Healthcare Access, Costs, and Treatment Dynamics: Evidence from 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  
University of Chicago  
Center for the Economics of Human Development

July 4, 2016

Abstract

We study public policies designed to improve access and reduce costs for In Vitro Fertilization (IVF). High out-of-pocket prices can deter potential patients from IVF, while active patients have an incentive to risk costly high-order pregnancies to improve their odds of treatment success. We analyze IVF’s rich choice structure by estimating a dynamic model of patients’ choices within and across treatments. Policy simulations show that insurance mandates for treatment or hard limits on treatment aggressiveness can improve access or costs, but not both. Insurance plus price-based incentives against risky treatment, however, can together improve patient welfare and reduce medical costs.

Keywords: In Vitro Fertilization, insurance mandates, healthcare access and cost, pricing risky behavior

JEL Codes: G22, I11, I13, I18, J13

*Authors’ contact information: hamiltonb@wustl.edu, jpanta@uchicago.edu, mcmanusb@unc.edu, jungheimx@wudosis.wustl.edu. 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

Healthcare policymakers struggle with the conflicting goals of providing ready access to medical care while simultaneously containing costs. A variety of approaches have been taken to “bend the cost curve.” For example, in the U.K. a medical treatment may not be covered by national insurance if it fails to meet a cost-effectiveness criterion. U.S. insurance plans have focused on alternative cost-sharing arrangements as a means of influencing patient choice, but the scope of such arrangements is limited due to relatively low out-of-pocket maximum payments (Feldstein and Gruber, 1995). Optimal health insurance design suggests that a patient desiring an expensive, technologically advanced treatment should face higher co-payments (see, e.g., Chernew et al., 2000). Unfortunately, such insurance policies are rarely found in practice and few empirical studies (e.g. Einav et al., 2016) evaluate the ability of treatment-dependent pricing to influence patient behavior and control costs.

In this paper, we study the impact of alternative policies on patient access to care, patient surplus, and healthcare costs in the context of the market for In Vitro Fertilization treatment (IVF) in the U.S. IVF use has increased substantially in recent years as the technology has matured, 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 annually and resulting in 1% of all births.\footnote{Treatment rates are even higher in some other countries; IVF accounted for over 3% of 2011 births in Israel, Belgium, Sweden, and Denmark. The aggregate treatment statistics reported in this paragraph for the U.S. 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). 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).} In addition, IVF patients now account for approximately 50% of all higher order births, which are very costly for the healthcare system.\footnote{The average cost of a singleton birth was $27,000 in 2012, while twin and triplet births cost $115,000 and $435,000 (Lemos et al (2013).} Despite the growth in demand for IVF, 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, coupled with within-treatment choices that may ameliorate them, require that infertile patients considering IVF solve a complex sequential decision problem. In this way, IVF resembles other medical treatments like those for heart disease or cancer, where dynamic decisions are made across a variety of treatment steps while health information is gradually revealed.\footnote{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...}
High out-of-pocket costs for IVF (typically $10,000 - $15,000 per cycle of treatment) and the high rate of expensive multiple births have led policymakers to consider a variety of interventions in the hopes of achieving two goals: improving IVF access in the population and reducing treatment risks and costs. One potential avenue to increase access is through subsidized treatment or mandatory insurance coverage. Insurance mandates, which have been implemented in ten U.S. states, can reduce the out-of-pocket price of each IVF attempt to $2,000-$3,000 and thus lead to higher utilization rates. Despite increased access, this policy also has the potential to contain costs. This might be the case if patients, aware that future insured attempts would also occur at a lower price, decide to engage in less aggressive treatment by taking fewer embryos. Transferring more embryos increases both the probability of pregnancy and the likelihood of a multiple birth. Another potential policy intervention is an explicit 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. While this policy nearly eliminates costly high-order births, it may have the effect of reducing access to IVF. When a patient can transfer only a single embryo, her odds of conceiving a child are greatly reduced, and this could lead patients to forego pursuing IVF altogether or abandoning treatment after unsuccessful initial attempts.

To evaluate the impact of alternative policies on access, consumer surplus, and costs, we specify a structural model of a forward-looking, infertile patients’ decisions to initiate and or/continue IVF treatment using fresh, non-donor eggs and embryos. Treatment choice dynamics for IVF patients reflect a number of mechanisms. First, patients must consider both the current and future price of treatment because outcomes are uncertain. Due to high potential future costs, patients may choose aggressive current treatment that raises both the probability of pregnancy and the likelihood of a multiple birth. As the medical costs associated with birth are generally paid by

---

4 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.

5 A large majority of IVF cycles use the patient’s own eggs (i.e. non-donor) and transfer embryos that were never frozen.

6 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. (2015) examine the dynamic implications of changes in out-of-pocket costs induced by exhaustion of deductibles.
insurers, treatment aggressiveness may be described as moral hazard. 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. For example, the success rate for patients in their early 40s falls to around 10% and twins are rare conditional on success. As treatment choices are affected by intertemporal variation in prices, health, and information, we expect that possible regulatory changes could have subtle impacts throughout the decision process.

We estimate the model using a novel dataset of the treatment histories of 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, and health 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 and the disutility of payments, yield a finite-horizon dynamic optimization problem that spans the reproductive years of (otherwise) clinically infertile women. The model delivers policy functions that specify: a) whether to initiate an IVF treatment, b) the treatment choices within an IVF cycle, and c) whether (and when) the patient re-_attempts to conceive through IVF. We estimate the empirical model in three steps. The first step is the estimation of the treatment “technologies” determining outcomes during the four stages of a given IVF cycle. The second step recovers the structural parameters of our within-clinic patient decision model by maximizing the likelihood of the observed treatment choice histories. 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. These estimates are necessary to assess how patients would respond to policies designed to reduce treatment aggressiveness. The third step captures potential patients’ decisions to ever initiate treatment.

We then conduct counterfactual policy experiments to investigate the impact of alternative policy proposals on individual patient actions, outcomes, and surplus. First, we estimate the effect
of extending insurance coverage to all women in the market. This policy’s primary impact is to substantially increase the number of women who ever initiate treatment, from 38% of clinically infertile 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 in principle could affect embryo transfer rates, we find only a small reduction in treatment aggressiveness because our estimates imply that patients receive about the same utility from singleton and twin births, implying little reason to transfer fewer embryos. Therefore, while this policy succeeds at increasing access, it is of no help in terms of cost control. Second, we explore the impact of restricting patients to transferring a single embryo during treatment. As expected, this policy has a clear effect in nearly eliminating multiple births. However, active patients are much less likely to conclude treatment with a child, and individuals are much less likely to initiate treatment. Consumer surplus shrinks to $700 per potential patient. While an embryo cap clearly helps control costs, it comes at the price of reduced access by leading many potential patients to turn away from treatment.

Finally, because universal insurance and embryo caps fail to deliver on the dual goals of access and cost control, we consider a pricing scheme closer to those suggested by theoretical models of optimal insurance design. In particular, we explore the impact of adding “top-up prices” for patients transferring more than one embryo. While current practice in IVF allows patients to increase the number of embryos at zero additional price, the full social cost of additional embryos is positive due to the increased medical costs of high-order pregnancies. We select top-up prices that cover the (expected) increase in medical expense relative to single-embryo transfers, and thus eliminate this potential source of moral hazard. Relative to the empirical patterns, we find that patients reduce but do not eliminate multiple-embryo transfers. As top-up 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 utilization 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 fact, total insurance payments are lower under this scenario and therefore insurance companies might be willing to extend universal coverage as long as it is

\footnote{Einav, Finkelstein, and Williams (2015) consider a closely related policy of top-up prices for medical care beyond a basic level covered by insurers.}
supplemented with our proposed top-up price design.

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 and our setting

2.1 IVF background

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.

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. 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. The medical costs calculated by Lemos et al. (2013) are typically not paid in full by the IVF patient herself, so they introduce the possibility of moral hazard to the embryo transfer decision.\(^8\)

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. More broadly, there exists some survey evidence (Hojgaard et al., 2007; Ryan et al., 2004) which suggests that patients may prefer a twin birth to a singleton. This preference could be motivated by a variety of factors, including declining fertility with age and the view that a twin birth is a cost-effective way to achieve a goal of multiple children.

### 2.2 Public policy and analysis of its impact

A key feature of the U.S. 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.\(^9\) Prior empirical research on insurance mandates has used a variety of approaches to examine issues related to IVF access, embryo transfer patterns, and birth outcomes. The distinguishing feature of the past research is that potential patients’ activity and outcomes are examined in aggregate, either at the population or clinic level. Schmidt (2005, 2007), Bitler (2008), Bundorf et al. (2008), and Buckles (2013) examine the impact of insurance mandates on population-level fertility rates, multiple birth rates, and child health. Hamilton and McManus (2011), Jain et al. (2002), and Henne and Bundorf (2008) investigate the impact of mandates on the number of patients served, embryo-transfer patterns, and birth outcomes at IVF clinics. Bitler and Schmidt (2006, 2012)

---

\(^8\)Some medical conditions associated with early births can require life-long medical care, but we do not consider that care’s costs in this paper.

\(^9\)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.
use survey data to examine similar issues. While patient-level decisions and characteristics play a central role in generating the aggregate activity studied in these papers, their empirical approaches do not allow a direct examination of choices, counterfactual experiments, or welfare effects.\footnote{Some studies exist in the medical literature that use patient-level IVF data (e.g. Malizia, Hacker, and Penzias, 2009), but their focus and approach are naturally different from our own.}

An additional policy proposal, motivated by the goal of reducing multiple births and their medical costs, is a single-embryo cap for IVF. 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. The prior literature on this policy has focused on the medical impact of single-embryo transfers, whether voluntary or by rule. For examples, see Ryan et al., 2007, Jungheim et al., 2010, Csokmay et al., 2011, or Velez et al. (2014). By contrast, our focus on individual utility-maximizing patients allows us to consider treatment initiation decisions and patient welfare in addition to the impact of an embryo cap on multiple birth rates and medical costs. Patient initiation and welfare are likely to be affected by a cap because additional (costly) cycles are needed, in expectation, to achieve a pregnancy, and because some patients may prefer twin outcomes to singletons births.

2.3 The clinic we study and its market

As we study a single IVF clinic, we assess how the clinic and market we study compares to the full U.S. IVF market (Centers for Disease Control, 2014). In Table 1 we present some statistics on clinic characteristics during the sample period of 2001 to 2009. The clinic we study was larger than the average U.S. clinic, with 370.2 cycles per year compared to the U.S. average of 217.1 cycles. This may be due to the clinic’s position in a large research-oriented hospital, which could both influence demand and facilitate the hiring of the clinic’s medical staff. The median patient age in our sample is 34 years old, which is slightly below the national median. The difference in patient age could be an equilibrium response to local insurance patterns, described below, which could encourage potential patients to begin IVF earlier following difficulty conceiving a child naturally. Patient health, as measured by the fraction of patients with multiple female infertility diagnoses and the fraction of cases with a male infertility diagnosis, are similar between the clinic we study and the national averages. Patient treatment choices over ICSI and embryos transferred reflect less aggressive treatment in the clinic we study relative to the U.S. averages. These treatment
patterns, along with a greater birthrate per cycle (33.5% at the clinic vs. 29.9% nationally), may be influenced by the difference in median patient age, which affects both treatment choices and outcomes.

Turning to the clinic’s market, we examine how some critical demographic features compare to the U.S. population. In our empirical analysis, 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. Our model accounts for age at first birth and wealth as measured through housing value. Fertility timing in the St. Louis market is nationally representative: within the age range we study (22-44) the median maternal age at first birth in 2001 was 31 in both the St. Louis metro area and nationally. We use zip code median home value as a proxy for wealth. In the St. Louis MSA this value was $103,900 in the year 2000, while nationally the value was slightly higher at $112,100. More specifically, 54.4% of St. Louis-area women lived in zip codes with median home values above $100,000. Nationally the number was 56.7%.

Next, we describe the Illinois insurance mandate, which affects a large number of potential patients in the market we study. For patients in our study residing in Illinois and working for an employer covered by the mandate, insurance plans are required to pay for up to 4 cycles of IVF if the woman has no children. This insurance coverage pays the majority of IVF costs, but about $3,000 in expenses remain for insured patients due to co-payments and deductibles. For patients paying out-of-pocket for IVF in our sample, the clinic charged about $11,000 per treatment cycle throughout the sample period; this figure includes all drugs and medical procedures that are part of a complete IVF cycle. 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 employees. The clinic has found that among those few patients from Missouri

---

11 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.
whose health insurance plan covers IVF (despite the absence of a mandate), coverage details are similar to those of Illinois patients with insurance plans under the mandate. The patient-level information on insurance status allows us to exploit two sources of price variation in this setting: cross-section variation is present through the locations of potential patients, while longitudinal variation arises as patients exhaust their insurance coverage over the course of multiple IVF cycles. We assume that St. Louis-area potential patients do not endogenously move or change employers in order to receive mandated insurance coverage under Illinois’s IVF insurance regulations.\footnote{The typical patient takes two IVF treatments, which would provide up to $16,000 in savings on out-of-pocket treatment expenses. Individuals who seek these savings, however, could experience substantial monetary or utility costs related to job search and switching, housing search and switching, and potentially suboptimal matches in the Illinois labor and housing markets.}

3 Model preliminaries

3.1 Timing

We consider two timing concepts in the model below. First, there are decision \textit{periods} when active patients choose whether to start or delay an IVF cycle. Second, there are four treatment \textit{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. The outside option, which we parameterize below, could include adoption, surrogacy methods of reproduction, or remaining childless. If the woman does not begin IVF at this critical moment, we assume she collects the value of the outside option and 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 \(a_{\text{min}}\). 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 $a^\text{max}$. At this age the IVF clinic will no longer treat the patient and her birth probability (via IVF or naturally) is zero.\textsuperscript{13} This allows a maximum of $4 \times (a^\text{max} - a_0 + 1)$ periods for a patient whose reproductive decisions start at age $a_0$. We set $a^\text{min} = 28$ and $a^\text{max} = 44$ 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 $a^\text{max}$ 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 ($\tilde{k}$), record of previous payments for IVF, and insurance status in the state vector $Z^D$.\textsuperscript{14} 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 ($\tilde{k} = 0$) at the treatment initiation decision, and then $\tilde{k}$ 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 ($\ell$) to four (the Illinois mandate value) for all patients

\textsuperscript{13}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.

\textsuperscript{14}Bitler and Schmidt (2006) document differences in infertility diagnoses and IVF usage across racial, ethnic, and socioeconomic groups. Unfortunately we do not have data on (potential) patients characteristics which would allow us to explore these issues directly.
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. 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
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 treat-
ment 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. We assume that the patient’s observed
biological state variables $Z^B$ and age fully capture her relevant fertility characteristics. As implied
by this assumption, we do not model additional unobserved health characteristics which the patient
and her doctor learn about through IVF treatment itself.

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 $Z_a = [Z_a^D, Z_a^B]$, 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 ($\bar{k}$) at the start of an IVF cycle and other personal characteristics.
Possible values of $k$ are in $\{0, 1, 2, 3\}$, and $\bar{k}$ takes values in $\{0, 1, 2\}$. (These values for $\bar{k}$ cover

\[15\text{ We abstract away from some of the complicated details of the Illinois insurance code, discussed above.}\]
98% of the patient population at the clinic.) We allow patients to have permanent unobservable heterogeneity, indexed by \( \tau \). Let \( U(k|\bar{k}, \tau) \) represent the lump-sum utility payoff from a treatment cycle that ends in \( k \) children conditional on \( \bar{k} \) and \( \tau \). As a normalization, we assume that treatment outcomes with \( k = 0 \) always result in \( U(k|\bar{k}, \tau) = 0 \) for all patients. While we treat \( U(k|\bar{k}, \tau) \) as a lump-sum benefit, we interpret it as the expected utility from \( k \) children, which may include health risks from high-order births.

To maintain tractability, we combine \( U(k|\bar{k}, \tau) \) additively with other payoff-related terms in an indirect intertemporal expected utility function. Patients undergoing treatment experience disutility, scaled by \( \alpha \), from paying positive prices. When a patient pays \( p \) within treatment, she has the immediate utility loss of \( \alpha p \). We allow the value of \( \alpha \) 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 \( \eta \) 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 starting versus delaying treatment. Relative to a baseline flow utility from delay that is normalized to zero, we assume that patients receive the flow benefit (or cost) of \( u_s(Z) \) during any period in which she begins treatment in stage 1. \( u_s(Z) \) can include any physical or psychological stress from undergoing IVF. Patients’ terminal payoffs are captured by the parameter vector \( u_T(Z_{a_{\text{max}}}) \). The patient receives \( u_T \) at age \( a_{\text{max}} + 1 \) regardless of whether she remains active in the model up until \( a_{\text{max}} \) or if her decision process ends due to \( k + \bar{k} \geq 3 \) at some earlier \( a \).

At each treatment node, the patient’s benefit from the available options includes an additional taste shock, \( \varepsilon \), which represents heterogeneity in patient’s circumstances and preferences. Following Rust (1987), for computational ease, we assume that \( \varepsilon \) 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 \( \beta \). 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 \( \beta^t \). We do not estimate \( \beta \) 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 \( \tau \), 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

\[
W_{y,j}(Z_a, \varepsilon_{y,j,a}) = \overline{W}_{y,j}(Z_a) + \varepsilon_{y,j,a}.
\]

(1)

\( E[W_j(Z_a, \varepsilon_{j,a})] \) 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

\[
W_j(Z_a, \varepsilon_{j,a}) = \max_{y \in Y_j} \{ W_{j,y}(Z_a, \varepsilon_{y,j,a}) \} = \max_{y \in Y_j} \{ \overline{W}_{y,j}(Z_a) + \varepsilon_{y,j,a} \}.
\]

(2)

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

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

(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 \(p_s(Z_a)\) out-of-pocket and begins a regimen of pharmaceuticals to promote egg production.

The value from starting a treatment cycle at age \(a\) is \(W_{s,1}(Z_a, \varepsilon_{s,1,a})\), and it includes the expected value from continuing to the second stage of treatment \(E[W_2(Z_a, \varepsilon_{2,a})]\); the utility normalization relative to delay, \(u_s\); the price of starting a treatment cycle, \(p_s(Z_a)\); 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

\[
W_{s,1}(Z_a, \varepsilon_{s,1,a}) = W_{s,1}(Z_a) + \varepsilon_{s,1,a} = u_s(Z_a) - \alpha(Z_a)p_s(Z_a) + \varepsilon_{s,1,a} + \sum_e E[W_2(e, Z_a, \varepsilon_{2,a})]f_e(e|Z_a).
\]

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

\[
W_{d,1}(Z_a, \varepsilon_{d,1,a}) = W_{d,1}(Z_a) + \varepsilon_{d,1,a} = 0 + \beta E[W_1(Z_{a+1}, \varepsilon_{1,a+1})] + \varepsilon_{d,1,a}.
\]

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 \(\beta E[W_1(Z_{a+1}, \varepsilon_{1,a+1})]\) accounts for the expectation of \(\varepsilon\), the payoffs in \(W_1\) 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

$$W_{nc,2}(e, Z_a, \varepsilon_{nc,2,a}) = W_{nc,2}(Z_a) + \varepsilon_{nc,2,a}$$

$$= 0 + \beta E[W_1(Z_{a+1}, \varepsilon_{1,a+1})] + \varepsilon_{nc,2,a}.$$

The value of continuing treatment includes an expectation taken over values of $r$ conditional on the realized signal $e$ and other patient characteristics:

$$W_{c,2}(e, Z_a, \varepsilon_{c,2,a}) = W_{c,2}(e, Z_a) + \varepsilon_{c,2,a}$$

$$= -\alpha(Z_a)p_c(Z_a) + \varepsilon_{c,2,a} + \sum_r E[W_3(r, Z_a, \varepsilon_{3,a})] f_r(r|e, Z_a).$$

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 ($zmf = 1$) 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. 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:

$$W_{m,3}(r, Z_a, \varepsilon_{m,3,a}) = W_{m,3}(r, Z_a) + \varepsilon_{m,3,a}$$

$$= -\alpha(Z_a)p_{m,3}(Z_a) + \varepsilon_{m,3,a} + \sum_X E[W_4(X, Z_a)] f_X(X|r, m, Z_a).$$
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

$$W_{x,4}(X,Z_a,\varepsilon_{x,4,a}) = \bar{W}_{x,4}(X,Z_a) + \varepsilon_{x,4,a}$$

$$= \eta(x,Z_a) + \varepsilon_{x,4,a} + f_k(0|x,Z_a)\beta E[W_1(Z_{a+1},\varepsilon_{1,a+1})] + \sum_{k>0} f_k(k|x,Z_a)\left\{U(k|\bar{k},Z_a) + \beta^k E[W_1(Z_{a+4},\varepsilon_{1,a+4})]\right\}.$$  

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

$$W(Z_{a_0}^D) = E[W_{s,1}(Z_{a_0})|Z_{a_0}^D] = \sum_{Z_B^a} W_{s,1}(Z_{a_0}^D, Z_B^B) f_{Z_B}(Z_B^B|a_0),$$

where we make the distinction between the state variables known prior to treatment ($Z_{a_0}^D$) and those learned after treatment begins. The potential patient compares $W(Z_{a_0}^D)$ to the utility from foregoing treatment, which we specify as $W^{OUT} = \mu + \nu$. The parameter $\mu$ captures the mean value of the outside option, and it is common across potential patients. $\nu$ 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) $\tau$ make different choices with respect to ever pursuing IVF.\(^{16}\) One possible interpretation of $\nu$ is that of a sunk utility cost that must be paid to pursue IVF. Some infertile women with strong objections to IVF, e.g. because of religious beliefs, may refuse treatment even if it is free and has no chance of failure.

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 initiate IVF treatment, and equal zero otherwise. Then

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

To make the decision problem more explicit in some of the analysis below, we sometimes add the policy index $g$ to our notation for the value of starting a very first cycle as $W_{s,1}(Z_{a_0}, g)$. We let $g = g_B$ represent the empirical “baseline” that we observe during the sample period. The index $g$ captures elements such as pricing, insurance, regulations, etc. Under alternative environments $g'$ the value of $W_{s,1}$ will change and therefore initiation decisions will be affected. In fact, all $W_{y,j}$ values change when the policy environment changes.

\(^{16}\)Note that there is no $\varepsilon_{1,s}$ for this very first cycle.
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.17 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-stage sample.” The treatment histories contain information on 1,027 initiated cycles in the main sample and 2,176 in the full first-stage sample. Beyond the initiation decision and its Peak E2 score realization, the main sample contains information on an additional 2,785 choices and stochastic health state outcomes from stages 2-4 of IVF cycles. In the full first-stage sample we have complete data on 4,779 stage 2-4 choices and health outcomes.

In Table 2 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 $100,000, which we use as a proxy for patient wealth ($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 2 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

---

17In 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.
experience at least one birth during their full treatment history. In Table 3 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. 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 3 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. Almost 60% of patients whose first cycle was unsuccessful returned for one or more additional cycles, while 16% of patients with a first-cycle singleton and zero patients with first-cycle twins returned for additional treatment. 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.

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.

18 While the options “full ICSI” and “partial ICSI” are separated in the data, we group them together in our model.

19 In 2009, 1.6% of all U.S IVF births from fresh non-donor cycles were triplets or more. Conditional on 3 or more embryos transferred, the triplet rate was 3.6% (Centers for Disease Control, 2014).
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 the clinic we study (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, while others 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 $f_{inf}(a)$ be an age-specific infertility rate, which increases with age. Together, $(N_{stl}, f_t(a), f_{inf}(a))$ provide the number of women, $\tilde{N}_a^{inf}$, of each age who realize that they are unable to conceive without IVF. These $\tilde{N}_a^{inf} = \sum_{a=a_{\text{min}}}^{a_{\text{max}}} \tilde{N}_a^{inf}$ 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 $\tilde{N}^{inf}$ to match the clinic’s market share as reported by the CDC. Our final estimate of $N^{inf} = 2146$. See Appendix B for additional details on the calculation of $N^{inf}$ and its relationship to $f_{ZD}(Z_{a_0})$, the distribution of patient characteristics.

We compute an empirical initiation share, $s^{\text{init}}$, for the clinic we study. We observe that $N^{clin} = 828$ new patients initiated treatment at the clinic during the period 2001-2007.\footnote{We use the 587 with complete data in estimation, but have records for 828 patients initiating treatment over this period.} Using our estimate
of $N_{inf}$, we calculate the share

$$s_{init} \approx \frac{N_{clin}}{N_{inf}} = \frac{828}{2146} = 0.39,$$

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 $W_{1,s}$.

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. We assume that the drug dosage is a function of the patient’s age and her values of $Z^B$. The woman’s characteristics and drug dosage 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 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:

\[
U(k|\{k\}) = u_k + \kappa \times 1\{k > 0\} + \zeta \times 1\{\tau = 2\}
\]

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 (\( \{k\} = 0 \)). 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 \( \kappa \) captures any difference in the marginal benefit
of a birth to patients with prior children; diminishing marginal utility from children would imply that \( \kappa \) 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 \((u_1, u_2, u_3, \kappa)\), while the remaining patients (with \( \tau = 2 \)) has, in addition, its utility payoff shifted by a scalar parameter \( \zeta \). A patient’s probability of being of type \( \tau = 2 \) depends on her state values at the time she initiated treatment, \( Z^{D}_{a0} \). Along with \( a_0 \), we allow the distribution of \( \tau \) to depend on a measure of her wealth level \((z_w)\), her initial number of insurance-covered cycles \((a_0)\), and a dummy \((z_{asrm0})\) for the ASRM guideline regime when treatment started. We assume that the probability of a high type \( (\tau = 2) \) is

\[
\Pr(\tau = 2|Z^{D}_{a0}, I = 1, \rho) = \frac{\exp(\rho_0 + \rho_1 a_0 + \rho_2 z_w + \rho_3 t_{a0} + \rho_4 z_{asrm0})}{1 + \exp(\rho_0 + \rho_1 a_0 + \rho_2 z_w + \rho_3 t_{a0} + \rho_4 z_{asrm0})}.
\]

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 \( \rho \) represent a column vector of \( \{\rho_j\}^4_{j=0} \) values. In addition, we write \([1, Z^{D}_{a0}]\) as a vector containing 1 and an individual patient’s row vector \( Z^{D}_{a0} \), and we let \( \Lambda \) represent the logistic distribution function so that \( \Lambda([1, Z^{D}_{a0}]|\rho) = \Pr(\tau = 2|Z^{D}_{a0}, I = 1, \rho) \).

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 \( \alpha_0 \) 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 \( \nu \) is distributed according to a continuous distribution \( F(\nu) \) in the population of potential patients. The realizations of \( \nu \) are i.i.d. Moreover, we assume \( \nu \) is independent of infertility problems and other observables in our model, so \( F(\nu|\tau, Z_{a0}^D) = F(\nu) \). We assume \( \nu \sim \text{Logistic} \) so we have \( F(\nu) = \Lambda(\nu) = \frac{\exp(\nu)}{1+\exp(\nu)} \).

Let \( \varphi \) represent a vector of all of the parameters except \( \mu, \zeta, \) and \( \rho \), and define \( \theta = (\zeta, \varphi, \rho) \). We estimate \( \mu \) separately from \( \theta \) 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 combination of fertility-related state variables 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 \( f_{ZB}(Z^B|a_0) \) 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 \( \theta \) using data exclusively from the population of 587 patients who are observed within the clinic. In the final stage we estimate \( \mu \) using our estimates of \( E[W_{s1}(Z_{a0})|Z_{a0}^D] \) 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 \((\zeta, \varphi)\), and the distributional assumptions on \( \varepsilon \), we are able to calculate \( W_{y,j}(Z_a, \tau; \zeta, \varphi) \) and \( E[W_j(Z_a, \tau; \zeta, \varphi)] \) for each \( y \) and \( j \) at every \( Z_a \). We perform this calculation by backward recursion separately for each type \( \tau \). For each potential state that might be reached when the patient is age \( a_{\text{max}} \), we use \((\zeta, \varphi)\) to compute the terminal payoff, the values of \( W_{y,j}(Z_{a_{\text{max}}}, \tau; \zeta, \varphi) \) working backwards through treatment stages, and the logit inclusive value \( E[W_j(Z_{a_{\text{max}}}, \tau; \zeta, \varphi)] \) for each stage. We then move to age \( a_{\text{max}} - 1 \) and use the \( a_{\text{max}} \) expected utility values while constructing \( W_{y,j}(Z_{a_{\text{max}}-1}, \tau; \zeta, \varphi) \) and \( E[W_j(Z_{a_{\text{max}}-1}, \tau; \zeta, \varphi)] \). The procedure continues back to age \( a_{\text{min}} \).
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 \( W_{y,j}(Z_a, \tau; \zeta, \varphi) \) for all \( Z_a \) and \( \tau \) to compute choice probabilities for each observed decision in our data. Conditional on a patient’s type \( \tau \), calculating this probability is a straightforward task due the i.i.d. extreme value assumption for the \( \varepsilon \) 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:

\[
\Pr(d_{c,2,a,i} = 1|Z_a, \tau; \zeta, \varphi) = \frac{\exp[W_{c,2}(Z_a, \tau; \zeta, \varphi)]}{\exp[W_{c,2}(Z_a, \tau; \zeta, \varphi) + W_{nc,2}(Z_a, \tau; \zeta, \varphi)]}
\]  

The values of \( W_{c,2}(Z_a, \tau; \zeta, \varphi) \) and \( W_{nc,2}(Z_a, \tau; \zeta, \varphi) \) are relatively simple functions of the estimated transition \( \hat{f}_r \), price and its disutility parameter, and the calculated values of \( E[W_3(r, Z_a, \varepsilon_{3,a})] \) and \( E[W_1(Z_{a+1}, \varepsilon_{1,a+1})] \). 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, \( \tau \), 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 \( \tau \). The patient is observed starting in period \( t_{i,0} \) and ending in \( T_i \). Conditional on \( \zeta \) and \( \varphi \), the type-specific joint probability of observing patient \( i \)'s sequence of choices is:

\[
L_i(d_i; \tau, \zeta, \varphi) = \prod_{a=a_0}^{T_i} \prod_{j=1}^{Y_j} \prod_{y=1}^{Y_y} \Pr(d_{y,j,a,i} = 1; Z_a, \tau, \zeta, \varphi)^{d_{y,j,a,i}}.
\]

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

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

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 \( \theta \) by maximizing the value of \( \mathcal{L}(\theta) \). We compute standard errors following the “outer
product of the score" method for \( \theta \) 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 \( \hat{\theta} = (\hat{\zeta}, \hat{\rho}, \hat{\varphi}) \) as given. Under our assumptions about initiation, we may write

\[
\Pr (I = 1 | Z_{a_0}^D, \tau, \varphi, \zeta, g_B, \mu) = \Lambda \left( W \left( Z_{a_0}^D, \tau, g_B, \varphi, \zeta \right) - \mu \right), \tag{12}
\]

where the variable \( I \) indicates whether a patient with characteristics \( Z_{a_0}^D \) 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, \( \hat{s}^{\text{init}} (\mu; g_B, \hat{\theta}) \). These distributions, \( f_{Z^D}(Z_{a_0}^D) \) for observed characteristics and \( f_{\tau}(\tau|Z_{a_0}^D) \) 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 \( \mu \) by solving \( \hat{s}^{\text{init}} (\mu; g_B, \hat{\theta}) = s^{\text{init}} \) for the empirical policy setting, \( g_B \). Under alternative policies, \( W (Z_{a_0}^D, \tau, g, \varphi, \zeta) \) will change but \( \mu \) remains fixed.

We approach the distributions \( f_{\tau}(\tau|Z_{a_0}^D) \) and \( f_{Z^D}(Z_{a_0}^D) \) using different strategies. In Appendix A we show that our assumptions on: a) the distribution of unobserved types conditional on treatment, \( f_{\tau}(\tau|Z_{a_0}^D, I_i = 1, \rho) \); and b) the initiation decision, are sufficient to back out the unconditional distribution \( f_{\tau}(\tau|Z_{a_0}^D) \). We write the frequency of \( \tau = 2 \) within-clinic as \( \Lambda([1, Z_{a_0}^D]|\rho) \), 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 \( f_{\tau}(\tau|Z_{a_0}^D) \) and \( f_{Z^D}(Z_{a_0}^D) \) allow us to to derive \( s^{\text{init}} (\theta, \mu, g_B) \), 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 \( s^{\text{init}} (\mu; g_B, \theta) \), we integrate \( \Pr (I = 1 | Z_{a_0}^D, \theta, \tau, \mu) \) over the distribution of \( Z_{a_0}^D \) and \( \tau \) among potential patients

\[
\hat{s}^{\text{init}} (\mu; g_B, \hat{\theta}) = \Pr \left( I = 1 | \hat{\theta}, \mu, g_B \right) = \sum_{Z_{a_0}^D} \left[ \sum_{\tau} \Lambda \left( W \left( Z_{a_0}^D, \tau, g_B, \varphi, \hat{\zeta} \right) - \mu \right) f_{\tau}(\tau|Z_{a_0}^D, \mu, \hat{\theta}) \right] f_{Z^D}(Z_{a_0}^D) \tag{13}
\]

We then estimate \( \mu \) as the value that solves \( s^{\text{init}} = \hat{s}^{\text{init}} (\mu; g_B, \hat{\theta}) \).
We use the sampling distribution of $\hat{\theta}$ to construct a confidence interval on $\hat{\mu}$. We draw 400 times from $\hat{\theta}$’s distribution, and for each draw we calculate the value of $\mu$ that equates $s^{\text{init}}$ and $\hat{s}^{\text{init}}$. We then sort the individual estimates of $\mu$, and then use the 2.5th and 97.5th percentile values as the 95% confidence interval. The confidence interval on $\mu$ 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, $z_{aft}$. 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 4 we display our estimates of $U(k|\tilde{k})$, $\alpha$, $\delta_0$, $\gamma$, and $\eta$. 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 $\kappa$ indicates that patients with 1 or 2 prior children have their utility from births shifted downward substantially. For example, for a patient with $\tilde{k} > 0$ and $\tau = 1$, the estimated $\kappa$ implies that the patient would prefer no additional children. The taste shifter $\zeta$ associated with type 2, however, is sufficient to increase the utility from additional births to be positive for patients with $\tilde{k} > 0$.

Table 4’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. We recover a significantly negative estimate for the start/delay parameter $\delta_0$, which plays a large role in determining whether a patient returns for additional treatment cycles after her first. Our estimate of the parameter $\gamma$ 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 4 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 5 reports results on the distribution of $\tau$ within the treated population. We estimate that about half of the patient population has type $\tau = 2$ given their $Z$ values. To interpret the individual $\rho$ parameters, consider the case of patient wealth. The negative coefficient ($\rho_2$) 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 treated 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 $\hat{\mu}_0 = -0.76$. This value of $\mu$ 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 compare 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 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 $\delta_0$) 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

---

21The 95% confidence interval for $\mu$ is $[-1.54, 0.49]$
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 2. 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. These potential patients represent the portion of the St. 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 \(W \left( Z_{u0}^D, \tau, g, \tilde{\varphi}, \tilde{\zeta} \right)\) for each simulated woman. The values of \(W \left( Z_{u0}^D, \tau, g, \tilde{\varphi}, \tilde{\zeta} \right)\) differ across policy experiments. We then allow potential patients to elect whether to begin treatment by comparing \(W \left( Z_{u0}^D, \tau, g, \tilde{\varphi}, \tilde{\zeta} \right)\) to the population-wide utility parameter \(\tilde{\nu}\) and a simulated value for the potential patient’s taste shock \(\nu\). 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

\(22\) 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.
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 6-7. The Tables contain both point estimates of counterfactual outcomes and 95% confidence intervals.23 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 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 last 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 $g_{PI}$.

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

---

23 As described above, we draw 400 times from the sampling distribution of $\theta$, and then estimate a new value of $\mu$ for each draw. We use each pair $(\theta, \mu)$ to compute the full set of patient histories under each counterfactual policy described below. We construct confidence intervals using the 2.5th and 97.5th percentile of each outcome (across $(\theta, \mu)$ pairs) within a policy setting.
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 $\Delta_i = W\left(Z^D_{i_0}, \tau_i, g, \bar{p}, \bar{\bar{c}}\right) - (\bar{p} + \nu_i)$, 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 $\alpha_i$ 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.\footnote{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 estimates 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 (e.g., 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.\footnote{By contrast, Hamilton and McManus (2011) find that insurance mandates reduce embryo transfers. That result, however, is based on an earlier sample period (1995-2003) when IVF technology was more likely to fail and patients transferred an average of one additional embryo per treatment.}

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 7, 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.

\subsection*{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 $\varepsilon$ 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 6, around 3% of $N^{inf}$ experience a birth within the sample period. This is due to a combination of frequently-unsuccesful 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 7. 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 $\bar{Z}$ 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

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

This expression acknowledges that there is some multiple birth risk for patients transferring a single embryo, $f_k(k|1, \bar{Z})$, 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 ($i$). In our simulations, when a 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 ($\eta$) 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 6 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 7, 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

26 For example, a patient with $\bar{Z}$ 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.
other hand, falls slightly relative to the embryo cap due to small differences in the top-up prices patients face (calculated using $Z$) 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 $8.4M, 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. While consumer surplus falls under $g_{PI}$ relative to baseline ($11.4M vs. $7M), this decline could in principle be offset by the substantial savings in treatment and delivery costs ($20.7M vs. $42.2M). Note that alternative top-up prices could be designed so that consumer surplus remains constant and savings from treatment and delivery costs are still realized. For example, reducing the top-up price for the second embryo from $12,000 to $5,000 would leave consumer surplus virtually identical to the baseline level but still reduce total costs by approximately $5M relative to $g_B$. 

38
Healthcare policymakers face the challenge of designing policies that simultaneously improve access to care and contain costs. These challenges are particularly acute in the market for IVF, where treatment is very costly, outcomes are uncertain, and there is substantial variation in insurance coverage for the procedure. A variety of policies have been proposed to address these issues, ranging from universal insurance mandates to cover IVF to restrictions on treatment choice in the form of embryo caps. To investigate the impact of these and other policies on patient welfare, outcomes, and healthcare costs, 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, patients being treated at the clinic face very different financial incentives.

Counterfactual simulations from our model show that a universal mandate to cover IVF substantially increases patient welfare by increasing usage of IVF. However, the mandate also increases both treatment and birth costs. Embryo caps have been proposed as a way to reduce the relatively high rates of expensive multiple births associated with IVF. We find that a policy of single embryo transfer does indeed substantially lower costs by reducing multiple births, but with a sharp reduction in consumer surplus. This is because patients need more cycles of IVF to achieve a birth, on average. In addition, our estimates imply that patients receive positive utility from twin births, which are limited under the embryo cap.

Given that neither unrestricted universal insurance nor embryo caps achieves the dual goals of increased access and lower per-birth costs, we propose a new policy in the spirit of Einav et al. (2015) in which all patients receive insurance coverage for the transfer of a single embryo, but then must pay a “top-up” price of $12,000 - $19,000 if they wish to transfer additional embryos. These
prices are chosen to fully internalize the expected medical costs of more aggressive treatments. Incorporating top-up prices substantially reduces expected birth and treatment costs relative to an unrestricted universal insurance mandate, while generating higher patient surplus than a single embryo cap for individuals in our sample. The magnitude of the cost savings suggest that insurers could consider voluntarily extending universal coverage for IVF in conjunction with top-up prices for more aggressive treatments. In addition, there is room to reduce the top-up prices below their fully internalizing level in an effort to increase patient surplus while still achieving significant cost savings.

Appendix A: Distribution of Types Among Potential Patients

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

\[
\text{Pr} \left( \tau = 2 | Z^{D}_{a_0}, I = 1, \tilde{\rho} \right) = \Lambda \left( [1, Z^{D}_{a_0}] \tilde{\rho} \right). \tag{14}
\]

In addition, from the treatment initiation model we know, for each possible \( \mu \) and \( Z^{D}_{a_0} \), the initiation rate among potential patients of each type. That is, we know

\[
\Lambda \left( W \left( Z^{D}_{a_0}, \tau, \tilde{\varphi}, \tilde{\zeta} \right) - \mu \right) \text{ for } \tau = 1, 2. \tag{15}
\]

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

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

Similarly for type 2, we know:

\[
N_{Z^{D}_{a_0}, 2}^{clin} (\tilde{\rho}) = N_{Z^{D}_{a_0}}^{clin} \times f_{\tau} \left( \tau = 2 | Z^{D}_{a_0}, I = 1, \tilde{\rho} \right). \tag{17}
\]
Note that while \( N_{\text{clin}}^{\text{inf} Z_{a_0}} \) is just data, \([N_{\text{clin}}^{\text{inf} Z_{a_0,1}}(\hat{\rho}), N_{\text{clin}}^{\text{inf} Z_{a_0,2}}(\hat{\rho})]\) depends on \( \rho \), which is identified by the differential behavior of the two types in the (within-clinic) patient histories. Recall that \( \rho \) parameterizes the within-clinic distribution of types and is estimated in our second step along with \((\varphi, \zeta)\).

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

\[
N_{Z_{a_0,1}}^{\text{inf} D} = \frac{N_{\text{clin}}^{\text{inf} Z_{a_0,1}}(\hat{\rho})}{\Lambda \left(W \left(Z_{a_0, \tau=1}^D, \tau=1, \varphi, \zeta\right) - \mu\right)} = \frac{N_{\text{clin}}^{\text{inf} Z_{a_0}} [1 - \Lambda([1, Z_{a_0}^D | \hat{\rho}])]}{\Lambda \left(W \left(Z_{a_0, \tau=1}^D, \tau=1, \varphi, \zeta\right) - \mu\right)}, \quad (18)
\]

\[
N_{Z_{a_0,2}}^{\text{inf} D} = \frac{N_{\text{clin}}^{\text{inf} Z_{a_0,2}}(\hat{\rho})}{\Lambda \left(W \left(Z_{a_0, \tau=2}^D, \tau=2, \varphi, \zeta\right) - \mu\right)} = \frac{N_{\text{clin}}^{\text{inf} Z_{a_0}} \Lambda([1, Z_{a_0}^D | \hat{\rho}])}{\Lambda \left(W \left(Z_{a_0, \tau=2}^D, \tau=2, \varphi, \zeta\right) - \mu\right)} \quad (19)
\]

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

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

Note that for given \( \mu \), everything in the RHS is known, so \( f_{\tau}(\tau = 2|Z_{a_0}^D) \) is known and \( f_{\tau}(\tau = 1|Z_{a_0}^D) = 1 - f_{\tau}(\tau = 2|Z_{a_0}^D) \). If both types were to select into the clinic at the same rate (i.e. they did not really have different preferences for children), we would have \( W \left(Z_{a_0, \tau=1}^D, \tau=1, g_B, \theta\right) = W \left(Z_{a_0, \tau=2}^D, \tau=2, g_B, \theta\right) \) so \( \frac{\Lambda(W(Z_{a_0, \tau=1}^D, \tau=1, \varphi, \zeta) - \mu)}{\Lambda(W(Z_{a_0, \tau=2}^D, \tau=2, \varphi, \zeta) - \mu)} = 1 \) and the distribution of types within the clinic and among potential patients would be the same, \( f_{\tau}(\tau = 2|Z_{a_0}, I = 1) = f_{\tau}(\tau = 2|Z_{a_0}^D) \), which is not consistent with our estimates.

**Appendix B: Approximating \( N^{\text{inf}} \) 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 $f_{Z^D}(Z^D_{a_0})$ 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:

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

- **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:

$$Z^B \perp (\tau_{0}, z_w) \mid a_0. \tag{22}$$

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:

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

- **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, $\Pr(\tau_{a_0}, z_w|a_0)$ is the same regardless of whether a woman is a potential patient. We further assume that $\Pr(\tau_{a_0}, z_w|a_0) = \Pr(\tau_{a_0}, z_w)$ for all $a_0$.

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

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

$$f_{Z^D}(Z^D_{a_0}) = f_{a_0, \tau_{a_0}, z_w}(a_0, \tau_{a_0}, z_w) f_{asrm}(z_{asrm_{a_0}}). \tag{24}$$
To estimate \( f_{a_0, t_0, z_w} (a_0, t_0, z_w) = f_{t_0, z_w} (t_0, z_w) f_{a_0} (a_0) \) 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, \((t, z_w)\). To estimate \( f_{t, z_w} (t, z_w) \) we consider \( f_{t, z_w} (t, z_w) = f_t (t|z_w) f_{z_w} (z_w) \) and develop a strategy at the zip code level for estimating \( f_t (t|z_w) \) 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 \((t, z_w)\). 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

\[
f_{z_w} (z_w = 1) = \sum_{l \in STL} I \{ z_{w,l} = 1 \} \pi_l = \sum_{l: z_{w,l} = 1} \pi_l,
\]

where \( \pi_l \) 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 \( \pi_l \) easily. We estimate that 39% of the potential patients in the St. Louis region are wealthy.

To estimate \( \Pr (t|w) \) 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+) 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%.\(^{27}\) Therefore we use the following adjustment

\(^{27}\)See Table 1 in Moscarini and Postel-Vinay (2012)
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 $f_{IVF}(\iota_{l}^{IVF}) = f_{priv}(\text{priv}_l) \times \psi^{MO}$.

Regarding Illinois counties, we know that there is a mandate. But small employers ($< 25$ employees) and self-insured employers (regardless of size) are excluded.$^{28}$ According to a Kaiser Family Foundation (2007) report, 55% of workers nationally are covered by plans that are partially or fully self-insured.$^{29}$ 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: $f_{IVF}(\iota_{l}^{IVF}) = f_{priv}(\text{priv}_l) \times \psi^{IL}$.

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

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

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 $\psi^{MO}$ is very low relative to $\psi^{IL}$, if we pool zip codes together in the aggregation we would end up with a spurious negative correlation. Therefore we compute $\Pr (\iota = 4|z_w = 1)$ in the following way:

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

where

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

$^{28}$Under this alternative definition of small employer, we interpolate the numbers in Moscarini and Postel-Vinay and find that 21.5% of employment is accounted for by firms with less than 25 employees.

$^{29}$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.
\( \hat{\Delta}_s \) 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.

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

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

**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 \( \tilde{N}_{inf} \) 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 \( \bar{N}_a \). 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, \( \bar{N}_a \) succeed and have births recorded in Vital Statistics; the group that fails, \( N_{inf} \), becomes our set of potential patients. Therefore \( N_{a}^{stl} = \bar{N}_a + \tilde{N}_{a}^{inf} \). Then using infertility rates by age among women who are attempting to get pregnant, \( f_{inf}(inf|a) \), we can back out \( \tilde{N}_{a}^{inf} = \frac{f_{inf}(inf|a)}{1-f_{inf}(inf|a)} \bar{N}_a \).

According to Vital Statistics the larger counties in and around the St. Louis region have an average of \( \bar{N}_{28-44} = 1172 \) 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.\(^{30}\) So \( \bar{N}_{28-44}^{75m} = 1172 \times 1.104 = 1294 \). Using infertility rates by age and summing across ages, we can then determine that there are \( \tilde{N}_t^{inf} = 198 \times 1.104 = 219 \) new potential patients, on average, each quarter.\(^{31}\) Since there are 28 quarters between 2001 and 2007, the size of the potential patient pool for our sample period is then \( \tilde{N}_{2001-07}^{inf} = 28 \times \tilde{N}_{inf}^{2001} = 28 \times 219 = 6132 \).

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 \( \tilde{N}_{2001-07}^{inf} \) in a proportional way. Ultimately, the potential-patient population for our clinic

\(^{30}\)Births occurring 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 occurred 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.

\(^{31}\)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).
is $N^{inf} = 2146$.

References


Figure 1: IVF treatment stages

Stage 1 / Period start

If already initiated

Delay

Start $\varepsilon_{s1}$

$\varepsilon_{d1} + \beta EW_1(Z_{a+1})$

Learn Peak E2

If at initiation decision

Start

No initiation

$W_{OUT} = \mu + \nu$

Stage 2

Stage 3

Stage 4

Learn embryos $(X)$

Choose $x \leq X$

$\sum_{k=0}^{\infty} f(k|x)[U(k) + \beta^k EW_1(Z_{a+k})] + f_0(x)\beta EW_1(Z_{a+1}) + \varepsilon_{c4}$

Go to next period

No ICSI

ICSI

$\varepsilon_{m23} - \alpha p_{m23}$

$\varepsilon_{m13}$

$\varepsilon_{nc2} + \beta EW_1(Z_{a+1})$

Go to next period

$\varepsilon_{c2} - \alpha p_{c2}$

Cancel

Continue

$u_s - \alpha p_{s1}$

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

- Baseline vs. Univ. insurance
- Baseline vs. Embryo Cap
- Baseline vs. Top-up prices
- Baseline vs. Top-up + insur
Figure 12: Births Outcomes in Counterfactual Experiments

- Baseline vs. Univ. insurance
- Baseline vs. Embryo Cap
- Baseline vs. Top-up prices
- Baseline vs. Top-up + insur
Table 1: This Paper’s Clinic in Context

<table>
<thead>
<tr>
<th></th>
<th>This Paper’s Clinic</th>
<th>All U.S. Clinics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of cycles per year</td>
<td>370.2</td>
<td>217.1</td>
</tr>
<tr>
<td>Patient median age</td>
<td>34</td>
<td>35-37 (see note)</td>
</tr>
<tr>
<td>Patients with 2+ infertility diagnoses</td>
<td>11%</td>
<td>12%</td>
</tr>
<tr>
<td>Male partners with infertility diagnosis</td>
<td>18%</td>
<td>17%</td>
</tr>
<tr>
<td>ICSI use</td>
<td>50%</td>
<td>59%</td>
</tr>
<tr>
<td>Embryos per cycle</td>
<td>2.41</td>
<td>2.68</td>
</tr>
<tr>
<td>Birth rate per cycle</td>
<td>33.5%</td>
<td>29.9%</td>
</tr>
</tbody>
</table>

**Note:** We use our main data sample to calculate patient median age for the clinic we study. For all other statistics we use the annual summary data published by the CDC for the clinic we study and the U.S. as a whole. In the CDC’s data patient ages are reported in intervals.
Table 2: Patient-level Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Main sample N = 587</th>
<th>First-stage sample N = 1106</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. dev.</td>
</tr>
<tr>
<td><strong>Demographic state variables (Z^D)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient age at initiation</td>
<td>34.30</td>
<td>4.02</td>
</tr>
<tr>
<td>Insured at initiation? (Y = 1)</td>
<td>0.54</td>
<td>0.50</td>
</tr>
<tr>
<td>Wealthy zip code? (Y = 1)</td>
<td>0.82</td>
<td>0.39</td>
</tr>
<tr>
<td>Prior children at initiation</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Biological state variables (Z^B)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFC score</td>
<td>14.37</td>
<td>7.96</td>
</tr>
<tr>
<td>Female fertility problem? (Y = 1)</td>
<td>0.80</td>
<td>0.40</td>
</tr>
<tr>
<td>Male fertility problem? (Y = 1)</td>
<td>0.34</td>
<td>0.48</td>
</tr>
<tr>
<td><strong>Aggregate actions and outcomes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total cycles</td>
<td>1.75</td>
<td>1.02</td>
</tr>
<tr>
<td>Birth during sample period? (Y = 1)</td>
<td>0.53</td>
<td>0.50</td>
</tr>
</tbody>
</table>

The "Main sample" is used in second-stage estimation of patients’ choices. The "First-stage sample" is used to estimate treatment technologies.
Table 3: Actions and Outcomes within Treatment

<table>
<thead>
<tr>
<th>Stage 1-4 actions</th>
<th>Main sample</th>
<th>First-stage sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>Cancel treatment? (Y = 1)</td>
<td>1027</td>
<td>0.14</td>
</tr>
<tr>
<td>Fertilization method? (ICSI = 1)</td>
<td>879</td>
<td>0.60</td>
</tr>
<tr>
<td>Number of embryos transferred</td>
<td>875</td>
<td>2.29</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stage 1-3 outcomes</th>
<th>Main sample</th>
<th>First-stage sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>Peak E2 score</td>
<td>1027</td>
<td>16.82</td>
</tr>
<tr>
<td>Eggs retrieved</td>
<td>879</td>
<td>10.60</td>
</tr>
<tr>
<td>Embryos generated</td>
<td>881</td>
<td>6.11</td>
</tr>
<tr>
<td>4+ embryos? (Y = 1)</td>
<td>881</td>
<td>0.74</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stage 4 outcomes</th>
<th>Main sample</th>
<th>First-stage sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>Children born</td>
<td>848</td>
<td>0.51</td>
</tr>
<tr>
<td>Singleton birth? (Y = 1)</td>
<td>848</td>
<td>0.27</td>
</tr>
<tr>
<td>Twin birth? (Y = 1)</td>
<td>848</td>
<td>0.12</td>
</tr>
<tr>
<td>Triplet birth? (Y = 1)</td>
<td>848</td>
<td>0.00</td>
</tr>
</tbody>
</table>

The “Main sample” is used in second-stage estimation of patients’ choices. The “First-stage sample” is used to estimate treatment technologies.
### Table 4: Utility Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter Description</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utility of 1 birth ($u_1$)</td>
<td>5.147</td>
<td>0.932</td>
</tr>
<tr>
<td>Price sensitivity constant ($\alpha_0$)</td>
<td>0.311</td>
<td>0.071</td>
</tr>
<tr>
<td>Utility of 2 births ($u_2$)</td>
<td>5.967</td>
<td>1.690</td>
</tr>
<tr>
<td>Price sensitivity X wealth ($\alpha_w$)</td>
<td>-0.125</td>
<td>0.067</td>
</tr>
<tr>
<td>Utility of 3 births ($u_3$)</td>
<td>-14.063</td>
<td>4.372</td>
</tr>
<tr>
<td>Terminal payoff X Prev. payment ($\gamma_p$)</td>
<td>0.221</td>
<td>0.633</td>
</tr>
<tr>
<td>Utility shift when $\tilde{k} &gt; 0$ ($\kappa$)</td>
<td>-11.698</td>
<td>0.955</td>
</tr>
<tr>
<td>Utility shift from beginning treatment ($\delta_0$)</td>
<td>-4.917</td>
<td>0.098</td>
</tr>
<tr>
<td>Preference shifter ($\zeta$)</td>
<td>9.631</td>
<td>0.855</td>
</tr>
<tr>
<td>Penalty for violating ASRM embryo guidelines</td>
<td>-3.049</td>
<td>0.194</td>
</tr>
</tbody>
</table>

Standard errors are in parentheses.

### Table 5: Utility-Type Distribution Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter Description</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type distr. constant ($\rho_0$)</td>
<td>-1.318</td>
<td>0.556</td>
</tr>
<tr>
<td>Type distr. insurance ($\rho_3$)</td>
<td>-0.073</td>
<td>0.370</td>
</tr>
<tr>
<td>Type distr. age ($\rho_1$)</td>
<td>0.044</td>
<td>0.013</td>
</tr>
<tr>
<td>Type distr. ASRM regime ($\rho_4$)</td>
<td>0.442</td>
<td>0.339</td>
</tr>
<tr>
<td>Type distr. wealth ($\rho_2$)</td>
<td>-0.861</td>
<td>0.458</td>
</tr>
</tbody>
</table>

Standard errors are in parentheses.
### Table 6: Initiations, Cycles, and Outcomes across Policy Settings

<table>
<thead>
<tr>
<th>Policy setting (g)</th>
<th>Share initiating</th>
<th>N cycles if initiate</th>
<th>Share with birth</th>
<th>N births</th>
<th>N infants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0.384 (0.288, 0.449)</td>
<td>1.628 (1.524, 1.770)</td>
<td>0.263 (0.192, 0.314)</td>
<td>581.9 (431.2, 702.6)</td>
<td>805.9 (600.6, 972.4)</td>
</tr>
<tr>
<td>Universal insurance</td>
<td>0.561 (0.500, 0.682)</td>
<td>1.726 (1.566, 1.909)</td>
<td>0.380 (0.345, 0.452)</td>
<td>864.6 (785.5,1025.3)</td>
<td>1195.2 (1087.2,1421.2)</td>
</tr>
<tr>
<td>Embryo cap</td>
<td>0.104 (0.050, 0.180)</td>
<td>1.361 (1.217, 1.547)</td>
<td>0.033 (0.016, 0.056)</td>
<td>71.1 (34.7, 120.4)</td>
<td>80.7 (39.6, 137.3)</td>
</tr>
<tr>
<td>Top-up prices</td>
<td>0.154 (0.084, 0.288)</td>
<td>1.474 (1.322, 1.637)</td>
<td>0.073 (0.040, 0.155)</td>
<td>157.8 (86.8,338.4)</td>
<td>202.8 (111.8,451.5)</td>
</tr>
<tr>
<td>Top-up prices + insurance</td>
<td>0.336 (0.217, 0.416)</td>
<td>1.625 (1.476, 1.798)</td>
<td>0.158 (0.104, 0.228)</td>
<td>345.9 (230.6,503.5)</td>
<td>436.8 (292.1,653.9)</td>
</tr>
</tbody>
</table>

95% confidence intervals are in parentheses
Table 7: Surplus and Costs across Policy Settings

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

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