Who Do Innovations Reach?
The Influence of Training on Mental Health Treatments*

Alex Hoagland†

September 16, 2019

Abstract

As fields become more specialized, ineffective communication between innovators and practitioners can slow the diffusion of ideas. This paper examines the impact of continuing education in eating disorder treatment, comparing the take-up of (i) tangible innovations (psychopharmacology) and (ii) intangible innovations (psychotherapy) following professional conferences. I use a novel extension of an estimator proposed by Calvi, Lewbel, and Tommasi (2019) in an event study setting to overcome data limitations. I find very small responses among therapists for both kinds of innovations, suggesting that continuing medical education is not an important channel for treatment diffusion. Therapists respond more to education in pharmacology than psychotherapy, being about 3 percentage points more likely to write new prescriptions following a conference. This increase occurs mainly for adolescent patients being treated by non-psychiatrist prescribers. Response to purely psychotherapeutic innovations is limited to more academic-oriented specialists such as psychologists.

Keywords: Mental Health Care , Innovation, Diffusion of Ideas, Classification Error, Event Study

*I'm very grateful to Randall Ellis, James Feigenbaum, Raymond Fisman, Jihye Jeon, Timothy Layton, Arthur Lewbel, Albert Ching-to Ma, Daniele Paserman, Patrick Power, & Marc Rysman for helpful comments on this project. Also thanks to participants from the Lecznar Memorial Lecture and the BU Applied Microeconomics Reading Group. All remaining errors are mine.

†Department of Economics, Boston University. *Email:* alcobe@bu.edu *Website:* alex-hoagland.github.io

1 Introduction

Innovations rest at the heart of many endeavors, and their development, diffusion, and deployment pose critical questions across the spectrum of economic investigation. Generally, innovations studied in economic models are all treated alike, either as random shocks changing a technological process, or a simple event disrupting an equilibrium. In these senses, innovations can be evaluated as though they were policy changes, utilizing many of the simple causal inference tools popular in the field.

However, a more in-depth study of how innovations are discovered and proceed to sway equilibria requires an explicit differentiation of innovation types. Some innovations, for example, are mechanical, such as a software update to a technology that can improve performance for a one-time fixed cost. Others require a more hands-on approach, such as those that require learning-by-doing (Arrow, 1962) or similar methods. Innovations—like many other economic objects—are heterogeneous, and can take on a continuum of values in a potentially high-dimensional characteristic space.

One question that has yet to be asked is how these characteristics affects each innovation's success. It is reasonable that innovations with higher fixed costs, more variation in outcome, or other frictions may diffuse more slowly than innovations with a more straightforward one-time updating cost. Hence, especially as a landscape of innovation tends to the more intangible and artisanal, the spread of new ideas in a field may slow, resulting in gaps between the cutting edge of research and the use of these techniques in practice. Such a gap—commonly referred to as a research-to-practice gap (RPG)—constitutes an important problem in many areas of research, including healthcare (Glanz, Rimer, & Viswanath, 2008; Glasgow & Emmons, 2007; Wandersman et al., 2008) with particular emphasis on mental health (Jensen et al., 1999; Kazdin, 2011, 2017, 2018; Kazdin, Fitzsimmons-Craft, & Wilfley, 2017). Other important fields investigating RPGs include management practices (Bansal, Bertels, Ewart, MacConnachie, & O'Brien, 2012; Burke & Rau, 2010; Rynes, Colbert, & Brown, 2002), education (Coburn & Penuel, 2016; Strohman, 2014), and civil practices such as social work (Rountree & Pomeroy, 2010).

This project studies a RPG in mental health care. Mental health care is a burgeoning field of both research and practice, especially as mental health issues become more prominent in the United States (Olfson, Druss, & Marcus, 2015). Developments in mental health treatments are typically of two types: pharmacological (e.g., new drugs) or therapeutic (e.g., new models of psychotherapy). My aim is to exploit the differences in these innovations to examine a potentially differentiated rate of innovation take-up among practitioners. I exploit quasi-random attendance of professional trainings (in the form of professional conferences)

in both psychotherapy and psychopharmacology among mental health professionals, and assess the impact of each. I implement a panel event study design to assess changes in treatment patterns for therapists who are most likely to attend professional conferences in eating disorder treatments. I explore potentially differentiated responses by provider type and patient demographics, and conclude with an exploration of potential mechanisms for these responses and a validation of my treatment assigning algorithm.

I find muted response among mental health professionals to either kind of professional conference. While this may be the result of an overtaxing estimation process, it provides some suggestive evidence that continuing medical education is not the driver for changes in the treatment behaviors of therapists. Therapists did increase their use of olanzapine (an atypical antipsychotic occasionally prescribed for eating disorder treatments, discussed more in Section 2.2). Interestingly, this response occurred only among non-psychiatrists (e.g., psychiatric nurse practitioners) and was used on adolescent patients. However, therapists did not have a similar response to therapeutic education; in fact, when exploring the overall variation in a provider's treatment profile, I find suggestive evidence that a conference *discourages* experimentation.

For clarity, in this paper I make the (somewhat informal) distinction between *tangible* and *intangible* innovations. Tangible innovations are algorithmic in nature: while they may require specific skills and training to be able to implement, their implementation requires little creativity and varies little across implementations and practitioners. Many of the innovations that come easily to mind—new drugs, medical equipment, etc.—fall into this category. In contrast, intangible innovations depend more heavily on human capital, and therefore can vary widely based on who is implementing it (or even across cases with the same practitioner). The example of intangible innovation used in this project is psychotherapy, which is a rigorous and scientific medical treatment, but also requires a conscious cultivation of relationship between therapist and patient that is impossible to achieve algorithmically. While new therapeutic techniques can be proposed and validated by mental health researchers, the passing on of these guidelines from researcher to practitioner will inevitably leave room for practitioners to adapt the practice to their own treatment style, potentially altering the benefits of the development. Other examples of intangible innovations in health care include testing and prescription guidelines (Mullainathan & Obermeyer, 2019), as well as any other behaviors subject to clinician interpretation.

Of course, this distinction is a simplifying one, as nearly all innovations contain elements of both "art" and "science". For example, Graham, Lattie, and Mohr (2019) discuss the implementation of new digital mental health technologies, an ostensibly algorithmic innovation (e.g., a cell phone application) that requires specialist understanding of the mechanisms at

play in order to be successfully integrated into a treatment plan. While elements of artisanal and algorithmic innovations exist in almost every development (particularly in a field such as mental health), I have attempted to choose two key innovations that are as close to purely tangible and intangible as possible: psychotropic medication and psychotherapeutic techniques.

The contributions of this study are both methodological and practical. First, this study proposes a way to point identify dynamic treatment effects even in the presence of classification error. This extension of recent work (most notably, Calvi et al., 2019) increases researchers' flexibility to answer causally motivated questions in the presence of limited data, as well as suggesting ways predictive algorithms (such as machine learning techniques) could be used in causal designs. In addition, this paper discusses how an interpretation of these results might change when the necessary assumptions are implausible or hold only partially, and outlines how validation samples can be used to test the necessary assumptions.

From a clinical perspective, this project also contributes to a broad discussion on gaps between research and practice by highlighting one of the most common frictions in the diffusion of ideas: communication. Some papers find strong responses of medical professionals to randomized trials (Depalo, Bhattacharya, Atella, & Belotti, 2019), but the dissemination of this information is not always straightforward (Casper, 2007; Grimshaw et al., 2001). Continuing education is the most common method by which medical professionals receive information about new medical research (Church et al., 2010). However, as even medical conferences become more specialized, tailored either to academics[1] or professionals, continuing education has the potential to devolve into a "blind leading the blind" environment, where the trainers are as removed from medical research as the trainees. This, and many other factors, warrants an evaluation of continuing education as a potential source of research-to-practice gaps. This study contributes not only to a discussion on the uses of continuing education, but also a much larger literature on innovation diffusion in intangible settings.

This paper is also tangentially related to a burgeoning literature on the diffusion of ideas, a discussion on how intangible goods such as international ideals (Gilardi, 2012) and social movements (Rane & Salem, 2012). For example, Ash, Chen, and Naidu (2019) examine the spread of economic language among judges following a training program. Their particular type of policy evaluation (with dynamic treatment effects) is similar to the aims of this paper.

[1]For example, the Eating Disorders Research Society holds an annual conference limited only to its members. As a result, only academics attend, not professionals.

2 Background & Data

The diffusion of innovation into practice is a central issue for nearly every area of technological advancement. In simple cases, standard economic models predict that technologies that increase marginal benefit or decrease marginal cost will have quicker take-up by practitioners, becoming a new norm until further innovation disrupts the equilibrium again (Christensen, Baumann, Ruggles, & Sadtler, 2006; Christensen, Grossman, & Hwang, 2009; Christensen, Raynor, & McDonald, 2015). However, in the presence of frictions, the diffusion of innovations may depend on much more than their simple benefit/cost contributions, and standard models may be insufficient to predict how a field will develop. This is particularly true when innovations are intangible in nature, as this makes them particularly vulnerable to frictions.

2.1 Research to Practice Gaps

One friction that is particularly salient in the diffusion of intangible medical innovation is a growing divide between academics and professionals (Kazdin & Blase, 2011; Kazdin et al., 2017). With increasing specialization, a burgeoning field such as mental health care becomes split into two camps: one performing and reporting the results of clinical trials and other research, and a second that interprets and incorporates these results as they treat real patients. However, as this specialization progresses, the distance a new idea must travel from the laboratory to the patient increases, raising the chances that it will either not be adopted, or adopted in some stunted capacity.

Communication between these two groups—especially in the medical profession—is incentivized through continuing medical education (CME) programs for practitioners. These programs are motivated by the documented fact that physicians who have been practicing longer tend to stall in updating their practices, putting them at risk for delivering lower-quality care (Choudhry, Fletcher, & Soumerai, 2005). While the structure of CME programs tends to vary across states and facilities, a typical curriculum generally requires a mix of completing courses taught by state-approved providers, preparing and teaching courses to other professionals, and presenting at professional conferences, with additional options for research, publications, or media involvement. Licensures may be awarded following the completion of certain milestones in a CME program, allowing a mental health professional to advertise as "licensed" in an attempt to increase demand.

In recent years, CME programs have evolved to allow online learning through approved online classes, webinars, and presentations. This has been done largely to reduce the burden continuing education places on rural physicians (Curran, Fleet, & Kirby, 2006) and improve access more generally. In fact, Hugenholtz, de Croon, Smits, van Dijk, and Nieuwenhuijsen

(2008) have demonstrated that online continuing education is just as effective as traditional, in-person lectures. Despite this, most states still require at least some continuing education to be done in person. Because of this, professional conferences continue to be hubs for continuing education presentations, exams, and courses.

The potential benefits for professional conferences are inherent in the nature of the event, and tend to be highly favored by practitioners (Dysart & Tomlin, 2002). In fact, according to Dysart and Tomlin, professional conferences are attended with about the same frequency as all other continuing education events combined;[2] however, their work also highlights the difficulties associated with receiving education through expensive and travel-intensive methods such as conference attendance. Healthcare facilities are rarely generous in providing time off for conference attendance, and conference and travel fees are typically borne by the provider rather than the employer.

2.2 The Case of Eating Disorder Treatments

In an attempt to assess the quality of communication and training in inducing innovation takeup, the current project examines continuing education on practices in the treatment of eating disorders. These mental disorders centered around unhealthy relationships with food and eating. They include *anorexia nervosa*, typified by body dysmorphia and severe restriction of food intake; *bulimia nervosa*, characterized by purging excessive food consumption; *binge eating disorder*, a disease marked by superfluous food consumption (but no purging); and other unspecified diseases. This study will focus on patients with diagnoses of either anorexia nervosa or bulimia nervosa exclusively.[3]

These diseases are ideal for the current study for three principal reasons. First, these diseases have the highest mortality rate of any mental illness (Arcelus, Mitchell, Wales, & Nielsen, 2011), making them a pragmatically relevant area of focus. Second, treatment of eating disorders involves both algorithmic and intangible processes: for example, the refeeding process of severely malnourished anorexic patients is medically more straightforward than the psychotherapeutic aspects of treatment. However, as discussed in more detail below, many of the algorithmic treatment methods—such as pharmacological treatments—have much weaker empirical support than psychotherapies. Hence, in the absence of a research-to-practice gap biasing treatments towards algorithmic interventions, one should observe the use of intangible treatments (e.g., psychotherapy) dwarfing the number of pharmacological interventions. Finally, the study of eating disorders meshses well with available data. While

[2]Their study examined occupational therapists, rather than mental health professionals.

[3]Note that binge eating disorder did not have its own diagnosis code until the release of the ICD-10-CM Diagnosis Codes, which were used beginning in October 2015 (after my sample started).

it is a myth that they affect only female adolescents from the middle- and upper- classes (Mitchison, Hay, Slewa-Younan, & Mond, 2014), a substantial number of those suffering from this disease will have private insurance. Additionally, there are easily identifiable diagnosis codes for each eating disorder and treatment codes for the two treatments of interest (family-based therapy and olanzapine prescriptions, discussed below). Therefore, I have a clean identification of the populations and outcomes of interest.

While treatment patterns vary for each individual patient, treatment of eating disorders is recommended to follow a team-based model of care (American Psychiatric Association, 2006), with the team generally comprised of a principal psychotherapist, a dietitian (and other general medicine professionals if needed to deal with secondary effects of the disorder), a psychiatrist, and occassionally a social worker. Treatment proceeds in stages, with early stages focused on rectifying any secondary effects of an eating disorder (e.g., a refeeding or rapid weight gain process), and later stages focusing on mental health treatments. Hospitalizations—if any are required—typically take place in the first stages, with the latter stages largely taking place in an outpatient setting. It is this latter, mental health-oriented stage, with which this project is concerned. This stage typically consists of two major treatment modalities: psychotherapeutic and psychopharmacological.

Family-based therapies (FBT) are considered an optimal therapeutic intervention for the treatment of anorexia nervosa, bulimia nervosa, and eating disorders not otherwise specified (Loeb, Lock, Greif, & Le Grange, 2012). In this treatment, family members of a patient are integrated into a team of health professionals, as opposed to other psychological practices, which at best ostracize family members and at worst paint them as responsible for mental illnesses (Le Grange, Lock, Loeb, & Nicholls, 2010). FBT, developed at the Maudsley hospital in London by Dare and Eisler (2000) and manualized for anorexia nervosa by Lock and Grange (2015), currently boasts the strongest empirical support of any psycho-therapeutic intervention for treating anorexia nervosa, including hospitalization.[4] The two most recent meta-analyses—Couturier, Kimber, and Szatmari (2013) and Bulik, Berkman, Brownley, Sedway, and Lohr (2007)—each conclude that family-based treatments are more efficacious than many routine treatment methods, particularly for adolescents and youth. Importantly, the advantages of FBT are most notable in the long-term, with positive impacts 6–12 months after treatment that outweigh even the benefits of individual cognitive-based therapy (Couturier et al., 2013). As these authors write:

> "Family therapy focusing on symptom interruption of eating disordered behaviors should be recommended as the first line of treatment for adolescents with

[4] A complete list of RCTs evaluating the effectiveness of FBT for eating disorder treatments can be found here.

> eating disorders. Given the growing evidence base for FBT for adolescents with eating disorders, it would be prudent to study implementation strategies and effectiveness of this treatment in the community." (Couturier et al., 2013)

Family-based therapy is recommended by the American Psychiatric Association (American Psychiatric Association, 2006) and the National Institute for Health and Care Excellence in the UK (for Health and Care Excellence, 2017) as the main intervention for eating disorders. Despite this, however, the overall use of FBT in eating disorder treatments in the outpatient setting remains consistently low. Figure 1 shows the percentage of all eating disorder patients in the MarketScan data receiving any form of family-based treatment over time. The graph shows that only around 15% of the 23,000 patients in the sample (and around 26% of the 10,000 youth and adolescent patients) ever receive FBT in their treatment. Furthermore, the graph shows the publication dates of major RCTs and meta-analyses positively evaluating FBT, with little implied physician response shown as a result. This suggests that providers may already have sorted into those who provide FBT to their patients and those who do not, and that the current stream of ongoing research does not affect their decision to provide this treatment.

Of course, FBT will not be ideal for every eating disorder patient. Factors such as family instability, need for longer treatment, and co-morbid psychiatric disorder may influence a patient's lack of response to FBT (Lock, Couturier, Bryson, & Agras, 2006). Additionally, FBT has been proven more useful for adolescents than adults (Bulik et al., 2007). Finally, some specialists are able to use family-based techniques across a wide range of diagnoses outside of eating disorders; this may incentivize certain mental health practitioners to specialize in a family-based approach, allowing specific patients to seek out this treatment modality if they feel it may be a good match for their needs.

The second major treatment modality for eating disorders is pharmacological; however, the evidence base for this style of treatment is scant relative to that of therapeutic techniques. There are only two FDA-approved medications for eating disorder treatments: fluoxetine for bulimia nervosa (approved in 1994) and vyvanse for binge eating disorder (expanded approval to BED in 2015), both of which suppress purging behaviors. Additional medications—particularly SSRIs or other antidepressants—are typically prescribed to assist in mitigating co-morbid depression and/or anxiety symptoms (American Psychiatric Association, 2006). Overall, there are no good pharmacological treatments to handle a patient's relationship with food, making the therapeutic treatment arm essential.

Even without empirical support, an increasing number of prescribers have begun engaging in off-label experimentation in the treatment of eating disorders (Maglione & Hu, 2011). Much of this experimentation uses atypical antipsychotics—which are FDA approved

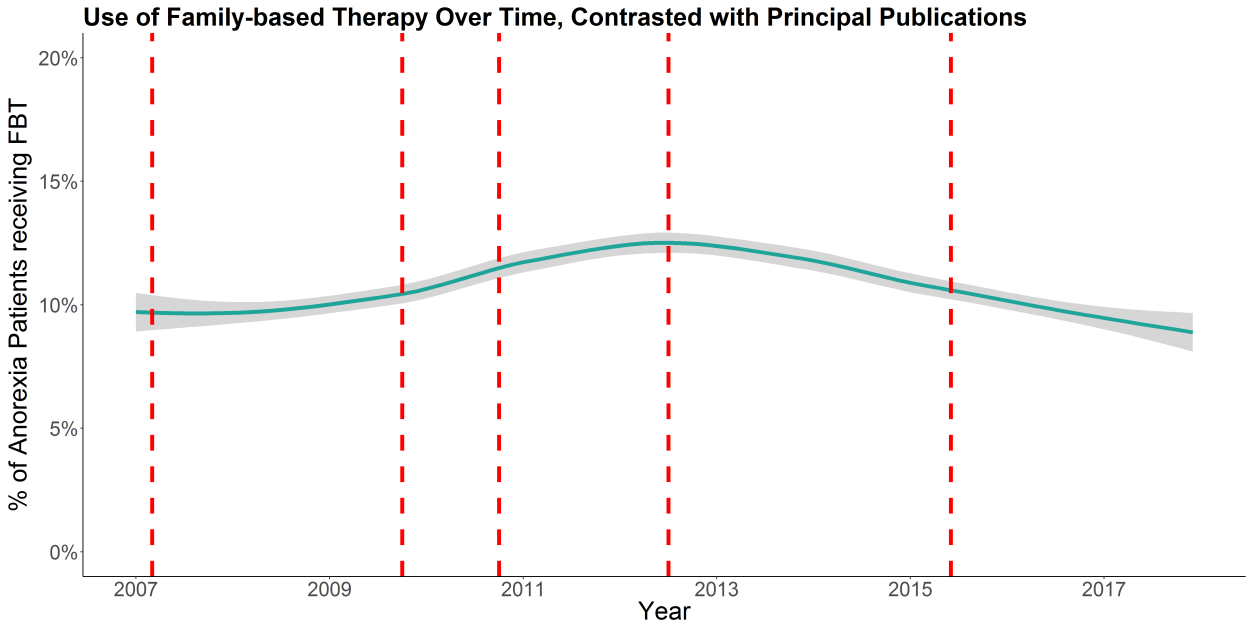

Figure 1
Therapist Response to Publications on Family-Based Therapy

and typically prescribed for schizophrenia and bipolar disorder—to manage weight gain. For example, olanzapine (the most commonly prescribed atypical antipsychotic for eating disorders) is known to induce weight gain as a common side effect, and hence has been viewed as potentially useful in anorexia nervosa treatments (Flament, Bissada, & Spettigue, 2012). While there have been some studies examining these medications (see Maglione and Hu (2011) for a meta-analysis), there is not enough conclusive evidence that these medications are effective in treating eating disorders to warrant a change in their FDA approval status presently; however, continuing education and professional conferences still include discussions of incorporating off-label drugs into psychopharmacological practice in an ED treatment profile.

2.3 This Project

This project focuses on a single potential friction between academic research and practice: the impact of professional education. Specifically, I focus on the implementation of FBT and prescription of olanzapine in eating disorder treatments, two innovations that embody different styles of innovation and may thus diffuse differently. The prescription of olanzapine, while an off-label practice with relatively little empirical support, has a straightforward implementation, and constitutes a more algorithmic innovation. However, the use of family-based therapies requires specialists to provide a higher level of care, and its implementation

therefore varies widely across therapists, in keeping with intangible innovation. This heterogeneous implementation of FBT in eating disorder treatments has been documented in Kosmerly, Waller, and Robinson (2015).

To evaluate this takeup, I use a list of about 70 conferences targeted at eating disorder professionals and clinicians. For each conference whose online program is available, I am able to ascertain if the conference ran any sessions or presentations on either FBT or olanzapine use in ED treatments, as well as creating a registry of the conference locations and times. Table 1 shows the organizations and conferences examined. Aside from conferences whose programs are not available, I have the universe of such professionally-oriented conferences.[5] I couple this with a sample of 4,476 therapists and professionals from the Truven MarketScan data to examine treatment profiles of specialists before and after conference attendance.

The main complication is that I have no data on who actually chose to attend each conference;[6] instead, I will estimate treatment status based on each specialist's cost (in travel) of attending a conference. By assuming that therapists are more likely to attend conferences that are low-cost to them, I am able to artificially assign specialists to treatment and control groups, as discussed in more detail in Section 3.1. Finally, I extend recent work on dealing with classification error in treatment effect models (Calvi et al., 2019) to approximate the local average treatment effect of attending these conferences.

Hence, this paper provides two distinct contributions. The first is methodological in nature, and presents a toolkit of econometric techniques to assist researchers in overcoming data limitations. Specifically, this paper introduces an instrumental variables technique for the event study approach, integrates predictive algorithms into a causal framework, and extends results that adjust these frameworks for classification errors. Due to the reasonably complicated procedure by which my results are derived, several sections of this paper are dedicated to the exposition of the algorithm and intuition behind its use.

Secondly, I present information detailing how medical professionals respond to continuing education in the form of professional conferences. I argue that these responses are potentially differentiated on the basis of which techniques or tools are being discussed, and examine heterogeneity by audience (specialist type) and population of interest (patient demographics). Ultimately, the results of this exercise provide little evidence that continuing education changes behavior in the aggregate, either for intangible or algorithmic innovations (psychotherapy and prescriptions, respectively). This finding warrants future research in

[5]Note that this excludes academic conferences which are limited to members of the academic organization only (for example, the Academy of Eating Disorders) as not all clinicians would have the opportunity to attend.

[6]Note that I have data on conference registration for a few conferences, which I will use in a validation exercise in Section 5.

Table 1
Professional Conferences on Eating Disorder (ED) Treatments Examined

Organization	Conference Name	Frequency	Total Programs	FBT	Olanzapine
Academy for EDs	International Conference on ED	Annual	9	7	4
American Academy of Child & Adolescent Psychiatry	AACAP Meetings	Annual	10	2	4
Annual Eating Recovery Foundation	ERF Conference	Annual	3	3	0
Center for Change	National ED Conference for Professionals	Annual	5	1	
International Association of ED Professionals	IAEDP Symposium	Annual	9	6	2
Maudsley Parents	One-Day FBT Conferences	Sporadic	3	3	0
Multi-service ED Association	MEDACon	Annual	3	2	1
National ED Association	NEDACon	Annual	4	3	0
Renfrew Center Foundation	Conference for Professionals, Seminar Series	Annual+	15	8	0
Center for ED at Sheppard Pratt	Professional Symposium	Annual	6	3	0
Summit for Clinical Excellence	National ED Conference	Sporadic	4	1	1
UCSD ED Treatment Center	Trainings for Professionals	Sporadic	2	1	0
		Total:	73	40	12

[1.] *Abbreviations:* ED = eating disorder; AACAP = American Academy of Child & Adolescent Psychiatry; ERF = Eating Recovery Foundation; IAEDP = International Association of Eating Disorder Professionals; FBT = Family-based therapy; UCSD = University of California at San Diego.

light of the severe data limitations and complex econometric procedure, which is taxing for the available data; however, if true, this finding suggests a need to better understand the optimal way to transmit information to practicing professionals.

3 Empirical Design

Dynamic treatment effects are at the heart of questions surrounding innovation adoption. I am ultimately interested in how professional conferences impacted the use of FBT and olanzapine *over time* for each specialist who attended. I have concrete data on each specialist's treatment profile for their subset of patients who are covered by an insurer in the MarketScan database; however, I do not have reliable data on conference attendance for these physicians. My empirical approach will (i) estimate treatment status for each medical professional and conference, (ii) estimate a dynamic treatment effect of professional education using an event study framework, and (iii) adjust for potential classification error in the first step.

This project combines various econometric approaches to attempt point identification of my dynamic treatment effect of interest. To fill in data gaps, I employ a predictive algorithm that infers who attends each conference; this suggests a place for more sophisticated machine learning techniques in causal research designs. To deal with the flaws inherent in any such

algorithm, I extend an estimator that is robust to measurement error in a treatment variable to an panel event study framework.[7] Using this estimator in tandem with a transformed IV approach allows me to approximate a Dynamic Local Average Treatment Effect (D-LATE) for the specialists in my sample who are induced to take-up treatment (the compliers).

Event study designs have become increasingly popular in recent years (see Abraham and Sun (2018); Borusyak and Jaravel (2017); and de Chaisemartin and D'Haultfoeuille (2019) for important reviews on the subject). These designs rely on variation in treatment time (with or without the presence of a control group to explore treatment effects in periods both leading up to and following the treatment period, as well as the presence of a control group to correctly control for time fixed effects (Hull, 2018). This design can flexibly be used to explore heterogeneous responses in an appealing way when the number of groups to compare is relatively small, as in Johannesen and Stolper (2017).

3.1 Estimating Treatment Categories

While my data are ideally suited for the study of a medical professional's treatment profile, they contain no information on continuing education or conference attendance. Hence, I use a predictive algorithm to infer each specialist's decision to attend a CME conference based on their travel costs. The algorithm is based off of the assumption that given that opportunities for continuing education are nearly ubiquitous, decisions to attend conferences for professionals will be driven largely by costs: an ED specialist in Boston is far more likely to attend conferences when they are held in New England than when they are held in California.

Details of this algorithm are relegated to Appendix B for brevity. In general, for each mental health specialist and each conference, I compute a measure of travel cost taking into account both the physical cost of travel and the opportunity cost of time. From this continuous measure, I infer a treatment group as the smallest η-percentile of specialists when ranked by their travel costs. This move from a continuous variable to a discrete one is motivated by the classification error framework laid out in the next subsection; by varying this threshold I change the probabilities of misclassifying a treated/control therapist in a near-monotonic fashion.[8] This is useful for the assumptions of the MR-LATE estimator discussed in Section 3.3. However, future research might explore the potential use of this continuous measure in a propensity-weighting framework, as well as how such a framework

[7] The Mismeasurement Robust LATE Estimator of Calvi et al. (2019), discussed in more detail in Section 3.3.1.

[8] That is, as η increases, I tend to increase the probability of classifying a control therapist as being treated, while decreasing the probability of classifying a treated therapist as part of the control group.

compares to that of Calvi et al. (2019). Additionally, future research could integrate more sophisticated machine learning techniques to improve prediction accuracy, providing a better approximation of the true LATE (as discussed in Section 3.3).

Figures 2 and 3 show an example of the algorithm's output for a sample conference for professionals that took place in September 2012 in Boston. Figure 2 shows the estimated travel cost to attend the conference for each specialist in the sample at that time, while Figure 3 shows the estimated distribution of travel costs, including various cities as reference points. Specialists in cities farther away from Boston incur greater travel costs to attending the conference, but those in distant *rural* areas (such as Mountain Home, Idaho) incur even greater travel costs. By selecting the lowest η-quantile of the distribution, different treatment groups are created, with differing levels of austerity in selecting the treatment (or control) groups[9]. Notice that these treatment groups are not merely centered around the conference location—indeed Atlanta, Georgia, which is a hub for major airlines, has a lower travel cost to a Boston 2012 conference than does New Haven, Connecticut. This illustrates that incorporating travel costs into the predictive algorithm may provide an improvement in prediction quality over a simple geographic distance calculation.

When repeated for all conferences, this procedure creates estimated treatment groups for each of the conferences in the sample (40 conferences for FBT trainings and 12 for olanzapine prescriptions). In order to conduct an event study analysis, it is important that treatment be an absorbing state for each therapist;[10] hence, each therapist is assigned a treatment date as the earliest time period for which it is estimated that they attended a conference on FBT or olanzapine. From this estimated treatment time, relative time dummies typical for an event study are created, completing the necessary data configuration.

3.2 Instruments in Event Study Designs

While the use of instruments in an event study is rare compared to their prevalence in other empirical designs, the generalization from the instrumented difference-in-differences design (DDIV) to an instrumented event study design (ESIV) is straightforward. As explained in Hudson, Hull, and Liebersohn (2017), the basic model for the DDIV is Equation 1:

$$y_{it} = \alpha_i + \tau_t + \beta D_{it} + \epsilon_{it}, \tag{1}$$

[9]The red line in Figure 3 illustrates a treatment group based on the lowest 10% of travel costs.

[10]That is, each specialist ought to be treated only once, and remain treated throughout the duration of the sample after that.

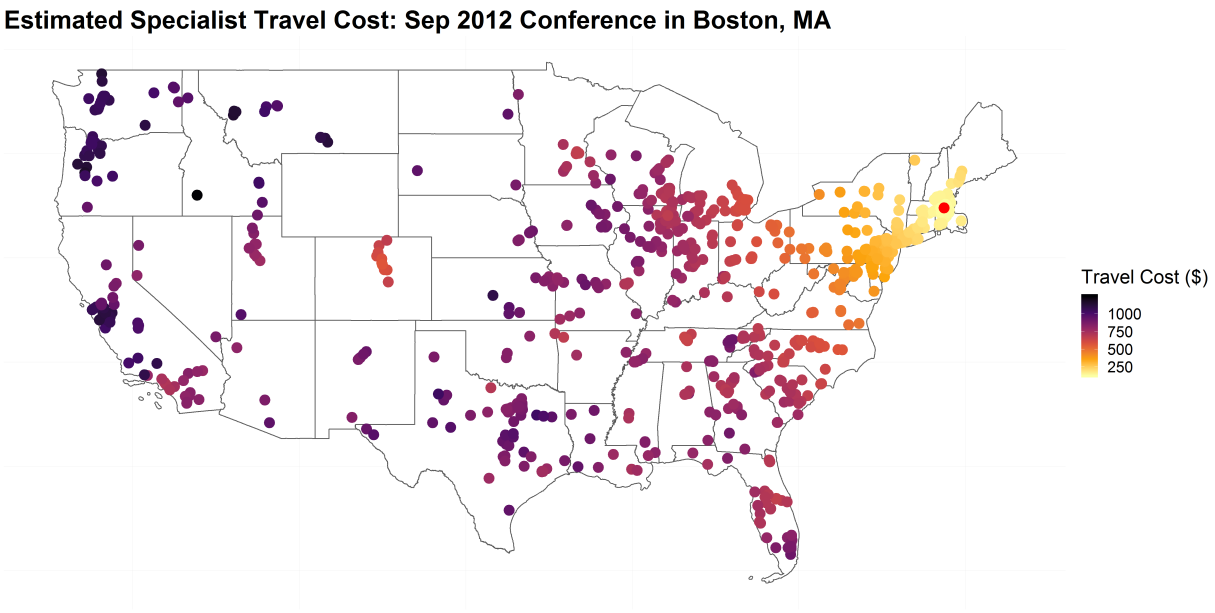

Figure 2
Estimated Travel Cost for all Specialists, Boston 2012 Conference

where α_i and τ_t represent individual and time fixed effects, and D_{it} is the binary (potentially endogenous) treatment status.[11] To deal with the endogeneity of treatment D_{it}, a binary instrument Z_{it} is used.

The event study framework generalizes this by mapping between a single treatment indicator D_{it} and a vector of *relative time dummies*, which indicate how much time has elapsed since the treatment event. For each individual i in a panel, the event is denoted as $E_i = \min_t\{D_{it} = 1\}$; given this, each period t can be assigned a value $K_{it} = t - E_i$. This essentially re-orders the time periods in a panel so that each individual appears to have been treated simultaneously. Once this is complete, the estimating equation can be written as Equation 2

$$y_{it} = \alpha_i + \tau_t + \sum_{k=-\infty}^{\infty} \gamma_k \mathbb{1}\{K_{it} = k\} + \epsilon_{it}. \tag{2}$$

In this setup, each parameter γ_i indicates the effect of the treatment event on the outcome variable i periods before or after the event itself. See Borusyak and Jaravel (2017), de Chaisemartin and D'Haultfoeuille (2019) for a more detailed discussion of the event study approach. Generally, applied researchers do not estimate the fully dynamic specification (where k ranges over all integers), but limit $k \in [-A, B]$ for two positive integers A and

[11] Note that additional controls can be added if desired. I ignore them in this section to simplify the exposition.

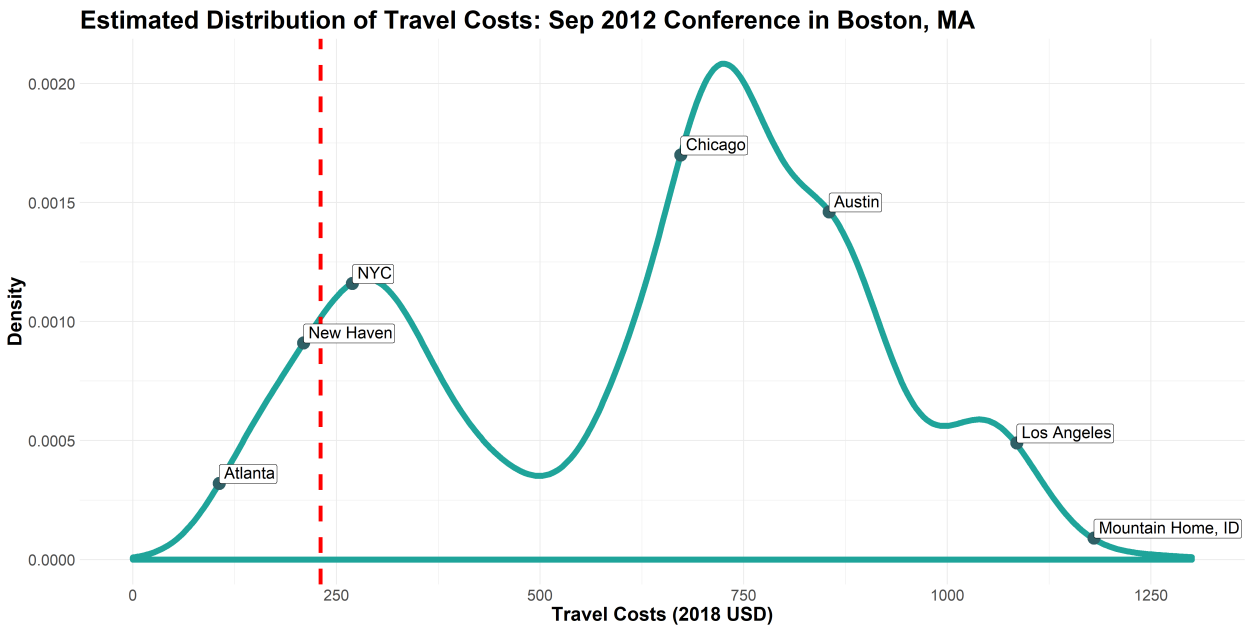

Figure 3
Estimated Distribution of Travel Cost for all Specialists, Boston 2012 Conference

B[12]. This establishes the vector parameters $(\gamma_0, \gamma_1, ..., \gamma_B)$ as the parameters of interest (sometimes referred to as the *dynamic treatment effect* parameters).

Suppose now that there exists a valid instrument Z_{it} for D_{it}. To transform this instrument to be a valid one for the event study approach, one need only follow the same procedure outlined above: for each individual i, define the instrumented event Z_i' as the point that is most likely to induce the treatment event, then define the relative time periods $Z_{it}' = t - Z_i'$ as before. Given that Z_{it} is correlated with D_{it}, the transformed instrument Z_{it}' will be correlated with K_{it}, ensuring that the procedure is valid.

My main instrumental variable is the presence of a "slow spell" for a therapist in the months leading up to a conference. Specifically, if a specialist's average patient volume 4–6 months prior to a conference is lower than their overall average volume, the binary instrument is given a value of 1 (and 0 otherwise). By indicating a potential decline in patients treated during a registration period for a conference, I hypothesize that this instrument will be positively correlated with true conference attendance. Additionally, since such a measure is uncorrelated with both (i) distance between therapist and conference and (ii) therapist treatment profiles, this measure is a valid instrument for the treatment. I therefore define Z_i' as the period with the lowest measured *lagged* patient volume for each specialist i.

[12]For identification, such an approach requires omitting a dummy as a reference group, which is typically chosen to be γ_{-1}.

3.3 Dealing with Classification Error

Given that I infer treatment status based on an imperfect proxy (travel costs), dealing with classification error is a first order concern in my estimation approach.[13] There is a small, but vibrant, literature on dealing with classification errors in applied microeconometric models. The most notable of these papers, Lewbel (2007b) point identifies the average treatment effect (ATE) in a simple treatment effects model. Other important papers extend this result to include covariates or discretized treatment levels (Hu, 2008; Lewbel, 2007a; Mahajan, 2006). Most recently, these researchers have turned to the problem of estimating Local Average Treatment Effects (LATEs) in the presence of potentially endogenous selection into treatment. This paper extends the recent work of Calvi et al. (2019), who identify a mismeasurement-robust estimator of the LATE (the MR-LATE) used for bias reduction in classification error problems.[14]

3.3.1 The MR-LATE Estimator of Calvi et al. (2019)

Calvi et al. (2019) propose an estimator that is "mismeasurement robust" in the sense that it can approximate the LATE under weak assumptions. In their framework, there is a true treatment status $D \in \{0, 1\}$, which is unobserved and cannot be consistently estimated. In addition, there exists a binary instrument Z such that the typical LATE assumptions of Imbens and Angrist (1994) are satisfied (as replicated in Assumption 1).

Assumption 1: LATE Assumptions. The outcome Y and true treatment status D, together with a binary instrument Z satisfy:

i. $0 < \mathbb{E}[D] < 1$, $0 < \mathbb{E}[Z] < 1$, and $Z \perp (Y_1, Y_0, D_1, D_0)$.
ii. (Y_1, Y_0, D_1, D_0, Z) are independent across individuals and have finite means.
iii. There are no defiers, hence $\mathbb{P}(D_0 = 1 \cap D_1 = 0) = 0$,

where subscripts are indicative of potential outcomes following the typical framework. While D cannot be consistently estimated, it is approximated by two imperfect measures $T^a, T^b \in \{0, 1\}$. These measures satisfy an extended set of the LATE assumptions in Assumption 2 (where compliers are denoted by C):

[13] Classification error refers to measurement error in a variable denoting treatment status. Ignoring this error—which by construction is nonclassical—can lead to serious problems in estimating a treatment effect, as discussed in detail in Millimet (2010). Kreider (2010) shows that even in a case of infrequent classification error—from 2% or less—can result in estimated effects whose confidence intervals do not overlap the true treatment effect, and may even suggest the opposite sign of the true ATE.

[14] There is another recent paper that tackles this issue (Yanagi, 2018), but this requires additional assumptions and applies to a more restricted class of circumstances.

Assumption 2: Mismeasured LATE Assumptions. T^i is such that the following conditions are satisfied for $i \in \{a, b\}$:

i. $Z \perp (Y_1, Y_0, D_1, D_0, T_1^i, T_0^i)$.
ii. $(T_1^i, T_0^i) \perp (Y_1, Y_0) | C$.
iii. $\mathbb{E}[T_1^i - T_0^i | C] \neq 0$.

That is, in addition to the typical unconfoundness assumption, Assumption 2-i assumes the instrument is independent of the potential measurement errors in T^i. The second part of the assumption indicates that the potential outcomes of each mismeasurement are independent of the potential outcomes of the dependent variable Y; combined with the first assumption, this asserts that any measurement errors are uncorrelated with outcome variables. Finally, Assumption 2-iii requires only that T provide some information about D.

Given these two assumptions, Calvi et al. (2019) apply the reasonable well-known fact that a transformed two-stage least squares (2SLS) regression of YT on T (using Z as the instrument) can be written as a mixture of the potential outcomes for compliers:

$$\frac{\text{Cov}(YT^i, Z)}{\text{Cov}(T^i, Z)} = \frac{\mathbb{E}(YT^i | Z = 1) - \mathbb{E}(YT^i | Z = 0)}{\mathbb{E}(T^i | Z = 1) - \mathbb{E}(T^i | Z = 0)} \tag{3}$$

$$= \mathbb{E}\left[qY_1 + (1 - q)Y_0 | C\right], \tag{4}$$

where q is a weight related to the probability of measurement errors in T given true treatment D. Given this result,[15] Calvi and coauthors define the MR-LATE estimator as the difference in two 2SLS estimators, given the two mismeasured treatments T^a and T^b:

$$\text{MR-LATE} \equiv \rho = \frac{\text{Cov}(YT^a, Z)}{\text{Cov}(T^a, Z)} - \frac{\text{Cov}(YT^b, Z)}{\text{Cov}(T^b, Z)}.$$

Using this definition and the result from their first theorem (Equation 4), it follows immediately that the MR-LATE is a multiple of the LATE, with the weighting $(q^a - q^b)$; Hence, the MR-LATE is equal to the true LATE when $(q^a - q^b) = 1$. A sufficient condition for this to hold is that of Assumption 3:

Assumption 3: Sufficient Condition for MR-LATE = LATE. T^a and T^b are such that the following two conditions are met:

i. $p_0^a = 0$. That is, among compliers, the mismeasured treatment T^a never mistakes the actually untreated as treated.

[15] This result is not unique to Calvi et al. (2019), but has been mentioned in earlier work, including Abadie (2002) and Ura (2018).

ii. $p_1^b = 0$. That is, among compliers, the mismeasured treatment T^b never mistakes the actually treated as untreated.

These restrictions—that one treatment group is strict in its definition of the treatment group, and the other in its definition of the control group—are related to the no-defiers assumption typical in a LATE framework. By eliminating certain combinations of D and Z, the no-defiers assumption allows for a clean interpretation of the local average treatment effect. In a similar vein, Assumption 3 rules out certain types of measurement errors, thereby eliminating extraneous cases wherein the MR-LATE would be different from the true LATE.

As in cases where the no-defiers assumption is violated, an MR-LATE estimator *approximates* the LATE in cases where Assumption 3's conditions are nearly met (meaning that $q^a - q^b$ is close to one). Judging the extent to which the conditions of these assumptions are met is typically impossible given the limitations of the data; however, I have obtained actual conference registration data from recent ED conferences held by the Academy for Eating Disorders, which I use as a validation sample in Section 5. With this new data, I am also able to address concerns about a lack of strong identification arising from an imprecise treatment group estimation.

3.3.2 This paper: The Dynamic LATE (D-LATE) Estimator

Extending Lewbel's work to the event-study setting is relatively straightforward. Theorem 1 below restates the necessary setup and assumptions for the MR-LATE to be identified for each parameter β_i of the dynamic treatment effect.

Theorem 1 *Let $\{Y, D, Z, T^a, T^b\}$ be such that Assumptions 1 and 2 are satisfied. Consider estimating an instrumented event study regression (equation 2) on the transformed variable T^iY using T as the treatment measure and Z as the instrument. Then, for any time period t relative to the treatment period, the dynamic treatment coefficient γ_t satisfies*

$$\gamma_t = \mathbb{E}[qY_1 + (1-q)Y_0|C], \tag{5}$$

for a q related to the probability of mismeasurement in the substitute treatment measure T^i.

See Appendix A for a proof of this theorem. This extension of the theorem relies on two facts: first, that an event study design is simply a transformation of the DDIV estimator into one with many dummy variables, as discussed in 3.2. Hence, estimating a LATE model with one instrument is equivalent to estimating a corresponding ESIV model with many instruments (one for each dummy). Second, as discussed in Angrist and Imbens (1995),

coefficients in two-stage least squares models with multiple instruments can be written as a linear combination of each instrument-specific LATE.

Given the results on Theorem 1, the corollary of Calvi et al. (2019) immediately implies that a dynamic version of the MR-LATE (which I call the dynamic MR-LATE or D-LATE for short) is equivalent to the true LATE under the conditions stipulated in Assumption 3. Hence, in order to resolve issues of classification error while still obtaining a dynamic treatment effect, I use two measures of treatment status—one that never misclassifies the treated, and another that never misclassifies the untreated—and the quasi-randomized instrument of patient volume during the conference registration period, as discussed in Section 3.2.

For the two mismeasured treatment estimates, I can use the travel cost algorithm described in the preceding subsection with varying thresholds. That is, I create two estimated treatment groups for each conference, one with a very strict threshold for attendance (e.g., only the lowest ventile of travel costs) and one with a very liberal threshold (e.g., the 95th percentile of travel costs). In this way, I ensure that one of the mismeasured treatments is unlikely to mistake a truly treated professional as a control member, and the other is unlikely to make the opposite mistake, thereby at least approximating the sufficient conditions for the D-LATE estimator to be equivalent to the LATE.

I therefore obtain estimates and standard errors of the D-LATE using the following procedure. First, I estimate two event study regressions (using equation 2) using T^iY as the dependent variable, T^i as the treatment status (that determines the dummy variables), and Z as the instrument. The MR-LATE estimator for each coefficient of interest γ_i is given by $\gamma_i^{MR} = \gamma_i^a - \gamma_i^b$. Finally, I obtain bootstrapped panel errors for each coefficient use the panel bootstrap method.[16]

4 Estimation Results & Heterogeneity

The D-LATE estimator was implemented to evaluate two sets of professional conferences: one targetting the use of family-based therapies (FBT) and another the prescription of atypical antipsychotics (olanzapine) in eating disorders. In both cases, I am interested in the effect these conferences have on individual therapist experimentation; I therefore measure short-term responses to a conference by the likelihood of utilizing an innovation in the first 6 months following the event.

The main results of the event study on FBT takeup can be seen in Figure 4. The point estimates suggest that in the month following conference attendance, FBT techniques were about 8 percentage points more likely to be employed. However, large bootstrapped standard

[16]See Kapetanios (2008) for an excellent review of this procedure.

errors and large pre-trend effects suggest that this result is more attributable to sampling variation than a true therapist response. Even if there is a short-term response, it quickly diminishes in the subsequent periods, suggesting a short period of experimentation without true adoption. As I will discuss in Section 4.1, this result is robust to multiple specifications.

A similar result holds for olanzapine prescriptions, as seen in Figure 5. The estimated coefficients for this treatment effect are much smaller, with at most a 0.4 percentage point increase in prescriptions after conference attendance. Overall, the results suggest little, if any, change in prescribing behavior. The fact that this response is less dramatic than responses to FBT is somewhat surprising, given my hypothesis about innovation types. I will discuss potential interpretation of these results in Section 6.

4.1 Robustness

The main results shown above are robust to multiple expressions of the regression specification. In particular, I compared results with binary and continuous dependent variables, the use of all prescriptions (compared to only olanzapine prescriptions), and the decision of whether to normalize the travel costs by specialist salary (as discussed in Appendix B). Figures showing how the estimated coefficients changed based on these varying approaches can be found in Appendix C.

In addition to these typical robustness checks, I also assessed how the results changed relative to my specification for the two mis-measured treatments. My specification uses cutoff thresholds in travel costs to assign treatment status to specialists; however, as discussed in Section 3.3.2, there is a tradeoff between satisfying the conditions of Assumption 3 and maximizing their correlation with the true treatment status (e.g., mitigating concerns of a weak instrument problem). I therefore repeat the estimation procedure using various treatment thresholds, which can also be viewed in Appendix C. The results are quite consistent—if anything, models estimated with more stringent treatment thresholds (smaller η) appear to detect larger estimates, but have larger standard errors as well. While future work may elaborate on the optimal decision of treatment threshold to balance the trade off inherent in its selection, this figure provides sufficient evidence that the choice of threshold contributes little to the overall result.

4.2 Heterogeneous Responses by Patient Age

While the overall results show little specialist response to professional conferences—whether targetting algorithmic or intangible innovations—a null result may mask interesting heterogeneous responses. To that end, I investigate potentially differentiated responses by patient

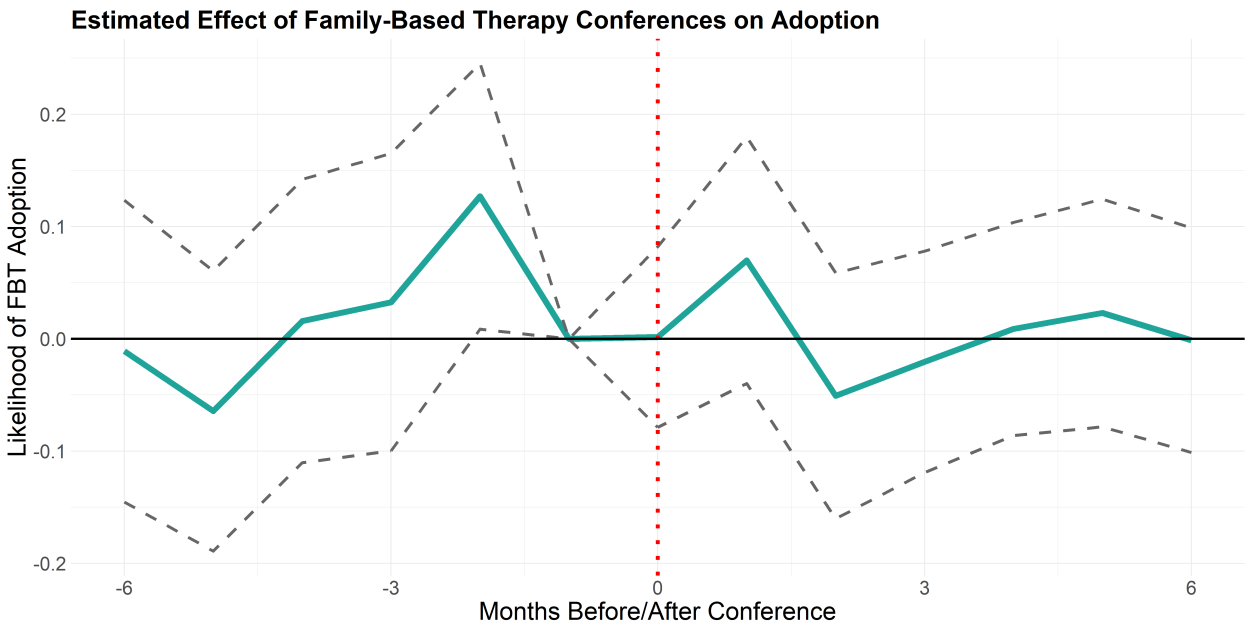

Figure 4
Therapist Response to Family-Based Therapy Conferences

and specialist type. First, specialists may respond to professional conferences selectively, choosing to implement new techniques on a subset of their patient pool. Particularly, family-based therapies have been shown to be more effective for adolescents and children, for whom family structure is a more integral social context (Couturier et al., 2013). On the other hand, pharmacological interventions may appear more tolerable for adult patients, especially those for whom FBT is not a viable option.

To explore potential heterogeneity along this dimension, I re-estimate the results on the subset of patients who are under 20 years old. Figures 6 and 7 show the results for the effect of FBT and olanzapine professional conferences on treatment profiles for youth and adolescents. The results for FBT use—a treatment which should ostensibly be easier to implement among adolescents and youth—are practically identical to those shown in Figure 4; however, the results for olanzapine use suggest a small, but more significant, increase in prescriptions for youth following pharmacological conferences. This suggests a certain degree of differentiated response among practitioners based on the type of patients they see, although not in the way one would generally hypothesize.

4.3 Heterogeneous Responses by Specialist Type

In addition to potentially heterogeneous response by patient types, specialists themselves may differ in their responses to professional conferences. For example, specialists who

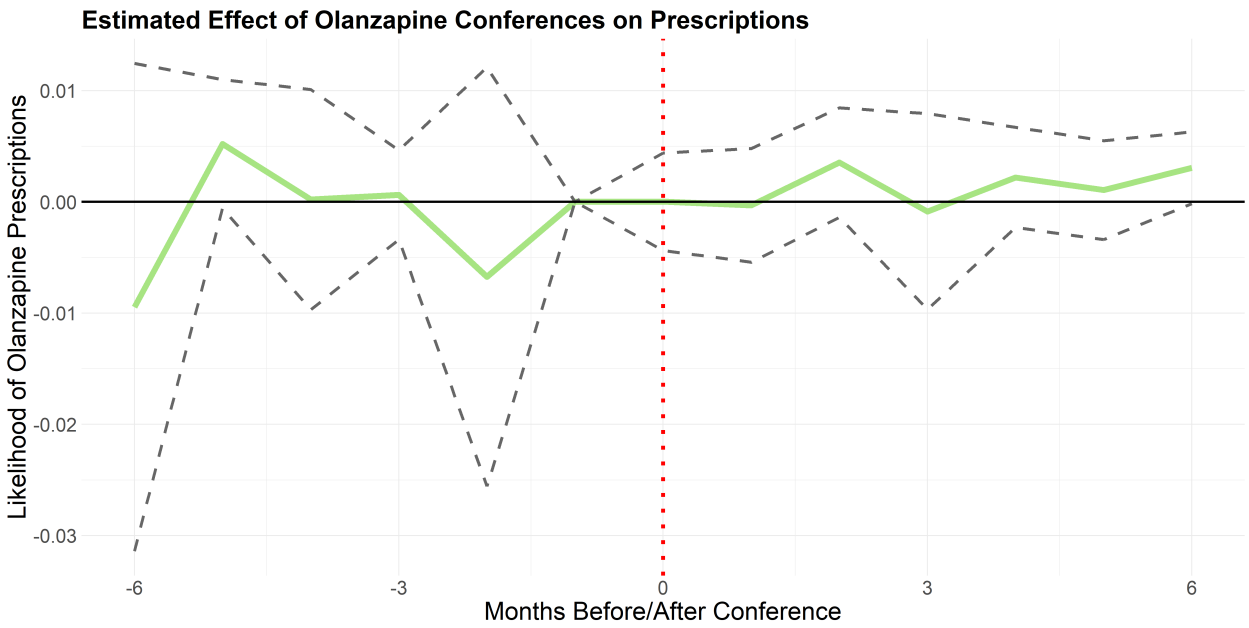

Figure 5
Prescriber Response to Olanzapine Conferences

hail from a more academic background (e.g., psychologists) may place a higher priority on evidence-based treatments, and may therefore be more likely to integrate FBT or olanzapine into their treatment profiles. To examine this question, I estimate an extended ESIV model using the D-LATE procedure, including interaction terms for specialist types. That is, I examine the specification in Equation 6:

$$y_{it} = \alpha_i + \tau_t + \vec{\gamma}\boldsymbol{T}_t + \vec{\delta}\left(\boldsymbol{T}_t \times \boldsymbol{s}_i\right) + \epsilon_{it}, \tag{6}$$

where $\boldsymbol{T}_t$ is the vector of relative time dummies used in the event study and $\boldsymbol{s}_i$ are the relevant specialty types examined in the regression. Then, the coefficients of interest are contained in the vector $\vec{\delta}$. Recent papers such as Johannesen and Stolper (2017) have used this approach as a simple way to explore potential heterogeneous treatment effects.[17]

To examine heterogeneous takeup of FBT, I compare psychologists and therapists to other mental health clinicians (family practice doctors, mental health facilities professionals, etc.); for olanzapine prescriptions, I compare psychiatrists to non mental-health prescribers (e.g., family practice doctors). Figure 8 shows the differentiated response for FBT takeup, while Figure 9 shows the same for prescribing. In each figure, the first panel illustrates the

[17]Notice that it isn't necessary to include level effects for each specialist type $s_i \in \boldsymbol{s}_i$, as these will be picked up by individual fixed effects (for the large majority of the individuals in the sample who don't switch provider types).

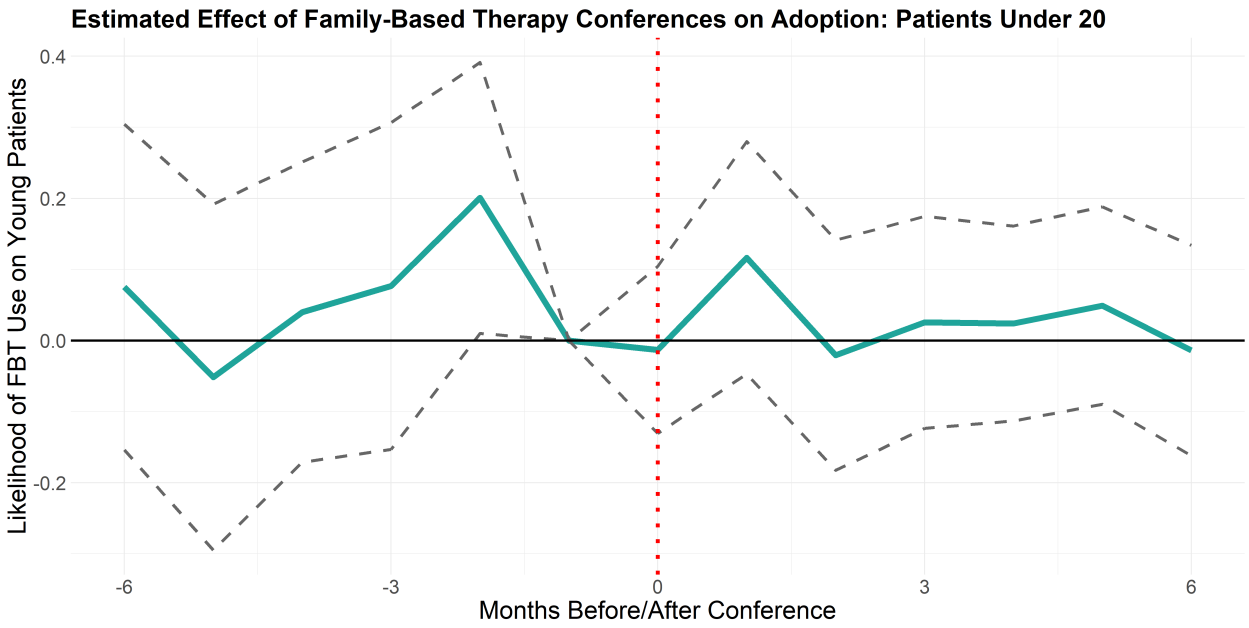

Figure 6
Therapist Response to Family-Based Therapy Conferences Among Patients Under 20

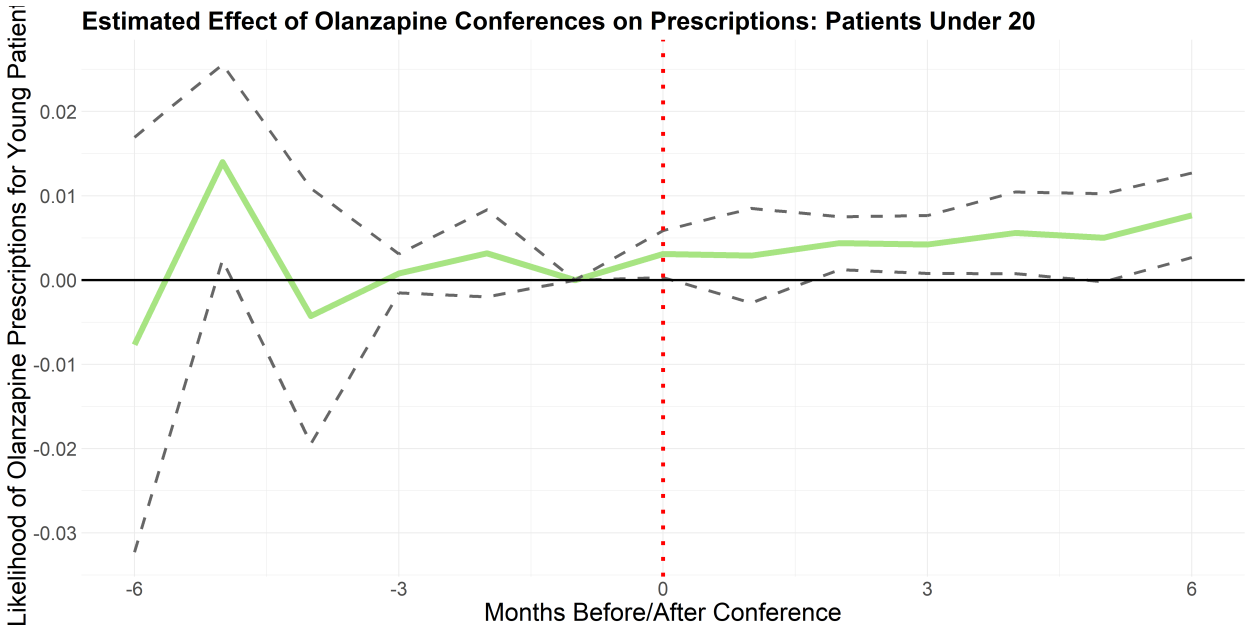

Figure 7
Prescriber Response to Olanzapine Conferences Among Patients Under 20

overall dynamic treatment effect (the vector γ in Equation 6), while the other panels are the relevant parts of the δ vector for each specialist type—therefore, these panels are interpreted as the *relative* differences in dynamic treatment effects for each group.[18] Psychologists in general appear to have a much higher fluctuation in the use of family-based therapy, but tend to use it overall more than their counterparts. The point estimates suggest a higher positive reaction to the use of FBT for them, but the pre-trends and large standard errors prevent any definitive conclusions. Other mental health professionals (including therapists) appear to have a more subdued response to professional conferences on family therapies.

The results for heterogeneity among prescribers are equally interesting. These estimates have greater power issues than others in this paper due to a smaller group of treated physicians. However, there is still a clear heterogeneous response among prescribers: mental health professionals who are *not* psychiatrists tend to respond positively to these conferences, with a small but significant (and lasting) increase in olanzapine prescriptions following the conference. Other prescribers show a less noticeable change in behavior; general practitioners do not respond at all, and psychiatrists respond for only a few periods following the conference. It may be that psychiatrists are better trained in understanding the risks of a pharmacological approach, or they may have more of an availability to engage in a psychotherapeutic intervention than another mental health prescriber (e.g., a psychiatric nurse practitioner). To the extent that either of these are true, professional conferences may reach those who have less time for therapeutic responses, a higher tolerance for pharmacological risk, or both.

4.4 Experimentation as a Possible Mechanism

Overall, the results suggest a limited and short-lived response to professional conferences. One potential explanation for this fact is that specialists return from conferences and experiment with new techniques, gauging their overall effectiveness and ease of use before integrating them into their regular treatment profile. But therapists who attempt to incorporate FBT, for example, may dislike the increased coordination cost or have a poor first experience with the treatment, which may lead them to revert to their original treatment methods.

To test this hypothesis, I explore the effect of these professional conferences on a specialist's likelihood to expand their treatment set. I measure this likelihood by computing each therapist's Herfindahl-Hirschman index (HHI) of their treatment profile, as measured by variation in their billed CPT codes. The HHI is calculated for each therapist i in period

[18] If one wanted to construct the dynamic treatment effect for psychologists, say, one would add the γ vector to the δ_{psych} vector. The standard errors would stay the same as those around $\delta_{\text{psych},t}$ for all points as this they are bootstrapped standard errors, which do not change under a linear shift.

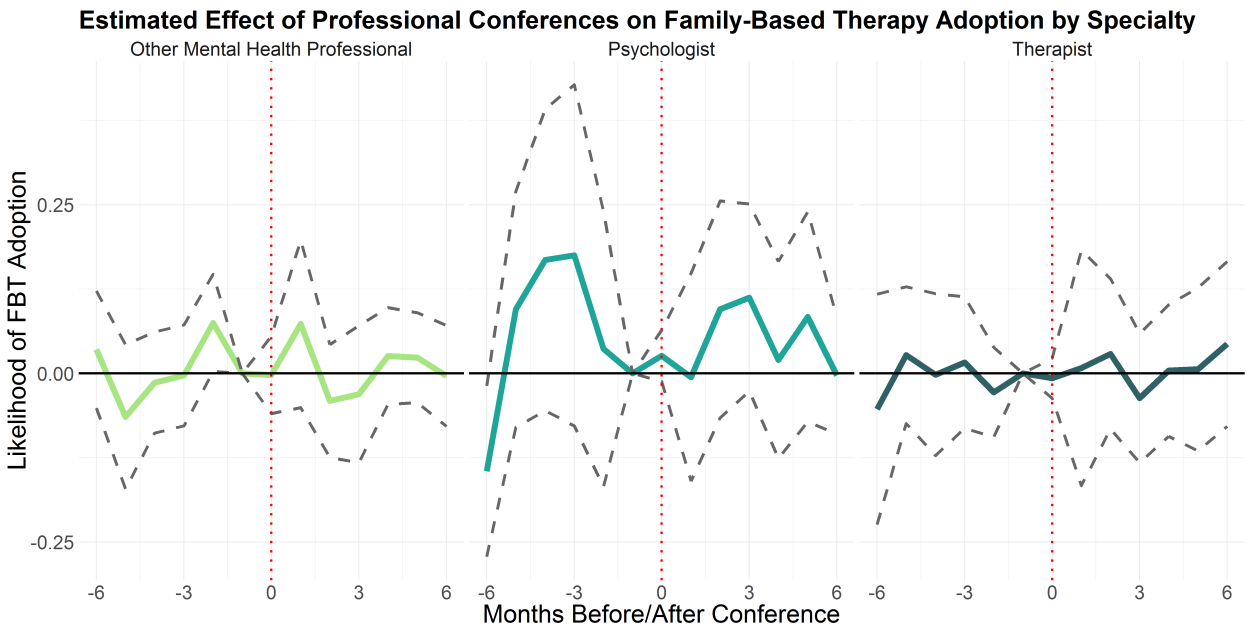

Figure 8
Professional Response to Family-Based Therapy Conferences by Specialty

t using the formula:

$$\text{HHI}_{it} = \sum_{j=1}^{n} s_{ijt}^2, \tag{7}$$

where s_{ijt} represents the fraction of provider i's treatments in time period t that are in the category j. I calculate the HHI using 9 categories, including individual, group, and family therapies amidst pharmacological interventions and other medical and administrative claims.[19]

To the extent that different CPT codes perfectly capture differences in utilized treatment,[20] this provides one measure of how specialized a therapist's treatments are. For example, if a therapist specializes exclusively in family-based therapies, there will be no variation in the treatment profile, leading to an HHI of 1; on the other hand, experimentation with different treatment methods will cause the HHI to *decrease*.

I re-estimate the dynamic treatment effects for these conferences using the calculated

[19]For reference, the 9 categories used are individual therapy, group therapy, family therapy, pharmacological interventions, evaluation and management, intake procedures, general consultations, hospitalization treatments, and other codes used rarely.

[20]There has been a recent discussion on how well physicians agree on the relevant CPT codes for given treatments (Bentley, Wilson, Derwin, Scodellaro, & Jackson, 2002; King, Sharp, & Lipsky, 2001) making this calculation an imperfect proxy of true specialization. However, I believe that (given the categories I've selected) disagreement about billing will be minimized in this case, thus making this a useful measure.

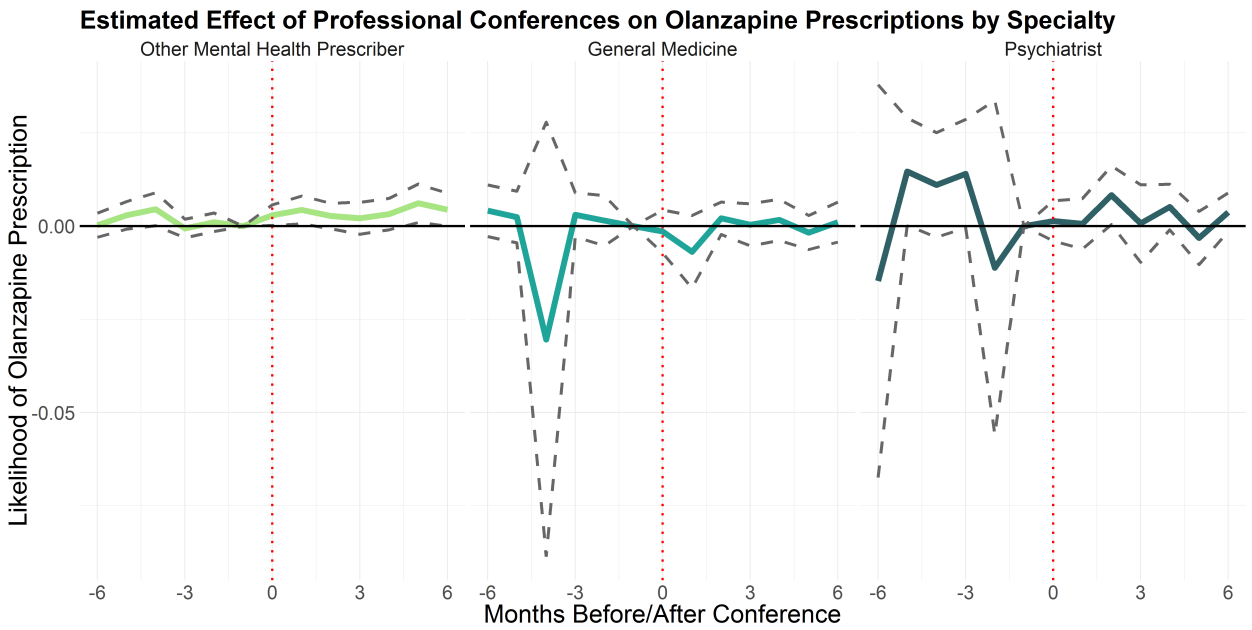

Figure 9
Professional Response to Olanzapine Conferences by Specialty

HHIs as the new dependent variable. That is, I measure to what extent professional conferences induced specialists to expand (or contract) their treatment methods, inducing experimentation or specialization respectively.

Figure 10 shows the effect that FBT conferences have on this measure of specialization. Again, there are no strong results, although there is a slight increase in specialization at the time of the conference (potentially lasting for a few periods). This may result from one of two potential causes: first, the conference itself may impose limitations on a therapist's time for treatment, requiring them to treat only the patients that they are specialized to treat. Second, it may also be the case that continuing education induces therapists to favor their own special skill sets more, as they feel more trained to implement their techniques. Either way, the effects do not suggest an increase in experimentation with new techniques after a conference, which would be indicated by a negative trend. A similar result for pharmaceutical conferences is relegated to Appendix C.

5 Travel Cost Validation

Critical to the interpretation of my results is the extent to which the D-LATE estimation technique approximates the true LATE. That is, I would ideally understand the probabilities that my treatment/control measures satisfy the conditions in Assumption 3. While I cannot verify this in my sample given the unobservability of true treatment, I have obtained a

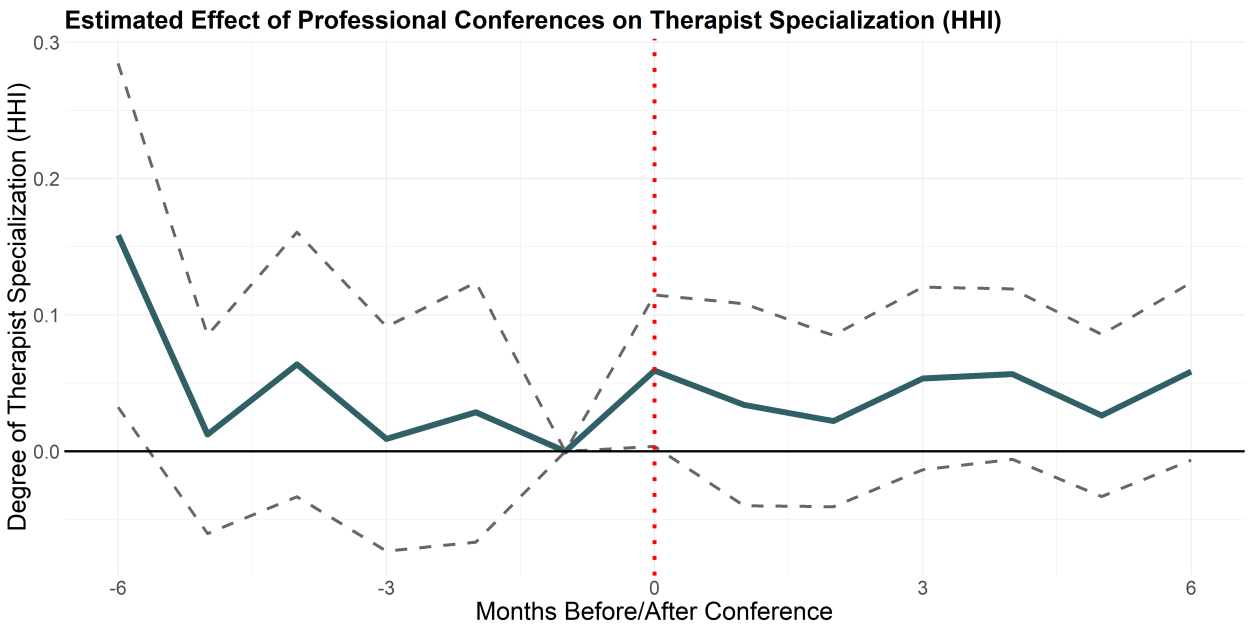

Figure 10
Effect of Professional FBT Conferences on Therapist Specailization (HHI)

validation sample of conferences from the Academy of Eating Disorders; this will allow me to obtain a sense of how well these assumptions might be satisfied in my main data.

For now, I have access to registration for the 2019 ICED Conference held in New York City, NY.[21] That is, I have records of each of the 612 unique US-based organizations which sent professionals to the conference, as well as their geocoded locations. Figure 11 shows the approximate home location of each attendee, with the conference location shown in red. Notice that, as expected, a large fraction of attendees live in close geographic proximity to the conference location. Interestingly, however, those who travel a greater distance to the conference appear to be based in metropolitan areas, which have greater proximity to an airport and subsequently lower travel costs.

I link this data to my Marketscan data in order to have some idea of a "true" treatment measure relative to a control group. To do so, I identify every therapist in my sample whose main location is within a 10 mile radius of an ICED attendee location as being "truly" treated. Next, to get an idea of how correlated my predicted treatment measure is with actual attendance, I estimate predicted travel costs for each of the therapists in my sample and the 2019 ICED conference. I then assign treatment groups using the same thresholds used throughout the paper, so that $\eta \in \{1, 2, 5, 10, 15\}$. These are hypothetical, as my sample does not extend to 2019, but will give an idea of how well the prediction algorithm

[21] I am in the process of widening this validation sample by working with other conference program directors.

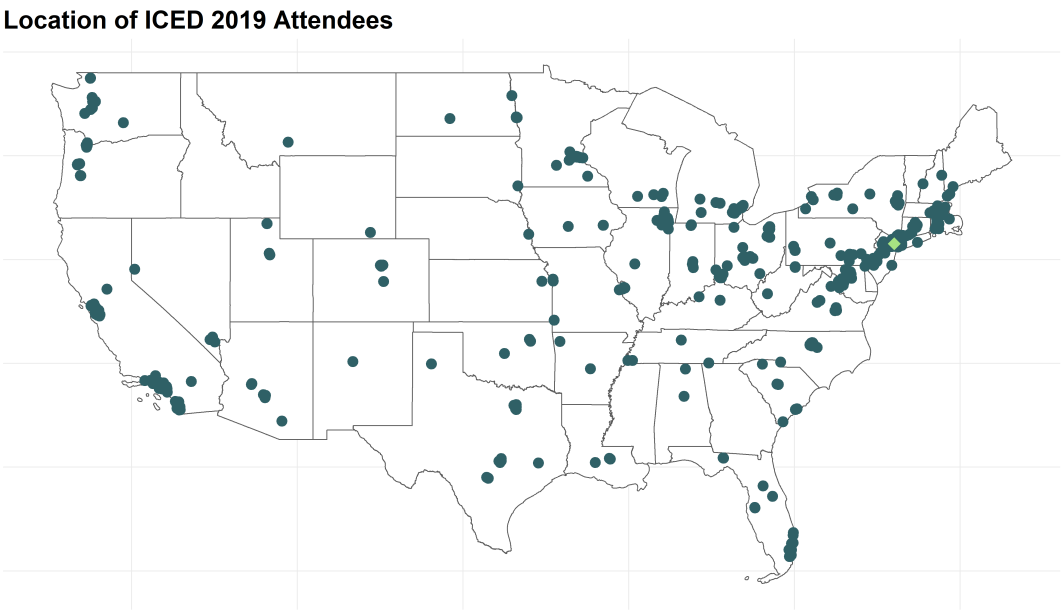

Figure 11
Actual Attendees of the 2019 ICED Conference in New York City, NY

does relative to the truth.

First, I verify the conditions listed in Assumption 3, which are sufficient for the D-LATE estimator to point identify the true LATE. Recall that for the two treatment measures T^a and T^b, one needs to assume that T^a never misclassifies the control group, and that T^b never misclassifies the treatment group. The probabilities of these misclassifications (labelled as p_0^a and p_1^b) are identified for each threshold in Table 2 for the unnormed treatment algorithm.

T^a Threshold	p_0^a	T^b Threshold	p_1^b
0.01	0.0000	0.99	0.0074
0.02	0.0000	0.98	0.0086
0.05	0.0172	0.95	0.0103
0.10	0.0645	0.90	0.0360
0.15	0.1155	0.85	0.0406

Table 2
Estimated misclassification probabilities using ICED 2019 data (unnormalized)

In general, these misclassification probabilities decrease with η, suggesting that results using smaller thresholds are closer to the true LATE. In fact, for this particular validation sample, the probability p_0^a decreases to exactly 0 after $\eta < 5$, and the corresponding probability p_1^b decreases to under 1%. This suggests that the conditions in the third assumption are

well approximated in my current sample, especially for the smallest two values of η (which include my preferred specification of $\eta = 2$).

Additionally, identification of the D-LATE estimator depends on a nonzero correlation between $T^i, i \in \{a, b\}$ and the true treatment status D. That is, the mismeasured treatments must give some information about true treatment status; without doing so (or with too small of a correlation), a problem similar to that of weak instruments arises. Although this cannot be verified in my sample of interest, I can again utilize the verification sample to assess this correlation. For brevity, the specific correlations are relegated to the Appendix; however, these correlations are strong for most measures and average around 0.15, suggesting little concern of weak identification.

5.1 Normed or Unnormed?

My travel cost algorithm assigned artificial treatment status based on two types of travel costs: a simple monetary measure (unnormed) and one measured in units of hourly salary (normed by salary). Thus, a simple question is to ask which of these measures best satisfies the assumptions needed for the D-LATE estimator to be meaningfully interpreted. In contrast to Table 2, Table 3 shows the misclassification measures the predicted travel cost for each therapist as a multiple of their expected hourly salary.

T^a Threshold	p_0^a	T^b Threshold	p_1^b
0.01	0.0004	0.99	0.0149
0.02	0.0062	0.98	0.0257
0.05	0.0315	0.95	0.0389
0.10	0.0645	0.90	0.0691
0.15	0.1019	0.85	0.0915

Table 3
Estimated misclassification probabilities using ICED 2019 data (normalized)

Overall, the algorithm performs significantly more poorly when normalizing by salary than when using a simple monetary measure. This is additionally advantageous because—as discussed in Appendix B—the normed travel cost measure appears to over-assign treatment status to those in the sample who make higher salaries (e.g, psychiatrists and family practice doctors) over those who stand to benefit the most from the professional conferences (e.g., therapists and mental health facility workers). Given both of these results, results using the non-normalized treatment algorithm should be taken as closer to the true LATE of interest.

6 Discussion & Conclusion

The methodology outlined above and the results arising from its application each have novel implications. In general, my project identifies ways that researchers can augment limited data with powerful statistical learning techniques to answer a broader range of questions than currently accessible, as illustrated by my analysis of professional conferences and mental health treatment behaviors.

6.1 Potential Uses of Methodology

A strong causal inference project typically requires rich data to be compelling. However, the set of questions researchers ask far eclipses the amount of adequate data available to them. The MR-LATE estimator of Calvi et al. (2019), as well as the D-LATE estimator proposed, discussed, and utilized here, offer ways researchers can incorporate imperfect data into an analysis without crippling it.

The estimator used in my project allows for the point identification of a dynamic local average treatment effect (D-LATE) under relatively mild assumptions. Current statistical learning techniques are more than capable of generating the mismeasured treatment assignments necessary for the estimator, and can use validation samples or other techniques to ensure that the assumptions are met at least approximately. Even in cases where the misclassification does not satisfy the conditions of Assumption 3, the D-LATE estimator can be looked at as a method of *bias reduction*, moving the estimated treatment effects closer to the truth by taking into account the possibility of misclassification.

This paper utilized a rather simple predictive algorithm to estimate treatment status of mental health professionals. Future research could integrate more advanced machine learning models in its place, thereby extending the set of questions the D-LATE estimator can answer. Additionally, future econometric research may explore the instrumental variables approach for event studies introduced here, identify more properties of the D-LATE estimator, and discuss how the work of Calvi et al. (2019) extends to other commonly used causal identification strategies.

6.2 Diffusion of Mental Health Treatments

This paper utilizes the methodology of the D-LATE estimator to allow imperfect data to shed light on an important problem in the healthcare industry: the diffusion of ideas. By examining how professionals respond to continuing medical education covering various types of innovations, I am able to assess to what extent research-to-practice gaps are developing

in mental health treatments.

The results are suggestive that therapists respond more to tangible innovations than to intangible ones. While there is no discernible response to conferences covering family-based therapies, there are situations in which providers are seen increasing their prescriptions of olanzapine following professional conferences on the subject. The identified heterogeneity discussed in Section 4 corroborates this finding; therapists from strong medical training and academic backgrounds (e.g., psychiatry and psychology) respond more positively to family-based therapy, which has a stronger evidence backing, while eschewing the somewhat weaker development of atypical antipsychotics. Interestingly, this increase in prescriptions occurs more among the younger patient population, despite the fact that adolescents and youth stand to gain the most from a family-based treatment approach rather than a pharmacological one.

Of course, future research is critical to confirming these findings. A crucial step will be extending this research beyond the Marketscan data, moving instead towards a holistic assessment of provider behavior amidst patients of various degrees of insurance coverage. Additionally, it will be important to gauge therapist response among other demographics, including experience, academic training, and clinic type. Replicating this project on a richer data set (such as all-payer claims data) can both confirm the validity of this estimator and its findings as well as identify with greater precision the subset of therapists who respond to professional conferences.

There are important questions outside of this domain that must be answered surrounding the impact of continuing medical education and research-to-practice gaps. This project overlooked the role that referring physicians and other members of the treatment team (e.g., dietitians) play in the decision to incorporate new treatments, either pharmacological or therapeutic. However, it may well be the case that these sidelined parties can induce innovation just as well—or better—than a CME program. Additionally, it may be useful to examine how provider payment mechanisms, network effects, and insurance coverage all dictate the decision to update or experiment with new treatments.

Finally, additional research may move beyond the communication problem of continuing medical education and into other frictions that exacerbate research-to-practice gaps. One might examine how the evolution of academic medical research may have siloed researchers into their own niche, and how this affects researcher involvement with practitioners at all. Additionally, projects might assess how researchers respond to other forms of media surrounding new treatments, including research articles and magazines. Finally, it will be useful to understand how implementing these new techniques affects the ultimate outcome for patients, especially those being treated by an intangible innovation.

Only by obtaining a more holistic picture of the different frictions and mechanisms can we hope to catch a glimpse at a solution to effectively incentivizing the diffusion of better mental health practices. Similarly, recognizing the manifold characteristics of individual innovations will allow a richer study of the economics of innovation. By doing so, future work can provide real solutions to gaps between academic research and real-world practice, as well as foster more efficient channels of communication in a broad spectrum of policy-oriented fields.

References

Abadie, A. (2002). Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models. *Journal of the American Statistical Association*, *97*(457), 284–292.

Abraham, S. & Sun, L. (2018). Estimating Dynamic Treatment Effects in Event Studies With Heterogeneous Treatment Effects. *SSRN Electronic Journal*.

American Psychiatric Association. (2006). Treatment of patients with eating disorders,third edition. American Psychiatric Association. *The American Journal of Psychiatry*, *163*(7 Suppl), 4–54.

Angrist, J. D. [Joshua D.] & Imbens, G. W. [Guido W.]. (1995). Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity. *Journal of the American Statistical Association*, *90*(430), 431–442.

Arcelus, J., Mitchell, A. J., Wales, J., & Nielsen, S. (2011). Mortality Rates in Patients With Anorexia Nervosa and Other Eating Disorders: A Meta-analysis of 36 Studies. *Archives of General Psychiatry*, *68*(7), 724–731.

Arrow, K. J. (1962). The economic implications of learning by doing. *The Review of Economic Studies*, *29*(3), 155–173.

Ash, E., Chen, D. L., & Naidu, S. (2019). Ideas Have Consequences: The Impact of Law and economics on American Justice. *NBER Working Paper*.

Bansal, P., Bertels, S., Ewart, T., MacConnachie, P., & O'Brien, J. (2012). Bridging the Research—Practice Gap. *Academy of Management Perspectives*, *26*(1), 73–92.

Bentley, P. N., Wilson, A. G., Derwin, M. E., Scodellaro, R., & Jackson, R. E. (2002). Reliability of assigning correct current procedural terminology—4 E/M codes. *Annals of Emergency Medicine*, *40*(3), 269–274.

Borusyak, K. & Jaravel, X. (2017). Revisiting event study designs. *Available at SSRN 2826228*.

Bulik, C. M., Berkman, N. D., Brownley, K. A., Sedway, J. A., & Lohr, K. N. (2007). Anorexia nervosa treatment: A systematic review of randomized controlled trials. *International Journal of Eating Disorders*, *40*(4), 310–320.

Burke, L. A. & Rau, B. (2010). The Research—Teaching Gap in Management. *Academy of Management Learning & Education*, *9*(1), 132–143.

Calvi, R., Lewbel, A., & Tommasi, D. (2019). Women's Empowerment and Family Health: Estimating LATE with Mismeasured Treatment. *Working Paper*.

Casper, E. S. (2007). The theory of planned behavior applied to continuing education for mental health professionals. *Psychiatric Services*, *58*(10), 1324–1329.

Choudhry, N. K., Fletcher, R. H., & Soumerai, S. B. (2005). Systematic Review: The Relationship between Clinical Experience and Quality of Health Care. *Annals of Internal Medicine*, *142*(4), 260.

Christensen, C. M., Baumann, H., Ruggles, R., & Sadtler, T. M. (2006). Disruptive innovation for social change. *Harvard business review*, *84*(12), 94–101, 163.

Christensen, C. M., Grossman, J. H., & Hwang, J. (2009). The Innovator's Prescription: A Disruptive Solution for Health Care.

Christensen, C. M., Raynor, M. E., & McDonald, R. (2015). What is disruptive innovation? *Harvard Business Review*, *93*(12), 44–53.

Church, E. A., Heath, O. J., Curran, V. R., Bethune, C., Callanan, T. S., & Cornish, P. A. (2010). Rural professionals' perceptions of interprofessional continuing education in mental health. *Health and Social Care in the Community*, *18*(4), 433–443.

Coburn, C. E. & Penuel, W. R. (2016). Research–Practice Partnerships in Education: Outcomes, Dynamics, and Open Questions. *Educational Researcher*, *45*(1), 48–54.

Couturier, J., Kimber, M., & Szatmari, P. (2013). Efficacy of family-based treatment for adolescents with eating disorders: A systematic review and meta-analysis. *International Journal of Eating Disorders*, *46*(1), 3–11.

Curran, V. R., Fleet, L., & Kirby, F. (2006). Factors influencing rural health care professionals' access to continuing professional education. *Australian Journal of Rural Health*, *14*(2), 51–55.

Dare, C. & Eisler, I. (2000). A multi-family group day treatment programme for adolescent eating disorder. *European Eating Disorders Review: The Professional Journal of the Eating Disorders Association*, *8*(1), 4–18.

de Chaisemartin, C. & D'Haultfoeuille, X. (2019). Two-way fixed effects estimators with heterogeneous treatment effects. *NBER Working Paper 25904*.

Depalo, D., Bhattacharya, J., Atella, V., & Belotti, F. (2019). When technological advance meets physician learning in drug prescribing. *NBER Working Paper*, (Working Paper 26202).

Dysart, A. M. & Tomlin, G. S. (2002). Factors Related to Evidence-Based Practice Among U.S. Occupational Therapy Clinicians. *American Journal of Occupational Therapy*, *56*(3), 275–284.

Flament, M. F., Bissada, H., & Spettigue, W. (2012). Evidence-based pharmacotherapy of eating disorders. *The International Journal of Neuropsychopharmacology*, *15*(02), 189–207.

for Health and Care Excellence, N. I. (2017). Eating disorders: Recognition and Treatment. *NICE*.

Gilardi, F. (2012). Transnational diffusion: Norms, ideas, and policies. In W. Carlsnaes, T. Risse, & B. Simmons (Eds.), *Handbook of International Relations* (pp. 453–477). Thousand Oaks. Sage Publications.

Glanz, K., Rimer, B. K., & Viswanath, K. (2008). *Health Behavior and Health Education: Theory, Research, and Practice* (4th ed.). John Wiley & Sons.

Glasgow, R. E. & Emmons, K. M. (2007). How Can We Increase Translation of Research into Practice? Types of Evidence Needed. *Annual Review of Public Health*, *28*(1), 413–433.

Graham, A. K., Lattie, E. G., & Mohr, D. C. (2019). Experimental Therapeutics for Digital Mental Health.

Grimshaw, J. M., Shirran, L., Thomas, R., Mowatt, G., Fraser, C., Bero, L., ... O'Brien, M. A. (2001). Changing provider behavior: An overview of systematic reviews of interventions. *Medical Care*, *39*(8), II–2.

Hu, Y. (2008). Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution. *Journal of Econometrics*, *144*(1), 27–61.

Hudson, S., Hull, P., & Liebersohn, J. (2017). Interpreting instrumented difference-in-differences. *Working Paper*.

Hugenholtz, N. I. R., de Croon, E. M., Smits, P. B., van Dijk, F. J. H., & Nieuwenhuijsen, K. (2008). Effectiveness of e-learning in continuing medical education for occupational physicians. *Occupational Medicine*, *58*(5), 370–372.

Hull, P. (2018). Estimating Treatment Effects in Mover Designs. *Working Paper*, 31.

Imbens, G. W. [G. W.] & Angrist, J. D. [J. D.]. (1994). Identification and estimation of local average treatment effects. *Econometrica*, *62*, 467–475.

Jensen, P. S., Bhatara, V. S., Vitiello, B., Hoagwood, K., Feil, M., & Burke, L. B. (1999). Psychoactive Medication Prescribing Practices for U.S. Children: Gaps Between Research and Clinical Practice. *Journal of the American Academy of Child & Adolescent Psychiatry*, *38*(5), 557–565.

Johannesen, N. & Stolper, T. (2017). The deterrence effect of whistleblowing: An event study of leaked customer information from banks in tax havens. *CESifo Working Paper No. 6784*.

Kapetanios, G. (2008). A bootstrap procedure for panel data sets with many cross-sectional units. *The Econometrics Journal*, *11*(2), 377–395.

Kazdin, A. E. (2011). Evidence-based treatment research: Advances, limitations, and next steps. *American Psychologist*, *66*(8), 685–698.

Kazdin, A. E. (2017). Addressing the treatment gap: A key challenge for extending evidence-based psychosocial interventions. *Behaviour Research and Therapy*. Expanding the impact of cognitive behaviour therapy: A special edition in honor of G. Terence Wilson Expanding the impact of cognitive behaviour therapy: A special edition in honor of G. Terence Wilson, *88*, 7–18.

Kazdin, A. E. (2018). Implementation and evaluation of treatments for children and adolescents with conduct problems: Findings, challenges, and future directions. *Psychotherapy Research*, *28*(1), 3–17.

Kazdin, A. E. & Blase, S. L. (2011). Rebooting Psychotherapy Research and Practice to Reduce the Burden of Mental Illness. *Perspectives on Psychological Science*, *6*(1), 21–37.

Kazdin, A. E., Fitzsimmons-Craft, E. E., & Wilfley, D. E. (2017). Addressing critical gaps in the treatment of eating disorders. *International Journal of Eating Disorders*, *50*(3), 170–189.

King, M. S., Sharp, L., & Lipsky, M. S. (2001). Accuracy of CPT Evaluation and Management Coding by Family Physicians. *Journal of American Board Family Practice*, *14*(3), 184–192.

Kosmerly, S., Waller, G., & Robinson, A. L. (2015). Clinician adherence to guidelines in the delivery of family-based therapy for eating disorders: Clinician Adherence to Guidelines in the Delivery of FBT. *International Journal of Eating Disorders*, *48*(2), 223–229.

Kreider, B. (2010). Regression coefficient identification decay in the presence of infrequent classification errors. *The Review of Economics and Statistics*, *92*(4), 1017–1023.

Le Grange, D., Lock, J., Loeb, K., & Nicholls, D. (2010). Academy for eating disorders position paper: The role of the family in eating disorders. *International Journal of Eating Disorders*, *43*(1), 1.

Lewbel, A. (2007a). A local generalized method of moments estimator. *Economics Letters*, *94*(1), 124–128.

Lewbel, A. (2007b). Estimation of average treatment effects with misclassification. *Econometrica*, *75*(2), 537–551.

Lock, J., Couturier, J., Bryson, S., & Agras, S. (2006). Predictors of dropout and remission in family therapy for adolescent anorexia nervosa in a randomized clinical trial. *International Journal of Eating Disorders*, *39*(8), 639–647.

Lock, J. & Grange, D. L. (2015). *Treatment Manual for Anorexia Nervosa, Second Edition: A Family-Based Approach* (2nd ed.). Google-Books-ID: PFuqCQAAQBAJ. New York, NY, US: Guilford Publications.

Loeb, K., Lock, J., Greif, R., & Le Grange, D. (2012). Transdiagnostic Theory and Application of Family-Based Treatment for Youth With Eating Disorders. *Cognitive and Behavioral Practice*, *19*(1), 17–30.

Maglione, M. & Hu, e. a. (2011). Off-Label Use of Atypical Antipsychotics: An Update. *Effective Health Care Review*, *43*, 437.

Mahajan, A. (2006). Identification and estimation of regression models with misclassification. *Econometrica*, *74*(3), 631–665.

Millimet, D. L. (2010). The elephant in the corner: A cautionary tale about measurement error in treatment effects models. *IZA Discussion Paper No. 5140*.

Mitchison, D., Hay, P., Slewa-Younan, S., & Mond, J. (2014). The changing demographic profile of eating disorder behaviors in the community. *BMC Public Health*, *14*(1), 943.

Mullainathan, S. & Obermeyer, Z. (2019). Who is tested for heart attach and who should be: Predicting patient risk and physician error. *NBER Working Paper 26168*.

Olfson, M., Druss, B. G., & Marcus, S. C. (2015). Trends in mental health care among children and adolescents. *New England Journal of Medicine*, *372.21*, 2029–2038.

Rane, H. & Salem, S. (2012). Social media, social movements and the diffusion of ideas in the Arab uprisings. *The Journal of International Communication*, *18*(1), 97–111.

Rountree, M. A. & Pomeroy, E. C. (2010). Bridging the Gaps among Social Justice, Research, and Practice. *Social Work*, *55*(4), 293–295.

Rynes, S. L., Colbert, A. E., & Brown, K. G. (2002). HR Professionals' beliefs about effective human resource practices: Correspondence between research and practice. *Human Resource Management*, *41*(2), 149–174.

Strohman, E. (2014). *An Examination of Research to Practice Gaps in Education: Whole Language, Learning Modalities, and Brain Gym* (Doctoral dissertation).

Ura, T. (2018). Heterogeneous treatment effects with mismeasured endogenous treatment. *Quantitative Economics*, *9*(3), 1335–1370.

Wandersman, A., Duffy, J., Flaspohler, P., Noonan, R., Lubell, K., Stillman, L., ... Saul, J. (2008). Bridging the Gap Between Prevention Research and Practice: The Interactive Systems Framework for Dissemination and Implementation. *American Journal of Community Psychology*, *41*(3-4).

Yanagi, T. (2018). Inference on local average treatment effects for misclassified treatment. *Working Paper*.

Appendix A: Proof of Theorem 1

Proof of Theorem 1. Suppose that we observe a panel $\{Y_{it}\}$ of outcomes for individuals $i = 1, 2, ..., N$ and periods $t = 1, 2, ..., T$. Additionally, suppose that there is a true binary treatment status $D_{it} \in \{0, 1\}$ for all i and t—for consistency with the event study framework, $D = 1$ is an absorbing state, so that $D_{is} = 1$ implies that $D_{it} = 1$ for all $t \geq s$. Additionally, there is a binary instrument $Z_{it} \in \{0, 1\}$ and two mis-measured treatments T^a, T^b satisfying Assumptions 1 and 2 above.

The event study is estimated as follows (using the notation from Abraham and Sun (2018)): for each individual, define the time of first treatment as $E_i = \min\{t : D_{it} = 1\}$ and the related treatment-time dummy variables $D_{it}^{\ell} = \mathbb{1}\{t - E_i = \ell\}$. The regression equation is as in 2 with D_{it}^{ℓ} in place of the indicator variables.

As the true treatment is unobserved, we are interested in the local cohort-specific ATE of a transformed regression of $T^j Y$ on T^j for $T^j \in \{T^a, T^b\}$. That is, we are interested in the vector $\vec{\gamma}$ resulting from estimation of:

$$T_{it}^j Y_{it} = \alpha_i + \tau_t + \sum_{\ell=-K}^{-} 2\gamma_\ell T_{it,j}^{\ell} + \sum_{\ell=0}^{L} \gamma_\ell T_{it,j}^{\ell} + \epsilon_{it}$$

In moving from a typical panel data analysis to an event study approach, we move from a single treatment T_{it}^j to a vector of dummy variables $\{T_{it}^{\ell,j}\}_\ell$. In the setting where the fully dynamic equation is not estimated, there are $L + K$ dummy variables to be concerned with (as we drop the period $\ell = -1$ to avoid any collinearity problems). We therefore similarly expand the set of instruments from Z_{it} to $\{Z_{it}^{\ell}\}$, again of size $L + K$. This vector of instrument dummies is created in the same way the vector of treatment dummies, and described in Section 3.2. By expanding to multiple instruments, however, each coefficient in a two-stage regression will be given by a weighted average of LATEs, as discussed in Angrist and Imbens (1995). That is, for each resulting coefficient γ_ℓ on any dummy $T_{it,j}^{\ell}$:

$$\gamma_\ell = \sum_{k=1}^{L+K} \beta_k \frac{\text{Cov}(T_{it}^j Y_{it}, Z_j}{\text{Cov}(T_{it,j}^{\ell}, Z_j},$$

where the weights β_k are in the interval $[0, 1]$ for all k and satisfy $\sum_k \beta_k = 1$. As in Calvi et al. (2019), define $q = \frac{p_1}{p_1 - p_0}$. Using their Theorem 1 and the result above:

$$
\begin{aligned}
\gamma_\ell &= \sum_{j=1}^{L+K} \beta_j \frac{\text{Cov}(Y_{it}T_{it}, Z_j)}{\text{Cov}(T_{it}^{\ell}, Z_j)} \\
&= \sum_{j=1}^{L+K} \beta_j \lambda_j \\
&= \sum_{j=1}^{L+K} \beta_j \mathbb{E}[q_j Y_1 + (1-q_j)Y_0 | C] \\
&= \sum_{j=1}^{L+K} \beta_j \mathbb{E}[q Y_1 + (1-q)Y_0 | C] \text{ (as each } q_j = q) \\
&= \mathbb{E}\left[\left(\sum_{j=1}^{L+K} \beta_j \right) q Y_1 + \left(\sum_{j=1}^{L+K} \beta_j \right) (1-q) Y_0 | C \right] \\
&= \mathbb{E}[q Y_1 + (1-q)Y_0 | C] \text{ (as weights sum to 1).}
\end{aligned}
$$

Hence, for any time period ℓ, Calvi et al.'s Theorem 1 applies. One can therefore use two mismeasured treatments T^a and T^b with the same properties as in their paper (so that $p_0^a = p_1^b = 0$) and construct the local cohort average treatment effect:

$$
\rho_\ell = \mathbb{E}\left[Y_{i,t+\ell}^e - Y_{i,t+\ell}^\infty | E_i = e, C \right] = \hat{\gamma}_\ell^a - \hat{\gamma}_\ell^b
$$

■

Appendix B: Details of the Travel Cost Algorithm

The travel cost algorithm is used to infer the earliest date each medical professional was exposed to a continuing education event targetting either FBT use or **drug** prescribing in eating disorder treatments.

The first step in the algorithm is to assign a location to each specialist. While MarketScan does not have specifically geotagged locations for their physicians, they do have information on the Metropolitan Statistical Area in which the main enrollee on each insurance plan resides. Hence, each claim in a physician's treatment profile is tagged to one of these MSAs. By taking the bulk of patients seen in a given month and taking a geocentric average of their home MSAs (taking as each patient's location the global midpoint of their MSA), I can assign a specific location to each specialist-month observation. To avoid large errors, I discard specialist-month observations that treat patients from larger than a 100-mile radius.

Once a specific location has been assigned to a specialist-month, I can compute the travel costs between therapists and a given conference. The algorithm allows a specialist to travel to the conference either by car directly, or by any network of flights. To estimate driving time and costs, I allow for different average driving speeds in urban areas and on freeways/interstates, and estimate the price of gas using data from the United States Bureau of Labor Statistics (BLS). To estimate flying time and costs, I incorporate data on airport locations, flight availability, and airfare from the United States Bureau of Transportation. Then, I construct a network between a specialist's origin point (their home location) and their destination (the conference location) that allows them to (i) drive to any of the 5 airports closest to their home, (ii) take any network of flights from that airport to any of the 5 airports closest to their destination, and (iii) drive from that airport to the conference location. Once this network is completed, I assume that travelers will choose the cheapest option (in terms of both airfare and opportunity cost of travel).

Opportunity cost of time is calculated using BLS wage data. I estimate this opportunity cost in two ways: assigning each specialist in my data set the same hourly wage (the average median wage for all specialist groups in the sample), or assigning each specialist group their own median wage. For example, psychiatrists would be assigned a median wage of $105.95 per hour while mental health clinicians would be assign a median wage of $21.46 per hour. Under the second method, travel costs are reported as a percentage of each specialist's average salary, to keep units consistent.

While there is clearly a large amount of variation in specialist wages, this variation appears to be negatively correlated with true attendance. That is, those with the lowest median wages (e.g., treatment center workers, therapists) have a greater incentive to attend conferences on eating disorder treatments than general practice doctors or psychiatrists, who treat a larger range of diagnoses. However, when adjusting for different salaries, I am implicitly making the costs of travel (gas, airfare, etc.) less inhibitive for those with higher salaries, so the algorithm may be more likely to predict treatment for those who are *less* incentivized to truly attend.

As discussed in the paper, once this algorithm is complete, each specialist-conference pair is assigned a travel cost $c \in \mathbb{R}$. For the "unnormalized" option where each specialist is assigned a flat salary, this is a monetary measure in 2016 U.S. dollars; on the other hand, when different salaries are assigned to different specialists, this is measured as each specialist's

travel costs in salary hours. From this continuous measure, I form the dichotomous prediction of treatment and control groups using an artificial cutoff $\eta \in [0, 1]$, where any specialist with travel cost at or below the value $F(\eta)$ is considered to have attended the conference.

Appendix C: Additional Results

As mentioned previously, the results shown in the paper are robust to several iterations of estimation. I re-estimated the results with a continuous dependent variable instead of a binary one, and with a travel cost that was normalized to be in terms of each therapist's estimated salary (instead of a pure monetary measure). These sensitivity results are shown in Figure 12 for all 12 regression coefficients of the dynamic treatment effect. Note that throughout, I include the figures only for family-based therapies; the results are similar for olanzapine prescriptions.

Additionally, the results are re-estimated with various thresholds used in assigning treatment/control status. Each treatment measure assumes that those therapists with travel costs in the lowest η-percentile of all travel costs for a given conference attended it, and were thus treated. Each specification consists of a treatment measure using η as the percentile, and the other using $1 - \eta$. For example, the specification of choice uses the "strict" treatment measure as those whose travel costs are in the bottom 2%, while the "liberal" one applies treatment to those in the bottom 98%. I also test specifications where $\eta \in \{1, 2, 5, 10, 15\}$. The resulting variation in estimated coefficients is illustrated in Figure 13.

The results are generally quite consistent—if anything, models estimated with more stringent treatment thresholds (smaller η) appear to detect larger estimates, but have larger standard errors as well. While future work may elaborate on the optimal decision of treatment threshold to balance the trade off inherent in its selection, this figure provides sufficient evidence that the choice of threshold contributes little to the overall result.

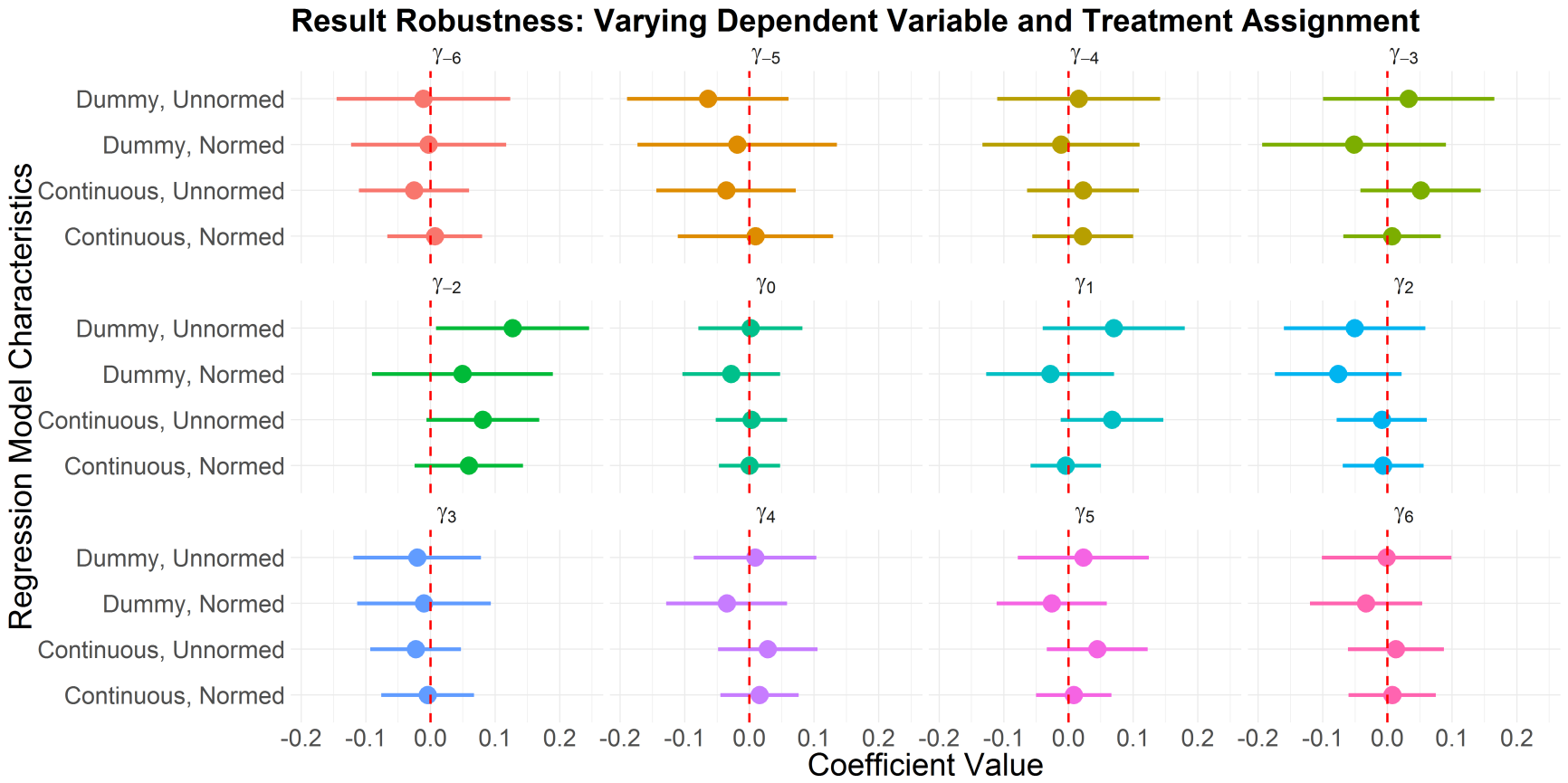

Figure 12
Response to Family-Based Therapy Conferences, Sensitivity by Estimation Strategy

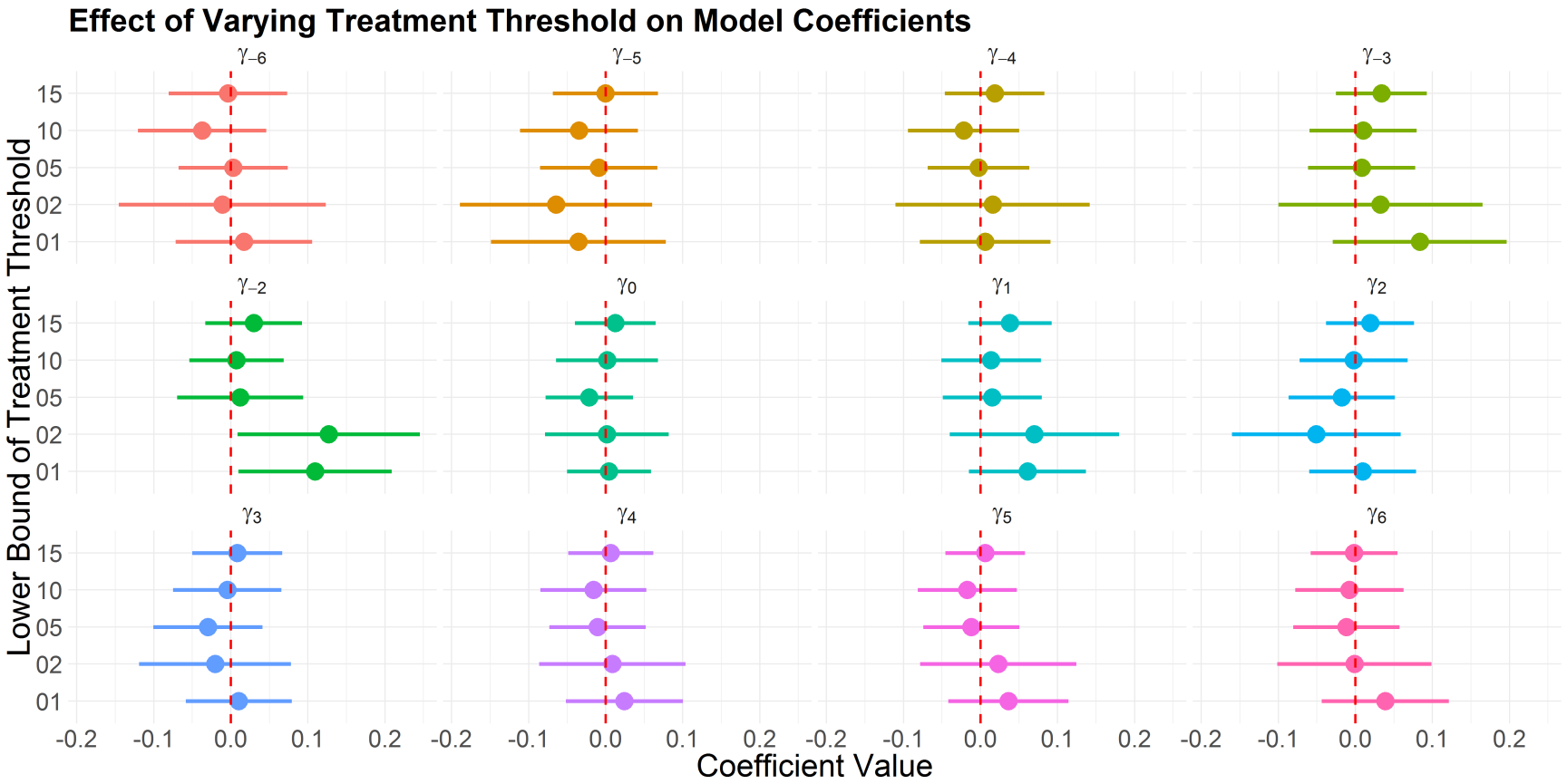

Figure 13
Response to Family-Based Therapy Conferences, Sensitivity by Treatment Thresholds