Topics in Machine Learning for Health Services Research

HARVARD UNIVERSITY
Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Committee on Higher Degrees in Health Policy

have examined a dissertation entitled

Topics in Machine Learning for Health Services Research

presented by Ian Nason

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature _____

Typed name: Prof. Mary Beth Landrum

Signature _____

Typed name: Prof. Richard Frank

Signature _____

Typed name: Prof. Thomas McGuire

Date: April 27, 2022

Topics in Machine Learning for Health Services Research

A dissertation presented

By

Ian Nason

To

The Committee for the PhD in Health Policy

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In the subject of

Health Policy (Methods for Policy Research)

Harvard University

Cambridge, Massachusetts

April 2022

Dissertation Advisor: Mary Beth Landrum Ian Nason

Topics in Machine Learning for Health Services Research

Abstract

This dissertation explores topics where machine learning can be used to improve or expand the scope of health services research.

Chapter 1: This study demonstrates the viability of using deep learning to identify people who are likely to benefit from osteoporotic or fragility fracture risk screening using chest radiographs. Previous work has shown that deep learning algorithms can identify osteoporotic individuals across a wide range of imaging modalities. However, the vast majority of studies have focused on patients that have been screened for osteoporosis and the study populations are almost entirely comprised of patients who are already recommended for screening. We develop and validate an algorithm on a large dataset of chest radiographs consisting of 59,737 individuals that is significantly more diverse across both age and sex than the data used in previous work. Using an ensemble of image classification models and conformal prediction, our algorithm was able to identify individuals without an osteoporosis or osteopenia diagnosis that were 1.93 (95% CI: 1.66, 2.26) times more likely to have experienced any fracture and more than twice as likely to experience common osteoporotic fractures, such as hip and pelvis fractures (Est: 2.19, 95% CI: 1.67, 2.88) or spine and rib fractures (Est: 2.49, 95% CI: 1.99, 3.11), after adjusting for age, sex, and BMI. Approximately 45% of individuals identified were not currently recommended for screening. We further outline how conformal prediction can be used to adjust the size of the flagged patient population to account for important implementation factors, such as a health

centers' capacity to screen additional patients or requirements for a higher standard of evidence when making recommendations that fall outside of current clinical guidelines. The size of the flagged patient population could also be set to match the projected cost-benefit tradeoff of screening additional people.

Chapter 2: Early entry by firms in markets for generic pharmaceuticals is paramount for price declines and increasing consumer welfare. The first-mover advantage (FMA), defined as the additional market share a firm earns by entering first compared to entering later, is considered one of the most important incentives for prompt entry into generic markets. Inherent in this definition of FMA is a notion of causality, where, in an ideal but infeasible experiment, we could observe *the same* firm entering at different times in *the same* market and compare market shares across different entry timing decisions. This paper will exploit unique characteristics of generic drug markets and advances in doubly robust methods for causal inference to estimate the FMA. Our findings suggest entering first results in a significant advantage when compared to if that same firm entered later. This advantage is mainly accrued through sales in periods where competition is limited and, to a lesser extent, higher market shares during the first two years after the start of competition between generics. The latter point contradicts a number of previous studies that found the FMA is sustained for up to six years after the start of competition between generics. We further characterize the impact that important regulatory features, such as exclusivity periods and the presence of authorized generics, have on the market shares of first entrants and provide evidence to suggest our estimates are unlikely to be explained by unobserved confounding.

Chapter 3: COVID-19 interrupted delivery of mental health care in the US. Symptoms associated with various mental disorders increased in prevalence at the same time. The expectation

is that treatment would increase with measured need. Departures from that expectation serve to index the degree of disruption in the delivery of mental health care to the US population. We conducted a retrospective observational analysis using prescription claims data covering 89 percent of all prescriptions in the US that compared observed new-starts of common psychotropic medications to forecasted new-starts. Forecasts were generated using the Prophet forecasting model. During the initial course of the COVID-19 pandemic new starts of antidepressants declined by 7.5 percent, anxiolytics by 5.6 percent, and antipsychotics by 2.6 percent compared with expected levels. Declines were more pronounced among children and adolescents, with declines in new starts ranging from 20 to 30% over the same period for the three drug classes. Our findings suggest that there was a large unmet need for mental health treatment in the US attributable to COVID-19 over this period.

TABLE OF CONTENTS

List of Tables and Figures

Chapter 1:

Chapter 2:

Chapter 3:

Acknowledgements

Thank you to my dissertation committee, Richard Frank, Mary Beth Landrum, and Thomas McGuire. I am immensely appreciative of your guidance and support over the course of my PhD. Your work and mentorship have had an immense impact on health policy and health economics, and it was truly an honor to work together.

To my mentors and friends in the economics department at Mount Allison, Frank Strain, Craig Brett, Niels Anthonisen, Rebecca Dafoe, and Rosie Cockshutt, my love of research started in the 'econ zone' and all of you helped cultivate a sense of curiosity about the world that led me here today. To Bacchus Barua, my time at the Fraser Institute was short but thank you for your mentorship and for giving me an introduction to the world of health economics. To my mentors and friends at the London School of Economics, Lou Delacoste, Alistair McGuire, Rishub Keelara, Mhairi Mckenzie, and Inna Thalmann, thank you for your help, support, and tolerance as I transitioned to being the 'weird machine learning guy'. To all my friends and mentors in the Health Policy PhD program, Micah Aaron, Rebecca Dafoe, A J Holmgren, John Giardina, Noemi Sportiche, Colleen Yout, Deborah Whitney, Anna Zink, and countless others, the people are what made this experience fulfilling and I am indebted to all of you for your friendship over the last five years. To all of the friends that have kept me grounded and sane throughout my academic career, some going back all the way to middle school, Robert Dryden, Ian Herzog, Thomas Kostelnik, Andrew Schoer Cam Teschuk, and Tyler Wills, it is difficult to express how important you have all been on this journey. Thank you.

Finally, to my family, Dad, Mom, Amelia, and Maria, I love you and would not be the person I am today without you all. Your support (and more recently pictures of Annie and Parker!) is what made this possible.

This dissertation is dedicated to my Mom and Dad.

Chapter 1:

Identification of Individuals at High Risk of Osteoporotic Fracture Using Deep Learning and Conformal Prediction.

Abstract

This study demonstrates the viability of using deep learning to identify people who are likely to benefit from osteoporotic or fragility fracture risk screening using chest radiographs. Previous work has shown that deep learning algorithms are capable of identifying osteoporotic individuals across a wide range of imaging modalities. However, the vast majority of these studies have focused on patients that have already been screened for osteoporosis and their study populations are almost entirely comprised of patients who are recommended for screening. We develop and validate an algorithm on a large dataset of chest radiographs consisting of 59,737 individuals that is significantly more diverse across both age and sex than the data used in previous work. Using an ensemble of image classification models and conformal prediction, our algorithm was able to identify individuals without an osteoporosis or osteopenia diagnosis that were 1.93 (95% CI: 1.66, 2.26) times more likely to have experienced any fracture and more than twice as likely to experience common osteoporotic fractures, such as hip and pelvis fractures (Est: 2.19, 95% CI: 1.67, 2.88) or spine and rib fractures (Est: 2.49, 95% CI: 1.99, 3.11), after adjusting for age, sex, and BMI. Approximately 45% of individuals identified were not currently recommended for screening. We further outline how conformal prediction can be used to approximately adjust the size of the flagged patient population to account for important implementation factors, such as a health centers' capacity to screen additional patients or requirements for a higher standard of evidence when making recommendations that fall outside of current clinical guidelines. The size of the flagged patient population could also be set to match the projected cost-benefit tradeoff of screening additional people.

Introduction

Low bone density, reflected clinically as osteopenia and osteoporosis, is associated with a high fracture risk. These fragility or low trauma fractures contribute to significant morbidity (Lundren et al., 2021) and mortality (Leboime et al., 2010). Low bone density is under diagnosed because of the fact that individuals are generally asymptomatic until a fracture occurs (Siris et al, 2014; Pouresmaeili et al., 2018). A major goal in the management of this disease process is to identify patients at risk for fragility fractures and institute preventative measures or pharmacological treatment. Fracture risk is assessed with a combination of bone mineral density studies (BMD) and assessment of clinical risk factors using a validated risk assessment tool such as FRAX (Marques et al., 2015; US Preventive Services Task Force, 2018).

Low bone density is a silent disease that is under diagnosed. Current recommendations support screening for low bone density in women who are older than 65 and postmenopausal women who are under 65 and are at an increased risk of osteoporosis. Estimates suggest less than 30% of eligible women are screened (Curtis et al., 2008; Leweicki et al., 2012). Preventative screening is generally not recommended in men of any age (Wilson et al, 2015; US Preventative Services Task Force, 2018). This has spurred a number of researchers to consider "opportunistic screening", where various types of medical images taken for unrelated purposes are used to identify individuals suffering from or at risk of osteoporosis (Pickhardt et al., 2013; Yamamoto et al., 2020; Yasaka et al., 2020; Zhang et al., 2020; Fang et al., 2021; Ho et al., 2021; Hsieh et al., 2021; Jang et al., 2021). However, the vast majority of previous studies have focused on relatively small patient populations where the majority or entirety of the sample consisted of older woman that are already recommended for screening (Ferizi, Honig, and Chang, 2019).

This study explores the viability of using chest radiographs to identify individuals for screening for fracture risk using deep learning and conformal set prediction in a large patient population that is significantly more diverse across age and sex when compared to related studies. The chest x-ray is the most common radiographic procedure in the world (Lundren et al., 2021) and provides a clear image of the spine, a key area for assessing bone quality, making it a good candidate for "opportunistic screening" (Black et al., 1992; Kelsey and Samelson, 2009; Leweicki et al., 2012). Naively using unrelated radiographs in electronic health records to guide screening highlights several issues when using deep learning algorithms in a clinical setting, such as an algorithm's potential to amplify existing biases when trained on historical data and overemphasis on predicting final disease labels, in lieu of creating algorithms with outputs and validation metrics that are more aligned with a clinical endpoint (Alder-Milstien et al., 2021; Esteva et al.,2021; Liu et al., 2019; Chen et al.,2020; Ancker et al.,2017; Jacobs et al., 2020; Seyyed-Kalantari et al., 2020) . This study also exhibits potential solutions to these problems and other implementation related issues when deep learning algorithms are used in the diagnostic pathway for osteoporosis and other related diagnoses. We use conformal prediction, a method that integrates uncertainty when making final predictions, as a means to approximately control the size of the patient population flagged for further screening. In theory, the flagged patient population could be restricted to a subset of individuals where screening for osteoporosis is cost-effective. To validate our algorithm, we focus on a set of statistical metrics that are atypical for medical deep learning studies but align more closely with the type of evidence used to justify treatment decisions in medicine.

Methods

Data Sources

To carry out our analysis we use data from the Nightingale Open Science 'Predicting Fractures and Pain' dataset (Lungren et al., 2021). The dataset is built on top of the Stanford CheXpert collection of chest x-ray images (Irvin et al., 2019). The sample includes 224,316 chest radiographs from 65,420 unique patients that sought care between 2002 and 2017. In addition to the radiologist interpretations available in the CheXpert dataset, each patients' images were linked to ICD9 codes related to fractures, basic demographic information such as age and sex, and patient characteristics such as height, weight, and body temperature derived from electronic health records in the Stanford Medicine system. We omit patients where fractures were visible in the x-ray image of their earliest included study, given that the indication for imaging for these patients is already more likely be related to osteoporotic fracture and an indication for fracture risk screening (Majumdar et al., 2005). For patients with multiple x-ray studies, studies where no fracture was visible that occurred prior to a study where a fracture was visible were included. In addition, only the earliest study in the sample period was included for each patient. The final sample included 198,981 images from 59,737 patients. The mean patient age was 60 and 45% of the study population are women. The majority of images (85.6%) in the dataset were frontal chest radiographs. The remaining images are taken from the lateral view.

Outcome Variable Definitions

Outcomes of interest were defined using ICD9 codes from electronic health records. The main outcome used to develop models was whether or not a patient was diagnosed with low bone density. For a given patient, all images were labeled as osteoporotic if the patient was diagnosed at any point over the sample period. The main outcome used for algorithm validation, whether or

not a patient suffered any type of fracture, was defined similarly. Results are reported for individual fracture locations that are split into groups based on the probable diagnosis related to the fracture: osteoporotic or fragility fractures and unrelated fractures. Whether or not a fracture was osteoporotic is defined based on Mai and colleagues (2019), who studied the fraction of fractures that are attributable to low bone density, Warriner and colleagues (2011), who convened a multidisciplinary panel of osteoporosis experts to rank fractures in terms of their likelihood of being osteoporosis-related, and Kelsey and Semelson (2009), who investigate variation in the risk factors associated with different fracture locations. In general, fractures around the hip, pelvis, spine, and ribs are the most likely to be related to osteoporosis or low bone density, whereas fractures around the skull, face, feet, and phalanges are the least likely to be related. Low bone density increases risk of all fractures to some extent.

Selecting Outcome Variables

Using different outcomes for model training and model validation is justified for several reasons. The goal of our algorithm is to identify people who would benefit from screening for osteoporosis. In an ideal world, we would conduct a prospective study where our algorithm is applied to chest radiographs taken of individuals without an osteoporosis diagnosis who are followed and periodically screened for osteoporosis overtime. An algorithm that performs well in this setting would prospectively identify individuals who either have osteoporosis, go on to suffer from osteoporosis, or that eventually experience a fragility fracture. The data in our study is retrospective and only includes osteoporosis labels derived from electronic health records. Given that osteoporosis is underdiagnosed, there are likely many individuals in our sample who are osteoporotic but have never been screened and do not have a recorded diagnosis. A ‘perfect’

model, designed to classify 100% of osteoporosis cases correctly, would miss the entire population that is currently underdiagnosed, perpetuating the status quo and potential biases.

Using osteoporosis-related outcomes for validation, as opposed to osteoporosis diagnoses, helps to abate this issue. Namely, we evaluate how well the algorithm identifies individuals at a higher risk of fracture in a population of individuals without an osteoporosis or osteopenia diagnosis and further exploit variation in the likelihood of a fracture in a given location being attributable to osteoporosis. Not only are fractures less likely to be underdiagnosed because of their symptoms, fractures located around the hip, pelvis, spine, and ribs tend to be more attributable to osteoporosis than fractures of the skull, face, or toes (Kelsey and Samelson, 2009; Warriner et al., 2010; Mai et al., 2019). An algorithm that is able to identify individuals that are at increased risk of fractures relative to the general population is likely identifying good candidates for screening. If the risk of osteoporosis-related fractures is higher than the risk of less related fractures in the identified population, it further suggests the algorithm is identifying osteoporotic individuals, rather than individuals who are at a higher risk of fractures in general, for whom treatment recommendations likely differ. Evaluating bone fractures in areas that are not visible on a chest radiograph also allows us to comment on the degree to which our algorithm identifies individuals with poor overall bone quality, rather than simply identifying individuals with issues that are directly visible in a chest x-ray.

It is also valid to question why we do not directly predict osteoporotic fracture outcomes, rather than predict osteoporosis diagnosis and then validate on fracture outcomes, given that the fracture outcomes are less likely to be underdiagnosed. Predicting fracture outcomes directly introduces another form of measurement error and, more importantly, complicates interpretability, which is an important factor for algorithm adoption. Although osteoporosis is a

strong predictor of future fracture occurrence, estimates suggest that less than 20 percent of fractures are attributable to osteoporosis in women and less than 13 percent of fractures are attributable to osteoporosis in men (Mai et al., 2019). Even for fractures that are most likely to be attributable to osteoporosis, such as hip and vertebral fractures, less than 41 percent of fractures in women and less than 21 percent of fractures in men are actually related to osteoporosis (Mai et al., 2019). Developing an algorithm that perfectly predicted fracture occurrence would overwhelmingly identify individuals who experience fractures for reasons unrelated to osteoporosis, which have different prevention and treatment recommendations compared to osteoporosis. Relatedly, predicting fractures directly may nudge any deep learning algorithm used toward identifying features in chest radiographs that predict fracture risk that are unrelated to bone quality. Although using whether or not an individual is diagnosed with osteoporosis is imperfect, it increases the likelihood that a deep learning algorithm is basing its predictions on bone quality.

Sample Definition and Algorithm Development

To develop and validate the algorithm, the dataset (N= 198,981 images) was divided into four segments. First, a balanced dataset was generated, which included all individuals with an osteoporosis or osteopenia diagnosis and an equivalent number of randomly sampled individuals without an osteoporosis or osteopenia diagnosis (N1 = 28,388 images). The balanced dataset was then randomly split into a model training set (70%), model validation set (15%), and a test/calibration set (15%), comprising the first three segments. The final segment that was saved for algorithm validation (N2= 170,593 images), was comprised entirely of patients without an osteoporosis or osteopenia diagnosis, which we will refer to as the remainder set.

Algorithm development was split into two parts. First, image classification models were used to generate the predicted probability that an image was labelled as diagnosed with osteoporosis. We used popular neural network architectures: ResNet50, Densenet-161, and VGG-16 to generate numerical representations of image features and a linear classifier to estimate logits for each class. Each classifier was trained using the binary cross-entropy (BCE) loss function for 50 epochs, a batch size of 16, and the Adam optimizer with a learning rate and weight decay of 5e-4. Predicted probabilities for each image were generated using a softmax function with temperature scaling after ensembling the predicted logits from each model by taking a simple average. Base predictions were made by predicting the class with the highest scaled logit value. For patients with multiple images in the same study, the maximum predicted probability across all images was used.

Final algorithm recommendations are made using a split conformal prediction procedure (Shafer and Vovk, 2008). First, we compute a class specific uncertainty metric for each image in the calibration set and the remainder set: the predicted probability of the class of interest, to be used as the conformal score. Higher scores are considered to be less uncertain. Using scores in the calibration set, we can then define a threshold value for each class. For example, we can choose the threshold value to be the class specific median conformal score in the calibration set. If the estimated score for a given class in the remainder set is greater than the threshold value calculated in the calibration set, we include that class in the prediction set. Accordingly, it is possible for the prediction set to include zero, one, or all classes, depending on the chosen threshold value. As the threshold value increases, fewer instances are labeled as a certain class, but there is a higher certainty that each instance is labeled correctly. The final algorithm output

was recorded as “recommend for screening” if the conformal prediction set included the osteoporosis label as the sole label in the prediction set, and as “do not flag”, otherwise.

Evaluating Algorithm Performance

To evaluate performance, we assess the algorithms ability to identify individuals who are at a high risk of fracture. We also evaluate whether or not the algorithm is identifying individuals who are not already recommended for screening and explore whether fracture risk in flagged individuals is higher for osteoporotic fractures compared to the risk of fractures less likely to be related to osteoporosis. For all evaluation procedures, we use the remainder set that is comprised entirely of individuals without a recorded osteoporosis diagnosis. The focus on individuals without an osteoporosis or osteopenia diagnosis is motivated by the fact that we are interested in how the algorithm performs in terms of identifying individuals who would benefit from screening, not how well it can identify individuals who are already diagnosed. We estimate fracture risk using a logistic regression of whether or not an individual is flagged as osteoporotic on individual fracture outcomes, controlling for age, sex, and body-mass-index. For our main results, we report fracture risks in the base case, where all individuals who are predicted to be osteoporotic by the deep learning model are flagged, and in two restricted cases, where the algorithm flags a maximum of 11 and 5 percent of individuals using conformal prediction. Eleven percent was selected given that is the fraction of individuals in our sample currently eligible for screening. Fracture risk was also estimated among those who are currently recommended for screening by clinical guidelines as another baseline comparison.

In this study the sample is comprised of non-randomly selected individuals who sought health care and are likely to be sicker and at a higher risk of fractures, causing estimates of fracture risk to be underestimates of risk compared to the general population, which may be a

more relevant measure to guide screening. To approximately adjust for this, we also report results after restricting the comparison group to individuals with a predicted osteoporosis risk below the median risk in the calibration set. This procedure provided approximate estimates of fracture risk relative to the general population under the assumption that those identified as less likely to be osteoporotic have a health status that is closer to that of the general population. We calculate all confidence intervals and assess statistical significance at the 95 percent confidence level.

We also compare our results to a procedure where potentially high-risk individuals are randomly flagged for screening. We define high risk as being a woman over 65 in accordance with screening guidelines. Specifically, we plot a distribution of adjusted odds ratio estimates, where we compare our observed results for a given fraction of people flagged by our algorithm to a distribution of estimates derived from randomly assigning the same number of high-risk people to be flagged. For example, in the case where our algorithm flags 10% of individuals, we can calculate 1000 different estimates after randomly assigning the same number of high-risk individuals to be flagged and evaluate where our actual estimate falls in the distribution.

To assess whether the algorithm is identifying individuals who are likely to benefit from a screening recommendation, we report the fraction of underserved individuals that are flagged by the algorithm. We define underserved as those not recommended for screening by the US Preventive Services Task Force (USPSTF). We define any women under 65 and any man as underserved, given that we do not have access to or knowledge of an individual's osteoporosis risk score. This information would also likely not be available to individuals taking chest radiographs for reasons unrelated to osteoporosis.

Results

In our sample, 10.8% of patents are women over the age of 65 that are currently captured by screening recommendations. We estimate that individuals currently recommended for screening are 1.21 (95% CI:1.09,1.34) times more likely to have experienced any fracture when compared to those who are not flagged in the validation population consisting entirely of individuals without an osteoporosis or osteopenia diagnosis. Our deep learning model flagged 20.6% of individuals with low bone density. We found that individuals who are flagged are 1.44 (95% CI: 1.31,1.59) times more likely to have experienced any fracture when compared to those who are not flagged, after adjusting for age, sex, and BMI (Table 1.1). We also found that the adjusted odds for osteoporotic fracture outcomes were greater than the adjusted odds for less likely to be related fractures (Table 1.1). When using conformal prediction to restrict the number of people the algorithm flagged to 11% and 5%, the adjusted odds ratios for experiencing any fracture increased to 1.53 (95% CI: 1.36, 1.73) and 1.93 (95% CI: 1.66, 2.66), respectively. We observe much larger increases in adjusted odds ratio for osteoporotic fractures after restricting the number of people flagged when compared to less likely to be related fractures. For example, when restricting the algorithm to flag approximately 5% of individuals, the adjusted odds of having experienced a hip or pelvis fracture are 2.19 (95% CI: 1.67, 2.88), whereas they are 1.24 (95% CI: 0.68, 2.25) for skull and face fractures. In Figure 1.1, we plot adjusted odds ratios as a function of the fraction of the population flagged to further highlight this pattern. In Table 1.2 and Figure 1.2, we reproduce the same sets of result after restricting the comparison group to those with a below median estimated risk of osteoporosis. Adjusted odds ratios for all outcomes were higher when defining the sample this way. Notably, we continue to observe the same pattern of higher relative increases in the odds of experiencing osteoporotic fractures compared

to less likely to be related fractures. For the scenario where the algorithm flags approximately 5% of individuals, the adjusted odds ratio of having experienced a hip or pelvis fracture is 3.32 (95% CI: 2.27, 4.85), whereas it is 1.56 (95% CI: 0.80, 3.05) for skull and face fractures.

In the base case where the deep learning algorithm flags 21% of individuals, 60% of individuals were from an underserved group. In the restricted case, 51% and 45% of individuals were from an underserved group when the algorithm flagged 11% and 5% of people respectively. We also compare our algorithm to a simulation where we randomly recommend 10% of people currently defined as high-risk for screening and compare them to individuals that are not selected. Exhibit A1 in the appendix plots a distribution of 1000 adjusted odds ratio estimates using this procedure and a red line indicating our algorithm's corresponding estimate. The median adjusted odds ratio of the simulated distribution was 1.29. The corresponding value when 10% of individuals were identified by the algorithm was 1.53, which falls at the 99th percentile of this distribution.

Figure 1.1 Adjusted Odds of Having Experienced Various Fractures as a Function of the Fraction of Individuals Flagged by the Algorithm – Full Sample Estimates

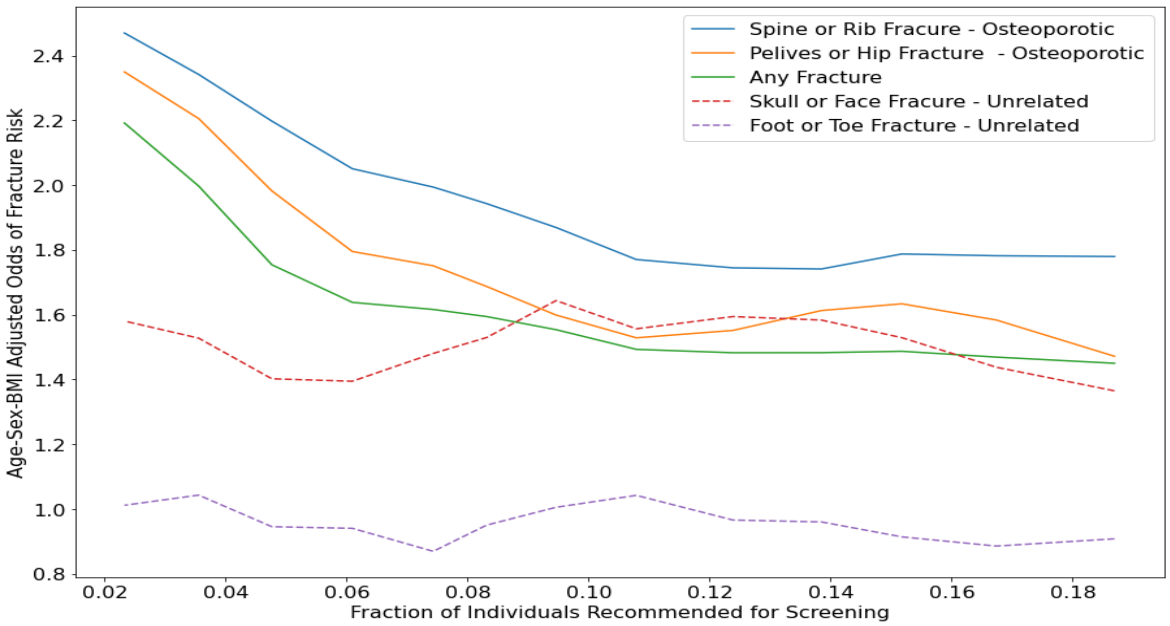


Figure 1.2 Adjusted Odds of Having Experienced Various Fractures as a Function of the Fraction of Individuals Flagged by the Algorithm – Full Sample Estimates

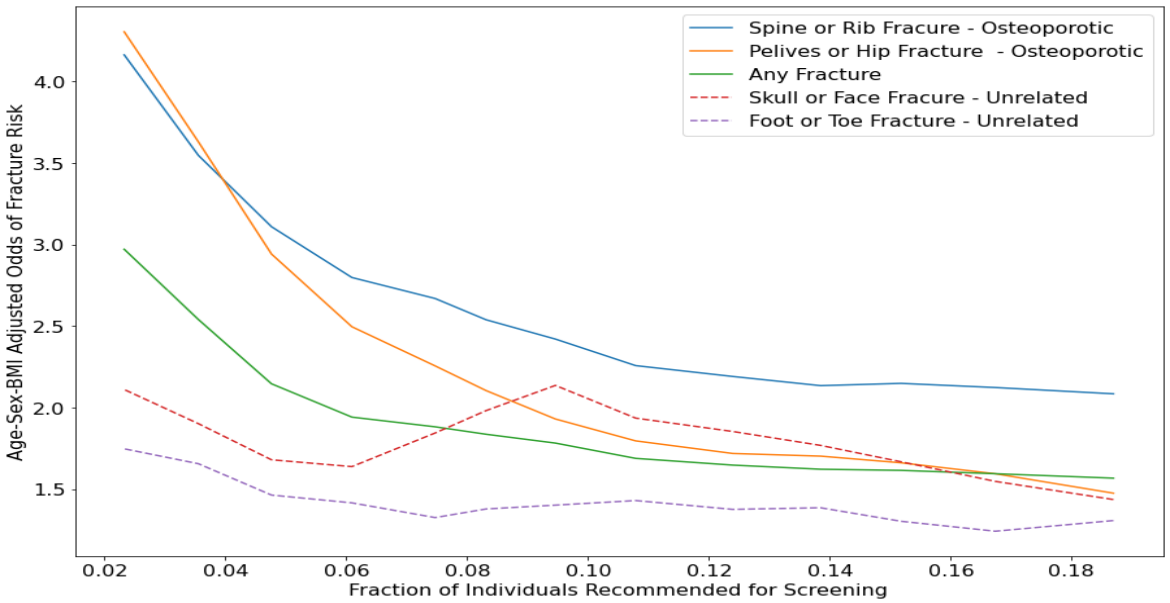


Table 1.1 Adjusted Odds of Having Experienced Various Fractures – Full Sample

Age-Sex-BMI Adjusted Odds Ratio Comparing Flagged Individuals to Not Flagged Individuals						
Outcome	**Fraction Flagged**	**Fraction Underserved**	**OR**	**Lower**	**Upper**	**P-Value**
Any Fracture	0.21	0.60	1.44	1.31	1.59	0.00
Any Fracture	0.11	0.51	1.53	1.36	1.73	0.00
Any Fracture	0.05	0.45	1.93	1.66	2.26	0.00
Spine or Rib Fracture	0.21	0.60	1.77	1.53	2.06	0.00
Spine or Rib Fracture	0.11	0.51	1.84	1.53	2.21	0.00
Spine or Rib Fracture	0.05	0.45	2.49	1.99	3.11	0.00
Hip or Pelvis Fracture	0.21	0.60	1.43	1.16	1.77	0.00
Hip or Pelvis Fracture	0.11	0.51	1.56	1.23	1.98	0.00
Hip or Pelvis Fracture	0.05	0.45	2.19	1.67	2.88	0.00
Foot or Toe Fracture	0.21	0.60	0.83	0.53	1.29	0.40
Foot or Toe Fracture	0.11	0.51	1.12	0.67	1.87	0.67
Foot or Toe Fracture	0.05	0.45	1.16	0.59	2.26	0.67
Skull or Face Fracture	0.21	0.60	1.38	1.00	1.90	0.05
Skull or Face Fracture	0.11	0.51	1.54	1.03	2.31	0.04
Skull or Face Fracture	0.05	0.45	1.24	0.68	2.25	0.49

Table 1.2 Adjusted Odds of Having Experienced Various Fractures – Restricted Sample

Age-Sex-BMI Adjusted Odds Ratio Comparing Flagged Individuals to Individuals at Below Median Estimated Risk of Osteoporosis						
Outcome	**Fraction Flagged**	**Fraction Underserved**	**OR**	**Lower**	**Upper**	**P-Value**
Any Fracture	0.21	0.60	1.56	1.38	1.77	0.00
Any Fracture	0.11	0.51	1.77	1.51	2.06	0.00
Any Fracture	0.05	0.45	2.37	1.96	2.87	0.00
Spine or Rib Fracture	0.21	0.60	2.08	1.73	2.50	0.00
Spine or Rib Fracture	0.11	0.51	2.37	1.88	2.99	0.00
Spine or Rib Fracture	0.05	0.45	3.48	2.63	4.62	0.00
Hip or Pelvis Fracture	0.21	0.60	1.45	1.10	1.91	0.01
Hip or Pelvis Fracture	0.11	0.51	1.92	1.38	2.67	0.00
Hip or Pelvis Fracture	0.05	0.45	3.32	2.27	4.85	0.00
Foot or Toe Fracture	0.21	0.60	1.21	0.66	2.19	0.54
Foot or Toe Fracture	0.11	0.51	1.48	0.74	2.95	0.27
Foot or Toe Fracture	0.05	0.45	1.70	0.73	3.95	0.22
Skull or Face Fracture	0.21	0.60	1.43	0.99	2.08	0.06
Skull or Face Fracture	0.11	0.51	2.06	1.29	3.28	0.00
Skull or Face Fracture	0.05	0.45	1.56	0.80	3.05	0.19

Discussion

This study demonstrates the viability of identifying people using chest radiographs who are likely to benefit from further assessment for actual fracture risk, which can be used to guide treatment (Tosteson et al., 2008). This analysis was tailored to address commonly cited issues with the use of deep learning algorithms in clinical settings such as an emphasis on predicting final disease labels and reporting statistical accuracy metrics. This study, in contrast, creates algorithms with outputs and validation metrics that are better tailored to clinical decision making. In addition, we use conformal prediction to address potential issues surrounding the size of the flagged population, an important factor when implementing algorithms in clinical practice.

In a population comprised entirely of individuals without a recorded osteoporosis or osteopenia diagnosis, we were able to use a deep learning algorithm to flag a subset of individuals who were much more likely to have experienced a fracture, after adjusting for age, sex, and BMI. Approximately half of the flagged individuals were from underserved populations and are currently not recommended for screening. Flagged individuals had higher adjusted odds ratios for osteoporotic fractures than for fractures less likely to be related to osteoporosis, suggesting that the algorithm is identifying individuals who are at higher risk of osteoporotic fractures as opposed to simply having a higher risk of fractures in general. Relatedly, the algorithm is able to detect individuals at increased risk of common osteoporotic fractures in locations that are not visible on a chest x-ray, such as the hip and pelvis, suggesting that the algorithm is identifying issues with overall bone quality and not simply issues visible in the chest x-ray itself.

Deep learning algorithms trained on historical data have the potential to amplify existing biases and often place an emphasis on predicting final disease labels, in lieu of creating

algorithms with outputs and validation metrics that are more aligned with a clinical endpoint. Previous studies have used chest radiographs to predict bone density (Pickhardt et al., 2013; Yamamoto et al., 2020; Yasaka et al., 2020; Zhang et al., 2020; Fang et al., 2021; Ho et al., 2021; Hsieh et al., 2021; Jang et al., 2021) with a high degree of accuracy with the goal of identifying individuals for screening for osteoporosis. Using low bone density alone to identify osteoporosis is problematic on several fronts. The majority of fragility fractures occur in individuals with bone minerals density in the osteopenic range (Unnanuntana et al., 2010). In addition, people undergoing bone mineral density measurement are not randomly selected. Algorithms trained and validated using this data consists of individuals that are already much more likely to be screened and the data samples are overwhelmingly or entirely comprised of older women (Ferizi, Honig, and Chang, 2019). Current screening recommendations place as much or more emphasis on clinical assessment with tools such as FRAX (Marques et al., 2015) in conjunction with bone mineral density studies to identify fragility fracture risk, which are then used to select patients for therapeutic interventions (Tosteson et lal., 2008). Algorithms identifying patients with the disease label osteoporosis are thus likely to underdiagnose patients with significant fracture risk when applied in a wider population.

The outcome of significance in screening for low bone density is fragility fracture risk. In this study we evaluated how well our algorithm identified individuals at a higher risk of fracture in a population of individuals without an osteoporosis or osteopenia diagnosis and further exploited variation in the likelihood of a fracture in a given location being attributable or not to osteoporosis. An algorithm that is able to identify individuals that are at increased risk of fractures relative to the general population, as observed in this study, is likely identifying good candidates for screening. Measures of association, such as the odds ratio used in this work, also

align more closely with the type of evidence used to justify treatment decisions than statistical measures of algorithm accuracy.

Many machine learning-based decision support tools are not used because of implementation issues (Jacob et al., 2021). Given the ubiquity of chest x-rays, approximately 110 million per year in the United States (Mettler et al., 2020), screening all individuals flagged by an algorithm may not be feasible because of capacity constraints. In addition, repeated flagging of individuals already recommended for screening could result in alert fatigue, which can make clinicians less likely to act on the algorithm alerts (Ancker et al., 2017). Algorithms that make recommendations that are outside current clinical guidelines may also require a higher standard of evidence (Jacob et al, 2021) to gain trust in the algorithm. Accordingly, any algorithm developed to guide screening would need to have a high sensitivity and a demonstrated ability to identify a group that is likely to benefit from a screening recommendation. To address these potential concerns, this work demonstrates how conformal prediction can be used to identify individuals who are most likely to benefit from recommendations for screening. Conformal prediction provides a means of integrating uncertainty into final predictions (Kompa et al., 2021), a way to control the size of the flagged patient population, and a way to alter algorithm outputs such that they better reflect what is useful for a given task (Liu et al., 2021). When combining our deep learning algorithm with a conformal prediction procedure, we were able to identify smaller subsets of the population that were at an even higher fracture risk than those identified when solely using the deep learning algorithm. The approximate fraction of people flagged could be adjusted such that it reflects a cost-effective increase in screening wherever the algorithm is being implemented.

In summary the algorithm presented could be integrated into a decision support tool that flags individuals who do not currently fall under the screening recommendations and have a high risk of fracture. The tool also integrates justifications for screening recommendations using metrics, such as the adjusted odds ratio for fracture risk, which more closely resembles evidence generally used, such as actual fracture risk, to guide treatment decisions. Finally, the model presented uses conformal prediction to control the size of the flagged population by identifying individuals with the highest likelihood of benefiting from opportunistic screenings for fragility fracture risk.

References

Adler-Milstein, J., Chen, J. H., & Dhaliwal, G. (2021). Next-Generation Artificial Intelligence for Diagnosis: From Predicting Diagnostic Labels to "Wayfinding." *JAMA*, *326*(24), 2467–2468. https://doi.org/10.1001/jama.2021.22396

Ancker, J. S., Edwards, A., Nosal, S., Hauser, D., Mauer, E., Kaushal, R., & with the HITEC Investigators. (2017). Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. *BMC Medical Informatics and Decision Making*, *17*(1), 36. https://doi.org/10.1186/s12911-017-0430-8

Black, D. M., Cummings, S. R., Genant, H. K., Nevitt, M. C., Palermo, L., & Browner, W. (1992). Axial and appendicular bone density predict fractures in older women. *Journal of Bone and Mineral Research*, *7*(6), 633–638. https://doi.org/10.1002/jbmr.5650070607

Chen, I. Y., Pierson, E., Rose, S., Joshi, S., Ferryman, K., & Ghassemi, M. (2021). Ethical Machine Learning in Health Care. *Annual Review of Biomedical Data Science*, *4*(1), 123–144. https://doi.org/10.1146/annurev-biodatasci-092820-114757

US Preventive Services Task Force, Curry, S. J., Krist, A. H., Owens, D. K., Barry, M. J., Caughey, A. B., Davidson, K. W., Doubeni, C. A., Epling, J. W., Kemper, A. R., Kubik, M., Landefeld, C. S., Mangione, C. M., Phipps, M. G., Pignone, M., Silverstein, M., Simon, M. A., Tseng, C.-W., & Wong, J. B. (2018). Screening for Osteoporosis to Prevent Fractures: US Preventive Services Task Force Recommendation Statement. *JAMA*, *319*(24), 2521. https://doi.org/10.1001/jama.2018.7498

Curtis, J. R., Carbone, L., Cheng, H., Hayes, B., Laster, A., Matthews, R., Saag, K. G., Sepanski, R., Tanner, S. B., & Delzell, E. (2008). Longitudinal trends in use of bone mass measurement among older americans, 1999-2005. *Journal of Bone and Mineral Research: The Official*

Journal of the American Society for Bone and Mineral Research, *23*(7), 1061–1067. https://doi.org/10.1359/jbmr.080232

Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., Liu, Y., Topol, E., Dean, J., & Socher, R. (2021). Deep learning-enabled medical computer vision. *Npj Digital Medicine*, *4*(1), 1–9. https://doi.org/10.1038/s41746-020-00376-2

Fang, Y., Li, W., Chen, X., Chen, K., Kang, H., Yu, P., Zhang, R., Liao, J., Hong, G., & Li, S. (2021). Opportunistic osteoporosis screening in multi-detector CT images using deep convolutional neural networks. *European Radiology*, *31*(4), 1831–1842. https://doi.org/10.1007/s00330-020-07312-8

Ferizi, U., Honig, S., & Chang, G. (2019). Artificial Intelligence, Osteoporosis and Fragility Fractures. *Current Opinion in Rheumatology*, *31*(4), 368–375. https://doi.org/10.1097/BOR.0000000000000607

Ho, C.-S., Chen, Y.-P., Fan, T.-Y., Kuo, C.-F., Yen, T.-Y., Liu, Y.-C., & Pei, Y.-C. (2021). Application of deep learning neural network in predicting bone mineral density from plain X-ray radiography. *Archives of Osteoporosis*, *16*(1), 153. https://doi.org/10.1007/s11657-021-00985-8

Hsieh, C.-I., Zheng, K., Lin, C., Mei, L., Lu, L., Li, W., Chen, F.-P., Wang, Y., Zhou, X., Wang, F., Xie, G., Xiao, J., Miao, S., & Kuo, C.-F. (2021). Automated bone mineral density prediction and fracture risk assessment using plain radiographs via deep learning. *Nature Communications*, *12*(1), 5472. https://doi.org/10.1038/s41467-021-25779-x

Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K., Seekins, J., Mong, D. A., Halabi, S. S., Sandberg, J. K., Jones, R., Larson, D. B., Langlotz, C. P., Patel, B. N., Lungren, M. P., & Ng, A. Y. (2019). CheXpert: A Large

Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. *ArXiv:1901.07031 [Cs, Eess]*. http://arxiv.org/abs/1901.07031

Jang, S., Graffy, P. M., Ziemlewicz, T. J., Lee, S. J., Summers, R. M., & Pickhardt, P. J. (2019). Opportunistic Osteoporosis Screening at Routine Abdominal and Thoracic CT: Normative L1 Trabecular Attenuation Values in More than 20 000 Adults. *Radiology*, *291*(2), 360–367. https://doi.org/10.1148/radiol.2019181648

Jacobs, M., He, J., F. Pradier, M., Lam, B., Ahn, A. C., McCoy, T. H., Perlis, R. H., Doshi-Velez, F., & Gajos, K. Z. (2021). Designing AI for Trust and Collaboration in Time-Constrained Medical Decisions: A Sociotechnical Lens. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3411764.3445385

Kelsey, J. L., & Samelson, E. J. (2009). Variation in Risk Factors for Fractures at Different Sites. *Current Osteoporosis Reports*, *7*(4), 127–133.

Kompa, B., Snoek, J., & Beam, A. L. (2021). Second opinion needed: Communicating uncertainty in medical machine learning. *Npj Digital Medicine*, *4*(1), 1–6. https://doi.org/10.1038/s41746-020-00367-3

Leboime, A., Confavreux, C. B., Mehsen, N., Paccou, J., David, C., & Roux, C. (2010). Osteoporosis and mortality. *Joint Bone Spine*, *77 Suppl 2*, S107-112. https://doi.org/10.1016/S1297-319X(10)70004-X

Lewiecki, E. M., Laster, A. J., Miller, P. D., & Bilezikian, J. P. (2012). More bone density testing is needed, not less. *Journal of Bone and Mineral Research*, *27*(4), 739–742. https://doi.org/10.1002/jbmr.1580

Liu, X., Faes, L., Kale, A. U., Wagner, S. K., Fu, D. J., Bruynseels, A., Mahendiran, T., Moraes, G., Shamdas, M., Kern, C., Ledsam, J. R., Schmid, M. K., Balaskas, K., Topol, E. J., Bachmann, L. M., Keane, P. A., & Denniston, A. K. (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. *The Lancet Digital Health*, *1*(6), e271–e297. https://doi.org/10.1016/S2589-7500(19)30123-2

Lu, C., Lemay, A., Chang, K., Hoebel, K., & Kalpathy-Cramer, J. (2022). Fair Conformal Predictors for Applications in Medical Imaging. *ArXiv:2109.04392 [Cs, Eess]*. http://arxiv.org/abs/2109.04392

Lungren M., Kim J., Bogdan S., Lane W., Risley J., Haynes K., and Obermeyer Z. (2021). Predicting fractures and pain using chest x-rays. *Nightingale Open Science Dataset*. doi:10.48815/N5RP44

Mai, H. T., Tran, T. S., Ho-Le, T. P., Center, J. R., Eisman, J. A., & Nguyen, T. V. (2019). Two-Thirds of All Fractures Are Not Attributable to Osteoporosis and Advancing Age: Implications for Fracture Prevention. *The Journal of Clinical Endocrinology & Metabolism*, *104*(8), 3514–3520. https://doi.org/10.1210/jc.2018-02614

Marques, A., Ferreira, R. J. O., Santos, E., Loza, E., Carmona, L., & da Silva, J. A. P. (2015). The accuracy of osteoporotic fracture risk prediction tools: A systematic review and meta-analysis. *Annals of the Rheumatic Diseases*, *74*(11), 1958–1967. https://doi.org/10.1136/annrheumdis-2015-207907

Majumdar, S. R., Kim, N., Colman, I., Chahal, A. M., Raymond, G., Jen, H., Siminoski, K. G., Hanley, D. A., & Rowe, B. H. (2005). Incidental Vertebral Fractures Discovered With Chest Radiography in the Emergency Department: Prevalence, Recognition, and Osteoporosis

Management in a Cohort of Elderly Patients. *Archives of Internal Medicine*, *165*(8), 905–909. https://doi.org/10.1001/archinte.165.8.905

Mettler, F. A., Mahesh, M., Bhargavan-Chatfield, M., Chambers, C. E., Elee, J. G., Frush, D. P., Miller, D. L., Royal, H. D., Milano, M. T., Spelic, D. C., Ansari, A. J., Bolch, W. E., Guebert, G. M., Sherrier, R. H., Smith, J. M., & Vetter, R. J. (2020). Patient Exposure from Radiologic and Nuclear Medicine Procedures in the United States: Procedure Volume and Effective Dose for the Period 2006–2016. *Radiology*, *295*(2), 418–427. https://doi.org/10.1148/radiol.2020192256

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science (New York, N.Y.)*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

Overman, R. A., Farley, J. F., Curtis, J. R., Zhang, J., Gourlay, M. L., & Deal, C. L. (2015). DXA utilization between 2006 and 2012 in commercially insured younger postmenopausal women. *Journal of Clinical Densitometry : The Official Journal of the International Society for Clinical Densitometry*, *18*(2), 145–149. https://doi.org/10.1016/j.jocd.2015.01.005

Pickhardt, P. J., Pooler, B. D., Lauder, T., del Rio, A. M., Bruce, R. J., & Binkley, N. (2013). Opportunistic screening for osteoporosis using abdominal computed tomography scans obtained for other indications. *Annals of Internal Medicine*, *158*(8), 588–595. https://doi.org/10.7326/0003-4819-158-8-201304160-00003

Pouresmaeili, F., Kamalidehghan, B., Kamarehei, M., & Goh, Y. M. (2018). A comprehensive overview on osteoporosis and its risk factors. *Therapeutics and Clinical Risk Management*, *14*, 2029–2049. https://doi.org/10.2147/TCRM.S138000

Seyyed-Kalantari, L., Zhang, H., McDermott, M. B. A., Chen, I. Y., & Ghassemi, M. (2021). Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-

served patient populations. *Nature Medicine*, *27*(12), 2176–2182. https://doi.org/10.1038/s41591-021-01595-0

Shafer, G., & Vovk, V. (2008). A Tutorial on Conformal Prediction. *Journal of Machine Learning Research*, *9*(12), 371–421.

Siris, E. S., Adler, R., Bilezikian, J., Bolognese, M., Dawson-Hughes, B., Favus, M. J., Harris, S. T., Jan de Beur, S. M., Khosla, S., Lane, N. E., Lindsay, R., Nana, A. D., Orwoll, E. S., Saag, K., Silverman, S., & Watts, N. B. (2014). The clinical diagnosis of osteoporosis: A position statement from the National Bone Health Alliance Working Group. *Osteoporosis International*, *25*(5), 1439–1443. https://doi.org/10.1007/s00198-014-2655-z

Tonekaboni, S., Joshi, S., McCradden, M. D., & Goldenberg, A. (2019). What Clinicians Want: Contextualizing Explainable Machine Learning for Clinical End Use. *ArXiv:1905.05134 [Cs, Stat]*. http://arxiv.org/abs/1905.05134

Tosteson, A. N. A., Melton, L. J., Dawson-Hughes, B., Baim, S., Favus, M. J., Khosla, S., Lindsay, R. L., & National Osteoporosis Foundation Guide Committee. (2008). Cost-effective osteoporosis treatment thresholds: The United States perspective. *Osteoporosis International: A Journal Established as Result of Cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA*, *19*(4), 437–447. https://doi.org/10.1007/s00198-007-0550-6

Warriner, A. H., Patkar, N. M., Curtis, J. R., Delzell, E., Gary, L., Kilgore, M., & Saag, K. G. (2011). Which Fractures Are Most Attributable to Osteoporosis? *Journal of Clinical Epidemiology*, *64*(1), 46–53. https://doi.org/10.1016/j.jclinepi.2010.07.007

Willson, T., Nelson, S. D., Newbold, J., Nelson, R. E., & LaFleur, J. (2015). The clinical epidemiology of male osteoporosis: A review of the recent literature. *Clinical Epidemiology*, *7*, 65–76. https://doi.org/10.2147/CLEP.S40966

Yamamoto, N., Sukegawa, S., Kitamura, A., Goto, R., Noda, T., Nakano, K., Takabatake, K., Kawai, H., Nagatsuka, H., Kawasaki, K., Furuki, Y., & Ozaki, T. (2020). Deep Learning for Osteoporosis Classification Using Hip Radiographs and Patient Clinical Covariates. *Biomolecules*, *10*(11), E1534. https://doi.org/10.3390/biom10111534

Yasaka, K., Akai, H., Kunimatsu, A., Kiryu, S., & Abe, O. (2020). Prediction of bone mineral density from computed tomography: Application of deep learning with a convolutional neural network. *European Radiology*, *30*(6), 3549–3557. https://doi.org/10.1007/s00330-020-06677-0

Zhang, B., Yu, K., Ning, Z., Wang, K., Dong, Y., Liu, X., Liu, S., Wang, J., Zhu, C., Yu, Q., Duan, Y., Lv, S., Zhang, X., Chen, Y., Wang, X., Shen, J., Peng, J., Chen, Q., Zhang, Y., … Zhang, S. (2020). Deep learning of lumbar spine X-ray for osteopenia and osteoporosis screening: A multicenter retrospective cohort study. *Bone*, *140*, 115561. https://doi.org/10.1016/j.bone.2020.115561

Chapter 2:

Dynamics of the First Mover: Evidence from the United States Generic Pharmaceutical System

Abstract

Early entry by firms in markets for generic pharmaceuticals is paramount for price declines and increasing consumer welfare. The first-mover advantage (FMA), defined as the additional market share a firm earns by entering first compared to entering second or later, is considered one of the most important incentives for prompt entry into generic markets. Inherent in this definition of FMA is a notion of causality, where, in an ideal but infeasible experiment, we could observe *the same* firm entering at different times in *the same* market and compare market shares across different entry timing decisions. This paper will exploit unique characteristics of generic drug markets and advances in doubly robust methods for causal inference to estimate the FMA. Our findings suggest entering first results in a significant advantage when compared to if that same firm entered later. This advantage is mainly accrued through sales in periods where competition is limited and, to a lesser extent, sustained higher market shares during the first few years after the start of generic competition. The latter point contradicts many previous studies in this area. We further characterize the impact that important regulatory features, such as exclusivity periods and the presence of authorized generics, have on the market shares of early entrants and provide evidence to suggest our estimates are extremely unlikely to be explained by unobserved confounding.

Introduction

Entry decisions by firms in generic pharmaceutical markets have been studied extensively and it is well established that early entry in a market is paramount for price declines and increasing consumer welfare (Scott Morton, 1999; Berndt and Aitken, 2010). The incentives for early entry into a market varies across industries and few markets are as intricate as generic markets where firms navigate a complex regulatory system (Reiffen and Ward, 2005). The first-mover advantage (FMA), defined as the additional market share a firm earns by entering first compared to entering second or later, is considered one of the most important incentives for prompt entry into generic markets (Appelt, 2015; Shajarizadeh, Grootendorst, and Hollis, 2015; Yu and Gupta, 2015). Inherent in this definition of FMA is a notion of causality, where, in an ideal but infeasible experiment, we could observe *the same* firm entering at different times in *the same* market and compare market shares across different entry timing decisions (Lieberman and Montgomery, 2013). First-entrants to generic markets in the United States are often afforded an extremely profitable 180-day period of exclusivity where regulators restrict additional entry into a market after successfully challenging a weak patent (Hemphill and Lemley, 2011). In other cases, generic firms may also sell during a period where competition is limited after receiving authorization from the branded firm (Berndt et al., 2007; Federal Trade Commission, 2011; Appelt, 2015; Bokhari, Mariuzzo, and Polanski, 2020; Peelish, 2020). First entrants also tend to maintain higher market shares than competitors that entered later, well after there is open competition between all approved generics, potentially providing an additional benefit to early entry (Federal Trade Commission, 2011; Shajarizadeh, Grootendorst, and Hollis, 2015; Yu and Gupta, 2015).

This paper will exploit unique characteristics of generic drug markets and advances in doubly robust methods for causal inference to estimate the FMA. Our findings suggest entering first results in a significant advantage when compared to if that same firm entered later and that this effect is heterogenous across important market characteristics. This advantage is mainly accrued through sales in periods where competition is limited and, to a lesser extent, sustained higher market shares during the first few years after the start of generic competition. The latter point contradicts many previous studies in this area. We further characterize the impact that important regulatory features, such as authorized generics, have on the profit of early entrants.

Any empirical assessment of the FMA needs to overcome endogeneity issues, wherein observed difference in market shares between early and late entrants may simply reflect systematic differences in the firms that enter early and the firms that enter late. To address endogeneity in our study, we focus on a specification that compares firms with first entrant status to firms that entered in the first year after the start of competition between generic firms. This removes unobserved confounding associated with the fact that firms that enter possibly years after loss of exclusivity may look very different than those that enter earlier. We also control for observed differences between the restricted set of early entrants using doubly robust methods for causal inference that are able to leverage machine learning in order to relax assumptions around functional form, while allowing for a possibly large number of potential confounders and complex confounding relationships (Chernozhukov et al., 2018). Given there are likely still unobserved differences between first and late entrants, we explore how sensitive our estimates are to unobserved confounding using a method outlined in Cinelli and Hazlett (2019). Since firms can achieve first entrant status in generic markets in a variety of different ways, such as by

winning exclusivity periods or by marketing authorized generics, we also provide estimates of the first mover advantage and how it varies overtime within industry relevant subgroups.

To estimate the FMA, I use firm level data of US pharmaceutical markets that lost patent from 2010-2012 during the so-called "Patent-Cliff". We estimate the FMA using two distinct outcomes. In each case, we use the difference in market shares between the first entrant and the predicted market share of a firm that entered in the first year after the start of generic competition as a proxy for the additional profit a firm accrues because of entering first. In our main specification, we define a firms' market share using all generic sales accrued from the start of a generic market to four years after the start of competition between different generics. This measure is higher for firms who sell product during more profitable periods where there is limited competition and higher prices. To evaluate whether or not a firm's advantage is sustained overtime, we also provide estimates where market shares are calculated using sales restricted to the first, second, third, and fourth year after the start of competition between generics. Although sustained higher market shares provide incrementally less profit as the market ages and prices drop, evaluating the dynamics of how the FMA varies overtime provides insight into the mechanisms through which an advantage may arise.

In our main specification, we estimate that first entrants have a market share that is 17 percentage points higher on average than if that firm had entered in the first year after the start of generic competition. After restricting the sample to the time period after the start of competition amongst generics, we estimate that first movers are afforded a 10, 3, 0, and -2 percentage point difference in market shares in the first, second, third, and fourth year after the start of competition between generics, respectively. The estimates in year three and four were not statistically different from zero. These results suggest that there is a large FMA, however, first

entrant market shares diminish overtime relative to later entrants, which is a key deviation from previous work. We further provide evidence to suggest that it is extremely unlikely that these results are driven by unobserved confounding. Even if an unobserved confounder that was unrelated to any of the included controls in our model exists and has the same explanatory power as a set of manufacturers fixed effects, we find that the lower bound of our main estimate would be 7 percentage points.

There is significant heterogeneity in the effect estimates across market characteristics. Smaller markets and markets where exclusivity periods are present afford first entrants much larger additional expected market shares. We further find that authorized generics reduce the FMA afforded to other first entrants. Yet, given that authorized generics are more likely enter into large markets with limited competition, the remaining potential entrants still have a large incentive to enter early and authorized generics are unlikely to have a large impact on market competitiveness in the short-term.

Industry Background

Unique features of generic markets complicate study of a FMA. Early on in a market, there is only a single manufacturer of a mol-form, the brand company claiming patents on the product (Scott-Morton and Kyle, 2011). Competition is restricted by a patent and, before generic entry, can only stem from manufactures of distinct molecules that share a therapeutic endpoint, often referred to as therapeutic competition (Regnier and Ridley, 2016). Generic entry into a market requires the approval of an Abbreviated New Drug Application by the FDA. Approval of a new generic product is based on therapeutic equivalence. The marketing of generic drugs is regulated by the Drug Price Competition and Patent Term Restoration Act of 1984 (The Hatch-

Waxman Act). Under the Hatch-Waxman Act, a generic firm stating its intent to enter a market can prompt a patent infringement lawsuit prior to entry, allowing firms earlier resolutions of issues of patent validity and the ability to prove therapeutic equivalence prior to the start of generic sales (Milne and Cairnes, 2003). Firms can enter a market with a license from the brand, "at-risk" prior to resolution of the patent suit, or, after all relevant patents expire pending FDA approval (US Food & Drug Administration, 2017). To stimulate early entry into a market, the first generic manufacturer or set of generic manufacturers to successfully challenge a patent or set of patents may be awarded a 180-day exclusivity period. In addition, branded firms can introduce an authorized generic at any point during a market's life cycle. Authorized generics are product's marketed by a branded firm through a contract with a supply partner or subsidiary (Appelt, 2015). Given that generic drugs are most profitable during the 180-day exclusivity period, authorized generics are often introduced alongside initial generic entrants over this period (Federal Trade Commission, 2011, Appelt, 2015, Peelish, 2020). Notably, many authorized generics are marketed by a limited set of generic manufacturers through arrangements with a branded firm (Federal Trade Commission, 2011; FDA Listing of Authorized Generics, 2021). After the branded product loses exclusivity and other forms of exclusivities are exhausted, competition is open to all firms with final ANDA approval from the FDA (Frank and Salkever, 1997; Scott-Morton and Kyle, 2011).

The unique competition structure in generic markets outlined above has several important implications for estimating the FMA. First, the first mover is not the first manufacturer to market a given product. Firms can observe characteristics of a given market and are able to select into markets that are most profitable to them (Scott-Morton, 1999). Second, firms are not usually sole first movers. The 180-day exclusivity structure or expiry of all exclusivities can result in multiple

firms entering a market first simultaneously and authorized generics are often introduced alongside the first generic option (Bokhari, Mariuzzo, and Polanski, 2020; Peelish, 2020). Third, a pool of manufacturers specialize in production and marketing generic drugs and the market is dominated by several major players. Prior to 2017, over 50% of all ANDAs were granted to just ten firms (Berndt. Conti, Murphy, 2017). Fourth, the highly regulated nature of pharmaceutical markets governs whether and when a firm is able to enter a market first and the extent of competition the first entrant faces. First entry in a generic market is expensive and requires navigation of a complicated legal process where branded firms often attempt to delay generic entry or shift sales to similar products that remain under patent (Berndt. Conti, and Murphy, 2017; Congressional Research Service, 2020; Peelish, 2020). The culmination of which is increased uncertainty for firms opting to enter generic markets that are not established. Lastly, firms tend to enter early in different ways, either through attempting to obtain 180-day exclusivity agreements or through winning agreements with branded manufactures to market authorized generics (Federal Trade Commission, 2011). In our sample, we observe that several major firms dominate overall entry into markets, and they are also more likely to enter earlier than other firms.

The structure of generic pharmaceutical purchasing further complicates how market share is won by a firm in generic markets. Regulations at the federal and state level, in combination with common practices in health insurance coverage and procurement, generally result in high levels of generic substitution as soon as a generic becomes available. In 2018, over 90% of all prescription drugs prescribed were generic. After restricting to drugs with a generic available, over 97% of drugs were dispensed as generic (IQVIA, 2019). Generic pharmaceutical sales notably do not flow directly from manufacturer to health care practices. Over 87% of all generic

pharmaceutical sales flowed through wholesalers and this increased to 95% by 2019 (Fein, 2020). The wholesale market is also highly concentrated. Three major players (AmerisourceBergen, Cardinal Health, McKesson Corporation) have made up over 75% of the wholesale market over the past decade. These wholesalers have further formed buying groups with major retailers in order to negotiate contracts with lower acquisition costs from generic manufacturers (IBIS World Report, 2018). For example, in 2013 Cardinal Health and CVS formed the joint-buying venture, "Red Oak Sourcing". By 2018, Red Oak Sourcing was responsible for one third of all generic purchases in the United States (Fein, 2019). Once in a market, a firm's market share is also determined by its ability to win contracts from large buying groups, through which the vast majority of sales flow (Federal Trade Commission, 2011; Berndt, Conti, and Murphy, 2017).

Previous Literature

Studies have examined the FMA using data on pharmaceuticals that lost patent prior to 2007, suggesting that there is a large and sustained difference in market share for early versus late entrants in generic markets. Using data on generic molecule-formulations (mol-forms) that lost patent from 2003 to 2007, Shajarizadeh, Grootendorst, and Hollis (2015) found that being the first entrant in Canadian markets is associated with higher market shares up to six years after loss of exclusivity when compared to entering third or later. Yu and Gupta (2008), using data from the US on mol-forms that lost patent from 1992-2000, also find that first entrants are afforded higher market shares sustained over several years of competition when compared to firms entering third or later. Several older studies have also found large and sustained FMA in generic markets (Grabowski and Vernon, 1992; Hollis, 2002).

In a 2011 report, the Federal Trade Commission completed an analysis of the short and long-term impacts of authorized generic entry on generic firm market shares in the United States using a sample of pharmaceuticals that lost patent between 2003-2008. Authorized generics are product's marketed by a branded firm through a contract with a supply partner or subsidiary (Appelt, 2015). Authorized generics are frequently launched alongside the first generic competitor in large pharmaceutical markets and the majority of 180-exclusivity periods feature competition from an authorized competitor (Bokhari, Mariuzzo, and Polanski, 2020; Peelish, 2020). The Federal Trade Commission report states that first entrants tend to have market shares that are higher than one may expect after the entry of potentially many other generic competitors. They also find that the revenues of first entrants are approximately 50-60% lower in markets where an authorized generic is introduced.

Each of the previously described studies comment on the potential mechanisms for why a large and sustained FMA is observed. There are clear reasons why a firm that enters first is likely to be more profitable in the early stages of a generic market, particularly in the United States. It is well established that marketing a generic in a period where competition is limited is extremely profitable and many first entrants enjoy periods of limited competition through winning 180-exclusivity periods or by marketing authorized generics (Scott-Morton and Kyle, 2011; Bokhari, Mariuzzo, and Polanski, 2020). Prices and associated profits per unit sold are higher during periods with limited competition and generally do not reach a stable minimum until one to two years after generic manufacturers enter a market (Berndt and Aitken, 2011; Frank, McGuire, and Nason, 2021). The observation that early entrants enjoy higher market shares for an extended period of time even after prices have stabilized is interesting, given that products are theoretically identical, and one may expect that price competition would be the main determinant

of a firm's market share (Scott-Morton and Kyle, 2011; Shajarizadeh et al., 2015; Yu and Gupta, 2015). Shajarizadeh and colleagues suggest that differencses in market shares in the long run are largely driven by 'transaction costs' where pharmacies have preferences for purchasing from the same manufacturers overtime given the small potential costs of switching. Yu and Gupta argue that patient preferences for a given pill aesthetic (e.g., shape and color) could result in consumer preferences for a given manufacturer. The Federal Trade Commission report remarks that it may be caused by early firms establishing themselves as a reliable supplier during a period with reduced competition.

Issues of Endogeneity

A valid explanation for why first movers may enjoy higher market shares than firms that enter later on is that the observed difference in market shares between early and late entrants may simply reflect systematic differences in the firms that enter early and the firms that enter late. If the characteristics of early and late entrants differ, a simple comparison of market shares between first and later entrants may conflate these systemic differences with the FMA. Indeed, there are many observed differences between early and late entrants in markets for generic drugs. Whether or not a given firm enters first is related to attributes of the firm, the market, and possibly complex interactions between the two that also determine a firm's expected market share (Cha and Yu, 2014). For example, our data show that large generic firms such as Mylan or Teva are more likely to be the first entrants in a given market and to command a higher market share than other entrants in the same market. These firms tend to have large product portfolios and higher manufacturing capacities than smaller firms, which likely affords them contracting advantages (Federal Trade Commission 2011; Cha and Yu, 2014). In addition, larger firms likely

have advantages navigating complex legal processes and barriers, such as 'patent thickets," created by brand firms to impair generic entry (Congressional Research Services, 2020). A generic firm's expected market share may also vary across therapeutic classes or identity of the branded firm, which gives rise to the possibility of confounding through a complex interaction between manufacturer and other market attributes (Scott Morton, 1999). For example, large firms can partner with branded manufacturers to market authorized generics, which potentially allows them to enter markets earlier than would be possible otherwise (Berndt et al., 2007; Peelish, 2020; FDA Listing of Authorized Generics, 2021). In our sample, just under half of the authorized generics sold are marketed by a firm that predominantly markets generic pharmaceuticals through Abbreviated New Drug Applications (ANDAs). In addition, several branded pharmaceutical firms market their own generic products or have a subsidiary that does so, and these firms almost exclusively enter markets early on (Federal Trade Commission, 2011; Appelt, 2015; Peelish, 2020). Given the complex nature of market entry in this setting, it is easy to conflate a FMA with larger market shares afforded to early entrants that result from some other factor than order of entry.

Previous work pertaining to off-patent pharmaceutical markets has leveraged the FDA Abbreviated New Drug Application (ANDA) review process that governs entry into a market to address endogeneity concerns when estimating the impact of firm entry on prices (Reiffen and Ward, 2007; Appelt, 2015; Frank et al., 2021). It is tempting to use a similar strategy when studying the FMA. A review process for a novel generic market often involves several firms applying for ANDAs prior to a market's existence without any knowledge of the number and identity of other entrants. The order of entry in a market in such cases is partially determined by the timing of the FDA review process and is less likely to be associated with attributes of a given

firm, making first-mover status more likely to be exogenously determined amongst earlier applicants. However, there are likely large differences between the initial set of applicants where this logic applies and those firms who do not apply until later. It is also common to have a firm amongst the earliest set of marketers that skirted the review process altogether. Branded firms or subsidiaries of branded firms tend to exclusively market authorized generics when they enter a generic market, and many authorized generics are marketed by large generic manufacturers (Federal Trade Commission, 2011; FDA Listing of Authorized Generics, 2021).

Methods

Data

To examine the impact of being the first entrant in a generic market on a firm's expected market share we combine data from several sources. Data from the National Sales Perspective Database (NSP) included quarterly data from 2009 (Q1) to 2018 (Q2) on all generic pharmaceutical retail and non-retail sales for drugs that lost patent protection during the so-called "patent-cliff" that occurred between 2010 and 2012. The main units of analysis were individual manufacturers within combined-molecule-product-form (mol-form-strength) markets (e.g. Teva in a market for Atorvastatin Oral Solids 10mg), which was the market definition used in previous papers studying the FMA (Shajarizadeh, Grootendorst, and Hollis, 2015; Yu and Gupta, 2015). We also report results at the combined-molecule-product-form level, which has been used in other studies focusing on US generic drug markets in appendix A (Frank, McGuire, and Nason, 2021). The NSP includes sales data for individual manufacturers based on the average invoice price of pharmaceuticals purchased from manufacturers or wholesalers. For each market, the data also included information on the marketing category (e.g. branded, branded beneric, or generic) of each manufacturer and the loss of exclusivity date for each product. Data on the number of ANDAs filed within each market and the number of ANDAs filed by each manufacturer were derived from the FDA Orange Book. Publicly available data on the FDA website was used to collect whether or not an authorized generic was present in a given market and to determine the identity of the authorized generic manufacturer. In instances where the authorized generic manufacturer name did not match any manufacturer listed in the FDA Orange Book, authorized generic manufacturer identity was determined using branded manufacturer press releases.

Information on whether a 180-exclusivity period was in place was extracted from the Paragraph IV 180-day exclusivity tracker created by the FDA Law Blog.

Sample Selection and Variable Definitions

The initial sample included 144 combined-mol-form markets, which consisted of 108 combined molecules that lost exclusivity between January 2010 and December 2012. To be included in the analysis sample for this paper, a market must have at least two unique non-branded entrants entering in sequence. Markets were also excluded if there was no second entrant within 12 months of the first entrant. Markets where no subsequent entrant was observed for until 12 months after the first entrant tended to be markets with smaller pre-loss of exclusivity sales values or markets where an authorized generic was marketed for a long period of time prior to the entry of an ANDA marketing firm. Moreover, we restricted our data to markets where sales data was continuously observed for 4 quarters prior, and 12 quarters post the start of generic competition. Repackagers were removed from the sample, which is common in the generic drug literature (Berndt, Conti, and Murphy 2017). Repackagers, who do not have FDA authorization to actually manufacture a given generic drug, were excluded by excluding firms that never appear in the FDA Orange Book and are not an authorized generic manufacturer. To address concerns surrounding unobserved systemic differences in firm characteristics between early and late entrants, we further restricted the sample to the firms that entered first or second in a given market. We present results where all manufacturers are included in appendix B. The final sample consisted of 218 mol-form-strength markets across 82 mol-form markets (see Table 2.1 for summary statistics). The final sample is predominantly large-volume oral solid product forms.

Table 2.1 Summary Statistics Table For Included Molecule-Formulation Markets

Summary Statistics				
Indicator Variables	**N= 82**			
Firm and Exclusivity Characersitics	**%**			
New Chemical Entity	0.67			
180-Day Exclusivity Awarded	0.57			
Authorized Generic Present	0.13			
Therapeutic Class				
Cardiovascular Drug	0.22			
Neurological Drug	0.36			
Other Therapeutic Class	0.64			
Product Formulation				
Oral Solid Formulation	0.83			
Injectable Formulation	0.10			
Other Formulation	0.07			
Exclusivity Year				
Lost Exclusivity in 2010	0.27			
Lost Exclusivity in 2011	0.31			
Lost Exclusivity in 2012	0.42			
Continuous Variables	**Mean**	**s.d.**	**min**	**max**
Total Sales in Year Prior to Loss of Exclusivity ($)	70 Million	10 Million	800 Thousand	8 Billion
Size of First Mover Cohort	2.5	1.2	1	16

Outcome and Exposure Variables

We estimate the FMA using two distinct outcomes. In each case, we use the difference in market shares between the first entrant and the predicted market share of a firm that entered in the first year after the start of generic competition. In our main specification, we define a firms' market share using all generic sales accrued from the start of a generic market, defined as the first quarter in which a non-branded firm has positive sales, to four years after the start of competition between different generics. This measure better captures profit over our sample period, given that it will be higher for firms who sell product during the more profitable period where there is limited competition and higher prices. In addition, even in markets where there was no period of limited competition, prices decline rapidly but are still higher in the first year after a product loses exclusivity than in subsequent years where prices reach a more stable minimum (Frank, McGuire, Nason, 2021). To evaluate whether or not a firm's advantage is sustained overtime, we also provide estimates where market shares are calculated using sales restricted to the first, second, third, and fourth year after the start of generic competition, which excludes any sales from periods where competition was limited. Although sustained higher market shares provide incrementally less profit as the market ages and prices drop, evaluating the dynamics of how the FMA varies overtime aids in understanding the mechanisms from which an advantage may occur.

The exposure of interest is an indicator variable equal to one if manufacturer entered a first and zero otherwise. Manufacturer entry into a market was determined by the first quarter in which they were observed to have a positive sales value. Any manufacturer that had a positive sales value in the first quarter in which there were positive sales for a manufacturer other than the brand manufacturer were deemed to be a first mover. As a result, first-mover status is shared

in many cases, and we refer to the collection of first movers as the first-mover cohort. Any manufacturer that entered in a quarter after the first entrant but not longer than four quarters after the first entrant was considered a subsequent entrant. We also provide additional results where first movers are compared to firms that entered third or later that is comparable to previous literature on the FMA.

Confounders

In the main specification of interest, we control for the pre-loss of exclusivity market size defined as the total sales in the year prior to loss of exclusivity, the year of loss of exclusivity, whether or not the mol-form is a new chemical entity, whether or not a 180-day exclusivity period was granted, whether or not the firm is an authorized generic manufacturer, whether or not the market includes an authorized generic competitor, an indicator for whether a product is an oral solid or injectable or other product form, indicators for whether a product is a cardiology drug or neurological drug or other, and adjusted manufacturer fixed effects. The adjusted fixed effects include indicators for the twenty largest firms in our sample defined by the number of markets they participate in and the remaining manufacturers are used as the reference group. We use adjusted fixed effects because of the fact that smaller manufacturers enter relatively few markets in the sample and that we estimate interactions between manufacturer indicators and market characters. In specifications where we only consider sales after the start of generic competition, we further include the duration of time between the first and subsequent entrants as a control.

Empirical Strategy

To formalize the causal structure of the problem, let Y be a firm's generic sales market share over a time period of interest, D be an indicator for whether or not a firm enters first, F be a set of firm characteristics, and X be a set of market characteristics. We are interested in the causal effect of D on Y.

To obtain an estimate of the FMA, we first restrict the sample to firms that entered a market up to one year after the start of competition between generics. We observe that firms who enter any market first are unlikely to enter other markets more than one year after the start of generic competition and that firms who enter later in markets are less likely to ever be a first mover. This observation suggests that there are likely systemic differences between early entrants and firms that enter over one year after the start of competition between generics that result in different optimal entry timings for a given firm, making said firms a poor comparison group. To address any potential observed confounding that may remain after restricting the sample to potential first-movers, we control for a set of market and firm attributes that may impact a firm's likelihood of entering first and their expected market share.

We define the FMA in generic markets over a time period of interest, after accounting for measured confounding, as the expected difference in market share when a firm enters first relative to if they entered later:

$$FMA = E_{X,F}[\, E[Y \mid D = 1, \boldsymbol{X}, \boldsymbol{F}\,] - E[Y \mid D = 0, \boldsymbol{X}, \boldsymbol{F}\,]\,]$$

To estimate the FMA, this work uses a double/debiased machine learning procedure for estimating a fully interactive regression model (Chernozhukov et al., 2018). Double/debiased machine learning is a substitution-based doubly robust estimator for causal and predictive partial effects. This estimator allows for a variety of different machine learning algorithms to be used when estimating nuisance parameters, such as the propensity score. Machine learning algorithms may be able to better capture complex and possibly non-linear relationships among included covariates (Rose, 2013; Chernozhukov et al., 2018). In this setting, these methods can readily model complex interaction between firm indicators and market characteristics that would be difficult to specify a priori. Specifically, we consider ordinary least squares, a well-tuned random forest, and a LASSO model to estimate nuisance parameters. In each case, the algorithm with the best cross-validated predictive performance was chosen for the main estimates presented. Standard errors for each estimator were calculated using the cluster bootstrap procedure outlined in (Cameron and Miller, 2015), clustered at the market level.

In order to interpret this estimate as causal, it is necessary to assume positivity and weak ignorability. Positivity violations refer to when, conditional on the covariates, an observation has a probability of being treated that is not bounded away from 0 or 1. To address practical positivity violations, the propensity score was trimmed, which is common practice with doubly robust methods (Lee, Lessler, and Stuart, 2011). Restricting the sample to first and second movers also reduced the likelihood of positivity violations. As stated previously, many manufacturers that were excluded from the analysis because late entrants were very unlikely to enter first or second, in any market. The weak ignorability assumption, in which treatment is assumed to be as good as randomly assigned conditional on covariates, is inherently untestable. However, the case for ignorability is strengthened by our sample restrictions and the use of a

broad set of covariates, for which we are able to leverage flexible machine learning methods to control for potentially complex confounding relationships. In the following sections we further explore heterogeneity in our main effect estimate and outline a procedure that probes the sensitivity of our estimate to unmeasured confounding. If the weak ignorability assumption is violated, these estimated can be interpreted as the predictive partial effect of being a first mover on firm's expected market share relative to firms that entered in the first year after the start of generic competition.

Heterogeneity

To explore potential heterogeneity in the FMA, we estimate the Conditional Average Treatment Effects (CATE) for different market structures using the methods outlined in Semonova and Chernozhukov (2021). We explore how the effect varies across the size of the first-mover cohort, the presence of an authorized generic, whether or not a firm is marketing a product under an ANDA, whether or not a 180-day exclusivity period was awarded, and whether or not the mol-form is an oral solid. Investigating heterogeneity in this setting is particularly important given that firms enter markets first in a variety of different ways in generic markets because of unique regulatory features, such as 180-day exclusivity periods and the presence of authorized generics. It should also be noted that there is overlap between the subgroups. For example, markets where an exclusivity period is awarded are much more likely to feature a small first-mover cohort and to feature an authorized generic.

Sensitivity to Confounding

To probe the sensitivity of the main estimate to unmeasured confounding, we use the procedure outlined in Cinelli and Hazlett (2019) which does not require assumptions surrounding the functional form of treatment assignment or the distribution of unobserved confounders. Their procedures provide two main estimates: 1) the "robustness value", which determines the strength of unobserved confounding necessary to result in the estimate being statistically insignificant and 2) a way to derive bounds on the size of the causal effect for a given strength of unobserved confounding determined by the researcher. The strength of unobserved confounding is defined as the partial R-squared value of an unobserved confounder that is orthogonal with the observed covariates and has a relationship with the treatment and outcome of interest. For the second estimate, we use the R-squared of the adjusted manufacturer fixed effects, the strongest predictor in our data, regressed on the treatment and outcome as a reasonable upper bound on the impact of a potential unobserved confounder. In order to estimate the baseline treatment effect for their procedure, a well-tuned random forest was used to estimate the nuisance parameters and the main estimate was derived using the partialling-out procedure defined in (Chernozhukov et al., 2018) given that it produced similar estimates to our main results.

Results

Estimates for the basic fixed effects regression and each double debiased machine learning estimator are presented in Table 2.2. The results across all estimators imply a large positive impact of entering a market first relative to entering a market in the first year after the start of competition between generics. Our main estimate, where market shares are calculated using all sales from when the first entrant begins marketing to four years after the start of competition between generics, indicates that firms that enter first are afforded a market share that is 17.2 (SE: 0.013, 95% CI: 15.0 to 19.3) percentage points higher when compared to entering later all else held equal. Estimates that restrict the sample to the period where any manufacturer with an ANDA may enter show that first movers are afforded 10.1 (SE: 1.31, 95% CI: 8.2 to 12.7), 3.50 (SE: 1.2, 95% CI: 1.78 to 6.50), -0.1 (SE: 1.22, 95% CI: -2.64 to 2.91), and -1.71 (SE: 1.39, 95% CI: -4.15 to 1.18) percentage point difference in expected shares in the first, second, third, and fourth year after the start of generic competition, respectively.

Table 2.2 Main Estimates and Estimates By Year of the First Mover Advantage

Estimates of the First Mover Advantage Overall and By Year			
		95% CI*	
Full Sample	**Est.**	**Lower**	**Upper**
Linear Fixed Effects	0.20	0.17	0.22
DML Random Forest	**0.17**	**0.15**	**0.19**
DML Lasso	0.16	0.12	0.19
Year 1			
Linear Fixed Effects	0.13	0.10	0.15
DML Random Forest	**0.10**	**0.08**	**0.13**
DML Lasso	0.09	0.06	0.13
Year 2			
Linear Fixed Effects	0.07	0.04	0.10
DML Random Forest	**0.03**	**0.02**	**0.06**
DML Lasso	0.02	0.00	0.07
Year 3			
Linear Fixed Effects	0.02	-0.01	0.05
DML Random Forest	**0.00**	**-0.03**	**0.03**
DML Lasso	-0.07	-0.08	0.04
Year 4			
Linear Fixed Effects	0.02	-0.02	0.05
DML Random Forest	**-0.02**	**-0.04**	**0.01**
DML Lasso	0.01	-0.03	0.06

Full sample estimates were calculated using all sales from the start of generic competition to four years after the start of competition between generic manufacturers. Estimates for each year were calculated using sales after the start of generic competition. DML: Double Machine Learning. *95% confidence intervals were calculated using a cluster bootstrap procedure with 500 bootstrapped samples.

Heterogeneity

Estimates of the conditional average treatment effects (CATEs) that correspond to our main estimate of the FMA are reported in table 2.3. Estimates of the FMA are greatly attenuated when first-mover status is shared. Sole first movers are afforded additional market shares that are far greater (33% vs 1.9%) than when first-mover status is shared with four or more other manufacturers. Markets with first-mover cohorts that are larger than four manufacturers tend to be markets where no manufacturer marketed a product during a period of limited competition and where there were many simultaneous entrants at the point where exclusivities relevant to the brand expired. First-movers in markets where an authorized generic was present were found to have higher expected market shares, even when only considering non-authorized generic firms. A potential explanation is that authorized generics are often introduced in markets where they will be most profitable. For example, authorized generics in our sample were more likely to be introduced in markets where a 180-day exclusivity period was awarded and they can also benefit from selling a generic during a period where competition is limited, which is consistent with previous studies (Federal Trade Commission, 2011). Authorized generics themselves tend to enjoy larger FMAs than entrants who market product through an ANDA (36% compared to 16%). This is at least partially explained by the fact that authorized generics often enter markets where they can exploit periods of limited competition. Manufacturers that entered first in a market where a 180-day exclusivity period was awarded experience a FMA that is over twice as large as markets with no 180-day exclusivity period (22% vs 10%). Manufacturers entering first in Oral product form markets or in markets of generic drugs that treat neurological disorders have smaller FMA than other product forms or therapeutic classes. These results are consistent with results observed for different market size quintiles. Markets for oral products and for

neurological drugs in our sample tend be larger markets. We observe that the FMA for firms that enter a market in the first quintile of size in terms of total dispensed treatments in the year prior to loss of exclusivity is over twice as large as when a firm enters first in market in the fifth quintile (23.9% vs 9.2%).

Results that focus on the first year after the start of competition between generics display a similar pattern to the heterogeneity observed in the estimates that include sales during periods with limited competition (Table 2.4). However, the FMA tends to decline year over year in each subgroup, and by the fourth year, we only observe a significant FMA in markets where a 180-day exclusivity period was awarded. We further observe that by year four there is potentially a disadvantage to moving first in markets where no exclusivity period was granted and when there is a large first-mover cohort. Firms in markets where an authorized generic was present also had significantly attenuated FMAs over this period when compared to markets where no authorize generic was marketed, particularly in later years. Although authorized generic firms themselves had higher FMAs early on, they do not appear to have significantly higher FMAs advantages after two years of competition between generics, suggesting their advantage is also not sustained overtime.

Table 2.3 Subgroup Estimates of the First Mover Advantage

Subgroup Estimates of the First Mover Advantage			
		95% CI*	
Full Sample	Est.	Lower	Upper
First Mover Cohort Size			
Sole first mover	0.33	0.29	0.37
Two first movers	0.25	0.22	0.28
Three first movers	0.18	0.13	0.23
Four or more first movers	0.02	0.00	0.04
Firm and Exclusivity Characteristics			
Market with no authorized generic	0.16	0.14	0.19
Market with an authorized generic	0.24	0.21	0.27
ANDA Marketing Firm	0.17	0.15	0.19
Authorized Generic Firm	0.36	0.27	0.46
Market with no 180-day exclusivity	0.10	0.07	0.14
Market with a 180-day exclusivity	0.22	0.19	0.24
Product Characteristics			
Oral Product	0.17	0.14	0.19
Injectable Product	0.27	0.16	0.40
Other Product Type	0.24	0.13	0.37
Cardiovascular Drug	0.22	0.19	0.26
Neurological Drug	0.12	0.09	0.16
Other Therapeutic Category	0.20	0.17	0.25
Market Size			
Market Size Quintile 1	0.24	0.18	0.29
Market Size Quintile 2	0.19	0.15	0.27
Market Size Quintile 3	0.19	0.13	0.23
Market Size Quintile 4	0.16	0.12	0.21
Market Size Quintile 5	0.09	0.05	0.14

Estimates were calculated using all sales from the start of generic competition to four years after the start of competition between generic manufacturers within the subgroup of interest. 95% confidence intervals were calculated using a cluster bootstrap procedure with 500 bootstrapped samples.

Table 2.4 Subgroup Estimates of the First Mover Advantage by Year

Subgroup Estimates of the First Mover Advantage by Year												
	Year 1			Year 2			Year 3			Year 4		
Full Sample	Est.	Lower	Upper	Est.	Lower	Upper	Est.	Lower	Upper	Est.	Lower	Upper
First Mover Cohort Size												
Sole first mover	0.14	0.09	0.19	0.11	0.06	0.15	0.06	0.00	0.13	-0.01	-0.07	0.06
Two first movers	0.19	0.16	0.24	0.07	0.02	0.12	0.04	-0.02	0.10	-0.01	-0.05	0.04
Three first movers	0.18	0.12	0.24	-0.08	-0.25	0.07	-0.20	-0.37	-0.05	-0.08	-0.29	0.11
Four or more first movers	0.00	-0.02	0.02	-0.02	-0.04	0.00	-0.05	-0.07	-0.02	-0.03	-0.06	0.00
Firm and Exclusivity Characteristics												
Market with no authorized generic	0.09	0.07	0.12	0.03	0.01	0.06	0.00	-0.03	0.03	-0.01	-0.03	0.02
Market with an authorized generic	0.17	0.12	0.21	0.13	0.04	0.23	-0.01	-0.07	0.05	-0.07	-0.11	-0.03
ANDA Marketing Firm	0.05	0.01	0.08	0.04	0.01	0.06	-0.08	-0.14	-0.03	-0.11	-0.16	-0.06
Authorized Generic Firm	0.13	0.11	0.16	0.04	-0.01	0.09	0.04	0.02	0.07	0.03	0.01	0.06
Market with no 180-day exclusivity	0.10	0.07	0.12	-0.06	-0.10	-0.01	0.00	-0.03	0.02	-0.02	-0.04	0.01
Market with a 180-day exclusivity	0.28	0.19	0.37	0.09	0.07	0.11	-0.02	-0.18	0.16	-0.01	-0.07	0.04
Product Characteristics												
Oral Product	0.11	0.09	0.13	0.04	0.01	0.06	-0.01	-0.04	0.02	-0.02	-0.05	0.00
Injectable Product	0.01	-0.16	0.17	0.00	-0.10	0.11	0.07	-0.01	0.16	0.04	-0.09	0.16
Other Product Type	0.04	-0.15	0.23	0.11	-0.07	0.30	0.09	-0.09	0.26	0.03	-0.08	0.15
Cardiovascular Drug	0.09	0.05	0.13	0.05	0.02	0.09	0.04	0.00	0.08	0.03	-0.01	0.07
Neurological Drug	0.16	0.12	0.20	0.05	0.00	0.10	-0.04	-0.12	0.02	-0.10	-0.16	-0.04
Other Therapeutic Category	0.08	0.05	0.11	0.01	-0.02	0.05	-0.02	-0.06	0.03	0.00	-0.05	0.04
Market Size												
Market Size Quintile 1	0.14	0.06	0.19	0.05	-0.02	0.11	0.05	-0.02	0.12	-0.06	-0.15	0.01
Market Size Quintile 2	0.12	0.08	0.19	0.03	-0.04	0.11	0.00	-0.09	0.08	0.04	-0.04	0.11
Market Size Quintile 3	0.08	0.04	0.13	0.03	-0.01	0.10	-0.03	-0.08	0.05	-0.02	-0.05	0.04
Market Size Quintile 4	0.13	0.08	0.18	0.08	0.02	0.12	0.01	-0.03	0.06	0.01	-0.02	0.04
Market Size Quintile 5	0.04	0.00	0.09	-0.01	-0.06	0.04	-0.06	-0.09	0.00	-0.07	-0.10	-0.02

Estimates for each year were calculated using sales after the start of generic competition. Subgroup estimates were derived from estimates calculated using the double machine learning random forest procedure. *95% confidence intervals were calculated using a cluster bootstrap procedure with 500 bootstrapped samples.

Sensitivity to Confounding

We did not find evidence that our results were sensitive to unobserved confounding. For our main specification, an unobserved confounder orthogonal to the included covariates would have to explain over 23% of residual variation in both the treatment and outcome in order to drive the estimate to be not statistically different from zero. This implies there would have to be an unobserved confounder that has more than double the explanatory power of the adjusted manufacturer fixed effects, which is the strongest observed confounder, explaining approximately 10% of the variation in both the treatment and outcome. Using the partial R-squared of adjusted manufacturer fixed effects as a plausible upper bound on the impact of an unobserved confounded on both the treatment and outcome, we find that plausible lower and upper bounds on the impact of entering first on market share in the first four years after loss of exclusivity are 7 and 27 percentage points, respectively. Furthermore, given that it is unlikely there is a confounder completely orthogonal to the included covariates that is as strong a predictor as the manufacturer fixed effects, we further estimate the bounds on the estimated effect if there are unobserved confounders has half the explanatory power of the manufacturer fixed effects, which are 12 and 22 percentage points. We believe this is convincing evidence that the FMA observed is unlikely to be driven by entirely by endogeneity.

Additional Results

Comparing the results of the doubly robust estimators to a fixed effects regression similar to what was used in previous studies, we find that the 'best' doubly robust estimate is generally smaller than the fixed effect estimates (Appendix Exhibit A1). Moreover, we find that the difference between estimates in terms of the relative effect sizes is larger the further the market is

from the start of competition between generics. Given that we hypothesized that confounding would likely result in effect sizes that are larger than the true effect size, it seems likely that the divergence in results is driven by observed factors that are not properly controlled for in the fixed effects estimates. Moreover, these patterns are consistent with the notion that although there is a large FMA early on, differences in market share that persist are more likely to be attributable to differences between early and late entrants. Comparing the results to estimates where product-form-molecule markets were considered in lieu of product-form-molecule-strength markets did not significantly impact the interpretation of results. Estimates using this alternative market definition were generally smaller in magnitude but not meaningfully different. Unsurprisingly, estimates that compared first entrants to those that enter third or later are larger than those that compare first entrants to those that enter second. However, it is worthwhile to note that the pattern where there is a large FMA that dissipates overtime is still observed.

Discussion

These results have several important implications. First, this work confirms that manufacturers are rewarded with significantly higher market shares when entering first compared to if they had entered later, even when they are not granted a 180-day exclusivity period or the right to market an authorized generic. We also find that firms are granted additional benefit over and above any exclusivity granted in the early stages of the market. However, in contrast to previous studies, this advantage appears to dissipate and is generally gone after two years of generic competition. Given that prices are higher during periods of limited competition and during the earliest stages of competition between generics, higher market shares afforded to entrants in the early stage of a market provide a large incentive to enter a market early, even if advantages are shorter lived than previous studies would suggest.

Previous work that demonstrated the ability of first-movers to obtain sustained higher market shares relative to later entrants suggest that this pattern can be explained by pharmacy unwillingness to incur transactions costs associated with switching manufacturers, patients pressuring pharmacies to carry pharmaceuticals with a given aesthetic, and early firms establishing themselves as a reliable supplier for pharmacies during a period with reduced competition. Our observation that firms are not afforded sustained higher market shares as the market ages call into question explanations that hinge on transaction costs at the pharmacy level. Moreover, given the prominence of several large manufacturers in generic markets, the majority of markets where there is competition between generics will feature multiple large and established manufacturers. It is unlikely that wholesalers have significant preferences for one established generic manufacture over another. It is more likely that previous studies were

conflating systematic differences in the market share of firms that enter early and firms that enter late.

We believe the pattern observed in our study, where firms are afforded a large advantage early on that dissipates rapidly, is more likely to be driven by contracting practices in generic drug markets. In general, wholesalers are likely indifferent between manufacturers of a drug and preferences are likely for the manufacturer that can offer the lowest price for a given drug or menu of drugs. In the early stages of a market, wholesalers have limited options for contracting and manufacturers are able to more easily secure higher market shares over the short term. As the market ages, wholesalers simply prefer the manufacturer that can provide the lowest price.

It is also important to question the degree to which a sustained FMA provides an important incentive for manufacturer entry. In the early stages of a market, prices and the associated profit per unit sold are higher. Thus, additional market share early on in a products life cycle is much more valuable than it is later on. Although studying the dynamics of the FMA is important for understanding potential mechanisms for why it exists, it is unclear how materially important a sustained higher market share is for manufacturers when making entry timing decisions.

The FMA was also heterogenous across important market characteristics. We found that FMAs are generally larger in markets where an authorized generic is present, for both the authorized generic itself and for ANDA marketing firms. This finding could assuage some concern that authorized generics may dampen an important incentive for early entry into a market. Authorized generics certainly reduce the expected share of sales afforded to any first entrant by increasing the size of the first-mover cohort and by potentially reducing the amount of time spent as a sole first mover. Yet, authorized generics generally enter into markets where it is

most profitable to enter early and the remaining potential entrants likely still have a large incentive to do so. These results are consistent with previous work, which highlighted that, in addition to lowering prices for consumers, authorized generics do not generally have a negative impact on the number of competitors or relative shares of different manufacturers in the long run-in a given market (Berndt et al., 2007; Federal Trade Commission, 2011; Appelt, 2015). However, it is important to highlight that our results are focused on markets with higher sales volumes and markets that feature competition between generics, which may not generalize to smaller markets or markets with limited competition.

References

Appelt, S. (2015). Authorized Generic Entry prior to Patent Expiry: Reassessing Incentives for Independent Generic Entry. *The Review of Economics and Statistics*, *97*(3), 654–666. https://doi.org/10.1162/REST_a_00488

Berndt, E. R., & Aitken, M. L. (2011). Brand Loyalty, Generic Entry and Price Competition in Pharmaceuticals in the Quarter Century after the 1984 Waxman-Hatch Legislation. *International Journal of the Economics of Business*, *18*(2), 177–201. https://doi.org/10.1080/13571516.2011.584423

Berndt, E. R., Mortimer, R., Bhattacharjya, A., Parece, A., & Tuttle, E. (2007). Authorized Generic Drugs, Price Competition, And Consumers' Welfare. *Health Affairs*, *26*(3), 790–799. https://doi.org/10.1377/hlthaff.26.3.790

Berndt, E. R., Conti, R., & Murphy, S. (2017). *The Landscape of Us Generic Prescription Drug Markets, 2004-2016* (SSRN Scholarly Paper ID 3011139). Social Science Research Network. https://papers.ssrn.com/abstract=3011139

Bokhari, F. A. S., Mariuzzo, F., & Polanski, A. (2020). Entry limiting agreements: First-mover advantage, authorized generics, and pay-for-delay deals. *Journal of Economics & Management Strategy*, *29*(3), 516–542. https://doi.org/10.1111/jems.12351

Cameron, A. C., & Miller, D. L. (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, *50*(2), 317–372. https://doi.org/10.3368/jhr.50.2.317

Cha, M., & Yu, F. (n.d.). *Pharma's first-to-market advantage | McKinsey*. Retrieved November 6, 2021, from https://www.mckinsey.com/industries/life-sciences/our-insights/pharmas-first-to-market-advantage

Cinelli, C., & Hazlett, C. (2020). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *82*(1), 39–67. https://doi.org/10.1111/rssb.12348

Congressional Research Service. (2020). *Drug Pricing and Pharmaceutical Patenting Practices*. https://www.everycrsreport.com/reports/R46221.html

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, *21*(1), C1–C68. https://doi.org/10.1111/ectj.12097

Elias, M., Monique, M., & Tom, W. (2004). *Regulating Pharmaceuticals In Europe: Striving For Efficiency, Equity And Quality: Striving for Efficiency, Equity and Quality*. McGraw-Hill Education (UK).

FDA. (2022). FDA List of Authorized Generic Drugs. https://www.fda.gov/drugs/abbreviated-new-drug-application-anda/fda-list-authorized-generic-drugs

FDA Law Blog. (n.d.). FDA Law Blog. Retrieved November 6, 2021, from https://www.thefdalawblog.com/

Federal Trade Commission. (2011). *Authorized Generic Drugs: Short-Term Effects and Long-Term Impact*.

Fein, A. J. *Six Crucial Trends Facing U.S. Drug Wholesalers*. Retrieved November 6, 2021, from https://www.drugchannels.net/2020/12/six-crucial-trends-facing-us-drug.html

Fein, A. J. *The Big Three Generic Drug Mega-Buyers Drove Double-Digit Deflation in 2018. Stability ahead?* Retrieved November 6, 2021, from https://www.drugchannels.net/2019/01/the-big-three-generic-drug-mega-buyers.html

Frank, R. G., & Salkever, D. S. (1997). Generic Entry and the Pricing of Pharmaceuticals. *Journal of Economics & Management Strategy*, *6*(1), 75–90. https://doi.org/10.1111/j.1430-9134.1997.00075.x

Frank, R. G., McGuire, T. G., & Nason, I. (2021). The Evolution of Supply and Demand in Markets for Generic Drugs. *The Milbank Quarterly*, *99*(3), 828–852. https://doi.org/10.1111/1468-0009.12517

Frank, R. G., Shahzad, M., Kesselheim, A. S., & Feldman, W. (2022). Biosimilar competition: Early learning. *Health Economics*, *31*(4), 647–663. https://doi.org/10.1002/hec.4471

Generic Entry and the Pricing of Pharmaceuticals—Frank—1997—Journal of Economics & Management Strategy—Wiley Online Library. (n.d.). Retrieved March 26, 2022, from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1430-9134.1997.00075.x

Grabowski, H. G., & Vernon, J. M. (1992). Brand Loyalty, Entry, and Price Competition in Pharmaceuticals after the 1984 Drug Act. *The Journal of Law & Economics*, *35*(2), 331–350.

Hemphill, C. S., & Lemley, M. A. (2011). EARNING EXCLUSIVITY: GENERIC DRUG INCENTIVES AND THE HATCH-WAXMAN ACT. *Antitrust Law Journal*, *77*(3), 947–989.

Hollis, A. (2002). The importance of being first: Evidence from Canadian generic pharmaceuticals. *Health Economics*, *11*(8), 723–734. https://doi.org/10.1002/hec.698

IBIS World Report. (2018). *Generic Pharmaceutical Manufacturing in the US*. https://my-ibisworld-com.ezp-prod1.hul.harvard.edu/us/en/industry/32541b/about

IQVIA. (2019). *Medicine Use and Spending in the U.S.*

Lee, B. K., Lessler, J., & Stuart, E. A. (2011). Weight Trimming and Propensity Score Weighting. *PLoS ONE*, *6*(3), e18174. https://doi.org/10.1371/journal.pone.0018174

Lieberman, M. B., & Montgomery, D. B. (2013). Conundra and Progress: Research on Entry Order and Performance. *Long Range Planning*, *46*(4), 312–324. https://doi.org/10.1016/j.lrp.2013.06.005

Milne, C.-P., & Cairns, C. (2003). Generic Drug Regulation in the US Under the Hatch-Waxman Act. *Pharmaceutical Development and Regulation*, *1*(1), 11–27. https://doi.org/10.1007/BF03257362

Peelish, N. (n.d.). *Antitrust and Authorized Generics*. Stanford Law Review. Retrieved April 7, 2022, from https://www.stanfordlawreview.org/print/article/antitrust-and-authorized-generics/

Regnier, S., & Ridley, D. B. (2015). *Forecasting Market Share in the US Pharmaceutical Market* (SSRN Scholarly Paper ID 2763659). Social Science Research Network. https://papers.ssrn.com/abstract=2763659

Reiffen, D., & Ward, M. R. (2005). Generic Drug Industry Dynamics. *The Review of Economics and Statistics*, *87*(1), 37–49. https://doi.org/10.1162/0034653053327694

Research, C. for D. E. and. (2021). FDA List of Authorized Generic Drugs. *FDA*. https://www.fda.gov/drugs/abbreviated-new-drug-application-anda/fda-list-authorized-generic-drugs

Rose, S. (2013). Mortality Risk Score Prediction in an Elderly Population Using Machine Learning. *American Journal of Epidemiology*, *177*(5), 443–452. https://doi.org/10.1093/aje/kws241

Scott Morton, F. (1999). *Entry Decisions in the Generic Pharmaceutical Industry* (SSRN Scholarly Paper ID 164431). Social Science Research Network. https://papers.ssrn.com/abstract=164431

Scott Morton, F., & Kyle, M. (2011). Chapter Twelve—Markets for Pharmaceutical Products11The authors thank Tom McGuire for editorial advice and improvements and participants at the Handbook conference for helpful suggestions. In M. V. Pauly, T. G. Mcguire, & P. P. Barros

(Eds.), *Handbook of Health Economics* (Vol. 2, pp. 763–823). Elsevier. https://doi.org/10.1016/B978-0-444-53592-4.00012-8

Semenova, V., & Chernozhukov, V. (2021). Debiased machine learning of conditional average treatment effects and other causal functions. *The Econometrics Journal*, *24*(2), 264–289. https://doi.org/10.1093/ectj/utaa027

Shajarizadeh, A., Grootendorst, P., & Hollis, A. (2015). Newton's First Law as Applied to Pharmacies: Why Entry Order Matters for Generics. *International Journal of the Economics of Business*, *22*(2), 201–217. https://doi.org/10.1080/13571516.2015.1045746

US Food & Drug Administration. (2017). *Approved Drug Products with Therapeutic Equivalence Evaluations.* US Department of Health and Human Services.

Yu, Y., & Gupta, S. (2008). *Pioneering Advantage in Generic Drug Competition* (SSRN Scholarly Paper ID 925346). Social Science Research Network. https://doi.org/10.2139/ssrn.925346

Chapter 3:

Decline in New Starts of Psychotropic Medications During The COVID-19 Pandemic

Written with

Dorit Stein

Richard Frank

Murray Stein

Formatted for review at Health Affairs

Abstract

COVID-19 interrupted delivery of mental health care in the US. Symptoms associated with various mental disorders increased in prevalence at the same time. The expectation is that treatment would increase with measured need. Departures from that expectation serve to index the degree of disruption in the delivery of mental health care to the US population. We conducted a retrospective observational analysis using prescription claims data covering 89 percent of all prescriptions in the US that compared observed new-starts of common psychotropic medications to forecasted new-starts. Forecasts were generated using the Prophet forecasting model. During the initial course of the COVID-19 pandemic new starts of antidepressants declined by 7.5 percent, anxiolytics by 5.6 percent, and antipsychotics by 2.6 percent compared with expected levels. Declines were more pronounced among children and adolescents, with declines in new starts ranging from 20 to 30% over the same period for the three drug classes (ages 0 to 18). Our findings suggest that there was a large unmet need for mental health treatment in the US attributable to COVID-19 over this period.

Introduction

The coronavirus disease 2019 (COVID-19) pandemic and the associated economic and social shocks have resulted in a spike in mental health conditions and disrupted delivery of mental health care in the US. The prevalence of symptoms related to anxiety disorder increased from 8.1 percent in the second quarter of 2019 to 25.5 percent in June 2020 and 36.9 percent in December 2020 (Czeisler et al., 2020; Vahratian et al., 2021). The prevalence of major depressive disorder symptoms increased from 6.5 percent to 24.3 percent to 30.2 percent over the same period (Czeisler et al., 2020; Vahratian et al., 2021). The pandemic also curtailed contact with health care professionals at exactly the time that stressors increased (Mehrotra et al., 2020). This may have limited the ability of the health care delivery system to respond to the treatment needs of the population.

In this analysis we found significant declines in the provision of psychotropic prescriptions over the first five months of the pandemic. For example, there were approximately 14.4 percent (597199) fewer new starts of antidepressants over the initial lockdown period (March to May 2020) and 2.2 percent (114880) fewer from May to August 2020 (exhibit 1). This is consistent with reports that outpatient visits to all providers declined dramatically at the beginning of the pandemic and rebounded, while remaining below baseline levels across many specialties, as states lifted pandemic restrictions. This pattern was especially true for visits to behavioral health providers. Total behavioral health visits, including increases in telehealth visits, were 15 percent lower than the pre-pandemic baseline at the end of July 2020 (Mehrotra et al., 2020). Beyond the direct impact of the pandemic, loss or shifts in insurance may have played some role in disrupting access to mental health care. Between March and September, the Kaiser Family Foundation estimates that between two and three million individuals lost employer

sponsored health insurance and that Medicaid enrollment increased by roughly 4.3 million people (McDermott et al., 2020).

We looked at the provision of psychotropic prescriptions to the US population during COVID-19 relative to one measure of the pre-COVID-19 norm. We used data on new prescriptions for psychotropic medications to measure the change in new starts of antidepressants, anxiolytics, and antipsychotics from March 13th to August 8th, 2020. This change in new prescription starts is an indicator of whether the mental health care delivery system increased the provision of pharmacotherapy in the face of increased population need (Czeisler et al., 2020; Vahratian et al., 2021). The expectation is that treatment would increase with measured need. Departures from that expectation serve to index the degree of disruption in the delivery of mental health care to the US population.

Methods

We conducted a retrospective observational analysis using IQVIA Longitudinal Prescription Claims data collected from January 1st, 2018 to August 8th, 2020. The IQVIA claims represent 89 percent of all prescriptions from retail, mail, and long-term care pharmacies in the US. Claims cover prescription purchases across all payers, such as Medicare, Medicaid, Employer and out-of-pocket. The data was provided by IQVIA to support health systems research on the indirect impacts of COVID-19.

The number of new prescription starts in a given week was defined as the total number of individuals who filled a psychotropic prescription and did not receive a prescription for a psychotropic medicine in the same therapeutic class in the previous three months. Using this definition, we compared the observed new starts in 2020 to a forecast of new starts estimated using Prophet, a method for time series forecasting (Taylor and Letham, 2018). The forecast model was developed using data prior to March 2020, and the final models were chosen using cross-validation. Our primary outcome was the cumulative sum of the difference between forecasted new starts in a given week and the observed number of new starts in that week. This comparison is likely an underestimate of the gap in new starts because our forecasting model does not take account of the increase in symptom prevalence.

To understand potential drivers of changes in new starts, we repeated this exercise within key sub-populations defined by age, gender, and treating clinician specialty (that is, primary care versus mental health specialty). In the online appendix, we also provide comparisons to observed new-starts in 2019, which can be interpreted as a lower bound on the expected number of prescriptions given that there is no adjustment for increases in prescriptions over time (See Chapter 3 Appendix).

Limitations

This analysis has several limitations. The estimated counterfactual level of prescriptions is based on pre-pandemic data and does not account for the increased incidence of symptoms associated with psychiatric disorders attributable to COVID-19. This likely results in an underestimate of the gap between new prescription starts and the optimal level of treatment needed to respond to the increased incidence. Moreover, medications are only one form of treatment and we are unable to comment on whether or not individuals were more likely to substitute other forms of treatment during the pandemic. The pandemic brought a dramatic shift toward telemedicine, particularly among psychiatric specialists, and it is unclear how this impacted the composition of mental health care (Haque, 2021). Telehealth likely prevented even larger declines in unmet need than observed. Data were only available from March until August 2020, however, over the subsequent period, the pandemic reached unprecedented levels of new cases and hospitalizations. This likely resulted in further declines in new starts of psychotropic medications beyond those observed from March through August 2020. CDC pulse survey estimates suggest that the prevalence of depression and anxiety symptoms increased over the period from August 2020 to February 2021 alongside unmet need for mental health treatment (Vahration et al., 2021).

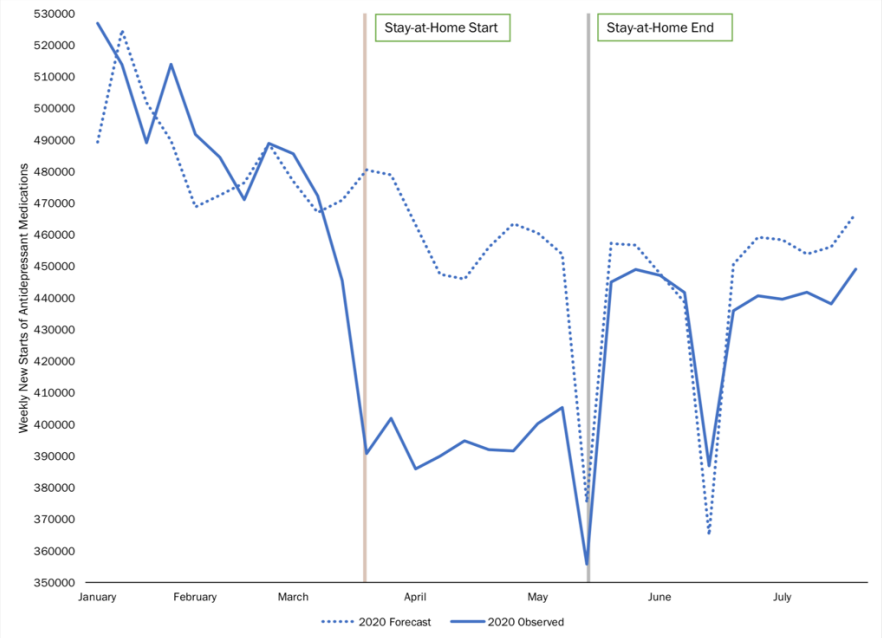
Results

New starts of all psychotropic medications—antidepressants, anxiolytics, and antipsychotics--fell dramatically during the initial stay-at-home order period, which began on March 13, 2020 and ended in mid to late May 2020 for most states (figure 3.1, 3.2, 3.3).8There were approximately 7.5 percent (712079) fewer new starts of antidepressants (table 3.1), 5.6 percent (465610) fewer new starts of anxiolytics (table 3.2), and 2.6 percent (474670) fewer new starts of antipsychotics (table 3.3) between March 13, 2020 and August 8, 2020 compared to the forecasted levels. The majority of the gap in new starts accrued in March and April 2020, but despite a substantial rebound, new starts remained below both forecasted and 2019 levels by the end of our sample period.

For all medications, declines in new starts were particularly pronounced for those under the age of eighteen. Compared to the forecast, there was a 34.6 percent reduction in antidepressant, 27.3 percent reduction in anxiolytic, and 22.2 percent reduction in antipsychotic new starts among those under the age of eighteen. Notably, new starts for individuals above age eighteen were much closer to expected levels during the May to August period compared to those below the age of eighteen.

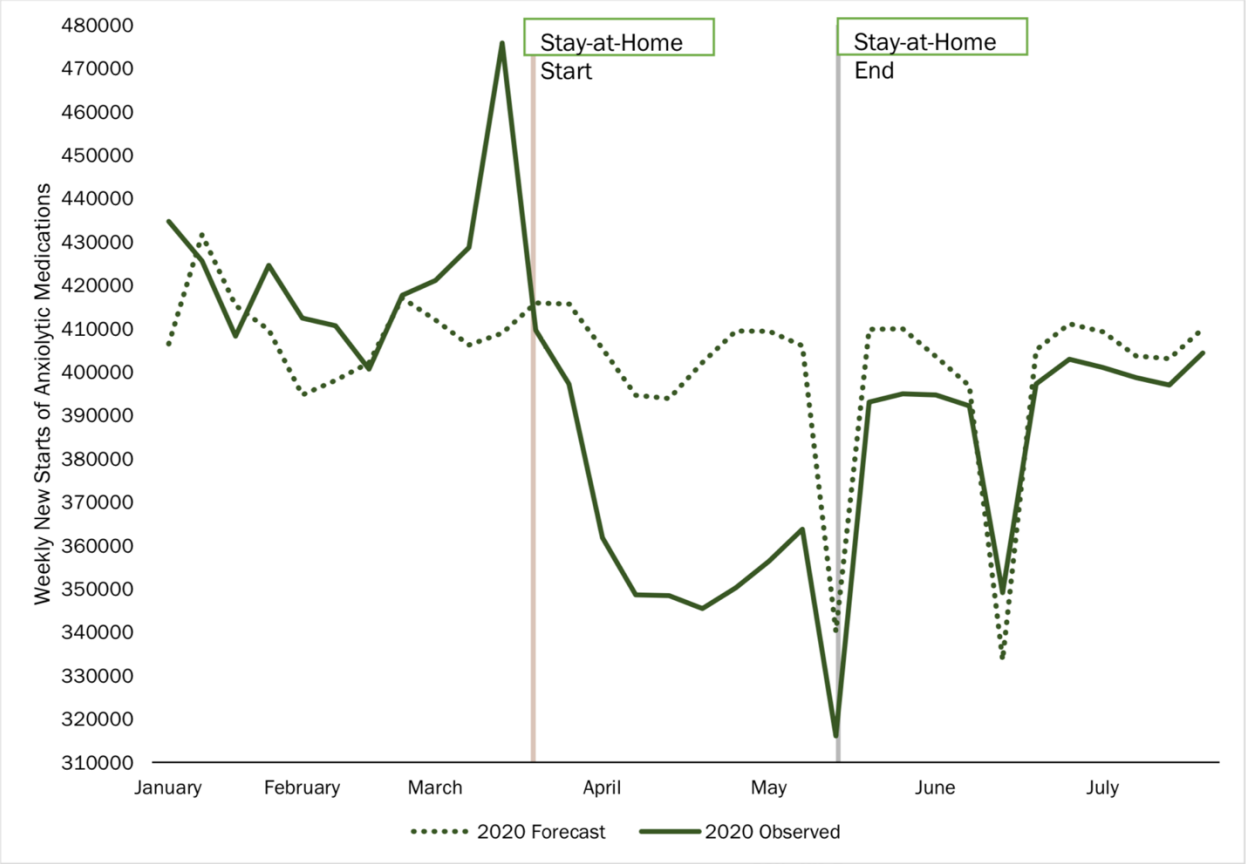
Reductions in new prescription starts tended to be greatest for males and although there were no clear patterns across therapeutic classes, non-physician prescribers tended to show larger declines when compared to other provider types. Similar, but muted patterns were observed when we compared new starts over the same period in 2020 to 2019 for the whole sample and in the majority of subgroups (See Chapter 3 Appendix).

Figure 3.1 Observed and forecasted new starts of antidepressant medications in 2020



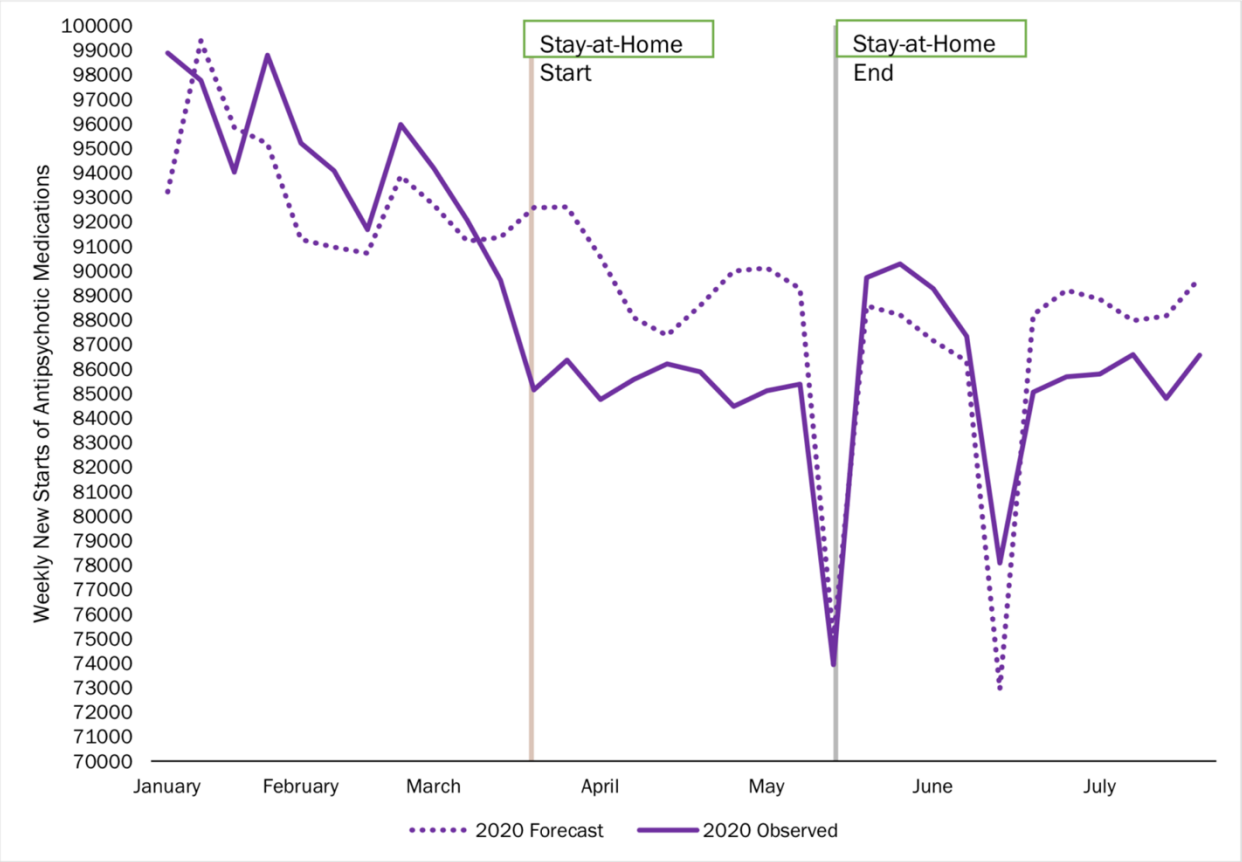
Notes Authors' analysis of data from IQVIA Longitudinal Prescription Data, 2020. Weekly new starts of antidepressant medications from January 1 to August 8. The shaded area represents the start (March 13, 2020) and approximate period that marked the end (May 15, 2020) of stay-at-home orders. Sharp changes in the trend reflect holiday weekends, where prescriptions tend to be significantly lower than the average day.

Figure 3.2 Observed and forecasted new starts of anxiolytic medications in 2020



Notes Authors' analysis of data from IQVIA Longitudinal Prescription Data, 2020. Weekly new starts of anxiolytic medications from January 1 to August 8. The shaded area represents the start (March 13, 2020) and approximate period that marked the end (May 15, 2020) of stay-at-home orders. Sharp changes in the trend reflect holiday weekends, where prescriptions tend to be significantly lower than the average day.

Exhibit 3.3 - Observed and forecasted new starts of antipsychotic medications in 2020



Notes Authors' analysis of data from IQVIA Longitudinal Prescription Data, 2020. Weekly new starts of antipsychotic medications from January 1st to August 8th. The shaded area represents the start (March 13, 2020) and approximate period that marked the end (May 15, 2020) of stay-at-home orders. Sharp changes in the trend reflect holiday weekends, where prescriptions tend to be significantly lower than the average day.

Table 3.1 Total cumulative difference in new starts of antidepressant medications for March 13–May 15, May 15–August 8, and March 13–August 8, 2020, compared with 2020 forecast overall and by subgroup

	March 13–May 15		**May 15–August 8**		**March 13–August 8**	
	N	**%**	**N**	**%**	**N**	**%**
Full sample	−597,199	−14.4	−114,880	−2.2	−712,079	−7.5
Age (years)						
0–18	−164,218	−43.0	−131,369	−27.8	−295,587	−34.6
18–30	−68,876	−9.3	16,491	1.7	−52,385	−3.1
30–65	−210,741	−9.7	−8,996	−0.3	−219,737	−4.4
65+	−158,905	−18.6	−15,070	−1.4	−173,975	−8.9
Sex						
Male	−271,877	−18.3	−120,418	−6.3	−392,295	−11.6
Female	−317,272	−12.0	12,680	0.4	−304,592	−5.0
Provider type						
Psychiatrist/CAP	−43,092	−8.9	−15,089	−2.6	−58,182	−5.4
Family medicine	−142,756	−13.1	2,483	0.2	−140,273	−5.7
NP/PA/other	−220,262	−16.9	−75,863	−4.5	−296,125	−9.9
Other physician specialist	−181,121	−14.3	−92,07	−0.6	−190,328	−6.6

Notes Authors' analysis of data from IQVIA Longitudinal Prescription Claims Data, 2020. The cumulative difference in new starts was calculated by taking the difference in cumulative observed new starts and cumulative forecasted new starts. Percentage change was calculated as the percentage difference in cumulative observed new starts and cumulative forecasted new starts. Individual forecasts were estimated for each subgroup. CAP is child and adolescent psychiatry. NP is nurse practitioner. PA is physician assistant. Other is nonphysician specialist.

Table 3.2 Total cumulative difference in new starts of anxiolytic medications for March 13–May 15, May 15–August 8, and March 13–August 8, 2020, compared with 2020 forecast overall and by subgroup

	March 13–May 15		**May 15–August 8**		**March 13–August 8**	
	N	**%**	**N**	**%**	**N**	**%**
Full sample	−371,021	−10.2	−94,589	−2.0	−465,610	−5.6
Age (years)						
0–18	−100,968	−38.1	−65,308	−19.0	−166,277	−27.3
18–30	−35,776	−7.5	19,559	3.0	−16,217	−1.4
30–65	−97,690	−4.9	−39,339	−1.5	−137,029	−3.0
65+	−140,020	−15.2	−21,131	−1.8	−161,152	−7.7
Sex						
Male	−181,662	−14.6	−62,085	−3.9	−243,747	−8.6
Female	−191,465	−7.9	−34,517	−1.1	−225,982	−4.1
Provider type						
Psychiatrist/CAP	−7,177	−2.5	−13,556	−3.7	−20,733	−3.2
Family medicine	7,028	0.8	−6,407	−0.6	621	0.0
NP/PA/other	−100,509	−9.5	−55,196	−3.9	−155,706	−6.3
Other physician specialist	−278,919	−18.8	−40,120	−2.1	−319,039	−9.4

Notes Authors' analysis of data from IQVIA Longitudinal Prescription Claims Data, 2020.The cumulative difference in new starts was calculated by taking the difference in cumulative observed new starts and cumulative forecasted new starts. Percent change was calculated as the percentage difference in cumulative observed new starts and cumulative forecasted new starts. Individual forecasts were estimated for each subgroup. CAP is child and adolescent psychiatry. NP is nurse practitioner. PA is physician assistant. Other is nonphysician specialist.

Table 3.3 Total cumulative difference in new starts of antipsychotic medications for March 13–May 15, May 15–August 8, and March 13–August 8, 2020 compared to 2020 forecast overall and by subgroup

	March 13–May 15		**May 15–August 8**		**March 13–August 8**	
	N	**%**	**N**	**%**	**N**	**%**
Full sample	−40,295	−5.0	−7,172	−0.7	−47,467	−2.6
Age (years)						
0–18	−25,455	−27.7	−18,274	−17.3	−43,729	−22.2
18–30	−1,318	−0.9	9,254	4.9	7,936	2.4
30–65	−7,639	−1.9	−6,994	−1.3	−14,633	−1.6
65+	−7,217	−4.3	7,272	3.4	56	0.0
Sex						
Male	−25,784	−7.2	−10,485	−2.3	−36,270	−4.4
Female	−14,047	−3.1	3,859	0.7	−10,188	−1.0
Provider type						
Psychiatrist/CAP	−16,624	−6.3	−4,317	−1.3	−20,941	−3.6
Family medicine	−5,615	−5.5	662	0.5	−4,953	−2.1
NP/PA/Other	−17,818	−6.3	−15,528	−4.2	−33,346	−5.1
Other physician specialist	−2,544	−1.5	8,353	3.9	5,809	1.5

Notes Authors' analysis of data from IQVIA Longitudinal Prescription Claims Data, 2020. The cumulative difference in new starts was calculated by taking the difference in cumulative observed new starts and cumulative forecasted new starts. Percent change was calculated as the percentage difference in cumulative observed new starts and cumulative forecasted new starts. Individual forecasts were estimated for each subgroup. CAP is child and adolescent psychiatry. NP is nurse practitioner. PA is physician assistant. Other is nonphysician specialist.

Discussion

We documented a significant decline in new starts of antidepressant, anxiolytic, and antipsychotic medications over the initial five-months of the COVID-19 pandemic. There was a significant rebound in new-prescription-starts after the end of stay-at-home orders in March 2020, however new-starts remained below expected levels across all therapeutic classes. Decreased treatment initiation for mental illnesses could be one factor contributing to increased emergency department visits for suicide and overdose in 2020 compared to 2019 (Holland et al., 2021). There was also considerable variation in the decline of new prescription starts across sub-populations. Substantial declines in new starts were observed for individuals under the age of eighteen in all medication classes. Since it is estimated that more than half of school-aged children who use mental health services receive some services in the school setting and more than one-third receive mental health services exclusively in the school setting, it is possible that school closures over the study period limited the opportunity to identify children in need of mental health care or otherwise resulted in limited access to care, especially for children from disadvantaged populations (Ali et al., 2019). The role of school as an access point for mental health care may partially explain why new prescription starts remained well below expected levels for children, while adults experienced levels of treatment much closer to what was expected by August 2020. Declines in new starts could also be a result of fewer outpatient mental health services being accessed among children (Centers for Medicare and Medicaid Services, 2020). This treatment gap may be contributing to the increased share of hospital emergency department visits that were for mental health needs of children between April and October 2020 (Leeb et al., 2020). This dramatic reduction in new prescription starts being provided to children and adolescents is particularly worrisome and bears further scrutiny.

Although health care use is rebounding as states ease restrictions and the number of vaccinations increases, there is still likely a large unmet need for mental health treatment in the US. Our findings suggest that numerous individuals have forgone or are currently forgoing treatment for mental health conditions. Providers and policy makers must work to increase access to treatment for psychiatric disorders, in addition to addressing the underlying causes of poor mental health outcomes during the pandemic.

References

Ali, M. M., West, K., Teich, J. L., Lynch, S., Mutter, R., & Dubenitz, J. (2019). Utilization of Mental Health Services in Educational Setting by Adolescents in the United States. *The Journal of School Health*, *89*(5), 393–401. https://doi.org/10.1111/josh.12753

Czeisler, M. É., Lane, R. I., Petrosky, E., Wiley, J. F., Christensen, A., Njai, R., Weaver, M. D., Robbins, R., Facer-Childs, E. R., Barger, L. K., Czeisler, C. A., Howard, M. E., & Rajaratnam, S. M. W. (2020). Mental Health, Substance Use, and Suicidal Ideation During the COVID-19 Pandemic—United States, June 24-30, 2020. *MMWR. Morbidity and Mortality Weekly Report*, *69*(32), 1049–1057. https://doi.org/10.15585/mmwr.mm6932a1

Fact Sheet: Service Use among Medicaid & CHIP Beneficiaries age 18 and Under during COVID-19 | CMS. (n.d.). Retrieved April 5, 2022, from https://www.cms.gov/newsroom/fact-sheets/fact-sheet-service-use-among-medicaid-chip-beneficiaries-age-18-and-under-during-covid-19

Haque, S. N. (2021). Telehealth Beyond COVID-19. *Psychiatric Services (Washington, D.C.)*, *72*(1), 100–103. https://doi.org/10.1176/appi.ps.202000368

Holland, K. M., Jones, C., Vivolo-Kantor, A. M., Idaikkadar, N., Zwald, M., Hoots, B., Yard, E., D'Inverno, A., Swedo, E., Chen, M. S., Petrosky, E., Board, A., Martinez, P., Stone, D. M., Law, R., Coletta, M. A., Adjemian, J., Thomas, C., Puddy, R. W., … Houry, D. (2021). Trends in US Emergency Department Visits for Mental Health, Overdose, and Violence Outcomes Before and During the COVID-19 Pandemic. *JAMA Psychiatry*, *78*(4), 372–379. https://doi.org/10.1001/jamapsychiatry.2020.4402

How Has the Pandemic Affected Health Coverage in the U.S.? (2020, December 9). *KFF*. https://www.kff.org/policy-watch/how-has-the-pandemic-affected-health-coverage-in-the-u-s/

Leeb, R. T. (2020). Mental Health–Related Emergency Department Visits Among Children Aged 18 Years During the COVID-19 Pandemic—United States, January 1–October 17, 2020. *MMWR. Morbidity and Mortality Weekly Report*, *69*. https://doi.org/10.15585/mmwr.mm6945a3

Moreland, A. (2020). Timing of State and Territorial COVID-19 Stay-at-Home Orders and Changes in Population Movement—United States, March 1–May 31, 2020. *MMWR. Morbidity and Mortality Weekly Report*, *69*. https://doi.org/10.15585/mmwr.mm6935a2

Taylor, S. J., & Letham, B. (2018). Forecasting at Scale. *The American Statistician*, *72*(1), 37–45. https://doi.org/10.1080/00031305.2017.1380080

The Impact of the COVID-19 Pandemic on Outpatient Care: Visits Return to Prepandemic Levels, but Not for All Providers and Patients. (2020, October 15). https://doi.org/10.26099/41xy-9m57

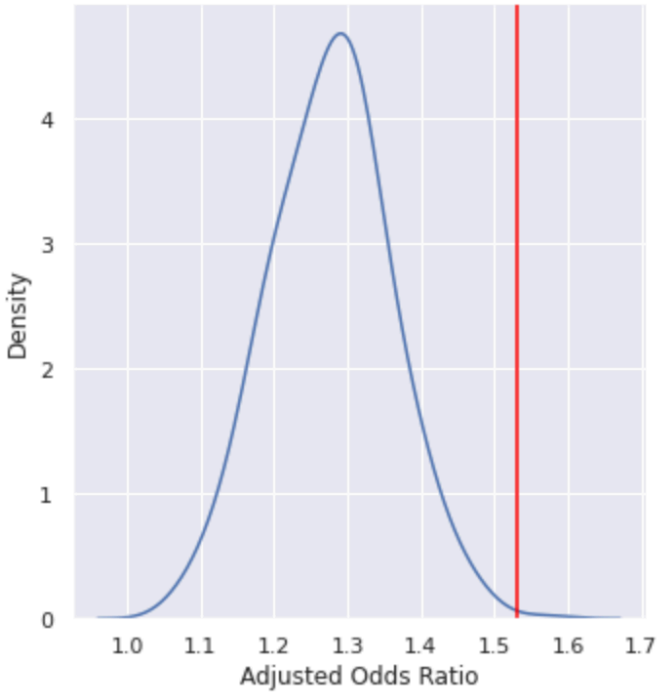
Vahratian, A., Blumberg, S. J., Terlizzi, E. P., & Schiller, J. S. (2021). Symptoms of Anxiety or Depressive Disorder and Use of Mental Health Care Among Adults During the COVID-19 Pandemic—United States, August 2020-February 2021. *MMWR. Morbidity and Mortality Weekly Report*, *70*(13), 490–494. https://doi.org/10.15585/mmwr.mm7013e2

Appendices

Appendices for Chapter 1

Exhibits

Exhibit A1 Distribution of Simulated Adjusted Odds Ratios of Experiencing Any Fracture



Appendices for Chapter 2

Exhibits

Exhibit A1 Title

Comparison of Estimates of the First Mover Advantage by Different Market and Treatment Definition								
Treatment Definition	**Comparison of First Movers to Second Entrants**				**Comparison of First Movers to Third or Later Entrants**			
Marker Defintion	**Product Form Molecule Strength**		**Product Form Molecule**		**Product Form Molecule Strength**		**Product Form Molecule**	
Full Sample	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**
Linear Fixed Effects	0.20	0.01	0.16	0.12	0.22	0.01	0.18	0.01
DML Random Forest	0.17	0.01	0.14	0.01	0.22	0.01	0.18	0.01
DML Lasso	0.16	0.02	0.17	0.03	0.22	0.01	0.21	0.02
Year 1								
Linear Fixed Effects	0.13	0.15	0.11	0.02	0.15	0.01	0.13	0.02
DML Random Forest	0.10	0.13	0.10	0.02	0.14	0.01	0.13	0.02
DML Lasso	0.09	0.13	0.14	0.03	0.13	0.01	0.16	0.03
Year 2								
Linear Fixed Effects	0.07	0.02	0.06	0.02	0.15	0.01	0.09	0.02
DML Random Forest	0.03	0.01	0.06	0.02	0.14	0.01	0.09	0.02
DML Lasso	0.02	0.02	0.12	0.07	0.14	0.01	0.14	0.05
Year 3								
Linear Fixed Effects	0.02	0.02	0.06	0.02	0.07	0.01	0.08	0.02
DML Random Forest	0.00	0.01	0.05	0.02	0.06	0.01	0.08	0.02
DML Lasso	-0.07	0.03	0.06	0.02	0.02	0.02	0.08	0.03
Year 4								
Linear Fixed Effects	0.02	0.02	0.06	0.03	0.04	0.01	0.07	0.02
DML Random Forest	-0.02	0.01	0.02	0.02	0.05	0.01	0.07	0.02
DML Lasso	0.01	0.02	0.02	0.04	0.05	0.01	0.04	0.05

Appendices for Chapter 3

Forecasting Approach

Previous studies used data from 2019 as a point of comparison for health systems metrics that are potentially impacted by the Covid-19 pandemic. However, given that the number of psychotropic prescriptions across all therapeutic classes in this analysis is increasing overtime, this would result in a substantial underestimate of the gap in new prescription starts. To forecast the number of new prescriptions starts we use Prophet, a method that leverages an additive model to flexibly fit non-linear trends that are seasonal and impacted by holidays. This flexibility is necessary for prescription data, given that the trend depends greatly on the time of year and there are substantial dips in the number of prescriptions on national holidays. The flexibility of the forecast is governed by several tuning parameters, such as the degree of seasonality and the impact of holidays. Given that higher degrees of flexibility may result in overfitting, the model was tuned using time-series cross-validation. Parameters were chosen such that the cross validated forecast error was minimized. The tuning process used data prior to March 2020.

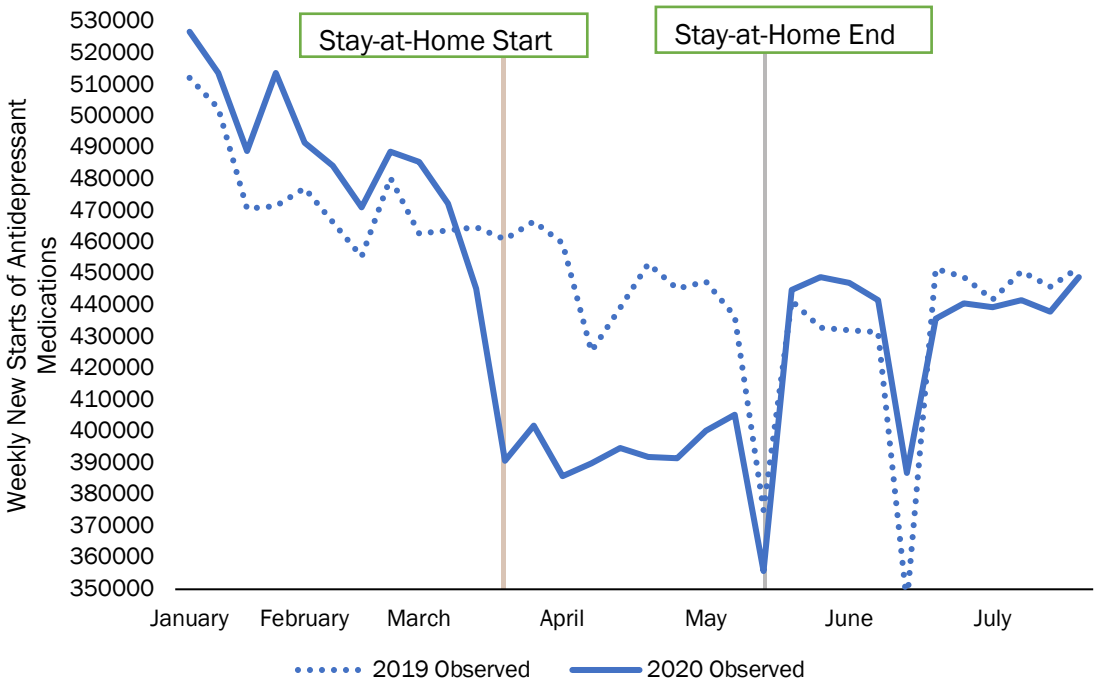
Comparison To 2019

To inspect the robustness of the observed declines in new-starts we further compare the observed number of new starts in 2020 to the observed number in 2019. Given that over a 'normal' period the number new starts increase overtime, a comparison to 2019 provides a convenient lower bound on the predicted number of prescriptions. Exhibits A1, A2, and A3 show that prior to March 2020 the number of new starts across all classes was strictly higher in 2020 than in 2019. Furthermore, Exhibits A4, A5, A6 replicate the tables from the main text. When compared to 2019, we see that there were still large declines in new-starts across therapeutic

classes, except for anti-psychotics. Those under the age of 18 continue to have the largest relative declines across all classes.

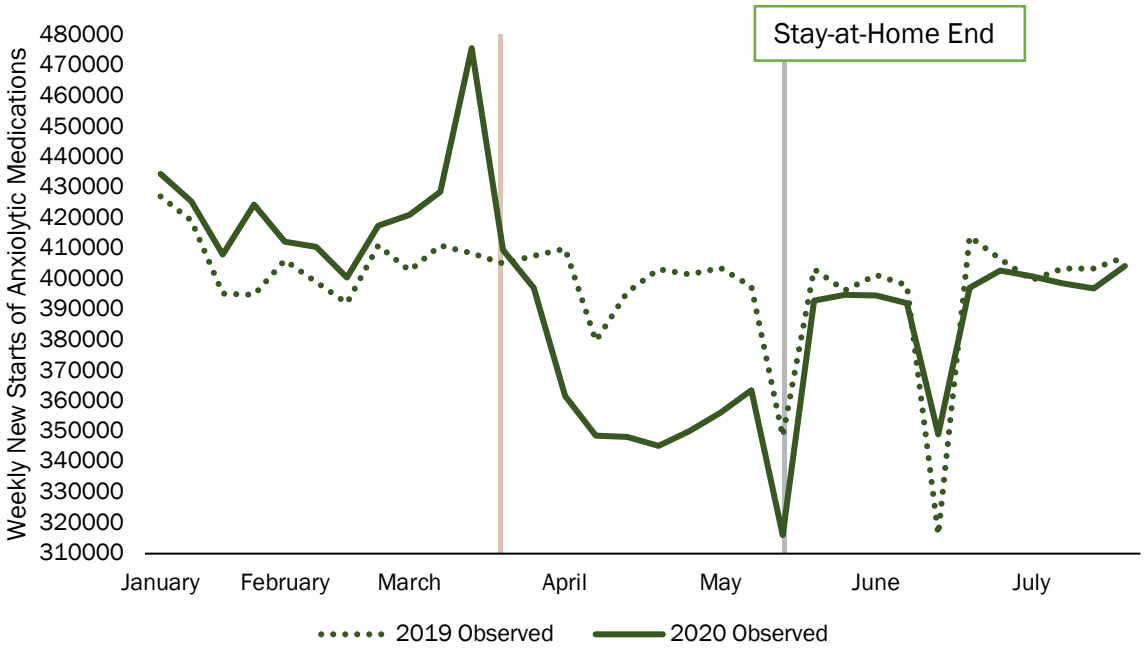
Exhibits

EXHIBIT A1. Observed New Starts of Antidepressant Medications in 2019 and 2020



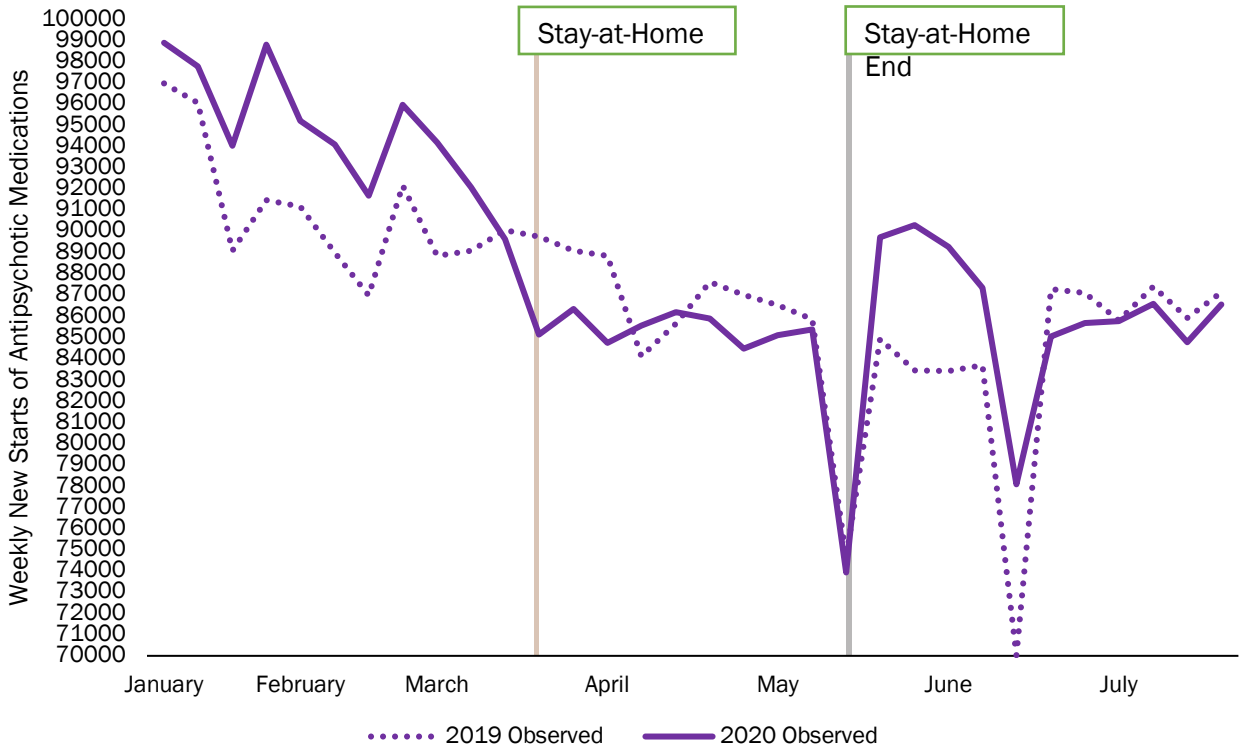
Notes Authors' analysis of data from IQVIA Longitudinal Prescription Claims, 2020. Weekly new starts of antidepressant medications from January 1st to August 8th. The vertical dashed lines represent the start (March 13, 2020) and approximate period that marked the end (May 15, 2020) of stay-at-home orders.

EXHIBIT A2. Observed New Starts of Anxiolytic Medications in 2019 and 2020



Notes Authors' analysis of data from IQVIA Longitudinal Prescription Claims, 2020. Weekly new starts of anxiolytic medications from January 1st to August 8th. The vertical dashed lines represent the start (March 13, 2020) and approximate period that marked the end (May 15, 2020) of stay-at-home orders.

EXHIBIT A3. Observed New Starts of Antipsychotic Medications in 2019 and 2020



Notes Authors' analysis of data from IQVIA Longitudinal Prescription Claims, 2020. Weekly new starts of antipsychotic medications from January 1st to August 8th. The vertical dashed lines represent the start (March 13, 2020) and approximate period that marked the end (May 15, 2020) of stay-at-home orders.

EXHIBIT A4. Total Cumulative Difference in New Starts of Antidepressant Medications from March 13th to May 15th, May 15th to August 8th, and March 13th to August 8th, 2020 Compared to 2019 Observed Overall and by Subgroup

	March 13th to May15th		May 15th to August 8th		March 13th to August 8th	
	N	%	N	%	N	%
Full Sample	-482906	-12.0%	20205	0.01%	-462701	-5.0%
Age						
0-18	-78973	-26.6%	4537	1.3%	-74436	-11.7%
18-30	-56143	-7.7%	32537	3.4%	-23606	-1.4%
30-65	-210309	-9.7%	-23125	-0.8%	-233434	-4.7%
65+	-137481	-16.5%	6256	0.6%	-131225	-6.9%
Sex						
Male	-197491	-14.0%	-21731	-1.2%	-219222	-6.8%
Female	-285415	-10.9%	41936	1.3%	-243479	-4.1%
Provider Type						
Psychiatrist/CAP	-65566	-13.0%	-48123	-7.7%	-113689	-10.1%
Family Medicine	-146651	-13.4%	-10031	-0.7%	-156682	-6.3%
NP/PA/Other	-102413	-8.7%	75345	4.9%	-27068	-1.0%
Other Physician Specialist	-168276	-13.4%	3014	0.2%	-165262	-5.8%

EXHIBIT A4 Continued.

Notes Authors' analysis of data from IQVIA Longitudinal Prescription Claims, 2020. The cumulative difference in new starts was calculated by taking the difference in cumulative observed new starts in 2019 and 2020. Percent change was calculated as the percent difference in cumulative observed new starts in 2019 and 2020. Individual forecasts were estimated for each subgroup. CAP = Child and Adolescent Psychiatry. NP = Nurse Practitioner. PA = Physician's Assistant. Other = Non-Physician specialist.

EXHIBIT A5. Total Cumulative Difference in New Starts of Anxiolytic Medications from March 13th to May 15th, May 15th to August 8th, and March 13th to August 8th, 2020 Compared to 2019 Observed Overall and by Subgroup

	March 13th to May15th		May 15th to August 8th		March 13th to August 8th	
	N	%	N	%	N	%
Full Sample	-323524	-9.0%	-57489	-1.2%	-381013	-4.6%
Age						
0-18	-73448	-30.9%	-24931	-8.2%	-98379	-18.2%
18-30	-24700	-5.3%	32711	5.2%	8011	0.7%
30-65	-102257	-5.1%	-60026	-2.3%	-162283	-3.5%
65+	-123119	-13.6%	-5243	-0.5%	-128362	-6.2%
Sex						
Male	-155452	-12.8%	-36895	-2.3%	-192347	-6.9%
Female	-168072	-7.0%	-20594	-0.7%	-188666	-3.4%
Provider Type						
Psychiatrist/CAP	-10492	-3.6%	-21267	-5.7%	-31759	-4.8%
Family Medicine	-144	0.0%	-21112	-2.0%	-21256	-1.1%
NP/PA/Other	-1642	-0.2%	70807	5.4%	69165	3.1%
Other Physician Specialist	-311246	-20.5%	-85917	-4.4%	-397163	-11.5%

EXHIBIT A5 Continued.

Notes Authors' analysis of data from IQVIA Longitudinal Prescription Claims, 2020. The cumulative difference in new starts was calculated by taking the difference in cumulative observed new starts in 2019 and 2020. Percent change was calculated as the percent difference in cumulative observed new starts in 2019 and 2020. Individual forecasts were estimated for each subgroup. CAP = Child and Adolescent Psychiatry. NP = Nurse Practitioner. PA = Physician's Assistant. Other = Non-Physician specialist.

EXHIBIT A6. Total Cumulative Difference in New Starts of Antipsychotic Medications from March 13th to May 15th, May 15th to August 8th, and March 13th to August 8th, 2020 Compared to 2019 Observed Overall and by Subgroup

	March 13th to May15th		May 15th to August 8th		March 13th to August 8th	
	N	%	N	%	N	%
Full Sample	-15617	-2.0%	22319	2.2%	6702	0.4%
Age						
0-18	-15518	-18.9%	-4559	-5.0%	-20077	-11.6%
18-30	1914	1.3%	13261	7.1%	15175	4.6%
30-65	-2729	-0.7%	-3559	-0.7%	-6288	-0.7%
65+	716	0.5%	17176	8.4%	17892	4.9%
Sex						
Male	-14552	-4.2%	3004	0.7%	-11548	-1.5%
Female	-1065	-0.2%	19315	3.5%	18250	1.8%
Provider Type						
Psychiatrist/CAP	-33819	-12.1%	-30506	-8.7%	-64325	-10.2%
Family Medicine	-2712	-2.7%	4521	3.6%	1809	0.8%
NP/PA/Other	18168	7.4%	32859	10.3%	51027	9.0%
Other Physician Specialist	2746	1.7%	15445	7.5%	18191	5.0%

EXHIBIT A6 Continued.

Notes Authors' analysis of data from IQVIA Longitudinal Prescription Claims, 2020. The cumulative difference in new starts was calculated by taking the difference in cumulative observed new starts in 2019 and 2020. Percent change was calculated as the percent difference in cumulative observed new starts in 2019 and 2020. Individual forecasts were estimated for each subgroup. CAP = Child and Adolescent Psychiatry. NP = Nurse Practitioner. PA = Physician's Assistant. Other = Non-Physician specialist.