at 90 days. These findings were published as Sacarny et al. (2016).

Subsequent Interventions

In the first quarter of 2015, we set to redesign the intervention for a second round. For this intervention, we used as a vehicle CMS's existing plans to contact the highest volume prescribers of quetiapine (branded Seroquel), an antipsychotic drug that can be inappropriately prescribed.

We made a series of changes to maximize the potential for the intervention to yield effects (we note that in so doing, we forewent the ability to pinpoint which change drove any differences in results). We speculated that the moral cost instrument alone may have been insufficient to drive behavior change. Referencing field experiments on tax and fee repayment, we added penalty-focused language about monitoring, audits, and investigations. Concerned that the letters were ignored by prescribers, we opted to send prescribers multiple letters rather than one. This approach aligned with energy conservation work that found stronger effects from more frequent messaging (Allcott and Rogers 2012). Prescribers below the extreme tail of the distribution may have been more likely to respond because they were less attached to prescribing the drugs. We more than tripled the sample size to move farther down the distribution. This change also enabled the detection of smaller effects.

CMS then identified high-volume prescribers of quetiapine in Medicare, calibrating the methodology to yield the sample size we requested. We randomized half of these prescribers to a treatment arm, and they were sent letters in April, August, and October 2015. The remainder received a placebo letter describing an unrelated Medicare regulation. Evaluation is now in progress.

A third round is also planned to target high-volume prescribers of opioid pain relievers. We plan to use the lessons of the first two rounds to intervene with this population. To attempt to activate the educational channel, both letters will include prescribing guidelines. The study will test two different letters, one focused on patient harms and another on penalties for prescribers. This design will provide evidence on the drivers of high-volume prescribing, as well as guidance on which messaging strategies change behavior for this group.

Discussion

Our study benefited from administrative data that was updated quickly and a willingness by CMS to respond to negative results with new approaches. We have engaged in a process of continuous improvement, iterating to find methods of communicating with prescribers that reflect the underlying drivers of their prescribing behavior.

Identifying inappropriate prescribing, both to select prescribers for targeting with letters and to measure impacts, presents a challenge. In the initial round, we selected and tracked prescribers based on their volume of Schedule II prescribing. A prescriber with many patients or patients who tended to need Schedule II drugs could be an outlier without engaging in inappropriate prescribing. But in certain cases, the far tail may be a reasonable approximation for questionable activity: for example, in the initial study round, more than one-fifth of prescribers had previously been investigated for fraud (we found null results for the previously investigated and not previously investigated subgroups). Moving forward, we are linking together CMS's rich administrative data to identify when prescribing is likely to conform or conflict with the clinical needs of patients.

As we have iterated on letter designs, we have also sought to consider the potential caveats of seemingly effective interventions. A penalty-focused letter could cause providers to respond by cutting back prescriptions to patients who benefited from the drugs or by substituting patients to less appropriate drugs. Watching for this behavior is a key focus of the study's second and third rounds. In addition, a letter with beneficial effects may be ineffective if used on other populations, targeted toward other drugs, or sent more frequently. If prescribers respond because they update their beliefs about the probability of detection, the power of warning messages must be treated as a finite resource.

Lessons and Conclusion

While randomized designs are common in clinical studies, health care delivery research rarely uses them (Finkelstein and Taubman 2015). Partnerships between academics and government are a promising avenue to promote randomization, since these studies often produce valuable evidence for both parties. Our work is a useful case study. CMS has been sending peer comparison letters to physicians since at least 2010 (the letters described billing of services, not prescribing) without rigorous evaluation. Now, CMS has access to an evidence base on different messaging strategies, and even though our initial intervention yielded a null finding, this result is useful as the agency seeks to improve its communications with health care providers moving forward.

We identify two factors that enabled this research. The first was the help of groups that facilitated the collaboration. The North America branch of the Abdul Latif Jameel Poverty Action Lab (J-PAL) connected academic researchers to CMS through its Health Care Delivery Initiative. J-PAL also provided financial and material support, including information on best practices in randomized evaluations. The OES, an interdisciplinary research team in the federal government, brought knowledge on promoting randomization and behavioral science in federal agencies. The group also acted as a conduit between academics and CMS.

Second, support from leadership at CMS was crucial to the study. Management was intrigued that through randomization, they could assess the efficacy of their work and, if it was successful, promote and expand the project. CMS also recognized that these evaluations might show that their existing messaging strategies were ineffective—and that a null result could provide the impetus for trying new approaches.

The factors that enabled our work with CMS are increasingly present at the state and local level. J-PAL North America is providing support for studies through its State and Local Innovation Initiative. Groups like The Lab @ D.C. and the Rhode Island Innovative Policy Lab are empowering government experimentation. The establishment of these groups by mayors and governors and the rapid expansion of their research portfolios are particularly encouraging signs.

Policymakers at all levels of government are in need of rigorous evidence on effective methods to stem prescribing that conflicts with the clinical needs of patients. We believe there is a bright future for randomized evaluations that bring together government and academics to address this issue and many others.

References

Allcott, Hunt and Todd Rogers. 2012. "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." w18492. Cambridge, MA: National Bureau of Economic Research.

Becker, Gary S. 1968. "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76 (2): 169–217. doi:10.1086/259394.

Castro, Lucio, and Carlos Scartascini. 2015. "Tax Compliance and Enforcement in the Pampas Evidence from a Field Experiment." *Journal of Economic Behavior & Organization* 116 (August): 65–82.

Fellner, Gerlinde, Rupert Sausgruber, and Christian Traxler. 2013. "Testing Enforcement Strategies In The Field: Threat, Moral Appeal And Social Information." *Journal of the European Economic Association* 11 (3): 634–60.

Finkelstein, Amy and Sarah Taubman. 2015. "Randomize Evaluations to Improve Health Care Delivery." *Science* 347 (6223): 720–22.

Fleming-Dutra, Katherine E., Adam L. Hersh, Daniel J. Shapiro, Monina Bartoces, Eva A. Enns, Thomas M. File, Jonathan A. Finkelstein, et al. 2016. "Prevalence of Inappropriate Antibiotic Prescriptions Among US Ambulatory Care Visits, 2010-2011." *JAMA* 315 (17): 1864–73.

Hallsworth, Michael, John List, Robert Metcalfe, and Ivo Vlaev. 2014. "The Behavioralist As Tax Collector: Using Natural Field Experiments to Enhance Tax Compliance." Working Paper 20007. National Bureau of Economic Research.

Levitt, Steven D., and John A. List. 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?" *Journal of Economic Perspectives* 21 (2): 153–74.

Mack, Karin A., Kun Zhang, Leonard Paulozzi, and Christopher Jones. 2015. "Prescription Practices Involving Opioid Analgesics among Americans with Medicaid, 2010." *Journal of Health Care for the Poor and Underserved* 26 (1): 182–98.

Paulozzi, Leonard J., Christopher M. Jones, Karin A. Mack, and Rose A. Rudd. 2011. "Vital Signs: Overdoses of Prescription Opioid Pain Relievers---United States, 1999--2008." *MMWR. Morbidity and Mortality Weekly Report* 60 (43): 1487–92.

Sacarny, Adam, David Yokum, Amy Finkelstein, and Shantanu Agrawal. 2016. "Medicare Letters To Curb Overprescribing Of Controlled Substances Had No Detectable Effect On Providers." *Health Affairs* 35 (3): 471–79.

Skinner, Jonathan. 2011. "Causes and Consequences of Regional Variations in Health Care." In *Handbook of Health Economics*, 2:45–93. Elsevier.

Figures

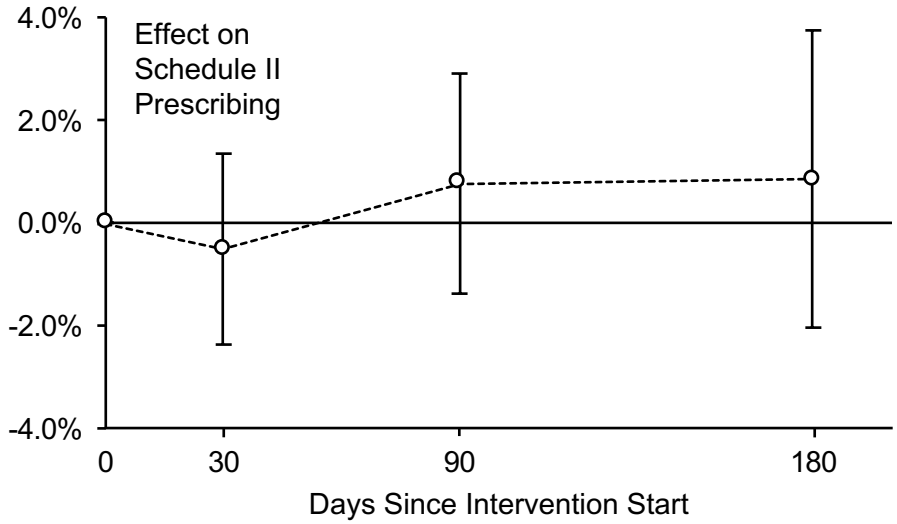

Figure 1. Effect of Initial Intervention on Prescribing

Note: Percent effect of initial intervention on prescribing of Schedule II controlled substances. Error bars are 95% confidence intervals. See Sacarny et al. (2016).