Policy Interventions When Medical Treatment Dynamics Matter: The Case of *In Vitro* Fertilization*

Barton H. Hamilton
Washington University in St. Louis
Olin Business School

Emily Jungheim
Washington University in St. Louis
School of Medicine

Brian McManus
University of North Carolina-Chapel Hill
Department of Economics

Juan Pantano
Washington University in St. Louis
Department of Economics

July 2015

Abstract

Decision-making dynamics are central to *In Vitro* Fertilization (IVF), an infertility treatment in which patients make a sequence of choices within and across treatment cycles. Patients' strategies are affected by their fertility characteristics, preferences, IVF's treatment technologies, and information gradually revealed during a cycle. Patients may forego treatment because of its expense, or may choose more aggressive treatment to reduce the substantial risk of IVF failure. Several policy interventions are possible to increase IVF access or encourage patients to take more conservative treatments, which reduce the chance of risky twin or triplet pregnancies. We evaluate potential policies after estimating a dynamic structural model of patients' choices within and across IVF treatments. The policies include insurance mandates, direct limits on or additional prices for aggressive treatment, and introducing improved treatment technologies. All policies have significant effects on patient choices and outcomes, but vary substantially in their welfare and cost consequences.

Keywords: In Vitro Fertilization, structural estimation of dynamic decision models, health policy analysis and counterfactuals

*We thank for comments and suggestions Jaap Abbring, Naoki Aizawa, Gary Biglaiser, Flavio Cunha, Liran Einav, Hanming Fang, George-Levi Gayle, Donna Gilleskie, Limor Golan, Pedro Mira, Bob Pollak, Jimmy Roberts, John Rust, Seth Sanders, Steve Stern, Xun Tang, Ken Wolpin, and participants at several conferences and workshops. All errors remain our own.

1 Introduction

Many medical ailments require patients and doctors to consider complex and lengthy treatment strategies. For example, cancer treatment may combine surgery, radiation, and chemotherapy in some sequence as the patient's response to each therapy is revealed. Similarly, treatment for heart disease may begin with pharmaceutical approaches and then progress through increasingly invasive types of surgery over time. Common themes across these therapies include decisions regarding how aggressively to treat the ailment (along with a consideration of associated side effects), uncertainty about treatment success, and the opportunity to dynamically update treatment strategies as information arrives. However, evaluation of new medical treatments and interventions is often static.[1] When a new public policy or technological advance is implemented that affects one or more treatment avenues for an ailment, we would expect that a patient's full treatment course could change rather than just an individual component. The extent of these changes (and their welfare impact) will depend on the precise details of treatment processes, the decision structure, and patients' preferences.

In this paper we study a technologically advanced infertility treatment, *In Vitro* Fertilization (IVF), in the institutional and policy context of the United States. IVF use has increased substantially in recent years as the technology has matured and demand for it has grown, driven in part by delayed fertility decisions by women. In the U.S., IVF has grown four-fold over the last 20 years, to 160,000 treatment cycles and resulting in 1% of all births.[2] Treatment rates are even higher in some other countries; IVF accounted for over 3% of 2011 births in Israel, Belgium, Sweden, and Denmark.[3] Despite this progress, a single treatment cycle is more likely to fail than succeed. Relatively young U.S. patients (under 35 years old) in 2012 achieved a birth after 40% of cycles, while patients just under 40 years old had success rates approximately half as large. These failure rates require individuals to consider both current and future treatment choices. Like

[1] Medical treatment guidelines themselves traditionally have been static or open-loop. Murphy (2003) introduced the notion of "optimal dynamic treatment regimes" to construct adaptive decision rules, and this framework has been applied to physicians treating disease. Murphy's approach builds on earlier work by Robins (1997) on dynamic treatment effects. Her approach, however, focuses on dose-response relationships and does not incorporate or estimate patient preferences. This inhibits the study of settings where patients exercise some discretion over treatment protocol, and limits the focus to objective outcomes (e.g. biological responses) rather than also allowing subjective ones (e.g. welfare). Abbring and Heckman (2007) describe this area of the statistics literature, and they contrast the literature's assumptions with structural econometric approaches that, as in our paper, rely on dynamic choice models.

[2] The aggregate treatment statistics reported in this paragraph are for fresh-embryo treatments during 2012, the most recent year for which U.S. data are available (Centers for Disease Control and Prevention (CDC), 2014).

[3] Statistics from European countries are from the European Society of Human Reproduction and Embryology's (2014) ART Fact Sheet. Israel's IVF birth share is reported by Simonstein et al. (2014).

therapies for cancer or heart disease, treatment choice dynamics for IVF patients reflect a number of mechanisms. First, forward-looking patients must consider both the current and future price of treatment if outcomes are uncertain. IVF can require substantial out-of-pocket payments, on the order of $10,000-$15,000 per attempt.[4] Due to high potential future costs, patients may choose aggressive current treatment that raises both the probability of pregnancy and the likelihood of an undesirable high-order birth; twin and triplet births entail medical costs many times as large as a singleton birth. Patients under age 35 received two embryos during treatment, on average, and a third of all births for these patients included twins or more. Second, patients may expect their medical condition to change over time. In the case of IVF, female fertility declines with age, particularly after age 35. Anticipation of future health declines affects current decisions regarding both the initiation and aggressiveness of treatment. Patients in their early 40s average about three embryos during treatment, but success rates fall to around 10% and twins are rare conditional on success. In general, treatment choice dynamics may also be influenced by learning about the severity of the health condition and effectiveness of earlier therapies.[5] In the case of IVF, most learning occurs at the initiation of treatment, implying that dynamic updating considerations are less important.[6] These mechanisms illustrate that under a constant regulatory and technological environment, treatment choice dynamics will reflect intertemporal variation in prices, health, and information acquisition.

Like the rest of the health care market, public policy plays a major role influencing IVF treatment. High out-of-pocket costs and the high rate of expensive multiple births have led policymakers in the U.S. and abroad to consider a variety of interventions to improve IVF access and reduce treatment risks. One potential avenue is through subsidized treatment or mandatory insurance coverage; publicly funded IVF is offered in several countries, and insurance mandates have been implemented in ten U.S. states. Insurance coverage for IVF can reduce the price of each attempt to $2,000-$3,000, which may lead patients to view failure as less expensive, since future treatment attempts can occur at a lower price. Previous studies at the population or clinic level have provided empirical evidence on the effectiveness of such policies.[7] Another potential policy intervention is a

[4] As discussed below, insured patients may be concerned about exhausting their benefits, which raises future prices for IVF. This is in contrast to other insurance scenarios in which patients initially experience high out-of-pocket costs for treatment and then face low prices once their insurance deductible is reached. Aron-Dine et al. (2012) examine the dynamic implications of changes in out-of-pocket costs induced by exhaustion of deductibles.

[5] See, for example, Crawford and Shum (2005), Chan and Hamilton (2006), and Dickstein (2014).

[6] As described below, at the initial IVF screening visit the doctor learns the patient's fertility characteristics.

[7] Schmidt (2007), Bitler (2008), and Bundorf, Henne, and Baker (2008) examine the impact of infertility mandates at the population level. Hamilton and McManus (2011), Jain et al (2002), and Henne and Bundorf (2008) investigate the impact of mandates on the number of patients served and birth outcomes at IVF clinics. See also Schmidt (2005),

cap on the number of embryos transferred during treatment. This restriction is imposed in some European countries, and accords with the U.S. medical community's sentiment that a singleton birth is the best possible outcome of treatment.[8] Finally, policies such as research grants and prizes can push forward technological progress, which affects treatment choices and outcomes.

We investigate the impact of these policies by specifying a dynamic structural model of the decisions made during IVF treatment by forward-looking patients. We estimate the model using a novel dataset of the treatment histories of 587 women undergoing IVF at an infertility clinic ("the clinic") between 2001 to 2009, as well as data on potential patients in the St. Louis, Missouri market where the clinic is located. This setting provides a valuable opportunity to understand how prices, preferences, health, and technology affect IVF treatment. The clinic serves patients from both Illinois, which mandates insurance coverage of IVF, and Missouri, which does not. Consequently, we are able to analyze the decisions of observationally equivalent patients facing vastly different prices (about $3,000 for covered patients versus $11,000 for those without insurance) undergoing the same procedure with the same physicians. Using highly detailed information on the fertility attributes of the patients and their treatment choices and outcomes, we estimate the various stochastic processes that determine success at each stage of an IVF treatment cycle. These processes, together with the specifications of patient preferences over children, delaying treatment, and the disutility of payments, yield a well-specified dynamic optimization problem for choices within and across IVF treatments. We then estimate the patients' preference parameters to maximize the likelihood of the observed treatment choice histories. Our model allows us to investigate the impact of alternative policies on individual patient actions, outcomes, and surplus in a variety of counterfactual settings. Critically, our preference estimates for singleton versus multiple births are a necessary component to assessing how patients respond to policies designed to reduce treatment aggressiveness.

We estimate the empirical model in three steps. The first step is the estimation of the stochastic processes which determine outcomes during the four stages of an IVF cycle; we refer to these processes as "technologies" in the discussion below. The second step recovers the structural parameters of our within-clinic patient decision model. These parameters indicate that patients prefer singleton and twin births to the more dangerous triplet births, and the utility from additional children falls in the number of children the patient already has. The model parameters predict patients' choices at various treatment stages, and we find that our estimates are able to reproduce the

Bitler and Schmidt (2006, 2012) and Buckles (2013).

[8] The Practice Committee for the Society for Assisted Reproductive Technology (2012) summarizes a number of studies on single embryo transfer and concludes that IVF clinics should promote elective single embryo transfer.

data's main moments fairly well. The final step captures potential patients' decisions to initiate treatment. We construct data on the local population of women "at risk" for infertility treatment, and we use these data together with observed treatment-initiation decisions to estimate a model that describes the willingness of potential patients to pursue IVF.

We use our structural estimates to evaluate a collection of counterfactual policy experiments. First, we estimate the impact of extending insurance coverage to all women in the market. This policy's primary impact is to substantially increase the number of women who initiate treatment, from 38% of "at risk" women under the baseline environment to 56% with universal insurance. Consumer surplus increases from \$5,300 to \$9,600 per potential patient when insurance is introduced. While insurance reduces the opportunity cost of failed treatment, which could affect embryo transfer rates, we find only a small reduction in treatment aggressiveness as insurance coverage becomes more common. Our estimates imply that patients receive about the same utility from singleton and twin births, so they see little reason to transfer fewer embryos. Second, we explore the impact of restricting patients to transferring a single embryo during treatment. While this policy has a clear effect in nearly eliminating multiple births, we find that active patients are much less likely to conclude treatment with a child, and they are also less likely to begin treatment at all. Treatment initiation falls from the 38% baseline to 10% in this experiment, and consumer surplus shrinks to \$700 per potential patient.

In a third experiment, we explore the impact of adding "top-up prices" for patients who transfer more than one embryo.[9] Current practice in IVF allows patients to increase the number of embryos at zero additional price. Moreover, patients are largely shielded from the increased medical costs of high-order pregnancies, since birth expenses are generally covered by their health insurance policies. We select top-up prices that cover the increased medical expense relative to single-embryo transfers, and we find that patients reduce but do not eliminate multiple-embryo transfers relative to observed practices. The top-up prices allow patients with stronger preferences for having children to pursue the outcomes they value most, while also internalizing the additional expected medical costs associated with twin or triplet births. Such a policy could generate additional interest beyond IVF, as it strikes a balance between insured treatment and out-of-pocket expenditures.[10] As top-up

[9] Einav, Finkelstein, and Williams (2015) consider a closely related policy of top-up prices for medical care beyond a basic level covered by insurers.

[10] When patients need to resolve a medical issue, their options may include treatments that vary by convenience, side effects, risk-reward tradeoffs, aesthetic effects, etc. Insurance programs are likely to be inefficiently rigid if they mandate that all patients take the same least-cost treatment. At the other extreme, when doctors and patients bear no costs for selecting treatment avenues that are most pleasing to the patients, efficiency-reducing moral hazard can result.

prices alone entail a large price increase for many patients, we specify an additional experiment in which top-up prices are imposed in combination with universal insurance coverage for single-embryo IVF treatment. The presence of insurance allows access rates to be approximately equal (34%) to the empirical baseline, while top-up prices result in substantially lower medical costs associated with births ($37,000 per delivery versus $68,000 under the baseline), due to patient payments against birth expenses and a reduction in multiple birth rates.

In a final experiment, we illustrate the impact of another potential avenue for improved treatment access and reduced multiple-birth risk: an improvement in treatment technology. We focus on the issue of embryo selection. Many embryos are not equipped to develop into viable pregnancies. Indeed, this is the reason why, despite a successful conception, many pregnancies fail during natural reproduction. Research is currently underway to better understand why some embryos develop into successful pregnancies and others do not.[11] Identification and selection of such embryos before transfer would significantly increase success rates. To capture this, we add an embryo screening stage to treatment, and we allow it to substantially reduce the uncertainty about whether any individual embryo will yield a successful pregnancy. As a result, more patients (48%) are willing to start treatment at the current prices (which we hold fixed), and treatments are more likely to end in a birth.

In addition to our main focus on patients' responses to policy changes, we contribute to literature on understanding responses to changes in medical care prices. The rapid rise in health care expenditures in the United States over the past three decades has generated substantial interest in this area. A growing literature using both experimental and observational data has attempted to empirically measure the relationship between the out-of-pocket price paid by the patient and the utilization and cost of health care. The primary focus of this literature is examining how alternative cost-sharing arrangements in a patient's health insurance contract (e.g., co-payment rates, deductibles) affect his or her total health care expenditures in a given year. Much less attention has been paid to how an individual's treatment choice for a particular ailment responds to changes in that treatment's full price. Consequently, little is known about how the composition of medical treatments may change in response to changes in their relative prices. These changes may be especially difficult to study when intensive-margin choices (i.e. the selection of specific treatments

[11] Most embryos do not develop into successful pregancies because they are genetically abnormal. Pre-implantation genetic testing can help identify embryos with good potential (Scott et al., 2013), although current technology in this area risks damaging embryos during testing. Another approach to embryo selection is based on time-lapse imaging of the developing embryo. See Chavez et al. (2012) and Wong et al. (2010).

options as opposed to extensive margin choice whether to treat or not) and intertemporal substitution are especially salient. We use our model to estimate separate elasticities for treatment initiation (-0.68) and the total number of cycles (-0.76) when patients pay the full treatment price. These elasticities are distinct from (and about twice the magnitude) of those describing the impact of changing out-of-pocket expenses for insured cycles.

The remainder of the paper proceeds as follows: Section 2 provides a preview of the four stages of an IVF treatment cycle, and describes state level policies governing insurance coverage of infertility treatment. Section 3 covers assumptions on model components, which are incorporated into our dynamic structural model of treatment choice developed in Section 4. In Section 5 we describe the data we obtained from the clinic, plus additional market-level data. Section 6 discusses the empirical specification of our model and Section 7 provides estimation details. Section 8 presents the parameter estimates and measures of model fit, and Section 9 contains the results from our counterfactual policy simulations. Conclusions follow.

2 IVF overview

A couple is defined to be medically infertile if they are unable to conceive after attempting to do so for 12 months. Initial treatment for infertility often includes the use of the drug clomiphene to induce ovulation, or the use of hormone shots with or without intrauterine insemination. While such treatments are relatively low cost, they are less effective than more technologically advanced treatments, more likely to lead to higher order pregnancies, and can be especially poorly suited to older patients and those with male factor infertility. Due to these limitations, couples may choose to directly undergo IVF. Others may eventually turn to IVF after failing to conceive through these less advanced treatments. As noted in the Introduction, the U.S. market for IVF has grown substantially in recent years, so that by 2012 65,000 children were born through IVF. However, a cycle of treatment is still more likely to fail than to succeed, with live birth rates ranging from 10% to 45% depending on the age of the woman and the health status of the couple.

Once a patient has decided to use IVF, the treatment cycle unfolds in stages. First, the woman takes drugs to stimulate egg production. The patient and doctor monitor the response to these drugs and may choose to cancel the cycle if the patient's response is not favorable; if a cycle is cancelled, the patient may start IVF again in the future. If the cycle is not cancelled, the eggs are retrieved during a minor surgical procedure and then fertilized in the laboratory. The doctor may

recommend the use of intracytoplasmic sperm injection (ICSI), in which a single sperm is injected into the egg. ICSI was initially used to address male-factor infertility problems, but has become more widely used. The patient then decides how many fertilized eggs (embryos) to transfer to the womb; this choice may be constrained by the number of embryos that develop. At this point the patient faces an important tradeoff: the probability of a live birth increases with the number of embryos transferred, but so does the likelihood of a potentially costly and medically risky multiple birth. Lemos et al. (2013) calculate that the average medical cost of a singleton IVF pregnancy and initial child medical care is $26,922, while twin and triplet births entail costs of $115,238 and $434,668, respectively.[12] The high costs of multiple births are due largely to shorter gestation periods, which can lead to newborns being admitted to neonatal intensive care units. If the IVF cycle does not result in a live birth, the patient then must decide whether (and when) to attempt another cycle of treatment. Because fertility declines with age, subsequent cycles are less likely to be successful, all else equal, and couples potentially incur substantial out-of-pocket cost if they try again. Patients whose treatments succeed may also try IVF again, if they want to add more children to their families.

2.1 Insurance and IVF

A key feature of the market for IVF is the presence of state-level mandates regarding whether and how insurers must offer coverage for infertility treatment, including IVF. During the period of our study, 2001-2009, seven states had mandates requiring some form of insurance coverage for IVF. Connecticut (after 2005), Illinois, Massachusetts, New Jersey, Rhode Island had the strongest mandates for IVF, requiring insurers to cover a certain number of IVF treatment cycles.[13] Prior research has found that these mandates increase the number of IVF treatment cycles at clinics in covered states, reduce the number of embryos transferred, and reduce multiple-birth rates.[14] These studies have generally examined data aggregated at the population or clinic level.

For patients in our study residing in Illinois and working for an employer covered by the mandate,

[12] These costs include the portion paid by patients (e.g. co-pays), so they may overstate the potential moral hazard costs of multiple births.

[13] See Schmidt (2005,2007). Maryland, Arkansas, Hawaii and Montana are also classified mandate-to-cover states where the mandate includes IVF. Texas has a mandate requiring insurers to offer plans that include IVF coverage. Nothing prevents insurers, however, from charging substantially higher prices for plans that include this coverage.

[14] When looking at multiple-birth outcomes, it is important to distinguish between changes in rates and levels. Among existing IVF patients insurance mandates reduce multiple birth rates by facilitating less aggressive treatment. But the overall number of multiple births may increase if enough new patients can pursue IVF treatment under the mandated coverage.

insurance plans are required to pay for up to 4 cycles of IVF if the woman has no children.[15] This insurance coverage pays the cost of the IVF procedure, but may not cover the full cost of drugs used during treatment due to co-payments and deductibles. These drugs have full prices of approximately $3,000. For patients paying out-of-pocket for IVF in our sample, the clinic charged about $11,000 (including drugs) per treatment cycle throughout the sample period. The Illinois mandate exempts firms with fewer than 25 employees and organizations such as the Catholic Church that may object to IVF for religious reasons. These individuals pay the full price of IVF.

Our study exploits the fact that the clinic draws patients from the greater St. Louis metro area, which includes both Missouri, which has no insurance mandate, and Illinois. However, a feature of our data is that some patients residing in Missouri have private insurance covering their IVF cycle, even in the absence of a mandate. Some employers may choose to offer IVF coverage as a means to attract and retain better employees. In addition, some firms operating in the St. Louis metro area have locations in both Illinois and Missouri. Rather than offer IVF coverage only to their Illinois employees, many of these firms choose to offer insurance coverage to all their workers in order to reduce administrative costs and eliminate inequality in benefits. The clinic has found that the insurance plan characteristics covering Missouri patients are very similar to those of plans under the Illinois mandate. The patient-level information on insurance status allows us to exploit both cross-sectional and longitudinal variation in the out-of-pocket prices faced by individuals in our sample.

2.2 Embryo caps for reducing multiple-birth risk

The relatively high rate of multiple births associated with IVF, along with the high costs of such births, has led some to argue that IVF clinics should adopt a policy of single embryo transfer (see, e.g., Ryan et al., 2007; Csokmay et al., 2011). By requiring patients to transfer only one embryo per IVF cycle, the risks of multiple births can be reduced.[16] While single embryo transfer is uncommon in the United States (only 10% of IVF cycles in 2009 involved a transfer of one embryo), it is widely practiced in Europe. For example, 69% of IVF cycles in Sweden transfer a single embryo and in Belgium it is required. While it does reduce multiple birth risk, an embryo cap policy can have drawbacks. Obviously, patients would require more cycles in order to get pregnant; the associated

[15] If the patient has already had a birth through IVF, the number of remaining covered cycles is set to 2. This implies that an Illinois resident can have as few as three or as many as six covered cycles, depending on when or whether she has a successful cycle.

[16] See Jungheim et al. (2010) for a discussion single embryo transfer policies.

increase in the expected cost of treatment may reduce access to IVF.[17] A more subtle cost of single embryo transfer is that some patients may prefer a twin birth to a singleton, given declining fertility with age and the desire for more than one child. Older women may view twins as a more efficient and effective way of achieving the goal of more than one child (Hojgaard et al., 2007; Ryan, 2004). Embryo transfer restrictions may thus reduce patient utility as well as increase some treatment costs. We investigate the importance of these issues in our counterfactuals presented in section 9.

3 Model preliminaries

3.1 Timing

We consider two timing concepts in the model below. First, there are decision *periods* when active patients choose whether to start or delay an IVF cycle. Second, there are four treatment *stages* during which patients in an IVF treatment cycle make one choice per stage. We index time periods with t, and we use j to index stages. Within each stage j, the patient selects an action, y_j, from the set Y_j.

We assume that potential patients' decisions begin with an exogenous event which prompts them to consider having children. Women who are able to reproduce naturally (or with less technologically advanced infertility treatments) are immediately removed from the process we study in this paper. The remaining women have reproductive difficulties that can be solved by IVF only. These women, who constitute our "at risk" population, evaluate the expected benefit of beginning IVF relative to an outside option, which we parameterize below. If the woman does not begin IVF at this critical moment, we assume she exits the model permanently.

We track patient's decisions in three-month periods (i.e. quarters). The exogenous event to consider reproduction begins when the patient is of age a_0, which we assume to be not smaller than a lower limit $a^{\min}$. In our data we observe patients with a_0 between their late twenties and early forties. If at that point she opts to pursue IVF treatment, the patient will continue to make decisions up to, possibly, the fourth quarter of age $a^{\max}$. At this age the IVF clinic will no longer treat the patient and her birth probability (via IVF or naturally) is zero.[18] This allows a maximum of $4 \times (a^{\max} - a_0)$ periods for a patient whose reproductive decisions start at age a_0. We set

[17] Velez et al. (2014) shows that multiple births declined almost 80% after Quebec introduced universal insurance combined with single embryo transfer (for certain age groups). However, the clinical pregnancy rate per cycle also fell by 38%.

[18] We assume this age upper bound for tractability. The clinic does not have a preset age limitation and, instead, evaluates each patient on a case by case basis.

$a^{\min} = 28$ and $a^{\max} = 44$ in implementing our model. Once a patient's total number of children reaches 3 (or more), she automatically stops making decisions within the model.

In addition to the age index, a time index (t) is useful for describing the data sample and econometric procedure. Let $t_{i,0}$ represent the period during which we first observe patient i. We see a patient for the last time in T_i, which might be equal to $a^{\max}$ or $\overline{T}$, the end of the sample period. We assume that all treatment stages that follow from a treatment starting in period t also occur in period t.

3.2 State variables and initial information

A patient who is considering treatment is aware of several personal characteristics that affect treatment effectiveness and utility. There are two types of state variables in the model. First, there are the state variables collected in the vector Z, which remain constant within periods but may transition between them. Second, there are state variables which are revealed during the stages of a treatment cycle, but do not carry over between periods. These variables include information about treatment progress and additional taste shocks that affect the value of each treatment option at a decision stage. We discuss these variables in detail below, when we introduce our model of IVF treatment behavior.

We divide the state vector Z into two parts. We track a patient's age, a measure of her wealth, number of prior children, record of previous payments for IVF, and insurance status in the state vector Z^D. These variables vary in how they evolve between periods. Age increases exogenously by three months every period. We treat wealth (z_w) as a permanent characteristic of patients, and we model this with zipcode-level data on housing values. We focus on patients with zero prior children ($\widetilde{k} = 0$) at the treatment initiation decision, and then $\widetilde{k}$ evolves endogenously according to treatment outcomes. Likewise, the patient's record of prior IVF payments (z_p) and remaining insured cycles evolve endogenously according to the patient's decisions within the model. We initialize the number of insured cycles (ι) to four (the Illinois mandate value) for all patients who ever use insurance, and this number falls by one whenever an insured patient advances to the second stage of treatment, when eggs are removed during surgery.[19] Most insured patients are from Illinois but not all; likewise most Illinois patients are insured but not all of them.

The second part of Z includes the patient's biological characteristics, Z^B. We assume that the patient learns her own Z^B if she initiates treatment. The characteristics in Z^B include: z_{afc}, the

[19] We abstract away from some of the complicated details of the Illinois insurance code, discussed above.

women's antral follicle count (AFC score), an indicator of her egg-producing ability; z_{ff}, whether she has one or more specific infertility diagnoses (e.g. endometriosis); and z_{mf}, whether her partner has male-factor infertility. At the treatment initiation decision, the patient considers the possible values of Z^B she may have using the population frequency of these characteristics conditional on her initiating age, $f_{Z^B}(Z^B|a_0)$.

Our assumptions on Z include a few simplifications that we impose to maintain tractability. First, we do not allow patients to receive a detailed fertility screening before deciding to initiate treatment, which could be used to reveal Z^B. While such screenings are feasible in actual treatment markets, we make this simplification in order to reduce the dimensions of potential patient heterogeneity prior to treatment. Second, we assume that patients (and their doctors) use no other biological data in choosing a treatment path for patients. Although fertility doctors often collect information on patients' pre-treatment follicle-stimulating hormone (FSH) and estradiol (E2) levels, we do not observe these items in our data. We effectively assume that the patient's observed biological state variables Z^B and age fully capture her relevant fertility characteristics.

Once the value of Z^B is realized we consolidate notation and refer to the state vector $Z = [Z^D, Z^B]$. In addition to acting as a state variable which influences treatment outcomes, patient age also functions as a time index for decision periods, so we add an 'a' subscript to Z where appropriate. During an arbitrary age, we have $Z_a = [Z_a^D, Z^B]$, and at treatment initiation the state variables have the value Z_{a_0}. We assume that the doctor knows how the variables in Z_a affect treatment outcome probabilities. Each patient receives this information from her doctor and also knows her preferences over treatment outcomes.

3.3 Patients' preferences

Patients have preferences over birth outcomes (k), and these preferences can depend on the patient's existing number of children ($\widetilde{k}$) at the start of an IVF cycle and other personal characteristics. Possible values of k are in $\{0, 1, 2, 3\}$, and $\widetilde{k}$ takes values in $\{0, 1, 2\}$. (These values for $\widetilde{k}$ cover 98% of the patient population at the clinic.) We allow patients to have permanent unobservable heterogeneity, indexed by τ. Let $U(k|\widetilde{k}, \tau)$ represent the lump-sum utility payoff from a treatment cycle that ends in k children conditional on $\widetilde{k}$ and τ. As a normalization, we assume that treatment outcomes with $k = 0$ always result in $U = 0$ for all patients.

In addition to payoffs through U which may be received at the end of IVF treatment, patients undergoing treatment experience disutility, scaled by α, from paying positive prices. When a

patient pays p within treatment, she has the immediate utility loss of αp. We allow the value of α to depend on a patient's demographic characteristics, so we write $\alpha(Z)$. Likewise, a patient's price depends on her insurance status, so we write $p(Z)$. An additional potential source of disutility is in a patient's choice to deviate from the American Society for Reproductive Medicine (ASRM) guidelines for embryo transfers. During our sample period the ASRM generally recommended against four-embryo transfers for all patients, and single-embryo transfers for older patients. We assume that a patient's utility falls by $\eta(x, Z_a)$ if she makes a choice outside of the guidelines, where x is the number of embryos transferred. We write η as a function of state variables to capture shifts in ASRM guidelines within our sample period and their dependence on patient age. We assume that all ASRM rule changes come as a surprise to decision-makers.

The remaining parts of patients' preferences concern a terminal value for patients and the value of delaying treatment. Relative to the delay baseline of zero, we assume that patients receive the flow benefit of $u_s(Z)$ during any period she begins treatment in stage 1. Patients' terminal payoffs are captured by the parameter vector $u_T(Z_{a^{\max}})$. The patient receives u_T at age $a^{\max}+1$ regardless of whether she remains active in the model up until $a^{\max}$ or if her decision process ends due to $k \geq 3$ at some earlier a.

At each treatment node, the patient's benefit from the available options includes an additional taste shock, ε, which represents heterogeneity in patient's circumstances and preferences. Following Rust (1987), for computational ease, we assume that ε is distributed i.i.d. type 1 extreme value across patients, time periods, treatment stages, and alternatives within each stage.

Finally, we assume that patients discount future decision periods by the factor β. We assume that all discounting occurs across periods, and not across treatment stages. Treatment options and outcomes that occur t periods into the future are discounted by β^t. We do not estimate β in this paper, so we set its value equal to $\beta = 0.97$.

3.4 Technology and prices

During each IVF stage, a patient makes a choice about treatment; possibly pays a price out-of-pocket; and anticipates the outcome of a random process, the results of which are revealed before the next choice occurs. We now review notation for these processes, i.e. the treatment technologies, and the prices patients pay. We assume that the technologies did not change during the sample period. This accords with the actual practice of IVF during the early 2000s.

For a patient who has committed to the first stage of IVF treatment, her personal characteristics

and drug regimen will yield a Peak E2 score (e) to be revealed at the start of stage 2. The score is a signal of the patient's success in generating eggs. During the first stage, however, the patient knows only the distribution over possible e values rather than the signal's realization. Let $f_e(e|Z_a)$ represent the probability of a patient with characteristics Z_a receiving a score with value e, which takes positive integer values. Moving to the second stage, we denote as $f_r(r|e, Z_a)$ the probability of successfully retrieving r eggs from a patient with Peak E2 score e and personal characteristics Z_a. A patient with a greater value of r is more likely to generate a large number of embryos during later stages. Once the patient reaches the third IVF stage, she observes her value of r and considers the distribution over possible numbers of embryos, denoted X, available for transfer, which will be realized following her decision on fertilization method (m). We write this distribution as $f_X(X|r, m, Z_a)$, and note that it may be shifted by r, m, and the patient's state variables. Finally, in stage four the patient considers the number of children (k) that will be born, which is affected by the number of embryos transferred (x out of the realized X) and the patient's Z values. The distribution over realizations of k is $f_k(k|x, Z_a)$.

We consider prices that consumers may pay at three treatment stages. The price of action y in stage j is $p_{y,j}(Z)$. In the first treatment stage, uninsured patients pay $p_{s,1} = \$3{,}000$ if they choose the action 'start' (s), while insured patients pay $p_{s,1} = \$1{,}000$. The positive price for insured patients is due to deductibles, co-payments, and co-insurance charges. Patients who continue (c) treatment in stage 2 pay $p_{c,2} = \$6{,}000$ if uninsured, and $p_{c,2} = \$2{,}000$ if insured. The third-stage option to use ICSI (m_2) carries a price of $p_{m_2,3} = \$2{,}000$ for uninsured patients, and a price of zero for insured patients. The final stage, embryo transfer, has zero price for all patients regardless of the number of embryos transferred.

4 Decision model and value functions

We now describe how patient preferences and IVF technology come together into a multi-stage decision process. Conditional on starting IVF treatment, a patient makes a series of choices regarding the aggressiveness of her treatment and whether the treatment continues at all. Along the way, the patient uses information that is known at the start of treatment (e.g. age, current number of children, basic fertility diagnoses) and information that is collected incrementally as treatment progresses (e.g. the numbers of eggs retrieved and embryos available for transfer). See Figure 1 for an illustration of the IVF treatment stages described below. The figure contains some notation on

utility payoffs that is introduced later.

Some notational conventions are common across stages. We write $W_{y,j}(Z_a, \varepsilon_{y,j,a})$ as the value of choice y during stage j of a treatment cycle. This function accounts for: expectations over future treatment outcomes, taste shocks in current and future stages, and optimal behavior in future stages. Patients' values of $W_{y,j}(Z_a, \varepsilon_{y,j,a})$ depend on τ, but we suppress this term and the i subscript for notational simplicity. Let $\overline{W}_{y,j}$ be the systematic component of $W_{y,j}(Z_a, \varepsilon_{y,j,a})$, i.e. $W_{y,j}$ net of the additive preference shock $\varepsilon_{y,j,a}$. We then have

$$W_{y,j}(Z_a, \varepsilon_{y,j,a}) = \overline{W}_{y,j}(Z_a) + \varepsilon_{y,j,a} \tag{1}$$

$E\left[W_j(Z_a, \varepsilon_{j,a})\right]$ represents the expected value from an optimal decision within treatment stage j, before observing the realization of the vector $\varepsilon_{j,a}$. The patient observes the realization of $\varepsilon_{j,a}$ before making her choice during stage j. The patient's value at the start of stage j is

$$W_j(Z_a, \varepsilon_{j,a}) = \max_{y \in Y_j} \{W_{j,y}(Z_a, \varepsilon_{y,j,a})\} = \max_{y \in Y_j} \{\overline{W}_{y,j}(Z_a) + \varepsilon_{y,j,a}\} \tag{2}$$

Due to the extreme value assumption for ε, we can write $E\left[W_j(Z_a, \varepsilon_{j,a})\right]$ with the inclusive value expression:

$$E\left[W_j(Z_a, \varepsilon_{j,a})\right] = \log \left\{ \sum_{y \in Y_j} \exp[\overline{W}_{y,j}(Z_a)] \right\}. \tag{3}$$

This is the patient's expected value from stage j before the values in $\varepsilon_{j,a}$ are realized.

We begin by focusing on the treatment stages that occur within IVF, after the patient has learned her value of Z^B. We return to the initiation decision later in this section.

4.1 Stage 1: Start treatment vs. delay

In all periods after the initiation decision, patients who began IVF previously will return to stage 1 and choose between the actions start (s) and delay (d). If the patient starts treatment, she pays the price $p_{s,1}(Z_a)$ out-of-pocket and begins a regimen of pharmaceuticals to promote egg production.

The value from starting a treatment cycle at age a is $W_{s,1}$, and it includes the expected value from continuing to the second stage of treatment ($E\left[W_2(Z_a, \varepsilon_{2,a})\right]$); the utility normalization relative to delay, u_s; the price of starting a treatment cycle, $p_{s,1}\left(Z_a\right)$; and a taste shock, $\varepsilon_{s,1,a}$. The value of the second stage depends on the realization of e (the Peak E2 score), but this is not known

during stage 1. The value from starting a treatment cycle at age a is then

$$\begin{aligned} W_{s,1}(Z_a, \varepsilon_{s,1,a}) &= \overline{W}_{s,1}(Z_a) + \varepsilon_{s,1,a} \\ &= u_s(Z_a) - \alpha(Z_a)p_s(Z_a) + \varepsilon_{s,1,a} + \sum_e E\left[W_2(e, Z_a, \varepsilon_{2,a})\right] f_e(e|Z_a) \end{aligned} \tag{4}$$

The value of delaying the IVF decision until the start of the next period is:

$$\begin{aligned} W_{d,1}(Z_a, \varepsilon_{d,1,a}) &= \overline{W}_{d,1}(Z_a) + \varepsilon_{d,1,a} \\ &= 0 + \beta E\left[W_1(Z_{a+1}, \varepsilon_{1,a+1})\right] + \varepsilon_{d,1,a} \end{aligned} \tag{5}$$

Changes in Z across periods, in this case, are due to the patient becoming older, which affects her fertility characteristics and the probability of a favorable outcome at any treatment stage. The discounted expected value $\beta E\left[W_1(Z_{a+1}, \varepsilon_{1,a+1})\right]$ accounts for the expectation of ε, the payoffs in $\overline{W}_1$ associated with starting or delaying IVF at age $a+1$, and patient's option to choose the optimal action. If the patient is already at age $a^{\max}$, however, she receives the terminal value $W_T(Z_{a^{\max}}) = u_T(Z_{a^{\max}})$ at the start of the next period and exits the model. This type of exit is also possible in stages 2 and 4, described below, but we do not list it explicitly.

4.2 Stage 2: Continue vs. cancel

The patient makes her next significant choice after the value of e is realized. A larger value of e is generally associated with a larger number of eggs (r) that are ready for retrieval from the patient's ovaries. During the second treatment stage, she considers e and her personal characteristics (Z_a) while deciding whether to continue (c) or cancel (nc) treatment, thus $Y_2 = \{nc, c\}$. If the patient cancels treatment, she pays no additional treatment fees, and she is able to consider starting treatment again in the future. If the patient continues treatment, she pays the additional fee $p_{c,2}(Z_a)$ and undergoes a surgical process in which eggs are retrieved.

If the patient decides to stop treatment, she receives the value

$$\begin{aligned} W_{nc,2}(e, Z_a, \varepsilon_{nc,2,a}) &= \overline{W}_{nc,2}(Z_a) + \varepsilon_{nc,2,a} \\ &= 0 + \beta E\left[W_1(Z_{a+1}, \varepsilon_{1,a+1})\right] + \varepsilon_{nc,2,a} \end{aligned} \tag{6}$$

The value of continuing treatment includes an expectation taken over values of r conditional on

the realized signal e and other patient characteristics:

$$\begin{aligned} W_{c,2}(e, Z_a, \varepsilon_{c,2,a}) &= \overline{W}_{c,2}(e, Z_a) + \varepsilon_{c,2,a} \\ &= -\alpha(Z_a)p_c(Z_a) + \varepsilon_{c,2,a} + \left(\sum_r E\left[W_3(r, Z_a, \varepsilon_{3,a})\right] f_r(r|e, Z_a) \right) \end{aligned} \tag{7}$$

The full value of the second stage is the maximum of these two options.

4.3 Stage 3: Fertilization

If treatment is not cancelled, the patient's eggs are retrieved and she observes the realized value of r. The patient's next choice is how to fertilize the eggs. The fertilization method is represented by the variable m, and the patient's options are: natural fertilization (m_1) or with ICSI (m_2). Thus $Y_3 = \{m_1, m_2\}$. The patient's characteristics (Z_a), her number of eggs (r), and her fertilization choice (m) determine the number of viable embryos generated for the patient. Couples with male factor infertility ($z_{mf} = 1$) are likely to receive the greatest benefits from fertilizing via ICSI ($m = m_2$). When $m = m_2$, the patient pays the additional price $p_{m,3}(Z_a)$.

Let X represent a possible realization for the number of embryos. Possible values of X are in $\{0, 1, 2, 3, 4+\}$. We cap the maximum value of X at 4 because this is the greatest number of embryos that we see transferred to patients during the final treatment stage. In practice, patients may choose to freeze excess embryos for potential later use, but we do not examine that decision. Frozen-embryo cycles account for only 12% of the clinic's treatments during the sample period. When making her choice over fertilization method, the patient considers the probability of receiving X embryos, $f_X(X|r, m, Z_a)$. We write the patient's choice-specific value from a third-stage action:

$$\begin{aligned} W_{m,3}(r, Z_a, \varepsilon_{m,3,a}) &= \overline{W}_{m,3}(r, Z_a) + \varepsilon_{m,3,a} \\ &= -\alpha(Z_a)p_{m,3}(Z_a) + \varepsilon_{m,3,a} + \left(\sum_X E\left[W_4(X, Z_a)\right] f_X(X|r, m, Z_a) \right) \end{aligned} \tag{8}$$

The patient selects the action, m, with the greater of two $W_{m,3}(r, Z_a, \varepsilon_{m,3,a})$ values.

4.4 Stage 4: Embryo transfer

At the start of the fourth and final treatment stage, the patient learns her number of viable embryos, X. The patient chooses x, the number of embryos to transfer during the final treatment stage, subject to $x \leq X$. We assume that the patient selects $x = 0$ only if $X = 0$. A patient's

treatment outcome is influenced by her number of embryos (x) and her personal characteristics (Z). As a result of treatment, k children are born with probability $f_k(k|x, Z_a)$. There is no price for this treatment stage. If treatment fails she moves to the start of the next period, but if treatment is successful she waits for three additional periods (i.e. 9 months) before making her next reproductive decision.

When the patient elects to transfer x embryos, she receives an expected benefit of

$$\begin{aligned} W_{x,4}(X, Z_a, \varepsilon_{x,4,a}) &= \overline{W}_{x,4}(X, Z_a) + \varepsilon_{x,4,a} \\ &= \eta(x, Z_a) + \varepsilon_{x,4,a} + f_k(0|x, Z_a)\beta E\left[W_1(Z_{a+1}, \varepsilon_{1,a+1})\right] \\ &\quad + \left(\sum_{k>0} f_k(k|x, Z_a)\left\{U(k|\widetilde{k}, Z_a) + \beta^4 E\left[W_1(Z_{a+4}, \varepsilon_{1,a+4})\right]\right\}\right) \end{aligned} \tag{9}$$

This expression includes the possibilities of failed treatment ($k = 0$) and successful treatment ($k > 0$). The future value of a patient's decision, $E\left[W_1(.)\right]$, will depend on the realization of the current treatment. If the treatment is successful, Z_a will evolve to a value Z_{a+4} which reflects that the patient is one full year older and has k additional children. Moreover, this future value is discounted at β^4. If treatment fails, then the next decision's value is discounted by β and Z_{a+1} reflects that the patient is just three months older.

4.5 Initiation decision

Now consider the decision of a potential patient at age a_0 who is deciding whether to start IVF for the very first time. This is somewhat different from the decision to begin a new cycle by an already-active patient. This potential patient does not yet know her values of Z^B, but she knows the population distribution of Z^B values conditional on age, $f_{Z^B}(Z^B_{a_0}|a_0)$. The potential patient's expected value from starting treatment is:

$$W\left(Z^D_{a_0}\right) = E\left[\overline{W}_{s,1}(Z_{a_0})|Z^D_{a_0}\right] = \sum_{Z^B_{a_0}} \overline{W}_{s,1}(Z^D_{a_0}, Z^B_{a_0}) f_{Z^B}(Z^B_{a_0}|a_0),$$

where we make the distinction between the state variables known prior to treatment ($Z^D_{a_0}$) and those learned after treatment begins. The potential patient compares $W\left(Z^D_{a_0}\right)$ to the utility from foregoing treatment, which we specify as $W^{OUT} = \mu + \nu$. The parameter μ captures the mean value of the outside option, and it is common across potential patients. ν is specific to each

potential patient and captures heterogeneity in the value of permanently foregoing IVF treatment and explains why potential patients with the same $Z_{a_0}^D$ and (unobserved) τ make different choices with respect to ever pursuing IVF.[20] One possible interpretation of ν is that of a sunk utility cost that must be paid to pursue IVF. Under these assumptions, the potential patient becomes a patient (i.e. enters the clinic to initiate her first IVF cycle) if the expected value of pursuing IVF is higher than the value of foregoing treatment, and she exits the model otherwise. We let the indicator I equal one whenever a potential patient enters the clinic to pursue IVF treatment, and equal zero otherwise. Then

$$I = 1 \Leftrightarrow W\left(Z_{a_0}^D\right) \geq \mu + \nu.$$

To make the decision problem more explicit in some of the analysis below, we write the value associated with starting a very first cycle as $\overline{W}_{s,1}\left(Z_{a_0}, g\right)$, where g indexes potential policy environments, and $g = g_B$ represents the empirical "baseline" that we observe during the sample period. The index g captures elements such as pricing, insurance, technology, regulations, etc. Under alternative environments g' the value of $\overline{W}_{s,1}$ will change and therefore initiation decisions will be affected. In fact, all $\overline{W}_{y,j}$ values change when the policy environment changes.

5 Data

5.1 Clinic data

Our primary data cover individual patient histories at the clinic during 2001-09. We observe all treatment cycles conducted during this period for patients who underwent their first IVF cycle between 2001 and 2007. While these data allow us to describe a patient's IVF history from the start of her treatments, we do not observe whether a patient returns to the clinic after 2009 or visits a different clinic after her final visit at the clinic. We handle this potential right-censoring by assuming that patients continue to make choices as described by our model, with no changes to the policy environment, prices, or technology.

The main data sample contains treatment histories for 587 patients who use only fresh embryos (i.e. not frozen) and have complete data on their personal characteristics and treatment details. We supplement these observations with data from an additional 519 patients for whom we have data on all state variables and most treatment choices. We refer to the expanded data as the "first-

[20] Note that there is no $\varepsilon_{1,s}$ for this very first cycle.

stage sample." In Table 1 we display some basic characteristics of the patients, their treatment choices, and their outcomes; we separately report statistics for the main sample of 587 patients and the 1106 patients in the full first-stage sample. The average patient in the main sample is 34 years old at the time of her first cycle in the clinic, and over half of all patients have insurance. Most patients' homes are in a zip code with a median house price above $100,000, which we use as a proxy for patient wealth (z_w). The patients in the main sample have no children when they initiated treatment, but some patients in the first-stage sample have prior children. The biological variables (Z^B) exhibit some minor differences between the main and first-stage samples, with the former set of patients displaying slightly worse fertility characteristics.

At the bottom of Table 1 we display patient-level statistics on treatment choices and outcomes. Patients in the main sample average 1.75 treatments during the sample period, and about half experience at least one birth during their full treatment history. In Table 2 we report summary statistics on choices and outcomes within treatment stages. Most patients at stage 2 choose to continue treatment, with only a 14% cancellation rate. Most patients (60%) fertilize their eggs with ICSI; this rate is closer to 90% when male-factor infertility is present.[21] Finally, patients take 2.3 embryos on average during a treatment. The embryo transfer choices are most often made with a choice set of 4+ embryos, due to over 6 embryos being generated during an average cycle. At the bottom of Table 2 we report treatment-level outcomes. To obtain the main sample's average of 0.51 children born per cycle, we include all stage 4 decisions with $x > 0$ and birth outcomes in $\{0, 1, 2, 3\}$. A singleton birth occurs in 27% of cycles, and twins occur in an additional 12%. While we observe no triplet births in the main sample, they occur at a rate of about 1% in the larger first-stage sample; this allows us to account for triplet risk when estimating the structural model.

Some correlations among patient characteristics and treatment sequences suggest the role of dynamics and the importance of the state variables in patients' decision-making. Conditional on having one successful cycle at the clinic, 8% of patients return to start another cycle. Patients who receive Peak E2 scores in the lowest quartile chose to cancel treatment in 40% of all cases, while patients with scores in the $25^{th} - 75^{th}$ percentile cancel only 4% of cycles. Patients who are 35 or older take an average of 2.6 embryos in their first cycle, while younger patients take 2 embryos on average. Uninsured patients take more embryos (2.4) during their first cycle than insured patients (2.2), but this difference shrinks in the second and third cycle, as insurance coverage is drawn down.

[21] While the options "full ICSI" and "partial ICSI" are separated in the data, we group them together in our model.

5.2 Market data

We use several pieces of market data to describe the set of potential patients for our clinic. These data are used in a separate estimation step to estimate a model of treatment initiation. We assume that potential patients are drawn from all zip codes with centroids within 75 miles of our clinic. The area includes the city of St. Louis, its surrounding suburbs, and some rural towns outside of the metro area. This area captures almost all of the patients who ever visit the clinic; a small number come from greater distances.

We first describe the various data sources for the market data, and then we describe how they are assembled into an estimate of the "at risk" population. We use the Centers for Disease Control and Prevention's (CDC) Vital Statistics database to construct the market's distribution of maternal age at first birth. This distribution, along with estimates of infertility rates by age from Dunson et al. (2004), allows us to construct an age distribution for women who may consider IVF. For zip-code level information on the population share with private IVF insurance, we use data from the 2012 American Community Survey (ACS) and combine it with other sources of information, which we describe in the appendix. We also collect data on the median home value for each zip code in the area, taken from 2000 Decennial Population Census. This allows us to provide an estimate for the distribution of patients' wealth. Combining the zip code level data on home values and IVF insurance coverage we can then construct an estimate of the joint distribution of IVF insurance coverage and our measure of wealth. Finally, we use data from the CDC on the number of cycles conducted at each infertility clinic in the market to assess how many in the pool of potential patients would rely on our clinic (rather than a different clinic), if they decided to pursue IVF.

Next, we describe some of the steps we use to construct the pool of potential patients. Assuming stationarity and stable cohort sizes, at any given point in time (quarter) there are N^{stl} couples in the St. Louis region who have optimal life cycle fertility plans that induce them to pursue their first pregnancy. Therefore, every quarter t there is a distribution of age at first (attempted) birth for these women $f_t(a)$. Some of them will succeed immediately, some will take more time. If, after 12 months of natural attempts, the woman does not get pregnant, the couple is diagnosed with clinical infertility. Let $f_{inf}(inf|a)$ be an age-specific infertility rate, which increases with age. Together $\left(N^{stl}, f_t(a), f_{inf}(inf|a)\right)$ provide the number of women of each age $\widetilde{N}_a^{inf}$ that realize that they are unable to conceive without IVF. These $\widetilde{N}^{inf} = \sum_{a=a^{\min}}^{a^{\max}} \widetilde{N}_a^{inf}$ women constitute the *risk set,* i.e. all women in the St. Louis region who may consider IVF treatment. In a final step, we obtain the

risk set for our clinic, N^{inf}, by deflating $\widetilde{N}^{inf}$ to match the clinic's market share as reported by the CDC. Our final estimate of $N^{inf} = 2146$. See Appendix B for additional detail on the calculation of N^{inf}.

We compute an empirical initiation share, s^{init}, for the clinic we study. We observe that $N^{clin} = 828$ new patients initiated treatment at the clinic during the period 2001-2007.[22] Using our estimate of N^{inf}, we calculate the share

$$s^{init} \approx \frac{N^{clin}}{N^{inf}} = \frac{828}{2146} = 0.39, \tag{10}$$

which means that 39% of the clinic's potential IVF patients decided to pursue treatment. The remaining 1318 potential patients could be induced to seek IVF treatment through large enough increases in $\overline{W}_{1,s}$.

6 Empirical specification

In this section we describe our assumptions regarding functional forms and how outcomes and utility may vary with patients' observable characteristics.

6.1 Treatment technologies

During each treatment stage, a patient makes her choice while considering a probability distribution over outcomes that will be realized at the stage's conclusion. We now describe the functional forms and data assumptions that describe the distributions.

In the first stage, a woman knows some basic facts about her fertility including Z^B, and takes drugs to stimulate egg production. While we observe drug dosage, we do not model the choice, so we assume that dosage is selected deterministically based on the patient's characteristics. The woman's characteristics and (unmodeled) drug decision affect a stochastic process that determines her Peak E2 score, e. We model the probability of a particular e with a multinomial logit model for $f_e(e|Z_a)$. In the data we observe e values between 0 and 10,196 pg/mL, with a mean and median around 1,600, and 99% of all values below 4,500. In the empirical implementation, we assume that the possible realizations of e are in discrete bins with values 0-500, 500-1000, 1000-1500, 1500-2000, 2000-2500, and over 2500. We use a multinomial logit model here rather than an ordered model

[22] We use the 587 with complete data in estimation, but have records for 828 patients initiating treatment over this period.

because especially high values of e can be seen as bad for the patient.

In estimating f_e, we include variables for a woman's age, the average of any AFC scores she receives over the entire treatment history, and her number of initially diagnosed fertility problems. The age variables we include are indicators for whether the patient's age is: 28, 29-31, 32-34, 35-37, 38-40, 41-43, or 44. (We exclude age 35-37 for the empirical implementation.) We separate the patient's AFC score (z_{afc}) into categories for scores from 1-5, 6-10, 11-15, 16-25, and 26+, with the highest category excluded for the empirical implementation. For patient fertility problems, we include an indicator for whether the patient has one or more distinct diagnosed issues ($z_{ff} = 1$).

In the second stage, the patient observes her realized value of e and considers the number of eggs, r, that might be retrieved if she continues treatment. The distribution of r depends on e and Z_a. In the data, r takes integer values from 0 to 38 with a mean of 10.6 and median of 10. The 90th percentile is at $r = 18$, and 99% of all r values are below 27. We use an ordered probit model for this distribution, with possible values of r as 0-4, 5-10, 11-20, and 21+. The variables that can affect the realization of r are: indicators for possible values of e, split as they are in the model for f_e; the same age categories in f_e; the AFC score categories from f_e; and the indicator for whether a patient has one or more documented fertility problems.

In the third stage, the patient observes her realized value of r and selects a fertilization method (m). The patient's number of transferable (cleavage-stage) embryos X will depend on r, m, and the patient's characteristics. We model the process determining X with an ordered probit. We include as regressors: the possible values of r as described in the model for f_r; the patient's age, AFC score, and fertility problems as described above; and the patient's choice of m plus the interaction of m with an indicator for male-factor infertility.

In the final stage of treatment, the patient is subject to the stochastic process f_k, which determines her number of live births. We model f_k as a multinomial logit, with the probability of each outcome determined by the number of transferred embryos, the patient's age, and the indicator for female fertility problems. Some patient and treatment characteristics, like AFC score or male factor infertility, are not relevant here because their role in determining outcomes is finished once the patient has her cohort of transferable embryos.

6.2 Utility assumptions

We must make functional form assumptions for several expressions that are relevant for patients' utility. In addition to the restriction that all patients have the payoff of $U = 0$ from zero-birth

outcomes, we assume that outcomes with $k > 0$ provide utility according to:

$$U(k|\widetilde{k}, \tau) = u_k + \kappa \times 1\{\widetilde{k} > 0\} + \zeta \times 1\{\tau = 2\}$$

The vector (u_1, u_2, u_3) contains parameters that (respectively) capture the lump-sum payoff from a singleton, twin, and triplet birth to a patient with no prior children ($\widetilde{k} = 0$). Given the health risks and other challenges for triplets, we anticipate that $u_3 < u_2$ and $u_3 < u_1$, but these parameters are unrestricted in estimation. The parameter κ captures any difference in the marginal benefit of a birth to patients with prior children; diminishing marginal utility from children would imply that κ is negative. For patients who deviate from ASRM guidelines, we assume a constant utility penalty $\eta(x, Z_a) = \eta_0 \times 1\{x, Z_a\}$, where $1\{x, Z_a\}$ is an indicator function that is equal to one when x embryos is outside of ASRM guidelines for a patient with state variables Z_a.

We assume a simple two-type structure for patients' permanent unobserved heterogeneity. A share of patients with type $\tau = 1$ has preferences for birth outcomes represented only by (u_1, u_2, u_3, κ), while the remaining patients (with $\tau = 2$) has, in addition, its utility payoff shifted by a scalar parameter ζ. A patient's probability of being of type $\tau = 2$ depends on her state values at the time she initiated treatment, $Z^D_{a_0}$. Along with a_0, we allow the distribution of τ to depend on a measure of her wealth level (z_w), her initial number of insurance-covered cycles (ι_{a_0}), and a dummy (z_{asrm_0}) for the ASRM guideline regime when treatment started. We assume that the probability of a high type ($\tau = 2$) is

$$\Pr(\tau = 2|Z^D_{a_0}, I = 1, \rho) = \frac{\exp(\rho_0 + \rho_1 a_0 + \rho_2 z_w + \rho_3 \iota_{a_0} + \rho_4 z_{asrm_0})}{1 + \exp(\rho_0 + \rho_1 a_0 + \rho_2 z_w + \rho_3 \iota_0 + \rho_4 z_{asrm_0})}.$$

During estimation we restrict $\rho_0 < 0$ for computational purposes, but this adds no real restrictions on the utility parameters. For notational convenience, we let ρ represent a column vector of $\{\rho_j\}_{j=0}^4$ values. In addition, we write $[1, Z^D_{a_0}]$ as a vector containing 1 and an individual patient's row vector $Z^D_{a_0}$, and we let Λ represent the logistic distribution function so that $\Lambda([1, Z^D_{a_0}]\rho) = \Pr(\tau = 2|Z^D_{a_0}, I = 1, \rho)$.

As the patient makes her choice between starting a treatment cycle or delaying, she considers the additional flow benefit u_s which she receives (or pays) when she begins a treatment cycle. We assume that $u_s = \delta_0$, a scalar parameter. The value of u_s will be identified, in part, by the frequency with which clinic patients return for additional treatment cycles following their first cycle.

The first three stages of IVF treatment include $\alpha(Z)$, the disutility from paying a price p for

some treatment component. We specify $\alpha(Z)$ so that it is allowed to vary with a patient's initial wealth: $\alpha(Z) = \alpha_0 + \alpha_w z_w$. Since the effect of price is subtracted from within-stage value functions above, we expect α_0 to be positive for consistency with downward-sloping demand. If wealthier patients are less price sensitive, this will be captured through $\alpha_w < 0$.

We assume that the terminal payoff u_T is a function of the patient's cumulative payments for treatment. Children born due to treatment are not included here because those benefits are included in U. We add the variable z_p as an indicator for whether a patient ever paid full price for a treatment cycle. We assume $u_T = \gamma_p z_p$, which includes the normalization $u_T = 0$ for patients who have never paid the full price of treatment.

At the initiation stage we specify that the individual-specific taste shock ν is distributed according to a continuous distribution $F(\nu)$ in the population of potential patients. The realizations of ν are i.i.d. Moreover, we assume ν is independent of infertility problems and other observables in our model, so $F\left(\nu|\ \tau, Z_{a_0}^D\right) = F(\nu)$. We assume $\nu \sim Logistic$ so we have $F(\nu) = \Lambda(\nu) = \frac{\exp(\nu)}{1+\exp(\nu)}$.

Let φ represent a vector of all of the parameters except μ, ζ, and ρ, and define $\theta = (\zeta, \varphi, \rho)$. We estimate μ separately from θ so it is convenient for us to distinguish between the two.

7 Estimation

We estimate the model in three stages. We estimate the treatment technologies, f_e, f_r, f_X, and f_k in the first stage. These models are easy to estimate using conventional statistics packages. We use the parameter estimates from this estimation step to characterize the stage-specific distributions of treatment outcomes for each possible unique value of the state vector and each possible stage-specific action a patient may take. We implicitly assume that we as econometricians have the same information on outcome probabilities as the patient and her doctor. Also within this stage we estimate the distribution of $f_{Z^B}(Z^B|a_0)$ non-parametrically using frequencies of Z^B realizations from within the population of women who initiate treatment. In the second stage we estimate the parameters in θ using data exclusively from the population of 587 patients who are observed within the clinic. In the final stage we estimate μ using our estimates of $E\left[\overline{W}_{s,1}(Z_{a_0})|Z_{a_0}^D\right]$ together with the market-level data.

7.1 Within-clinic choices

Given the estimated treatment technologies, a guess at the value of the structural parameters in (ζ, φ), and the distributional assumptions on ε, we are able to calculate $\overline{W}_{y,j}(Z_a, \tau; \zeta, \varphi)$ and $E\left[W_j(Z_a, \tau; \zeta, \varphi)\right]$ for each y and j at every Z_a. We perform this calculation by backward recursion separately for each type τ. For each potential state that might be reached when the patient is age $a^{\max}$, we use (ζ, φ) to compute the terminal payoff, the values of $\overline{W}_{y,j}(Z_{a^{\max}}, \tau; \zeta, \varphi)$ working backwards through treatment stages, and the logit inclusive value $E\left[W_j(Z_{a^{\max}}, \tau; \zeta, \varphi)\right]$ for each stage. We then move to age $a^{\max} - 1$ and use the $a^{\max}$ expected utility values while constructing $\overline{W}_{y,j}(Z_{a^{\max}-1}, \tau; \zeta, \varphi)$ and $E\left[W_j(Z_{a^{\max}-1}, \tau; \zeta, \varphi)\right]$. The procedure continues back to age $a^{\min}$.

Let $d_{y,j,a,i} \in \{0, 1\}$ represent patient i's binary choice whether to take action y in stage j while at age a. We write d_i as the patient's complete history of choices at the clinic. We use the calculated values of $\overline{W}_{y,j}(Z_a, \tau; \zeta, \varphi)$ for all Z_a and τ to compute choice probabilities for each observed decision in our data. Conditional on a patient's type τ, calculating this probability is a straightforward task due the i.i.d. extreme value assumption for the ε terms. For example, conditional on a patient reaching at age a a stage-2 decision over whether to continue (c) or cancel (nc) the current treatment cycle, her probability of continuing is:

$$\Pr(d_{c,2,a,i} = 1 | Z_a, \tau; \zeta, \varphi) = \frac{\exp[\overline{W}_{c,2}(Z_a, \tau; \zeta, \varphi)]}{\exp[\overline{W}_{c,2}(Z_a, \tau; \zeta, \varphi) + \overline{W}_{nc,2}(Z_a, \tau; \zeta, \varphi)]} \tag{11}$$

The values of $\overline{W}_{c,2}(Z_a, \tau; \zeta, \varphi)$ and $\overline{W}_{nc,2}(Z_a, \tau; \zeta, \varphi)$ are relatively simple functions of the estimated transition $\widehat{f}_r$, price and its disutility parameter, and the calculated values of $E\left[W_3(r, Z_a, \varepsilon_{3,a})\right]$ and $E\left[W_1(Z_{a+1}, \varepsilon_{1,a+1})\right]$. We calculate a probability like this one for each observed decision by each patient, including the implicit choices to delay further treatment attempts which occur during periods when the patient does not appear in the data despite starting treatment during some earlier period.

A patient's permanent unobserved type, τ, affects every period and stage of her decision problem. Let $\Pr(d_{y,j,a,i} = 1; Z_a, \tau, \zeta, \varphi)$ represent the predicted probability that patient i took her observed action $d_{y,j,a,i}$ if she were of type τ. The patient is observed starting in period $t_{i,0}$ and ending in T_i. Conditional on ζ and φ, the type-specific joint probability of observing patient i's

sequence of choices is:

$$L_i(d_i;\tau,\zeta,\varphi) = \prod_{a=a_{i,0}}^{T_i}\prod_{j=1}^{4}\prod_{y=1}^{Y_j}\Pr(d_{y,j,a,i}=1;Z_a,\tau,\zeta,\varphi)^{d_{y,j,a,i}}.$$

With i's true type τ unobserved, the likelihood of observing her choices requires integration over τ, which is simply

$$L_i(d_i;\theta) = \textstyle\sum_\tau L_i(d_i;\tau,\zeta,\varphi) f_\tau(\tau|Z_{i,a_0}, I=1,\rho).$$

The log-likelihood of observing the choices of all patients in the clinic data is

$$\mathcal{L}(\theta) = \sum_i \log[L_i(d_i;\theta)].$$

We estimate θ by maximizing the value of $\mathcal{L}(\theta)$. We compute standard errors following the "outer product of the score" method for θ only. In computing standard errors we do not account for potential sampling error in our first stage estimates.

7.2 Treatment initiation

We estimate the initiation decision in a third step, taking the within-clinic estimates from the second step $\widehat{\theta} = (\widehat{\zeta},\widehat{\rho},\widehat{\varphi})$ as given. Under our assumptions about initiation, we may write

$$\Pr\left(I=1|Z^D_{a_0},\tau,\varphi,\zeta,g_B,\mu\right) = \Lambda\left(W\left(Z^D_{a_0},\tau,g_B,\varphi,\zeta\right)-\mu\right), \tag{12}$$

where the variable I indicates whether a patient with characteristics $Z^D_{a_0}$ started treatment at age a_0. We integrate over observed and unobserved patient characteristics to compute the rate at which the clinic's potential patients actually become patients, $\widehat{s}^{init}\left(\mu;g_B,\widehat{\theta}\right)$. These distributions, $f_{Z^D}(Z^D_{a_0})$ for observed characteristics and $f_\tau(\tau|Z^D_{a_0})$ for unobserved heterogeneity, will in general be different in the full potential patient pool versus among patients who choose to initiate treatment. We estimate the remaining parameter μ by solving $\widehat{s}^{init}\left(\mu;g_B,\widehat{\theta}\right) = s^{init}$ for the empirical policy setting, g_B. Under alternative policies, $W\left(Z^D_{a_0},\tau,g,\varphi,\zeta\right)$ will change but μ remains fixed.

We approach the distributions $f_\tau(\tau|Z^D_{a_0})$ and $f_{Z^D}(Z^D_{a_0})$ using different strategies. In Appendix A we show that our assumptions on: a) the distribution of unobserved types conditional on treatment, $f_\tau\left(\tau|Z^D_{a_0},I_i=1,\rho\right)$; and b) the initiation decision, are sufficient to back out the unconditional distribution $f_\tau(\tau|Z^D_{a_0})$. We write the frequency of $\tau=2$ within-clinic as $\Lambda([1,Z^D_{a_0}]\rho)$,

and $\Lambda\left(W\left(Z_{a_0}^D, \tau, g_B, \varphi\right) - \mu\right)$ provides the probability of initiation for $\tau = 1, 2$. In Appendix B we describe our approach to constructing $f_{Z^D}(Z_{a_0}^D)$ using market data.

The estimates for $f_\tau\left(\tau | Z_{a_0}^D\right)$ and $f_{Z^D}\left(Z_{a_0}^D\right)$ allow us to to derive $s^{init}\left(\theta, \mu, g_B\right)$, the model-predicted fraction of potential patients who walk into the clinic (i.e. the percentage of potential patients who actually become patients). To obtain $s^{init}\left(\mu; g_B, \theta\right)$, we integrate $\Pr\left(I = 1 | Z_{a_0}^D, \theta, \tau, \mu\right)$ over the distribution of $Z_{a_0}^D$ and τ among potential patients

$$\begin{aligned} \widehat{s}^{init}\left(\mu; g_B, \widehat{\theta}\right) &= \Pr\left(I = 1 | \widehat{\theta}, \mu, g_B\right) = \\ & \sum_{Z_{a_0}^D} \left[\sum_{\tau} \Lambda\left(W\left(Z_{a_0}^D, \tau, g_B, \widehat{\varphi}, \widehat{\zeta}\right) - \mu\right) f_\tau\left(\tau | Z_{a_0}^D, \mu, \widehat{\theta}\right) \right] f_{Z^D}\left(Z_{a_0}^D\right) \end{aligned} \quad (13)$$

We then estimate μ as the value that solves $s^{init} = \widehat{s}^{init}\left(\mu; g_B, \widehat{\theta}\right)$.

We use the sampling distribution of $\widehat{\theta}$ to construct a confidence interval on $\widehat{\mu}$. We draw 400 times from $\widehat{\theta}$'s distribution, and for each draw we calculate the value of μ that equates s^{init} and $\widehat{s}^{init}$. We then sort the individual estimates of μ, and then use the 2.5^{th} and 97.5^{th} percentile values as the 95% confidence interval. The confidence interval on μ is not interesting in own right, but it plays a critical role in describing the precision of predicted treatment-initiation decisions, which we discuss below.

8 Results

8.1 Technology estimates

In this subsection we discuss our estimates of the four treatment stages' technologies. These technologies are dependent on a patient's characteristics, and a patient's knowledge of them is a crucial part of how she solves her personal dynamic optimization problem. Rather than providing parameter estimates for each treatment technology, we use a collection of figures to discuss the role each technology plays in the choice process. One of our overall goals is to emphasize the importance of allowing forward-looking dynamic behavior at each treatment stage.

During the first treatment stage, the patient decides whether to start or delay an IVF cycle. She is aware of her full state vector, Z, which includes her AFC score, z_{afc}. At this point in the decision process, she considers her probable peak estradiol score (e), which will be revealed in Stage 2 if she starts treatment. In Figure 2 we display probability distributions over e for two AFC score

categories. The figure shows that having an AFC score below 5 substantially shifts to the left the distribution of values of e that the patient can expect to realize at the beginning of stage 2.

The patient cares about her value of e because it affects outcomes in later stages. In Figure 3 we show that the realized value of e influences the distribution of the number of eggs that will be successfully retrieved (r) in stage 3. Indeed, if e is low (e.g. in the 500-1000 range) the mode of the distribution of eggs is 6-10 whereas if e is relatively high (2000-2500) the mode of the distribution of eggs is 11-20. Moreover, if e is high the probability of having a low retrieved egg count (1-5) is almost zero. This strong difference in r outcomes at different values of e justifies our treatment of e as a within-period state variable that is critical to continuation/cancellation decisions in stage 2.

In treatment stage 3, a patient chooses her fertilization method (m). This choice, interacted with the patient's state variables, influences the distribution of available embryos (X) in stage 4. In Figure 4 we display the distributions of X with (m_2) and without (m_1) ICSI for patients whose partners have male-factor infertility. The figure shows that the more technologically advanced fertilization method (ICSI) shifts the distribution to the right, increasing the probability of having 4 or more viable embryos and reducing the probability of having a small embryo count .

Once the patient has realized her value of X, she chooses the number of embryos (x) to transfer back into the uterus subject to $x \leq X$. In Figure 5 we display evidence on how x affects the distribution of births (k). Transferring 3 embryos instead of 2 reduces the chance of no birth from about 60% to under 50%, but the probabilities of twins and triplets increase. It is important to notice, however, that the probability of having no live births is fairly high regardless of whether 2 or 3 embryos are transferred. Finally, in Figure 6 we explore the effects of age. We focus on patients who transfer $x = 3$ embryos in stage 4. As expected the distribution for older (>35) women shifts to the left, noticeably increasing the odds of no live birth.

8.2 Utility parameters

Taking as inputs the technology parameters described above, we estimate the model's structural taste parameters. In Table 3 we display our estimates of $U(k|\widetilde{k})$, α, δ_0, γ, and η. Our estimates of u_1, u_2, and u_3 represent payoffs from different birth outcomes to patients with $\tau = 1$ and no prior children. These estimates show that patients receive a positive payoff from a singleton or twin birth, with the latter valued slightly more. Triplet births, by contrast, have a negative utility payoff for patients. The estimate of κ indicates that patients with 1 or 2 prior children have their utility from births shifted downward substantially. For example, for a patient with $\widetilde{k} > 0$ and $\tau = 1$,

the estimated κ implies that the patient would prefer no additional children. The taste shifter ζ associated with type 2, however, is sufficient to increase the utility from additional births to be positive for patients with $\widetilde{k} > 0$.

Table 3's results indicate that the baseline price disutility is significantly different from zero for all patients, but this disutility is smaller for patients in the top portion of the wealth distribution. (Recall that we subtract α from patient utility, so a negative α coefficient on z_w indicates reduced price sensitivity.) We recover a significantly negative estimate for the start/delay parameter δ_0, which plays a large role in determining whether a patient returns for additional treatment cycles after her first. The negative value of δ_0 may represent the physical or psychological stress in undergoing IVF. Our estimate of the parameter γ for a patient's terminal payoff u_T shows no significant difference between the utility of patients who have paid out-of-pocket for a treatment and those who have not. The final utility parameter on Table 3 is the utility shifter from selecting an x outside of ASRM embryo transfer guidelines. We recover a negative value for this parameter, indicating a penalty for deviating from the guidelines.

Table 4 reports results on the distribution of τ. We estimate that about half of the patient population has type $\tau = 2$ given their Z values. To interpret the individual ρ parameters, consider the case of patient wealth. The negative coefficient (ρ_2) on the wealth measure indicates that a high-wealth person selected from the treated population is less likely to have type $\tau = 2$ than a random low-wealth person. This accords with the intuition that treatment expenses are most likely to discourage low-wealth individuals with relatively small payoffs from having children through IVF.

Finally, in the third estimation step we recover $\widehat{\mu}_0 = -0.76$.[23] This value of μ ensures that the initiation model generates treatment initiation decisions such that, as estimated from our data, 39% of potential clinic patients indeed choose to become clinic patients and undergo at least one IVF cycle.

8.3 Model fit

We conduct two procedures to evaluate model fit. First, we contrast the estimated model's predicted choice probabilities to those we observe in the data. This provides a straightforward way to examine choice probabilities at the four stages of IVF treatment. Comparisons of the predicted and observed choice probabilities are displayed in Figures $7 - 10$. We omit the patient's initial choice to begin

[23] The 95% confidence interval for μ is $[-1.54, 0.49]$

her first cycle at the clinic. All predictions match the data fairly well. Start/delay decisions, which are observed most frequently in the data (and are assisted by the intercept term δ_0) have the tightest fit. Stage 2 and 3 predicted decisions also follow the data fairly closely but there are noticeable differences in the rate of treatment cancellations (stage 2) and ICSI use (stage 3). Some differences are to be expected, however, because these stages' fits depend on overall W_j values rather than individual parameters. Our predicted stage-4 choice succeeds in matching $x = 2$ as the most common choice, followed by $x = 3$. Transfers of 1 and 4 embryos are rare in the data (and model) because of the utility penalty for deviating from ASRM guidelines and the negative payoff from a triplet birth (in the case of $x = 4$).

In a second set of exercises, we evaluate the predicted choice and outcome histories for the population of 587 observed patients. These histories begin with the same state variables (Z) as the patients in the data, but then random draws on medical outcomes and taste shocks determine choices and outcomes over time. For each patient we repeat the process ten times, allowing for the realization of different taste shocks and stochastic medical outcomes. We average over patients and their individual simulated histories in computing the statistics we report below.

We focus on two critical measures of effectiveness and efficiency of IVF treatment. First we ask: What proportion of patients eventually succeed in delivering at least one live birth through IVF, regardless of the number of attempted cycles required to do so? We find that 59% of our simulated patient histories include a birth, which is reasonably close to the empirical value of 53% reported on Table 1. Second, we investigate how many cycles an individual patient receives at the clinic. In our simulation, 53% of patients are observed taking a single cycle, 28% undergo two cycles, and 19% receive three or more cycles. These results compare very well to the data, in which we see 54%, 27%, and 19% of patients receive one, two, or three or more cycles, respectively.

9 Counterfactual experiments

We use the model estimates to consider a set of counterfactual policy experiments which analyze potential IVF patients' responses to changes in their decision environment (g). Extensive-margin choices are crucial for this analysis, so we employ the full "at risk" population of $N^{inf} = 2146$ potential patients described above.[24] These potential patients represent the portion of the St.

[24] Women's human capital accumulation, career decisions, age at marriage, and age at first birth may be affected by some of the changes we implement in our counterfactuals experiments. These responses might have an effect on the size and composition of the pool of potential patients. We abstract from these considerations. Buckles (2005), Abramowitz (2014) and Gershoni and Low (2015) explore the existence and magnitude of such effects.

Louis market served by the clinic we study. While we do not discuss other clinics in the market, in our counterfactuals we implicitly assume that all clinics are subject to the same policies. When considering absolute magnitudes below (e.g., numbers of births, dollar values) these figures can be multiplied by about three to understand the impact of a policy on outcomes in the St. Louis market as a whole. In 2012, St. Louis clinics performed about 1% of all cycles in U.S. clinics.

For each potential patient, we draw age, wealth, insurance, and ASRM regime values that are consistent with the empirical distributions of these values. Along with the distribution of biological state variables (not yet revealed to potential patients), we use the estimated model to construct $W\left(Z_{a_0}^D, \tau, g, \widehat{\varphi}, \widehat{\zeta}\right)$ for each simulated woman. The values of $W\left(Z_{a_0}^D, \tau, g, \widehat{\varphi}, \widehat{\zeta}\right)$ differ across policy experiments. We then allow potential patients to elect whether to begin treatment by comparing $W\left(Z_{a_0}^D, \tau, g, \widehat{\varphi}, \widehat{\zeta}\right)$ to the population-wide utility parameter $\widehat{\mu}$ and a simulated value for the potential patient's taste shock ν. For all potential patients, we simulate initiation choices and decision histories in the same way described above for evaluating model fit, including repeating the process ten times for each potential patient in N^{inf}. Potential patients who do not start treatment at a_0 exit the model forever.

We assume that the N^{inf} simulated potential patients arrive at the fertility decision uniformly over the 2001-07 window during which the 587 observed clinic patients began treatment. As in the data used for estimation, the simulated patients' histories are followed from their initiation decision through 2009. To maintain consistency with our empirical model, we focus on counterfactual outcomes during 2001-09, and we continue to refer to this window as the "sample period."

Across all experiments we hold fixed the clinic's prices. While substantial changes in the policy environment may prompt the clinic to adjust its prices, we do not offer a model of how new equilibrium prices would be set. We note that during the full sample period the clinic elected to keep its prices fixed at the same level. The clinic is part of a large medical school's teaching hospital, so it is not clear what objective function is used to set prices.

We report our main results in Figures 11-12 and in Tables 5-6. The Tables contain both point estimates of counterfactual outcomes and 95% confidence intervals.[25] Because the figures and tables contain results from all experiments collected together, it is worthwhile to introduce them briefly and define terms. First, we calculate histories for N^{inf} potential patients under the observed

[25] As described above, we draw 400 times from the sampling distribution of θ, and then estimate a new value of μ for each draw. We use each pair (θ, μ) to compute the full set of patient histories under each counterfactual policy described below. We construct confidence intervals using the 2.5^{th} and 97.5^{th} percentile of each outcome (across (θ, μ) pairs) within a policy setting.

choice environment; the results of this simulation are labeled 'Baseline' and indexed as g_B. The first experiment, labeled 'Universal insurance' and indexed as g_I, extends Illinois-style insurance to all potential patients in the market. The second policy experiment is one which limits patients to a single embryo, and this is identified as 'Embryo cap' and with the index g_E. The next pair of experiments consider top-up prices in which we evaluate the impact of charging patients additional fees when they transfer two or more embryos. The first of the pair, labeled 'Top-up prices' and indexed as g_P, considers top-up prices alone, while the second in the pair combines the prices with universal insurance; this experiment is labeled 'Top-up prices + insurance' and is indexed as g_{PI}. The final policy experiment examines the impact of an improvement in the efficacy of embryo screening; this is labeled 'Technology shift' and indexed as g_T.

Before describing the individual policy experiments, we describe some of the results that come from our baseline scenario. Under the observed prices and constraints, we find that 38.4% of potential patients elect to begin treatment. This initiation rate, combined with success probabilities within the clinic, results in 26.3% of women in N^{inf} achieving at least one birth. (About 70% of simulated patients who begin IVF achieve a birth at some point in the treatment history; this is greater than the observed rate in the actual patient population because of differences in the distribution of initial Z^D values.) In total, the baseline simulations average 581.9 births across the 2146 patients in N^{inf}; these births deliver 805.9 infants to the population, implying an average number of infants per birth of 1.4.

For each patient who begins treatment we calculate $\Delta_i = W\left(Z^D_{i,a_0}, \tau_i, g, \widehat{\varphi}, \widehat{\zeta}\right) - (\widehat{\mu} + \nu_i)$, which is a measure of the net utility gain from initiating IVF above the outside option. Patients who elect to forego treatment receive $\Delta_i = 0$. We use α_i to obtain a patient-specific dollar-valued surplus measure, $CS_i(g_B) = \Delta_i/\alpha_i$. Across all potential patients in N^{inf}, including those who do not initiate treatment, the average $CS(g_B) =$ \$5,330. (Conditional on treatment, the average IVF patient has expected surplus of \$13,880 at initiation.) In total across the full potential patient population, the baseline scenario generates \$11.4 million in consumer surplus. If insurers must pay the difference between insured patients' prices and the full price, the baseline requires a total of \$2.9 million in payments from insurers to the clinic. Finally, we calculate the total medical costs of all pregnancies and births that occur under the baseline, using the cost estimations from Lemos et al. (2013) discussed above. Using these figures, the total pregnancy- and delivery-related medical cost of the baseline is \$39.3 million, or \$67,600 per birth.

We use the simulated population to calculate price elasticities as well. Prices paid by insured and

uninsured patients have different interpretations, so we calculate changes separately with respect to each price. When out-of-pocket prices for uninsured patients rise by 5%, we calculate that 3.4% fewer uninsured patients initiate treatment, implying an elasticity of -0.68 at the extensive margin. The same price increase has a slightly larger impact on the total number of uninsured cycles, which falls by 3.8% for an elasticity of -0.76. The elasticities values are different, in part, because patients who continue to initiate despite higher prices may choose to reduce their total numbers of cycles. We perform the same calculations with prices paid by insured patients (holding fixed uninsured prices), and we obtain elasticities that are smaller in magnitude. The impact of a 5% increase in out-of-pocket expenses for insured patients results in 1.5% fewer insured patients initiating and a total reduction of 1.7% in insured cycles. While elasticities above -1 are inconsistent with profit maximization, the clinic may have different objectives than a traditional firm. The elasticities we recover are comparable to others from the health care literature (Manning et al., 1987).

9.1 Expanding insurance coverage

Our first counterfactual considers a policy which endows all potential patients with 4 insured IVF cycles, as under Illinois' infertility insurance mandate. In the simulated population about half of potential patients have insurance in g_B, so this policy affects a large share of the population. The effective price reduction is about 70% for women who gain insurance under g_I.

We find that insurance leads to a substantial increase in the share of women who initiate treatment, which is 56% under g_I. The proportional change in treatment initiation is 46% greater than g_B, which is roughly in line with the uninsured-price elasticity described above. Despite a reduction in the price of treatment, the distribution of embryos transferred is very similar under g_B and g_I (Figure 11). Likewise, the distribution of births (Figure 12), shows little difference between g_B and g_I. This suggests that the extension of insurance benefits has a minimal impact on the multiple birth rate, whether through patient selection or the incentives of patients who would have received treatment even when paying full price. In our model, this is explained by the strong utility benefits that patients receive from twins, and the relatively low risk of triplets. Taken together, patients have little reason to reduce the aggressiveness of their embryo-transfer decisions. As might be expected, the widespread expansion of insurance leads to many more potential patients experiencing a birth through IVF (38%).

The patient surplus benefits of universal insurance are substantial, with an average $CS(g_I) =$ \$9,640 across all potential patients. For the full at-risk population, the difference in aggregate

consumer surplus is over \$9.2M between g_I and g_B. To account for the full impact of g_I, however, we must account for additional costs due to insurance payments and medical delivery costs. As we report on Table 6, insurance costs increase substantially, by \$11.6M or almost 400%. This difference between the change in consumer surplus and insurance costs is to be expected considering the traditional medical-demand "moral hazard" incentive of patients to take insured treatment when their willingness to pay is less than the price for uninsured patients. The expansion of insurance coverage, therefore, must be defended through arguments about fairness or equal access. While all potential patients benefit from universal insurance, we find that increases in access and surplus are greater for patients from lower wealth areas ($z_w = 0$). Overall medical delivery costs increase by an additional \$18.4M, although per-birth costs fall slightly.

9.2 Embryo transfer restrictions

We next explore the impact of restricting patients to transferring only a single embryo during treatment. To accomplish this we solve the model again at the estimated parameters but now impose the restriction $x \leq 1$ instead of $x \leq X$ in stage 4 (the embryo transfer stage). We also remove the utility penalty for single-embryo transfers for circumstances when these conflict with ASRM guidelines. We then use the new policy functions together with the same history of ε and medical technology shocks to simulate counterfactual patient histories under the one-embryo cap.

The restriction on embryo transfers entails a large utility penalty for patients considering IVF. The share of N^{inf} who initiate falls substantially relative to the baseline, from 38% under g_B to 10% under g_E. The cap has a very large mechanical effect on the distribution of embryos transferred (see Figure 11), which in turn yields a substantial shift in the distribution of births (Figure 12). Individual cycles fail to deliver a child in 73% of all treatments. The low birth probabilities of active patients translate into a low success rate for the overall at-risk population. As reported in Table 5, around 3% of N^{inf} experience a birth within the sample period. This is due to a combination of frequently-unsuccessful treatments and potential patients avoiding treatment altogether. The estimated stage-4 function $f_k(k|x, Z)$ implies a fairly high twin rate among single-embryo transfers at the clinic, so the number of delivered infants is 1.14 for every birth despite the single-embryo restriction.

The embryo cap leads to a large reduction in consumer surplus, reported in Table 6. For the overall population, the average $CS(g_E) = \$690$, a reduction of \$4,640 from the baseline value. Total consumer surplus in the patient population is only 13% of its level under the baseline. Insurer

costs for IVF treatment and medical delivery costs are greatly reduced as well, largely due to the reduction in number of treatments. Medical costs per birth fall by almost $30,000 relative to the baseline, down to $38,800 for the pregnant mother and infant(s). In summary, the embryo cap achieves its primary goal of reducing multiple births and their attendant medical costs, but this comes at substantial expense in terms of patient surplus and even single-birth outcomes.

9.3 Top-up prices for 2+ embryos

Given that a single embryo cap substantially reduces consumer surplus, in part because patients prefer a twin birth, we now consider a set of policies that internalize some of the medical costs of multiple births. These costs are largely borne by insurers rather than the patients who choose treatment aggressiveness, which implies a traditional form of moral hazard (distinct from the demand-related version associated with g_I) in which too much risky behavior occurs in equilibrium. The medical hazards of multiple births, which stimulate additional health care costs, are the motivating factor behind the embryo cap discussed above. A blunt policy like g_E, however, does not account for consumers' positive utility from some multiple birth outcomes.

To allow patients to trade-off directly between birth utility and medical costs, we construct prices that are paid when a patient transfers two or more embryos. Let c_k be the average medical cost of a delivery involving k infants. Further, let $\overline{Z}$ denote the state variable values for a median-age treated patient with no additional fertility problems. A patient who transfers $x > 1$ embryos pays the stage-4 top-up price of

$$p_{x,4} = \sum_{k>1} (c_k - c_1)[f_k(k|x, \overline{Z}) - f_k(k|1, \overline{Z})].$$

This expression acknowledges that there is some multiple birth risk for patients transferring a single embryo, $f_k(k|1, \overline{Z})$, but patients do not pay for this risk as part of the top-up price. Additionally, by including the cost difference $(c_k - c_1)$ we exclude the expense of a singleton infant. Given the values of c_k provided above (roughly $27,000, $115,000, and $435,000 for singletons, twins, and triplets respectively) and the probabilities in f_k, the top-up price for two embryos is about $12,000, and three or four embryos each entail top-up prices of roughly $19,000.[26] We assume that patients pay these prices for $x > 1$ regardless of their insurance status (ι). In our simulations, when a

[26] For example, a patient with $\overline{Z}$ who transfers 2 embryos experiences increases in her twin and triplet risks by 12 and 0.3 percentage points, respectively. Each change in risk is multiplied by the corresponding difference in medical cost relative to a singleton birth.

patient chooses to pay $p_{x,4} > 0$, we subtract this price from the summed medical costs of the full population's treatment (as if the accumulated top-up prices are saved in a fund to pay for medical expenses.) In calculating patient utility from any x, we remove all utility penalties (η) for transfers outside of ASRM guidelines.

In considering the results of the first top-up price policy – without the addition of universal insurance – the relevant comparisons are to the baseline (g_B) and embryo cap (g_E). In Table 5 we report that participation increases by 48% relative to g_E (from 10.4% to 15.4%), and more than double the share of potential patients (7.3%) eventually give birth. While multiple-embryo transfers are reduced substantially relative to the baseline (Figure 11), a significant share of patients choose $x > 1$. The average number of infants per delivery is 1.3, which is about halfway between the rates of g_B and g_E. Turning to the surplus and cost measures of Table 6, we find that total patient surplus under g_P is double that of g_E. Insurer costs of IVF treatments also increase by a relatively modest 26%, largely due to increased participation. Total medical costs double relative to g_E, but this is entirely due to the increased number of cycles and births. Medical cost per birth, on the other hand, falls slightly relative to the embryo cap due to small differences in the top-up prices patients face (calculated using $\overline{Z}$) and their actual multiple birth risks based on Z_i. In summary, the policy g_P performs just as well as the embryo cap in reducing individual births' costs while delivering greater patient surplus, but total patient surplus is substantially smaller than in the baseline scenario.

9.4 Top-up prices with insurance for single embryo transfer

The policies we describe above are able to achieve some success with regard to their motivating goals, but most policies have effects which compare unfavorably to the baseline. It is reasonable, therefore, to investigate how a combination of policies may achieve better results across a wide variety of outcomes. With this in mind, we construct a counterfactual policy (g_{PI}) which combines top-up prices with universal insurance coverage of single embryo transfer in the market.

We find that this policy is a substantial improvement on most outcome measures relative to g_P and g_E. More patients initiate treatment because they value the option to choose $x > 1$, and this is reflected in the aggregate patient surplus. About 16% of potential patients achieve one or more births during the sample period, which is substantially above the rates in g_P and g_E while below that of g_I. Insurance costs for IVF are \$8.3M, which is larger than in g_P, but still substantially lower than when universal insurance is offered without top-up prices. Medical delivery costs are again

substantially lower than under the baseline and universal insurance levels, largely because patients internalize (and pay for) their multiple birth risk. In fact, total insurance payments for treatment and delivery are substantially lower in g_{PI} than under the baseline, when only a fraction of the population has insurance coverage for IVF. Insurance firms may benefit, on net, from voluntarily extending treatment coverage while also including top-up prices.

9.5 Technology improvements

The risk of failure is a central motivation behind patients' choices to transfer multiple embryos under current IVF technology. In this counterfactual experiment we explore the impact of an improvement in IVF technology on patient choices and utility. Specifically, we alter IVF stages 3 and 4 to be roughly consistent with new advances in embryo screening technology (e.g., Chavez et al., 2012; Wong et al., 2010, Scott et al., 2013) currently in development. The technology shift requires two steps. First, we add a screening process to stage 3's technology, which generates a number of embryos, $X \in \{1, 2, 3, 4+\}$, for the patient given her number of retrieved eggs (r), the selected fertilization method (m), and her state variables (Z). We specify that some number of embryos, X^G, will be identified as "good" while the remaining embryos will have no chance of generating a successful pregnancy. We assume that each embryo has an independent probability, $p^G(a)$, of being good, where p^G is a declining function of patient age. Given X embryos, the probability $p^G(a)$, and the independence assumption, we can use the binomial distribution to calculate the probability of obtaining X^G. To implement this step we must address the possibility that a patient has $X > 4$ embryos, which we previously collected into the "4+" category. The primary determinant of a patients' total number of embryos is her number of retrieved eggs, r, which we track through 4 categorical variables. For patients in the $X = 4+$ group with the lowest realized value of r, we assume that $X = 4$. Patients in the remaining r categories are assigned $X = 6$, 9, or 13 in increasing order of their r categories. While we consider $X > 4$ in the pre-screening part of stage 3, the final collection of X^G values is again restricted to $\{1, 2, 3, 4+\}$. Ultimately, we are able to compute a distribution over X^G values as

$$f_{X^G}(X^G|r, m, Z_a) = \sum_X Bin(X^G|X, p^G(a)) f_X(X|r, m, Z_a),$$

where Bin is the binomial distribution function. In a final set of assumptions for stage 3, we assign $p^G(a) = 0.3$ for the youngest patients, $p^G(a) = 0.1$ for the oldest, and a uniform rate of decline for

age categories in between.

For the second step in altering the treatment technology, we adjust stage 4 to reflect the improved success probability for each good embryo. For women with $z_{ff} = 0$ (i.e., without other fertility problems), we assume that each embryo has an independent probability $p^B(z_{ff}) = 0.85$ of generating a successful singleton pregnancy. Women with $z_{ff} = 1$ have $p^B(z_{ff}) = 0.70$ for each embryo. We make use of the binomial distribution again to calculate the probability that a woman who transfers $x \leq X^G$ good embryos achieves k births. Together across stages 3 and 4, the probabilities p^G and p^B generate choice sets that are usually much smaller than the empirical ones (and contain zero good embryos fairly frequently), but patients proceed with the understanding that each transferred embryo is very likely to result in a child.

Technology improvements have a substantial impact on patients' choices and welfare, although we do not consider the cost of implementing g_T. 48% of potential patients now initiate treatment (Table 5), and single embryo transfers are now more common relative to g_B given the improved prospects for success with a single embryo (Figure 12) and the smaller choice sets that are sometimes realized.[27] The improved technology together with less aggressive embryo transfer choices lead to a substantial increase in singleton birth outcomes. Twin births are fairly common also, which is due to the positive utility value patients receive from twin deliveries. Across the full patient population, the share of women with at least one birth increases to 32%.

It is not surprising that the improved technology with constant prices leads to a substantial improvement in patient surplus. The average surplus in the full population, reported in Table 6, increases to \$9,000, or about \$3,700 greater than the baseline value. Over the full sample period, this is \$7.8M in additional surplus to the clinic's potential patients, which can be adjusted to account for the relative size of the full U.S. market (of which the observed clinic is about 0.4%). This additional potential patient surplus of two billion dollars can be compared to the likely expenses of scientific research that focuses on improving treatment technology. Other benefits accrue through reduced payments for IVF treatment and medical costs at delivery. Despite a larger number of initiating patients, a reduction in cycles per patient leads to a 9% reduction in insurance payments for IVF. Delivery-related medical costs increase by about 1% in total, but this follows a 50% increase in the total number of births relative to the baseline. The per-birth medical cost falls by over \$12,000 to \$54,800.

[27] The increase in zero-embryo transfers is due to the increased frequency of $X^G = 0$ outcomes in stage 3.

10 Conclusions

In this paper we structurally estimate a dynamic model of the treatment choices made by infertile women undergoing IVF. Our framework incorporates important mechanisms influencing these decisions, including patient preferences, the evolution of patient health, IVF treatment technologies, and financial incentives. In addition to the treatment initiation decision, our model highlights the key tradeoff faced by women undergoing IVF: More aggressive treatment choices increase the likelihood of a birth, and so reduce future treatment costs, but also increase the possibility of potentially undesirable higher-order births. We apply the model to a unique dataset of women undergoing IVF treatment at a major clinic in the St. Louis, Missouri, area between 2001 and 2009. The clinic is situated such that it draws clients from both the Illinois side of the St. Louis metro area, where IVF is covered under a mandated insurance benefit, and the Missouri side where it is not. Consequently, we observe the choices made by women with similar health characteristics being treated at the same clinic, but who face very different financial incentives. Our parameter estimates do indeed show that higher out-of-pocket costs discourage infertile women from both initiating IVF treatment and choosing to continue treatment after a failure.

Public policy toward IVF is motivated by the often conflicting goals of increasing patient access to treatment and reducing costly multiple births. We use our model estimates to assess a variety of policies in light of these objectives. Many advocate extending nationwide the insurance mandates for IVF currently found in a handful of states. Our counterfactual simulations show that such a universal mandate substantially increases the consumer surplus of patients in our sample by increasing access to IVF, but also increases both insurance and birth costs. Embryo caps have been proposed as a way to reduce the relatively high rates of multiple births associated with IVF. We find that a policy of single embryo transfer does indeed substantially reduce multiple births, but at the cost of a sharp reduction in consumer surplus. This is because patients need more cycles of IVF to achieve a birth, on average, plus our utility parameter estimates imply that patients have a slight preference for twin births.

Given that neither unrestricted universal insurance nor embryo caps achieves the dual goals of increased access and lower per-birth costs, we propose an alternative policy in the spirit of Einav et al. (2015) in which patients receive insurance coverage for the transfer of a single embryo, but then have to pay a "top-up" price of \$12,000 - \$19,000 if they wish to transfer additional embryos. Incorporating top-up prices substantially reduces expected birth and insurance costs relative to an

unrestricted universal insurance mandate, while generating higher consumer surplus than a single embryo cap for patients in our sample. Policymakers could adjust the top-up price up or down depending on the relative importance of access to IVF vs. insurance and expected birth costs.

A large literature debates the value of new medical technologies.[28] Because we carefully specify the IVF treatment technology in our framework, we are able to assess the potential value of recent advances in embryo screening that may improve the efficiency of IVF. Assuming that patient out-of-pocket costs do not change, our simulations show that these technological advances increase consumer surplus with a magnitude similar to that of a universal insurance mandate for IVF. At the same time, we predict that costs per birth decline. These findings highlight the idea that many of the goals of public policies toward IVF, such as insurance mandates or embryo caps, may in fact be accomplished through improvements in technology.

Our modeling framework can be adapted or extended to incorporate details that may be important in other medical markets. One example is the addition of gradual learning about disease severity, which is common in cancer treatment. A patient's dynamic strategy may begin with milder treatments, and, if these fail, progress to more aggressive actions as the patient updates his beliefs on his medical condition. A second example is technological progress within and beyond a study's sample period, which is common to many treatment areas. In this case, the patient's decisions to act versus delay will include expectations about future therapies or prices. We defer these issues to future research on medical markets well-suited to each example's particular challenges.

Appendix A: Distribution of Types Among Potential Patients

After estimating the model of decision-making within the clinic, we know $\widehat{\theta} = \left(\widehat{\varphi}, \widehat{\zeta}, \widehat{\rho}\right)$ and therefore $W\left(Z_{a_0}^D, \tau, g_B, \widehat{\varphi}, \widehat{\zeta}\right)$. We also know

$$\Pr\left(\tau = 2 | Z_{a_0}^D, I = 1, \widehat{\rho}\right) = \Lambda([1, Z_{a_0}^D]\widehat{\rho}). \tag{14}$$

In addition, from the treatment initiation model we know, for each possible μ and $Z_{a_0}^D$, the initiation rate among potential patients of each type. That is, we know

$$\Lambda\left(W\left(Z_{a_0}^D, \tau, \widehat{\varphi}, \widehat{\zeta}\right) - \mu\right) \text{ for } \tau = 1, 2. \tag{15}$$

[28] See, for example, Cutler and McClellan (2001) and Murphy and Topel (2006).

For each $Z_{a_0}^D$ we also know the total (i.e. unconditional on type) number of women with characteristics $Z_{a_0}^D$ who came into the clinic. Let this number be $N_{Z_{a_0}^D}^{clin}$. Together with $\Pr\left(\tau = 1|Z_{a_0}^D, I_i = 1,\widehat{\rho}\right)$ we then have an estimate of the number of patients of type 1 with characteristics $Z_{a_0}^D$ who came into the clinic, say $N_{Z_{a_0,1}^D}^{clin}\left(\widehat{\rho}\right)$, where

$$N_{Z_{a_0,1}^D}^{clin}\left(\widehat{\rho}\right) = N_{Z_{a_0}^D}^{clin} \times \left[1 - f_\tau\left(\tau = 2|Z_{a_0}^D, I = 1, \widehat{\rho}\right)\right]. \tag{16}$$

Similarly for type 2, we know:

$$N_{Z_{a_0,2}^D}^{clin}\left(\widehat{\rho}\right) = N_{Z_{a_0}^D}^{clin} \times f_\tau\left(\tau = 2|Z_{a_0}^D, I = 1, \widehat{\rho}\right). \tag{17}$$

Note that while $N_{Z_{a_0}^D}^{clin}$ is just data, $\left[N_{Z_{a_0,1}^D}^{clin}\left(\widehat{\rho}\right), N_{Z_{a_0,2}^D}^{clin}\left(\widehat{\rho}\right)\right]$ depends on ρ, which is identified by the differential behavior of the two types in the (within-clinic) patient histories. Recall that ρ parameterizes the within-clinic distribution of types and is estimated in our second step along with (φ, ζ).

Given μ, from the initiation model we know that $100 \times \Lambda\left(W\left(Z_{a_0}^D, \tau, \widehat{\varphi}, \widehat{\zeta}\right) - \mu\right)$ percent of potential patients with initial non-biological state $Z_{a_0}^D$ and type τ will choose to initiate treatment. We also know that there were $N_{Z_{a_0,\tau}^D}^{clin}\left(\widehat{\rho}\right)$ patients. Then it must be the case that the number of potential patients of each type is given by

$$N_{Z_{a_0,1}^D}^{inf} = \frac{N_{Z_{a_0,1}^D}^{clin}\left(\widehat{\rho}\right)}{\Lambda\left(W\left(Z_{a_0}^D, \tau = 1, \widehat{\varphi}\right) - \mu\right)} = \frac{N_{Z_{a_0}^D}^{clin}\left[1 - \Lambda([1, Z_{a_0}^D]\widehat{\rho})\right]}{\Lambda\left(W\left(Z_{a_0}^D, \tau = 1, \widehat{\varphi}\right) - \mu\right)}, \text{ and} \tag{18}$$

$$N_{Z_{a_0,2}^D}^{inf} = \frac{N_{Z_{a_0,2}^D}^{clin}\left(\widehat{\rho}\right)}{\Lambda\left(W\left(Z_{a_0}^D, \tau = 2, \widehat{\varphi}, \widehat{\zeta}\right) - \mu\right)} = \frac{N_{Z_{a_0}^D}^{clin}\Lambda([1, Z_{a_0}^D]\widehat{\rho})}{\Lambda\left(W\left(Z_{a_0}^D, \tau = 2, \widehat{\varphi}, \widehat{\zeta}\right) - \mu\right)}. \tag{19}$$

Then we can estimate the unconditional (i.e. not conditional on $I = 1$) prevalence of type 2 among potential patients with state $Z_{a_0}^D$ as

$$f_\tau\left(\tau = 2|Z_{a_0}^D\right) \approx \frac{N_{Z_{a_0,2}^D}^{inf}}{N_{Z_{a_0,1}^D}^{inf} + N_{Z_{a_0,2}^D}^{inf}} = \left(1 + \left[\frac{\frac{1-\Lambda([1,Z_{a_0}^D]\widehat{\rho})}{\Lambda\left(W\left(Z_{a_0}^D,\tau=1,\widehat{\varphi}\right)-\mu\right)}}{\frac{\Lambda([1,Z_{a_0}^D]\widehat{\rho})}{\Lambda\left(W\left(Z_{a_0}^D,\tau=2,\widehat{\varphi},\widehat{\zeta}\right)-\mu\right)}}\right]\right)^{-1}. \tag{20}$$

Note that for given μ, everything in the RHS is known, so $f_\tau\left(\tau = 2|Z_{a_0}^D\right)$ is known and $f_\tau\left(\tau = 1|Z_{a_0}^D\right) = 1 - f_\tau\left(\tau = 2|Z_{a_0}^D\right)$. If both types were to select into the clinic at the same rate (i.e. they did not really

had different preferences for children), we would have $W\left(Z_{a_0}^D, \tau = 1, g_B, \theta\right) = W\left(Z_{a_0}^D, \tau = 2, g_B, \theta\right)$ so $\frac{\Lambda\left(W\left(Z_{a_0}^D, \tau=1, \theta\right)-\mu\right)}{\Lambda\left(W\left(Z_{a_0}^D, \tau=2, \theta\right)-\mu\right)} = 1$ and the distribution of types within the clinic and among potential patients would be the same, $f_\tau\left(\tau = 2|Z_{a_0}^D, I = 1\right) = f_\tau\left(\tau = 2|Z_{a_0}^D\right)$, which is not consistent with our estimates.

Appendix B: Approximating N^{inf} and $f_{Z^D}\left(Z_{a_0}^D\right)$

To obtain a model-predicted IVF initiation rate among potential patients we must use an estimate of $f_{Z^D}\left(Z_{a_0}^D\right)$. Note that since the expected value of initiation depends on $Z_{a_0}^D$, the distribution of $Z_{a_0}^D$ among clinic patients will differ from that among potential patients. In particular, we expect women who we observe as patients at the clinic to be older, more likely to be covered by insurance and wealthier. To approximate $f_{Z^D}\left(Z_{a_0}^D\right)$ among potential patients we use the following assumptions:

- **Assumption 0 (Exogenous ASRM Guidelines).** The particular ASRM guidelines in place are independent of everything else in the model:

$$asrm \perp \left(Z^B, \iota_{a_0}, z_w, a_0\right). \tag{21}$$

- **Assumption 1 (Conditional Independence)**. Conditional on age at initiation, the 3 biological state variables related to infertility Z^B are independent of insurance and wealth:

$$Z^B \perp (\iota_0, z_w) \mid a_0. \tag{22}$$

  Note that Assumption 1 and the fact that the value of these 3 state variables only becomes observable after deciding to start a first cycle, imply that these variables will have the same conditional (on age) distribution in the risk set and in the clinic:

$$f\left(Z^B|a_0\right) = f\left(Z^B|\ a_0, I = 1\right). \tag{23}$$

- **Assumption 2 (Surprise).** Among the women attempting their first pregnancy at age a_0, finding out about the infertility problem is a surprise. Therefore, the joint distribution of wealth and insurance coverage among women of that age should be independent of whether they have any infertility problem (i.e. independent of whether they are among the potential

patients or not). Therefore, $\Pr(\iota_{a_0}, z_w|a_0)$ is the same regardless of whether a woman is a potential patient. We further assume that $\Pr(\iota_{a_0}, z_w|a_0) = \Pr(\iota_{a_0}, z_w)$ for all a_0.

Using these assumptions we can approximate the joint distribution of all state variables among potential patients as $f_Z(Z_{a_0}) = f_Z\left(Z^D_{a_0}, Z^B\right) = f_{Z^B}\left(Z_B|Z^D_{a_0}\right) f_{Z^D}\left(Z^D_{a_0}\right)$.

First note that by Assumptions 0 and 1, $f_{Z^B}\left(Z_B|Z^D_{a_0}\right) = f_{Z^B}(Z_B|a_0) = f_{Z^B}(Z_B|a_0, I=1)$ and we can easily construct an estimate $\widehat{f}_{Z^B}(Z_B|a_0, I=1)$ using patient data. So we only need to focus on $f_{Z^D}\left(Z^D_{a_0}\right)$,which is the critical input for the share-matching procedure described in Section 7.2. By Assumption 0,

$$f_{Z^D}\left(Z^D_{a_0}\right) = f_{a_0,\iota_{a_0},z_w}(a_0, \iota_{a_0}, z_w)\, f_{asrm}\left(z_{asrm_{a_0}}\right). \tag{24}$$

To estimate $f_{a_0,\iota_{a_0},z_w}(a_0, \iota_{a_0}, z_w) = f_{\iota_{a_0},z_w}(\iota_{a_0}, z_w)\, f_{a_0}(a_0)$ we rely on Assumption 2 which means that we do not need to restrict ourselves to the unobservable set of potential patients.

Distribution of age among potential patients. We first estimate $f_{a_0}(a_0)$ using data from the St. Louis region on (first) births and the maternal age associated with those births. Also because of Assumption 2, this gives us the distribution of age at first *attempted* birth (regardless of whether the attempt was successful or not) when combined with estimates of infertility rates by age. This provides the age distribution for our potential patients.

Joint distribution of IVF coverage and wealth. Finally we collect data on the joint distribution of IVF coverage and wealth, (ι, z_w). To estimate $f_{\iota,z_w}(\iota, z_w)$ we consider $f_{\iota,z_w}(\iota, z_w) = f_\iota(\iota|z_w)\, f_{z_w}(z_w)$and develop a strategy at the zip code level for estimating $f_\iota(\iota|z_w)$ and $f_{z_w}(z_w)$ using information from zip codes whose center is located within 75 miles from our clinic. To estimate $f_{z_w}(z_w)$ we assume patients from same zip code are homogenous regarding (ι, z_w). In particular, we know whether each zip code in the St. Louis area is considered "wealthy" or not by construction: we defined $z_{w,l} = 1$ if zipcode l's median home value is above \$100,000. This is consistent with the way we are defining a patient to be "wealthy" or not (i.e. whether she comes from a zip code where the median home value is above \$100,000. So within a zip code everyone is either wealthy or not wealthy. We can estimate $f_{z_w}(z_w)$ by

$$f_{z_w}(z_w = 1) = \sum_{l \in STL} I\{z_{w,l} = 1\}\, \pi_l = \sum_{l: z_{w,l}=1} \pi_l, \tag{25}$$

where π_l is a population weight that measures how important zip code l is within the St. Louis region in terms of population. We have the population by age for each zip code so we can construct

π_l easily. We estimate that 39% of the potential patients in the St. Louis region are wealthy.

To estimate $\Pr(\iota|w)$ we take the following steps. We have the percentage of population who have private insurance for each zip code l within the St. Louis area: $f_{priv}(priv_l)$. We obtain this from the 2012 American Community Survey (ACS) 5-year estimate. A 2005 Mercer Survey of Employer Health Insurance reported that 19 percent of those with large (500+ employees) employer-provided health insurance have IVF coverage (and 11 percent of those working for small employers do so). We assume the same rates apply to Missouri zip codes. We then use estimates from Census's Business Dynamics Statistics as reported by Moscarini and Postel-Vinay (2012) and estimate the employment share of large employers to be 48%.[29] Therefore we use the following adjustment factor $\psi^{MO} = 0.52 \times 0.11 + 0.48 \times 0.19 = 0.15$ to adjust the raw insurance coverage rates we obtain from ACS. Then the IVF coverage for each Missouri zip code l is given by $f_{IVF}\left(\iota_l^{IVF}\right) = f_{priv}(priv_l) \times \psi^{MO}$.

Regarding Illinois counties, we know that there is a mandate. But small employers (< 25 employees) and self-insured employers (regardless of size) are excluded.[30] According to a Kaiser Family Foundation (2007) report, 55% of workers nationally are covered by plans that are partially or fully self-insured.[31] So we adjust the raw county-level employer-sponsored health insurance coverage rate by the percentage of large employers and the percentage not self-funded and assume that no firm with less than 25 employees provides IVF coverage. We obtain the following adjustment factor for Illinois counties $\psi^{IL} = (0.215 \times 0 + 0.785 \times [0.45 \times 1 + 0.55 \times 0.19]) = 0.435$. Then the IVF coverage for each Illinois zip code l is given by: $f_{IVF}\left(\iota_l^{IVF}\right) = f_{priv}(priv_l) \times \psi^{IL}$.

We then compute the aggregate IVF coverage rate for the region conditional on wealth. First, we condition on $z_w = 0$ and compute

$$\Pr(\iota = 4|z_w = 0) = \sum_{l:w_l=1} f_{IVF}\left(\iota_l^{IVF}\right)\left(\frac{\pi_l}{\sum_{l:w_l=1}\pi_l}\right). \tag{26}$$

Regarding coverage conditional on high wealth ($w_z = 1$) we take a different approach. Since most of the wealthy zip codes are on the Missouri side but ψ^{MO} is very low relative to ψ^{IL}, if we pool zip codes together in the aggregation we would end up with a spurious negative correlation. Therefore

[29] See Table 1 in Moscarini & Postel-Vinay (2012)

[30] Under this alternative definition of small employer, we interpolate the numbers in Moscarini adn Postel-Vinay and find that 21.5% of employment is accounted for by firms with less than 25 employees.

[31] We assume that in these self-funded plans the same rate found in the Mercer survey (19%) for large employers applies. This is probably an upper bound because large employers here also include firms with 25 to 499 employees, not just those with 500+ as in the Mercer study definition.

we compute $\Pr(\iota = 4|z_w = 1)$ in the following way:

$$\Pr(\iota = 4|z_w = 1) = \Pr(\iota = 4|z_w = 0) + \widehat{\Delta},$$

where

$$\widehat{\Delta} = \left(\frac{\sum_{l:w_l=1, l\in IL} \pi_l}{\sum_{l:w_l=1} \pi_l}\right) \widehat{\Delta}_{IL} + \left(\frac{\sum_{l:w_l=1, l\in MO} \pi_l}{\sum_{l:w_l=1} \pi_l}\right) \widehat{\Delta}_{MO}. \tag{27}$$

$\widehat{\Delta}_s$ provides the estimated average increase in IVF coverage observed for state s when one moves from poor zip codes to wealthy zip codes within that state.

$$\widehat{\Delta}_s = \left[\sum_{l:w_l=1, l\in s} f_{IVF}(IVF_l)\left(\frac{\pi_l}{\sum_{l:w_l=1, l\in s} \pi_l}\right)\right] - \left[\sum_{l:w_l=0, l\in s} f_{IVF}(IVF_l)\left(\frac{\pi_l}{\sum_{l:w_l=0, l\in IL} \pi_l}\right)\right] \text{ for } s = \text{IL, MO.}$$

The results indicate that IVF insurance coverage rate depends of wealth. Among poor potential patients, 83% have $\iota = 0$ and 17% have $\iota = 4$. For wealthy potential patients, 75% have $\iota = 0$ and 25% have $\iota = 4$.

Size of Potential Patient Pool. In addition to the joint distribution of characteristics among potential patients, we need the size of the potential patient pool N^{inf}. We use $\widetilde{N}^{inf}$ to refer to all potential patients in the St. Louis area, and N^{inf} for the subset who might consider the clinic we study. We count the number of women of each age in the St. Louis region that give birth naturally to a first birth in any given quarter. Let this number be $\overline{N}_a$. We get this from Vital Statistics. The total number of women who attempt their first pregnancy at age a is N_a^{stl}. Of these, $\overline{N}_a$ succeed and have births recorded in Vital Statistics; the group that fails, N^{inf}, becomes our set of potential patients. Therefore $N_a^{stl} = \overline{N}_a + \widetilde{N}_a^{inf}$. Then using infertility rates by age among women who are attempting to get pregnant, $f_{inf}(inf|a)$, we can back out $\widetilde{N}_a^{inf} = \frac{f_{inf}(inf|a)}{[1-f_{inf}(inf|a)]}\overline{N}_a$. According to Vital Statistics the larger counties in and around the St. Louis region have an average of $\overline{N}_{28-44} = 1172$ first births each quarter distributed among mothers aged 28 to 44. To capture births occurring in the more rural areas, but still within our 75-mile radius area, we also estimate the births occurring in smaller counties within this area. An additional 10.4% of births come from these counties.[32] So $\overline{N}_{28-44}^{75m} = 1172 \times 1.104 = 1294$. Using infertility rates by age and summing

[32] Births occuring in smaller counties are combined and reported into a single residual county for each state in Vital Stats. So we know how many first births occured in these "residual" counties. We also know how important (in terms of number of households) the zipcodes belonging to small counties but located within the 75-mile radius are as a share of the each state specific residual county. Therefore we can augment the number of births in the relevant area by assuming that the same share of births comes from these zipcodes.

across ages, we can then determine that there are $\widetilde{N}_t^{inf} = 198 \times 1.104 = 219$ new potential patients, on average, each quarter.[33] Since there are 28 quarters between 2001 and 2007, the size of the potential patient pool for our sample period is then $\widetilde{N}_{2001-07}^{inf} = 28 \times \widetilde{N}^{inf} = 28 \times 219 = 6132$. While this pool of potential patients is valid for the full St. Louis area, our clinic has market share $s^{clin} < 1$. According to the CDC, the clinic we observe has market share of about a third, and we adjust $\widetilde{N}_{2001-07}^{inf}$ in a proportional way. Ultimately, the potential-patient population for our clinic is $N^{inf} = 2146$.

References

[1] Abbring, J.H., and J.J. Heckman (2007): "Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice and General Equilibrium Policy Evaluation," *Handbook of Econometrics*, Volume 6B, Chapter 72, Elsevier B.V.

[2] Abramowitz, J. (2014): "Turning Back the Ticking Clock: The Effect of Increased Affordability of Assisted Reproductive Technology on Women's Marriage Timing," *Journal of Population Economics* 27(2), 603-633.

[3] Aron-Dine, A., L. Einav, A. Finkelstein, and M. Cullen (2015): " Moral Hazard in Health Insurance: Do Dynamic Incentives Matter?" forthcoming in *Review of Economics and Statistics.*

[4] Bitler, M. P. (2008): "Effects of Increased Access to Infertility Treatment on Infant Health Outcomes: Evidence from Twin Births," mimeo, University of California-Irvine.

[5] Bitler, M. P., and L. Schmidt (2006): "Health Disparities and Infertility: Impacts of State-Level Insurance Mandates," *Fertility and Sterility* 85(4), 858-65.

[6] Bitler, M. P., and L. Schmidt (2012): "Utilization of Infertility Treatments: The Effects of Insurance Mandates," *Demography* 49(1), 124-149.

[7] Buckles, K. (2005): "Stopping the Biological Clock: Infertility and the Career-Family Tradeoff," University of Notre Dame working paper.

[33] To obtain age-specific infertility rates $f_{inf}(inf|a)$ we interpolate between ages 28 to 39 and extrapolate for ages 40-44 the 12-month infertility estimates reported in Dunson et al. (2004).

[8] Buckles, K. (2013): "Infertility Insurance Mandates and Multiple Births," *Health Economics* 22(7), 775-789.

[9] Bundorf, M.K., M. Henne, and L. Baker (2008): "Mandated Health Insurance Benefits and the Utilization and Outcomes of Infertility Treatments," NBER Working Paper #12820.

[10] Centers for Disease Control and Prevention (2014): "2012 Assisted Reproductive Technology National Summary Report," U.S. Dept. of Health and Human Services.

[11] Chan, T. Y. and B.H. Hamilton (2006): "Learning, private information, and the economic evaluation of randomized experiments," *Journal of Political Economy* 114(6), 997-1040.

[12] Chavez, S., K.E. Loewke, J. Han, F. Moussavi, P. Colls, S. Munne, B. Behr, and R.A. Reijo Pera (2012): "Dynamic Blastomere Behaviour Reflects Human Embryo Ploidy by the Four-Cell Stage," *Nature Communications* 3:1251, doi: 10.1038/ncomms2249.

[13] Crawford G. S. and M. Shum (2005): "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica* 73(4), 1137-1173.

[14] Csokmay, J.M., M.J. Hill, R.J. Chason, S. Hennessy, A.N. James, J. Cohen, A.H. DeCherney, J.H. Segars, M.D. Payson (2011): "Experience with a Patient-Friendly, Mandatory, Single-Blastocyst Transfer Policy: The Power of One," *Fertility and Sterility* 96(3), 580-584.

[15] Cutler D. and M. McClellan (2001): "Is Technological Change in Medicine Worth It?," *Health Affairs* 20(5), 11-29.

[16] Dickstein, M. (2014): "Efficient Provision of Experience Goods: Evidence from Antidepressant Choice," working paper, Stanford University.

[17] Dunson, D.B., D. Baird, and B. Colombo (2004): "Increased Infertility With Age in Men and Women," *Obstetrics and Gynecology* 103(1), 51-56

[18] Einav, L., A. Finkelstein, and M. Schrimpf (2010): "Optimal Mandates and the Welfare Cost of Asymmetric Information: Evidence from the U.K. Annuity Market," *Econometrica* 78(3), 1031–1092.

[19] Einav, L. A. Finkelstein, and H. Williams (2015): "Paying on the Margin for Medical Care: Evidence from Breast Cancer Treatments," forthcoming in *American Economic Journals: Economic Policy.*

[20] European Society of Human Reproduction and Embryology (2014): "ART Fact Sheet," http://www.eshre.eu/guidelines-and-legal/art-fact-sheet.aspx.

[21] Gershoni, N. and C. Low (2015): "The Impact of Extended Reproductive Time Horizons: Evidence from Israel's Expansion of Access to IVF," University of Pennsylvania working paper.

[22] Gleicher, N., and D. Barad (2006): "The Relative Myth of Elective Single Embryo Transfer," *Human Reproduction* 21(6), 1337-1344.

[23] Hamilton, B., and B. McManus (2011): "The Effects of Insurance Mandates on Choices and Outcomes in Infertility Treatment Markets," *Health Economics* 21(8), 994-1016.

[24] Henne, M. B., and M. K. Bundorf (2008): "Insurance Mandates and Trends in Infertility Treatments," *Fertility and Sterility* 89(1), 66-73.

[25] Jain, T. , B.L. Harlow, and M.D. Hornstein (2002): "Insurance Coverage and Outcomes of In Vitro Fertilization," *New England Journal of Medicine* 347(9), 661-666.

[26] Jungheim, E.S., G.L. Ryan, E.D. Levens, A.F. Cunningham, G.A. Macones, K.R. Carson, A.N. Beltsos, and R.R. Odem (2010): "Embryo Transfer Practices in the United States: A Survey of Clinics Registered with the Society for Assisted Reproductive Technology," *Fertility and Sterility* 94(4), 1432-36.

[27] Kaiser Family Foundation (2007): *Employer Health Benefits: 2007 Annual Survey.* Menlo Park, CA: Henry J. Kaiser Family Foundation.

[28] Lemos, E.V., D. Zhang, B.J. Van Voorhis, and X.H. Hu (2013): "Healthcare Expenses Associated with Multiple vs. Singleton Pregnancies in the United States," *American Journal of Obstetrics and Gynecology* 209(6), 586.e1-586.e11.

[29] Manning, W., J. Newhouse, N. Duan, E. Keeler, A. Leibowitz, and S. Marquis (1987): "Health and the Demand for Medical Care: Evidence from a Randomized Experiment," *American Economic Review* 77(3), 251-277.

[30] Moscarini, G. and F. Postel-Vinay (2012): "The Contribution of Large and Small Employers to Job Creation in Times of High and Low Unemployment," *American Economic Review* 102(6), 2509-2539.

[31] Murphy, S.A. (2003): "Optimal Dynamic Treatment Regimes," *Journal of the Royal Statistical Society* 65(2), 331-355.

[32] Murphy, K. and R. Topel (2006): "The Value of Health and Longevity," *Journal of Political Economy* 114(5), 871-904.

[33] Practice Committee of the Society for Assisted Reproductive Technology (2012): "Elective Single-Embryo Transfer," *Fertility and Sterility* 97(4), 835-842.

[34] Robins, J. M. (1997): "Causal Inference from Complex Longitudinal Data," in Berkane, M. (Ed.), *Latent Variable Modeling and Applications to Causality,* Volume 120 of *Lecture Notes in Statistics*, Springer-Verlag, New York, pp. 69-117.

[35] Rust, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica* 55(5), 999-1033.

[36] Ryan, G., S. Zhang, A. Dokras, C. Syrop, and B. Van Voorhis (2004): "The Desire of Infertile Patients for Multiple Births," *Fertility and Sterility* 81(3), 500-5004.

[37] Ryan, G.L., A.E.T. Sparks, C.S. Sipe, C.H. Syrop, A. Dokras, B.J. Van Voorhis (2007): "A Mandatory Single Blastocyst Transfer Policy with Educational Campaign in a United States IVF program Reduces Multiple Gestation Rates without Sacrificing Pregnancy Rates," *Fertility and Sterility* 88(2), 354-360.

[38] Schmidt, L. (2005): "Infertility Insurance Mandates and Fertility," *American Economic Review, Papers and Proceedings* 95(2), 204-208.

[39] Schmidt, L. (2007): "Effects of Infertility Insurance Mandates on Fertility," *Journal of Health Economics* 26(3), 431-446.

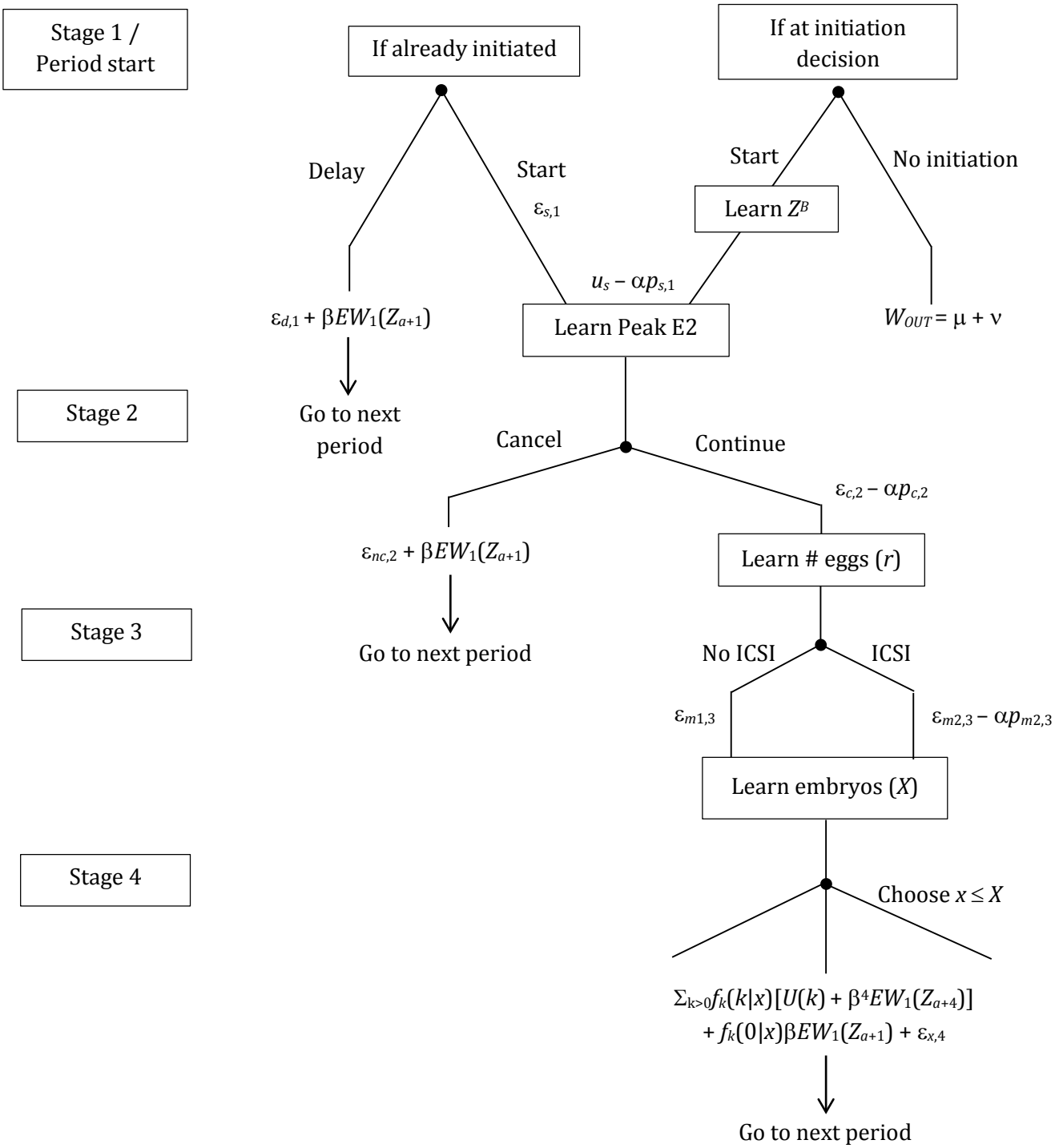
[40] Scott, R.T. Jr, Upham, K.M., Forman, E.J., Hong, K.H., Scott, K.L., et al. (2013) "Blastocyst biopsy with comprehensive chromosome screening and fresh embryo transfer significantly increases in vitro fertilization implantation and delivery rates: a randomized controlled trial." *Fertility and Sterility* 100(3), 697–703.

[41] Simonstein, F., M. Mashiach-Eizenberg, A. Revel, J.S. Younis (2014): "Assisted Reproductive Policies in Israel: A Retrospective Analysis of In Vitro Fertilization-Embryo Transfer," *Fertility and Sterility* 102(5), 1301-06.

[42] Velez, M., M. Connolly, I. Kadoch, S. Philips, and F. Bissonnette (2014): "Universal Coverage of IVF Pays Off," *Human Reproduction* 29(6), 1313-1319.

[43] Wong, C.C., K.E. Loewke, N.L. Bossert, B. Behr, C.J. DeJonge, T.M. Baer, and R.A. Reijo Pera (2010): "Non-Invasive Imaging of Human Embryos before Embryonic Genome Activation Predicts Development to the Blastocyst Stage," *Nature Biotechnology* 28(10), 1115-1121.

Figure 1: IVF treatment stages



Note: We display patients' immediate payoffs at each stage of the decision tree, and include expected future payoffs only where the patient has reached the end of a within-period decision sequence. We omit some notation to avoid clutter.

Figure 2: Distribution of Peak E2 Outcomes by AFC

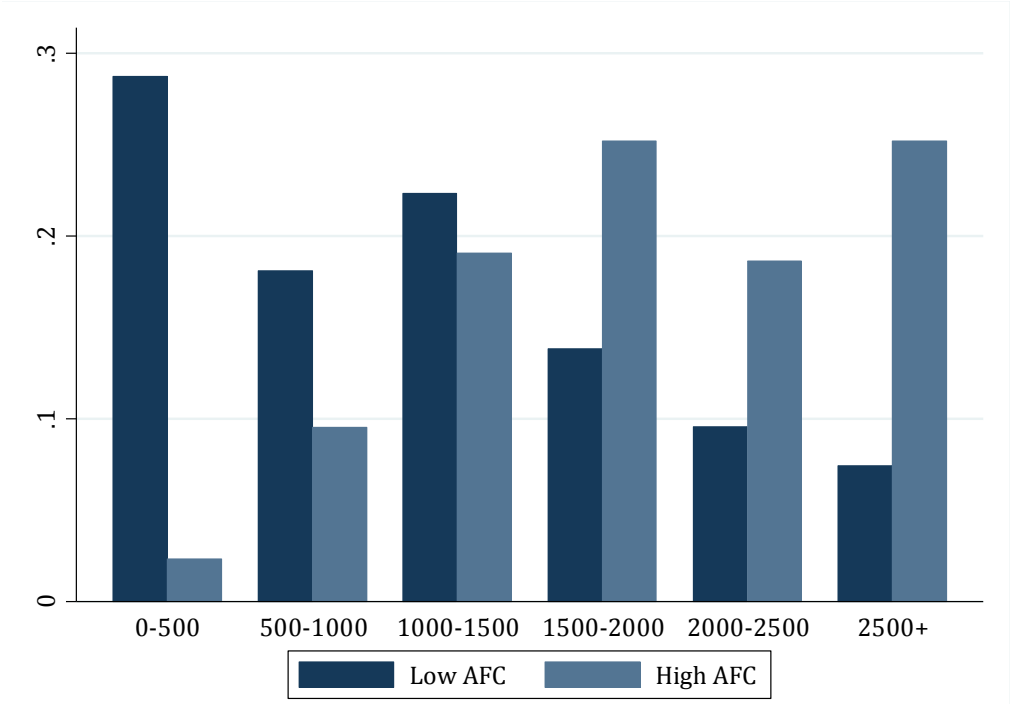


Figure 3: Distribution of Retrieved Egg Count by Peak E2

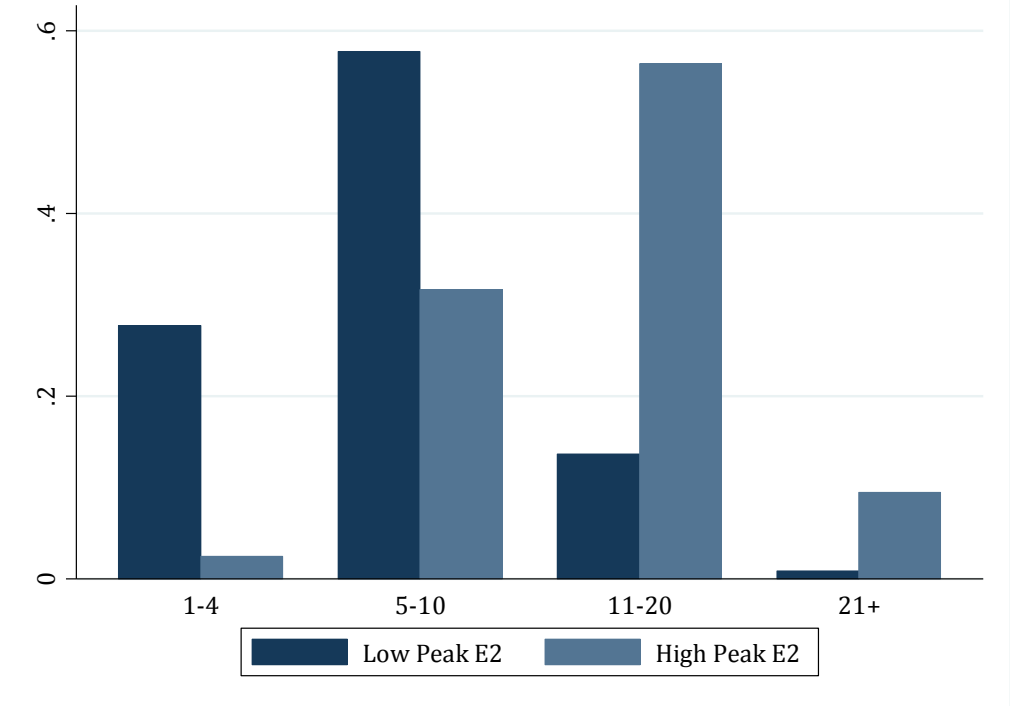


Figure 4: Distribution of Embryos Available by ICSI Use, Male Factor Patients

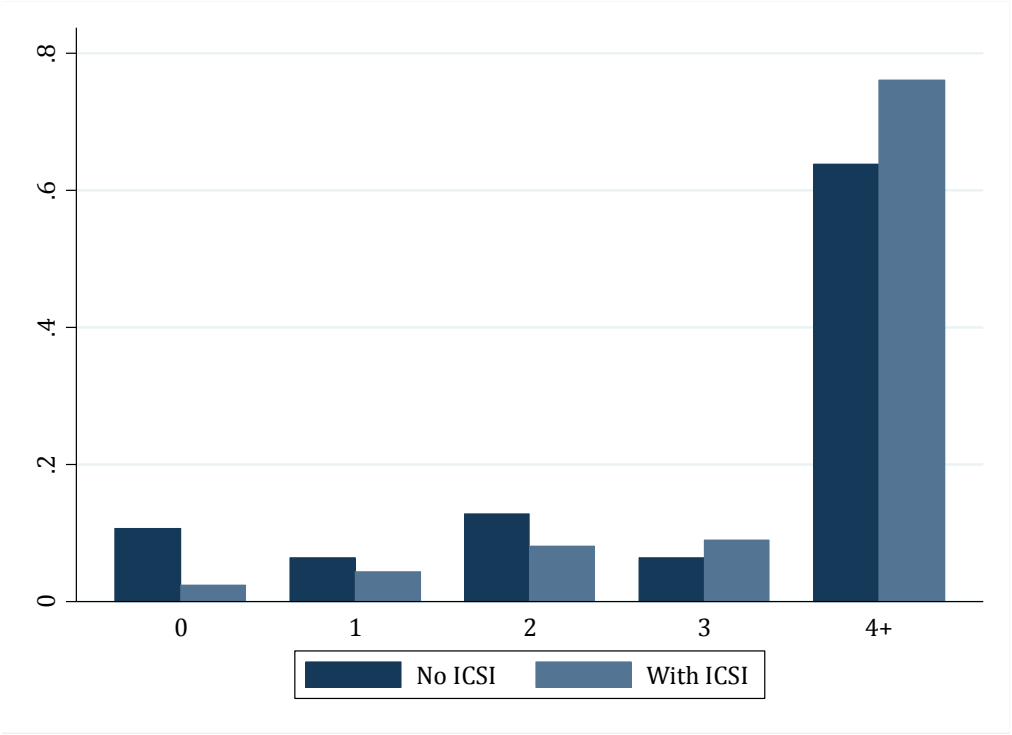


Figure 5: Distribution of Births by Embryos Transferred, Patient Age 34-36

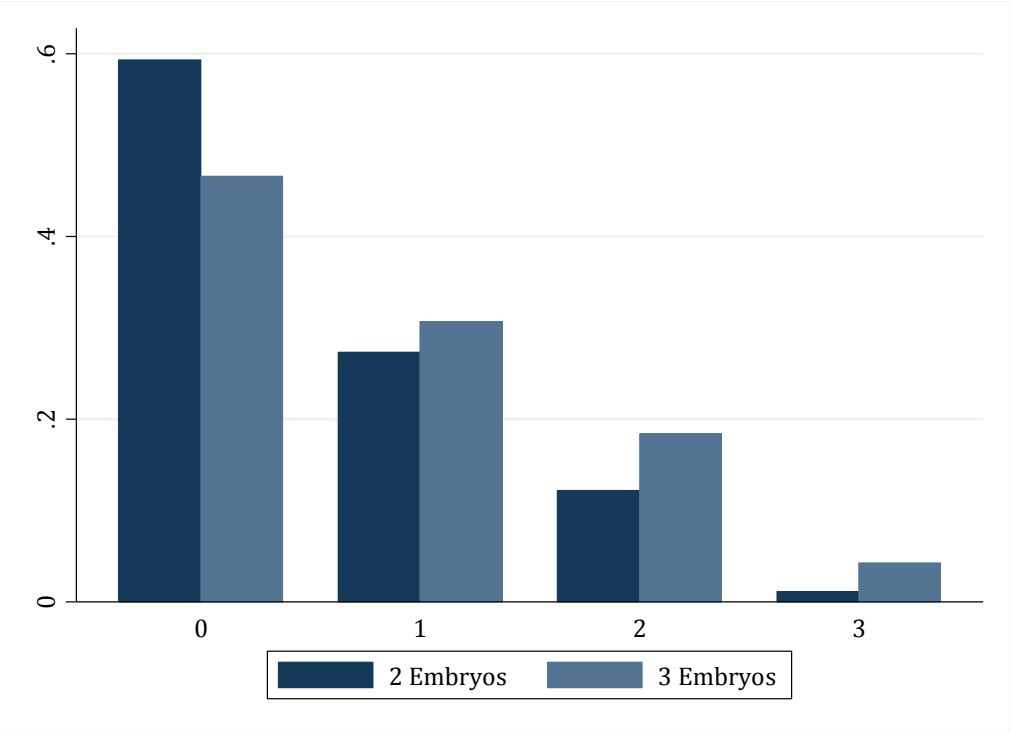


Figure 6: Distribution of Births by Patient Age, 3 embryos

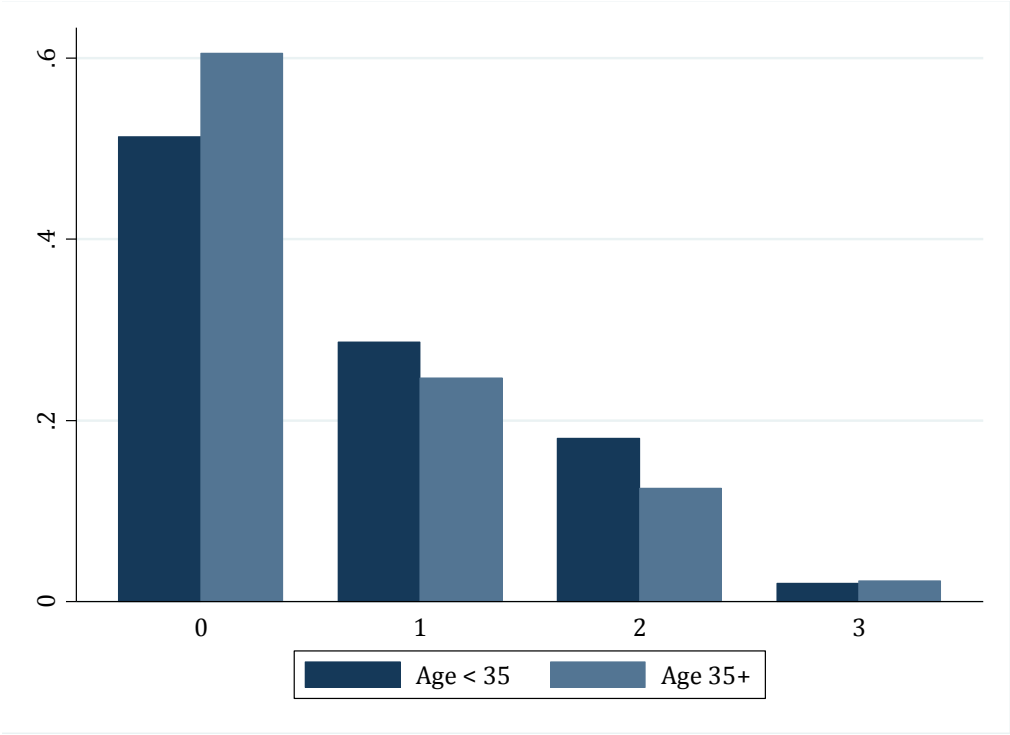


Figure 7: Predicted Stage 1 Decisions

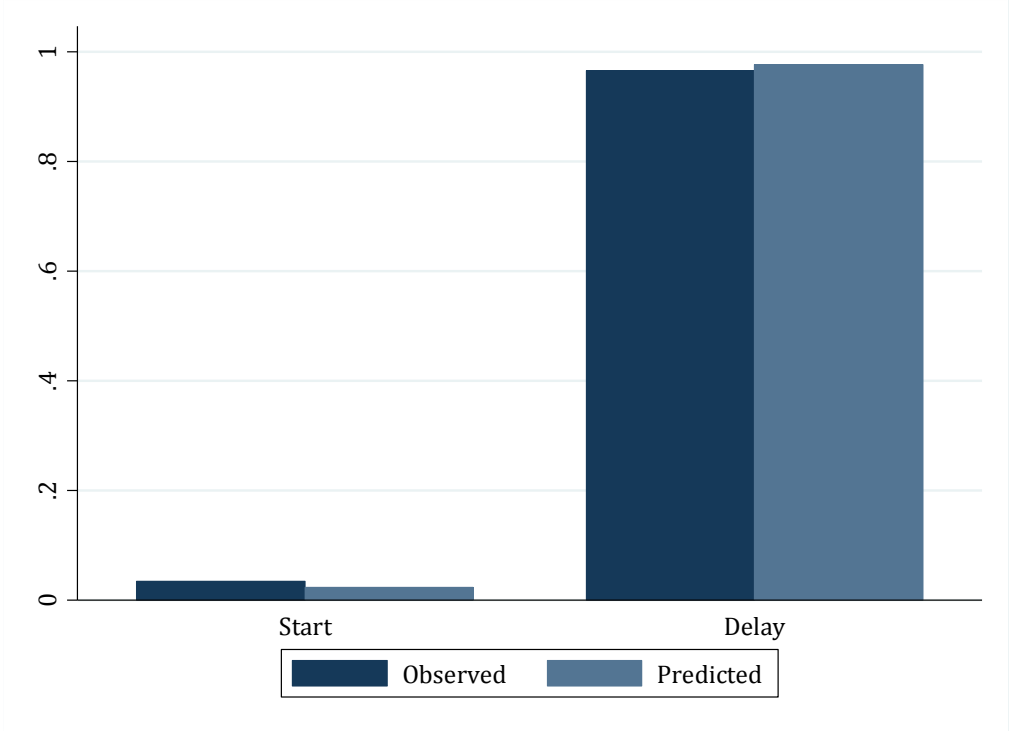


Figure 8: Predicted Stage 2 Decisions

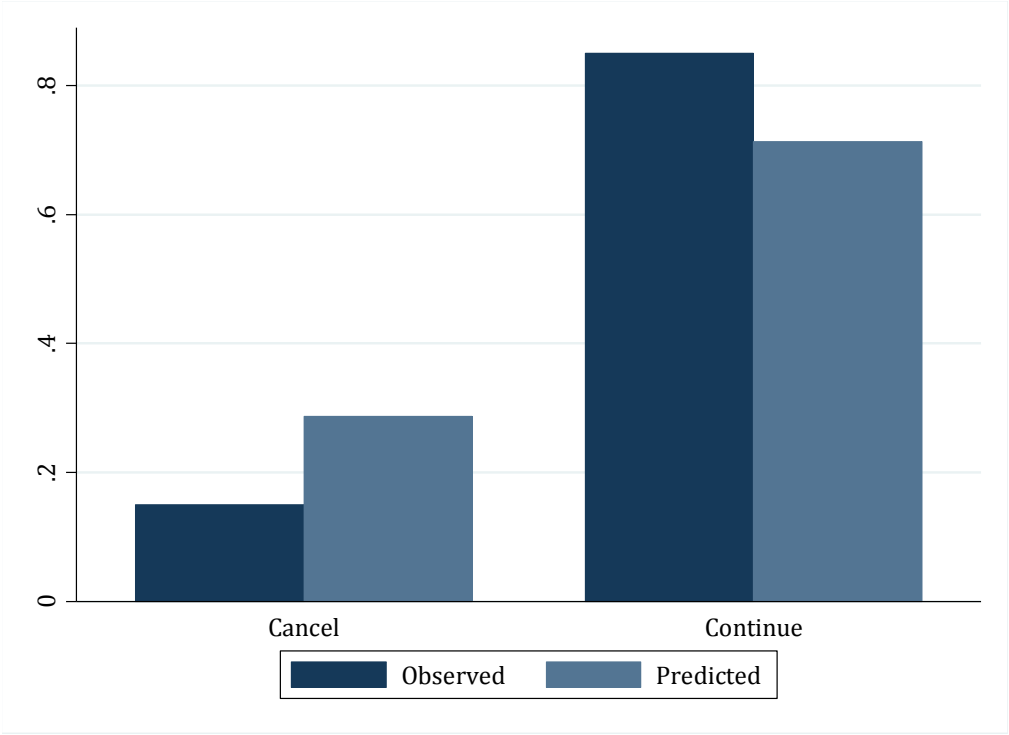


Figure 9: Predicted Stage 3 Decisions

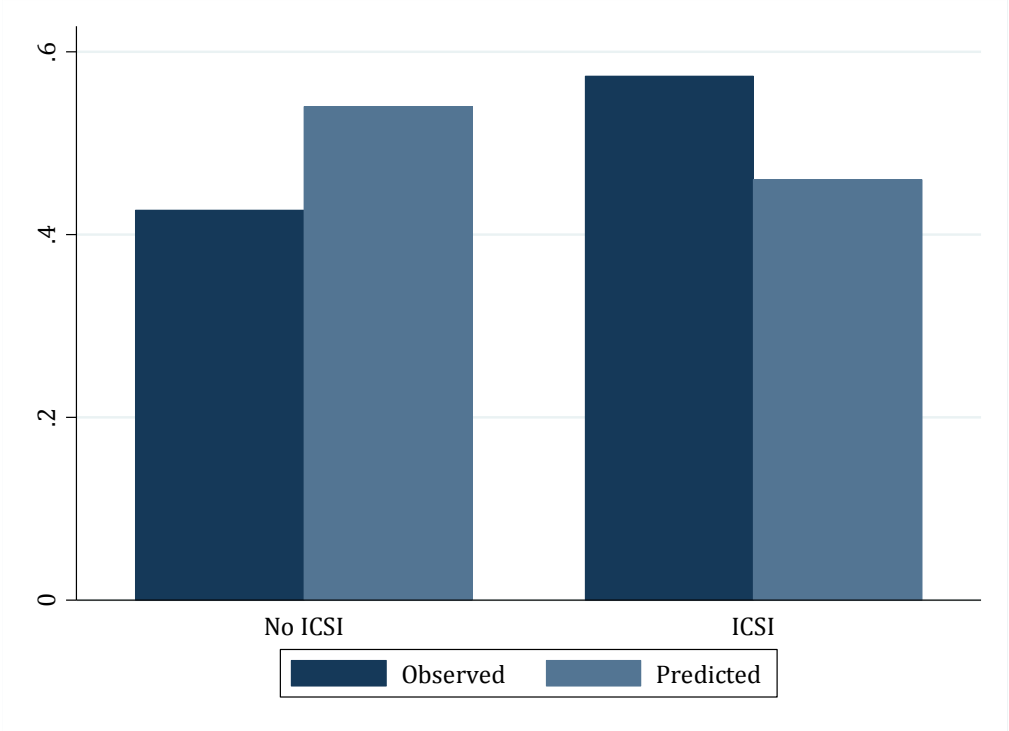


Figure 10: Predicted Stage 4 Decisions on Embryos Transferred

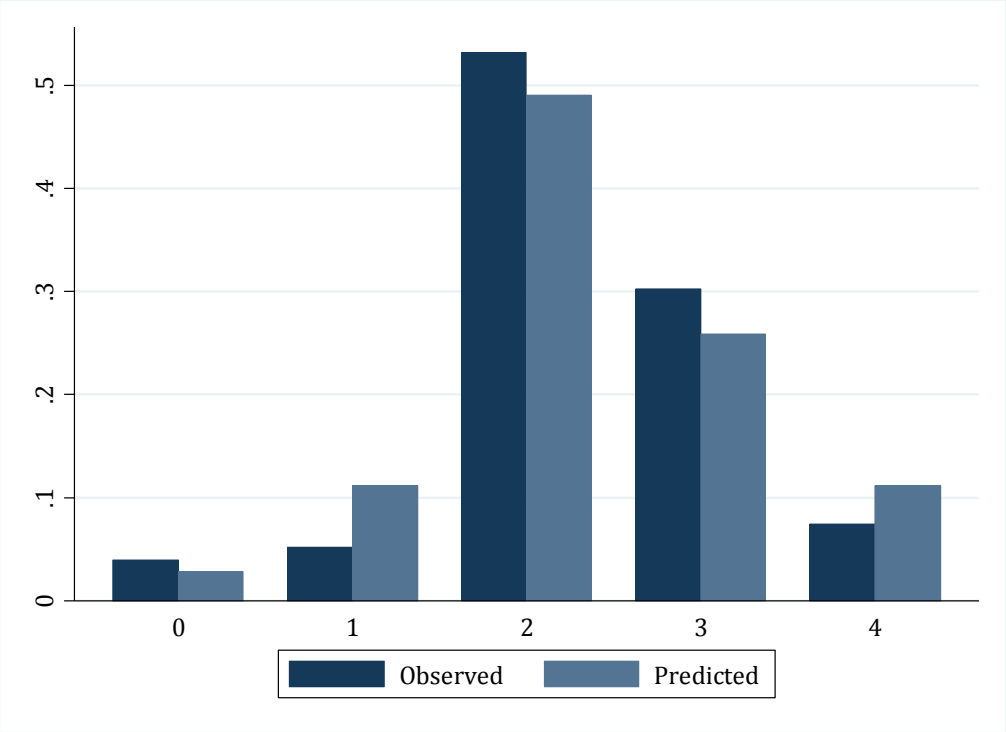


Figure 11: Embryo Transfer Choices in Counterfactual Experiments

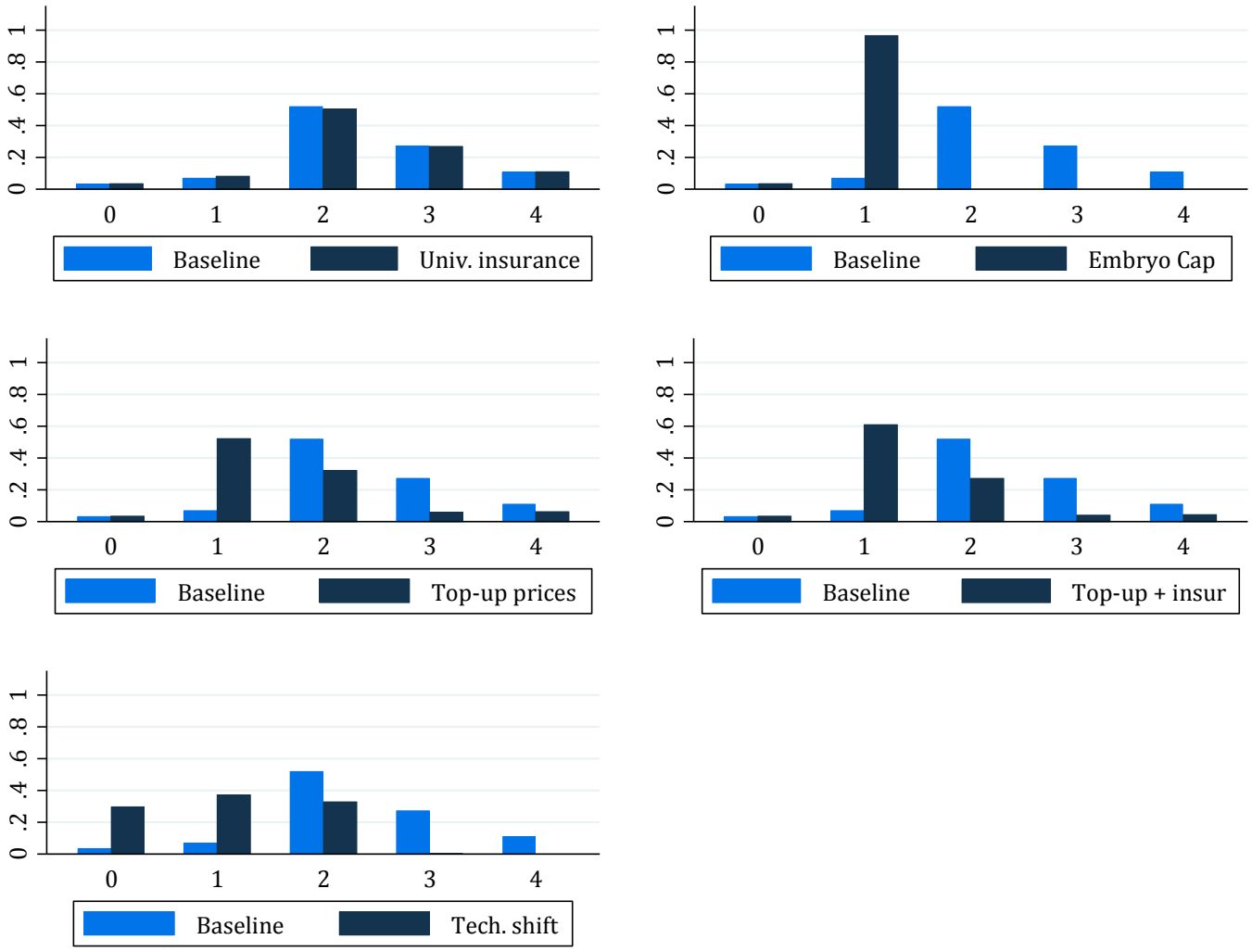


Figure 12: Births Outcomes in Counterfactual Experiments

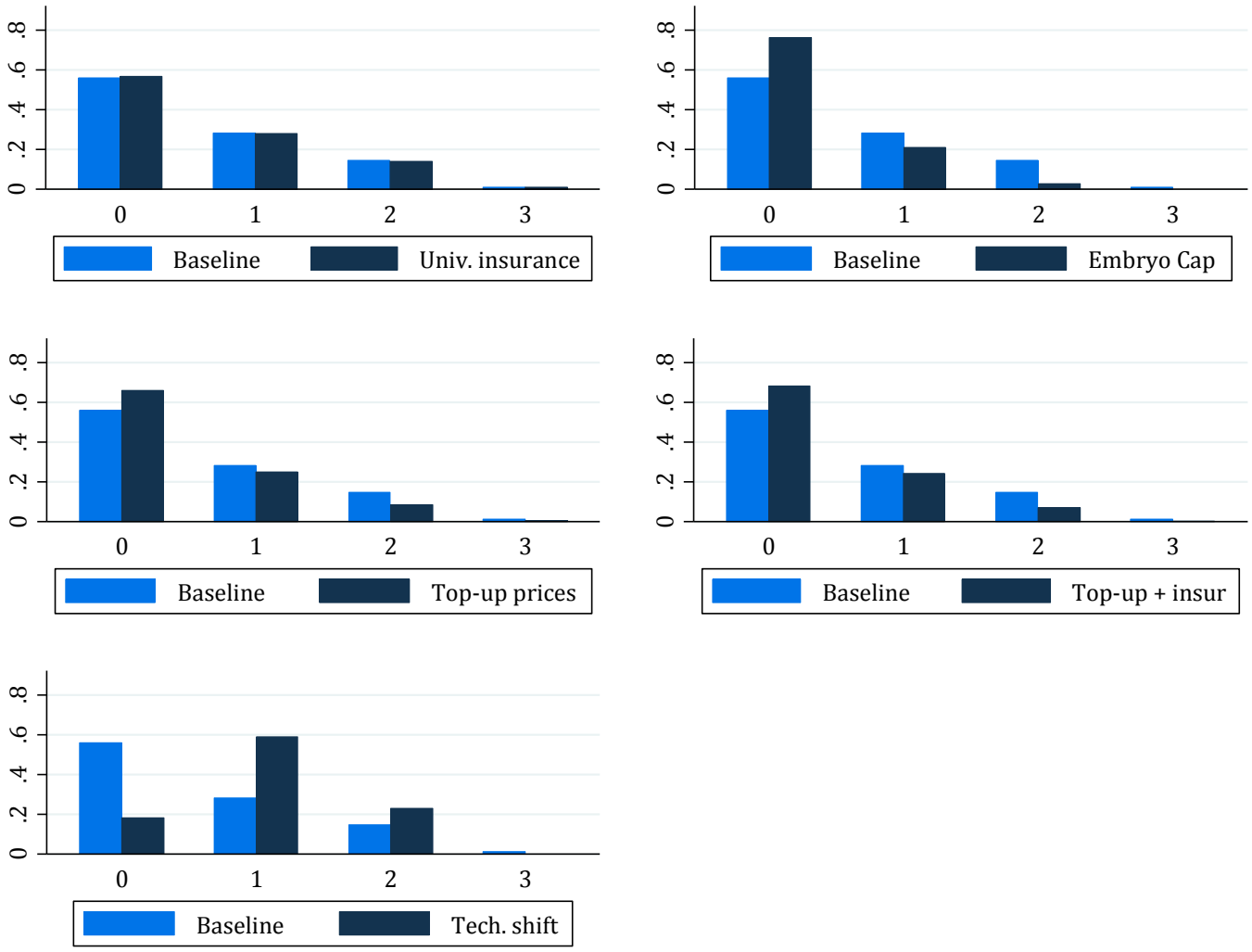


Table 1: Patient-level Characteristics

	Main sample N = 587		First-stage sample N = 1106	
	Mean	Std. dev.	Mean	Std. dev.
Demographic state variables (Z^D)				
Patient age at initiation	34.30	4.02	33.31	4.70
Insured at initiation? (Y = 1)	0.54	0.50	0.59	0.49
Wealthy zip code? (Y = 1)	0.82	0.39	0.79	0.41
Prior children at initiation	0.00	0.00	0.30	0.56
Biological state variables (Z^B)				
AFC score	14.37	7.96	14.61	8.13
Female fertility problem? (Y = 1)	0.80	0.40	0.69	0.46
Male fertility problem? (Y = 1)	0.34	0.48	0.30	0.46
Aggregate actions and outcomes				
Total cycles	1.75	1.02	1.97	1.21
Birth during sample period? (Y = 1)	0.53	0.50	0.55	0.50

The "Main sample" is used in second-stage estimation of patients' choices. The "First-stage sample" is used to estimate treatment technologies.

Table 2: Actions and Outcomes within Treatment

	Main sample			First-stage sample		
	N	Mean	Std. dev.	N	Mean	Std. dev.
Stage 1-4 actions						
Cancel treatment? (Y = 1)	1027	0.14	0.35	1859	0.14	0.35
Fertilization method? (ICSI = 1)	879	0.60	0.49	1597	0.59	0.49
Number of embryos transferred	875	2.29	0.81	1592	2.32	0.82
Stage 1-3 outcomes						
Peak E2 score	1027	16.82	9.73	1905	17.19	9.77
Eggs retrieved	879	10.60	5.46	1697	10.87	5.60
Embryos generated	881	6.11	3.75	1687	6.34	3.86
4+ embryos? (Y = 1)	881	0.74	0.44	1687	0.76	0.43
Stage 4 outcomes						
Children born	848	0.51	0.70	1632	0.55	0.75
Singleton birth? (Y = 1)	848	0.27	0.45	1632	0.27	0.45
Twin birth? (Y = 1)	848	0.12	0.32	1632	0.12	0.33
Triplet birth? (Y = 1)	848	0.00	0.00	1632	0.01	0.10

The "Main sample" is used in second-stage estimation of patients' choices. The "First-stage sample" is used to estimate treatment technologies.

Table 3: Utility Parameter Estimates

Utility of 1 birth (u_1)	5.147 (0.932)	Price sensitivity constant (α_0)	0.311 (0.071)
Utility of 2 births (u_2)	5.967 (1.690)	Price sensitivity X wealth (α_w)	-0.125 (0.067)
Utility of 3 births (u_3)	-14.063 (4.372)	Terminal payoff X Prev. payment (γ_p)	0.221 (0.633)
Utility shift when $\tilde{k} > 0$ (κ)	-11.698 (0.955)	Start/delay constant (δ_0)	4.917 (0.098)
Preference shifter ζ	9.631 (0.855)	Penalty for violating ASRM embryo guidelines	-3.049 (0.194)

Standard errors are in parentheses.

Table 4: Utility-Type Distribution Parameter Estimates

Type distr. constant (ρ_0)	-1.318 (0.556)	Type distr. insurance (ρ_3)	-0.073 (0.370)
Type distr. age (ρ_1)	0.044 (0.013)	Type distr. ASRM regime (ρ_4)	0.442 (0.339)
Type distr. weath (ρ_2)	-0.861 (0.458)		

Standard errors are in parentheses.

Table 5: Initiations, Cycles, and Outcomes across Policy Settings

Policy setting (g)	Share initiating	N cycles if initiate	Share with birth	N births	N infants
Baseline	0.384 (0.288, 0.449)	1.628 (1.524, 1.770)	0.263 (0.192, 0.314)	581.9 (431.2, 702.6)	805.9 (600.6, 972.4)
Universal insurance	0.561 (0.500, 0.682)	1.726 (1.566, 1.909)	0.380 (0.345, 0.452)	864.6 (785.5,1025.3)	1195.2 (1087.2,1421.2)
Embryo cap	0.104 (0.050, 0.180)	1.361 (1.217, 1.547)	0.033 (0.016, 0.056)	71.1 (34.7, 120.4)	80.7 (39.6, 137.3)
Top-up prices	0.154 (0.084, 0.288)	1.474 (1.322, 1.637)	0.073 (0.040, 0.155)	157.8 (86.8, 338.4)	202.8 (111.8, 451.5)
Top-up prices + insurance	0.336 (0.217, 0.416)	1.625 (1.476, 1.798)	0.158 (0.104, 0.228)	345.9 (230.6, 503.5)	436.8 (292.1, 653.9)
Technology shift	0.482 (0.446, 0.573)	1.215 (1.161, 1.293)	0.323 (0.296, 0.388)	724.3 (662.0, 882.7)	952.7 (827.5,1163.8)

95% confidence intervals are in parentheses

Table 6: Surplus and Costs across Policy Settings

Policy setting (g)	Avg surplus in risk population ($000)	Total surplus in risk population ($M)	Total IVF insurance cost ($M)	Total medical delivery cost ($M)	Delivery cost per birth ($000)
Baseline	5.33 (3.14, 9.69)	11.4 (6.7, 20.8)	2.9 (2.5, 3.6)	39.3 (29.5, 47.8)	67.6 (66.0, 70.1)
Universal insurance	9.64 (6.99, 14.87)	20.7 (15.0, 31.9)	14.5 (13.2, 17.3)	57.7 (52.6, 68.6)	66.8 (66.1, 69.2)
Embryo cap	0.69 (0.26, 1.60)	1.5 (0.6, 3.4)	1.4 (0.8, 2.0)	2.8 (1.4, 4.7)	38.8 (36.8, 40.9)
Top-up prices	1.38 (0.56, 3.80)	3.0 (1.2, 8.2)	1.8 (1.1, 2.4)	5.8 (3.3, 13.0)	36.9 (34.1, 40.4)
Top-up prices + insurance	3.26 (1.64, 6.99)	7.0 (3.5, 15.0)	8.4 (5.4, 10.5)	12.3 (8.4, 18.0)	35.7 (35.0, 38.8)
Technology shift	9.00 (6.27, 14.98)	19.3 (13.5, 32.1)	2.7 (2.3, 3.3)	39.7 (29.5, 50.2)	54.8 (39.8, 59.5)

Notes: "$M" indicates millions of dollars. 95% confidence intervals are in parentheses