# Empirical Strategy

To estimate the effect vertical integration has on opioid prescribing rates, we use Medicare Part D prescribing rates and map in group practice information from Physician Compare, using the broadest definition of group practice by provider[[1]](#footnote-2). We are unable to run a standard OLS regression as our dependent variable is a percent, meaning it is bound between 0 and 1, and face a two-fold selection problem with our sample. First, our prescribing rates are based on a providers’ prescribing patterns for their patients who are in Medicare Part D, limiting our sample to patients over the age of 65 which may not be wholly representative of a providers practice. Second, not all specialties who can prescribe do prescribe opioids (e.g. allergist), resulting in nearly 30% of the dependent datapoints having a zero value. These reasons make using a standard linear regression model inappropriate.

To correct the sample selection problem while simultaneously adjusting the model for the bounded dependent variable, we use a two-part model that is commonly used for datasets with similar characteristics (Borislava Mihaylova, Andrew Briggs, Anthony O’Hagan 2011; Belotti et al. 2015; Leung and Yu 1996; Duan et al. 1984). As the name suggests, our two-part model has two stages. First, we use a logistical regression to create a probability a provider prescribes an opioid. Second, we use these probabilities as a weight for each observation, and then use OLS with a logged dependent variable to calculate the coefficients (Duan et al. 1984; Belotti et al. 2015). The first stage of the model provides the extensive margin (logit model, if any opioid prescriptions) whereas the second stage provides the intensive marge (OLS, percentage of opioid prescribed, if provider prescribed at least one opioid.) To transform the coefficients back into the original units, we apply a Duan Smearing technique[[2]](#footnote-3). An added benefit of applying this technique is that it removes any distributional assumptions of the conditional error term (Duan et al. 1984). We will compare alternative models to the two-part model; the Heckman Selection model and a generalized linear model (GLM) with logit link and provide justifications for our selection.

The two-part model is similar to the Heckman selection model[[3]](#footnote-4) but has several key differences. The two-part model is designed to predict the actual outcome y and not the potential outcome, which is what the Heckman selection model predicts (Duan et al. 1984) One of the added benefits of the Heckman Selection model is that it is able to test and correct for potential biases created from non-random missingness in the dependent variable (Koné et al. 2019). Our dataset does not have missing dependent variables, as a zero represents a true zero, so these benefits do not apply. A shortcoming of the Heckman selection model is that it makes an additional assumption of correlation of error terms from the first and second part of the model, whereas the two-part model does not (Belotti et al. 2015) In addition to using less assumptions than the Heckman selection model, the two-part model has shown to outperform the former in terms of parameter squared error and perform comparable in terms of mean prediction bias and mean squared error (Leung and Yu 1996). Due to the two-part model using less assumptions and out performing the Heckman selection model we opted to use a two-part model over the Heckman selection model.

A variant of the two-part model is to use a GLM with a logit link in the second part instead of using OLS on a logged dependent variable. A benefit of this method is that by applying a GLM to the second part of the model, this maintains the original units of the dependent variable, therefore removing the need to transform the variables back to the original unit (Blough, Madden, and Hornbrook 1999). In addition, the GLM approach can be broadened to a quasi-likelihood estimation that removes the need for specifying the distribution of the dependent variable (Blough, Madden, and Hornbrook 1999). However, while this method is flexible and removes the need for the Duan Smearing transformation, it has shown to suffer efficiency losses and major losses in precision if a less than appropriate estimator is selected (Manning and Mullahy 2001). Applying a GLM with a logit links opens the study to precision errors if not executed correctly, and as our final dataset is large (n 400,000)[[4]](#footnote-5), the final estimates between the two-part model and the GLM with logit link would not differ significantly. For these reasons we selected to use the two-part model with OLS on a logged dependent variable for its reliability and flexibility.

### General Framework of model

We apply a two-part model for each of our regressions, all regressions have the same general framework but have varied data samples and explanatory variables. The first part of the model captures the extensive margin using a dummy variable indicating that a provider has given at least one opioid prescription. The second part captures the intensive margin using the log of the percentage of opioid prescriptions. The first part of the model is specified as logit:

#### First Part

(2.1)

(2.2)

) (2.3)

#### Second Part

The second part of the model is an OLS regression with a logged dependent variable

(3.1)

(3.2)

(3.3)

#### Complete Model

The full two-part model is calculated using the predicted probability from the first part and used to estimate the conditional mean to the second part:

(4)

When calculating the margins of the output, we transform the regressors from the log to raw value, using a Duan smearing transformation (Duan, 1984)

## Description of Model

Before we can estimate the impact vertical integration has on specialty prescribing rates, we first need to develop a baseline equation that modeled the effect specialty without controlling for any type of integration. We use this baseline equation to compare changes in coefficients by specialty groups for all subsequent equations:

where our dependent variable, Y, is a percentage of opioid prescriptions by provider. For our specialties, we use primary care as the reference category and compare the prescribing rates of pain management specialists (M), specialists (S), and nurse practitioners (N). All other control variables, including state indicator variables and county level characteristics are designed as Z. Lastly our error term is designated as

We find with high significance that primary care specialties have the lowest likelihood of prescribing an opioid followed by nurse practitioners, specialists, and then pain management specialists, these findings match those of previous literature **(Mark et al. 2019; Pletcher et al. 2008)** and unsurprisingly show that provider specialty has an impact on prescribing rates.

Now that we established our baseline impact specialty has on prescribing rates and rank order of specialties, our next question is how prescribing rates change when specialties work together. We hypothesize that through spillover effects of specialties with higher baselines working with specialties of lower baselines, an individual provider prescribing rate will actually be a weighted average of the types of specialties they work with. This leads us to our second equation, which builds on equation one by adding various forms of integration:

We add an inclusive set of group practices containing varied specialties to estimate the impact integration has on prescribing rates. In order to parse out vertical integration from horizontal integration, we divided group practices into either horizontally or vertically integrated classifications. Horizontally integrated group practices consist of single-specialty practices that may or may not have a nurse practitioner, these are represented by coefficients Mpractice and Spractice. Where M is single-specialty pain management practices and S is specialist practices, with single-specialty primary care practices as our reference group. Our vertically integrated group is Xpractice, which are any group practices that are some combination of M, S, and single-specialty primary care groups. Group X may or may not have nurse practitioners included

By incorporating types of group practice in the regression, we find that the magnitude of all specialties, in comparison to primary care, decrease. The ranked order of specialty remains the same, except for nurse practitioners whose prescribing rates become slightly negative. We find that providers who are in specialist group practices are the least likely to prescribe opioids, followed by primary care practices (reference category), vertically integrated practices (Group X), and pain specialist practices

The vertically integrated practices reside between our specialist, primary care, and pain management practices. This supports our hypothesis that prescribing rates are a weighted average of specialties that comprise the group practice.

To validate the robustness of our results from equation two, we wanted to test the impact vertical integration has on one specialty type. In selecting a specialty type, it is necessary for the specialty to commonly work in a variety of settings, as selecting a specialty that works heavily in one type of integrated group practice could skew the results. In this case, nurse practitioners[[5]](#footnote-6) were the logical choice as they operate in independent practices, under supervision of medical doctors, and a flexible specialty that can work with all other specialties. For equation three, we limit the sample to only nurse practitioners and keep all the control variables the same (by reducing the sample to nurse practitioner removes any other specialty):

We find that when nurse practitioners working only with other nurse practitioners prescribe the least amount of opioid prescriptions followed by, in increasing order, nurse practitioners who work in vertically integrated systems, primary care practices (reference category), specialist practices, and pain management practices.

By incorporating nurse practitioner practices, this slightly changes the ranking of the remaining group practices from regression two, switching the vertically integrated practices with the specialist practices.

This supports the results that were found in equation two – vertically integrated groups are a weighted average of the various specialties they interact with

1. For example, if a provider is listed in a primary care group practice and then again in a group practice of mixed specialties, they are indicated as a mixed specialty group practice [↑](#footnote-ref-2)
2. The Duan Smearing technique is a nonparametric method of retransforming the data to the original values **(duan 1983)** [↑](#footnote-ref-3)
3. Heckman selection model is also referred to as adjusted or generalized tobit (Belotti et al. 2015; **Amemiya 1985; Maddala 1983**) [↑](#footnote-ref-4)
4. Need to check for normality. [↑](#footnote-ref-5)
5. Nurse practitioners are able to practice independently with full scope of practice in certain states, work under physician supervision in other states, and most commonly specialize in primary care (CITE). These characteristics make them similar in specialty to primary care physicians but add a layer of complexity as they work with many other specialties [↑](#footnote-ref-6)