American Economic Review 2018, 108(12): 3725–3777
https://doi.org/10.1257/aer.20161014

Health Care Access, Costs, and Treatment Dynamics: Evidence from *In Vitro* Fertilization[†]

By Barton H. Hamilton, Emily Jungheim, Brian McManus, and Juan Pantano*

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. (*JEL* G22, I11, I13, I18, J13, J16)

Health care policymakers struggle with the conflicting goals of providing ready access to medical care while simultaneously containing costs. For example, in the United Kingdom a medical treatment may not be covered by national insurance if it fails to meet a cost-effectiveness criterion. US 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 (e.g., Chernew, Encinosa, and Hirth 2000). In this spirit, recently introduced "value-based" insurance plans tie patient co-pays to the clinical value of treatment (Chernew, Rosen, and Fendrick 2007). Spurred by the Affordable Care Act, value-based principles have been experimented with in primary care and pharmaceutical plans. However, such insurance policies are still uncommon and few

*Hamilton: Olin Business School, Washington University in St. Louis, MO (email: hamiltonb@wustl.edu); Jungheim: School of Medicine, Washington University in St. Louis, MO (email: jungheime@wudosis.wustl.edu); McManus: University of North Carolina, Chapel Hill, NC (email: mcmanusb@unc.edu); Pantano: Department of Economics, Stony Brook University, Stony Brook, NY (email: juan.pantano@stonybrook.edu). This paper was accepted to the *AER* under the guidance of Pinelopi Goldberg, Coeditor. 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. Jungheim practices medicine and performs IVF treatments at the clinic studied in this paper. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

[†]Go to https://doi.org/10.1257/aer.20161014 to visit the article page for additional materials and author disclosure statement(s).

empirical studies (e.g., Einav, Finkelstein, and Williams 2016) evaluate the ability of treatment-dependent pricing to influence patient behavior and control costs.[1]

In this paper, we study the impact of alternative policies on patients' access to care, patients' surplus, and health care costs in a context that embodies many of the features of more complicated medical treatment: the market for *in vitro* fertilization (IVF) in the United States. Over the last 20 years, IVF use has increased fourfold to 160,000 treatment cycles annually, driven in part by delayed fertility decisions by women.[2] Because individual cycles of IVF treatment often fail, potential IVF patients must solve a complex sequential decision problem under uncertainty about whether to initiate and/or continue therapy.[3] Patients (in consultation with their physicians) must also decide on the type of treatment, based on their preferences and the cost of therapy. More aggressive treatment increases the probability of pregnancy, thus reducing the need for additional IVF cycles, but it also increases the risk of costly higher-order births. IVF patients now account for approximately 50 percent of all higher-order births, which are generally 4 (twins) to 16 (triplets) times more expensive than a singleton birth. Consequently, 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, and patients weigh the direct effects of alternative treatment options against potential side effects (Chan and Hamilton 2006).[4]

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. To increase access, nine US states have mandated insurance coverage for IVF; these policies endow each potential patient with several covered treatment attempts for an out-of-pocket price of $2,000–$3,000. While mandates should lead to higher utilization rates, the 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.

[1] Baicker, Shepard, and Skinner (2013) simulate the impact of an alternative to Medicare in which all individuals are guaranteed basic benefits, but are provided the option to "top-up" by purchasing additional coverage for more expensive, less cost-effective treatments. Such a scheme may lead to more unequal health care spending but lower cost growth.

[2] IVF currently accounts for 1 percent of US births. Treatment rates are even higher in some other countries; IVF accounted for over 3 percent of 2011 births in Israel, Belgium, Sweden, and Denmark. The aggregate treatment statistics reported in this paragraph for the United States are for fresh-embryo treatments during 2012 (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).

[3] Relatively young US patients (under 35 years old) in 2012 achieved a birth after 40 percent of cycles, while patients just under 40 years old had success rates approximately one-half as large.

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

As the medical costs associated with birth are generally paid by insurers, treatment aggressiveness may be described as moral hazard.

Concern about moral hazard and the high rate of costly multiple births has led some European countries to limit the treatment options available to IVF patients in the form of an explicit cap on the number of embryos transferred during treatment. US medical experts have recently advocated for embryo caps due to their perspective that a singleton birth is the best possible outcome of treatment.[5] While single embryo transfer nearly eliminates costly high-order births, it also substantially reduces the odds of conception for each IVF attempt. Consequently, such restrictions will generally affect the optimal treatment paths selected by forward-looking patients. An embryo restriction effectively increases patients' expected discounted cost of conception, which can lead some patients to forgo IVF altogether or abandon treatment after initial failure, thus implicitly reducing access to IVF.

To evaluate the impact of these and other alternative policies, we specify a structural model of forward-looking, infertile patients' decisions to initiate and/or continue IVF treatment using fresh non-donor eggs and embryos.[6] 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.[7] 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. Second, patients may expect their medical condition to change over time. Female fertility declines with age, particularly after age 35, when many IVF patients receive treatment. Anticipation of future health declines affects current decisions regarding both the initiation choice and aggressiveness of treatment. 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.[8]

We estimate the model using a novel dataset of the treatment histories of women undergoing IVF at an infertility clinic ("the clinic") between 2001 and 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

[5] The Practice Committee of 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.

[6] A large majority of IVF cycles use the patient's own eggs (i.e., non-donor) and transfer embryos that were never frozen.

[7] 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. In a study of insurance's role in risk protection and moral hazard, Kowalski (2015) accounts for the impact of deductibles, co-insurance, and stop-losses in consumers' nonlinear budget sets.

[8] Fang and Gavazza (2011) provide another example of the dynamic interaction between government policy and health choices and outcomes. They provide evidence that the US employer-provided health insurance system provides disincentives to invest in health over the life cycle.

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 from 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: (i) whether to initiate an IVF treatment, (ii) the treatment choices within an IVF cycle, and (iii) whether (and when) the patient reattempts 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.

Due to the dynamic incentives and trade-offs highlighted by our model estimates, we conduct counterfactual experiments investigating the impact of alternative policy proposals on individual patient actions, outcomes, and surplus. We first consider the widely-discussed policies of mandated insurance coverage and treatment restrictions in the form of embryo caps. Relative to a benchmark case in which no women have insurance, introducing universal coverage for IVF results in more than a doubling of the share of infertile women who initiate treatment (24 percent versus 58 percent). This increase in access is accompanied by an increase in consumer surplus from \$2,700 to \$6,600 per potential patient. 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 little help in terms of cost control. In an effort to contain costs, we next evaluate the policy of a "European-style" treatment restriction combined with universal insurance. The strictest policy is a hard limit of one embryo per patient per cycle, which pushes treatment initiation down to its no-insurance level, but with greatly reduced total costs and aggregate births. In terms of costs per birth, however, the embryo cap is actually more expensive than the no-insurance benchmark due to low birth rates.

Given the cost growth induced by universal insurance, and the consumer surplus losses associated with treatment restrictions, we explore pricing schemes closer to those suggested by theoretical models of optimal insurance design. 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 higher-order pregnancies. Consequently, we evaluate the introduction of a value-based plan in which insured patients face "top-up prices" when transferring more than one embryo.[9] We select top-up prices that cover the (expected) increase

[9] Einav, Finkelstein, and Williams (2016) consider a closely related policy of top-up prices for medical care beyond a basic level covered by insurers.

in medical expense relative to single-embryo transfers to minimize this potential source of moral hazard. Under the top-up policy, patients reduce but do not eliminate multiple-embryo transfers, reflecting both a preference for twin births and the incentive to reduce lifetime treatment costs. The presence of insurance allows utilization (32 percent) to be greater than the no-insurance benchmark, but with significantly lower costs per birth ($63,100 per delivery versus $69,400 under the benchmark), due to patient payments against birth expenses and a reduction in multiple birth rates.[10] In fact, total insurance payments are lower with actuarially fair top-up prices and therefore insurance companies might be content with insurance mandates that allow top-up price designs similar to those we analyze.

While insurance coverage for IVF in the US is generally proposed as protection against the high costs of infertility treatment, Israel takes a different approach by insuring birth outcomes directly to encourage family formation. In particular, Israel allows unlimited IVF attempts until the patient has two or more children. Our final set of counterfactuals compares such outcome-based insurance designs with more traditional policies that provide a fixed number of treatments. We find that an outcome-focused policy has nearly identical patient benefits and costs as mandated universal insurance coverage found in US states such as Illinois. The Israeli design also highlights the dynamic impacts of IVF policies. We show that age-related limits on the Israeli policy leads to welfare and choice differences related to patients anticipating future reductions in policy generosity.

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

I. IVF Overview and Our Setting

A. *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.

[10] Greater IVF access can be achieved with smaller top-up prices, if desired.

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 canceled, the patient may start IVF again in the future. If the cycle is not canceled, 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 (cleavage-stage 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 trade-off: 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.[11] 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.[12]

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. Anticipation of future opportunities to receive IVF is an important part of a patient's decisions during the current IVF cycle. 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 (Højgaard 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.

B. *Public Policy and Analysis of Its Impact*

A key feature of the US 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, and Rhode Island had the strongest mandates

[11] Lemos et al. (2013) obtained a sample from MarketScan of 437,924 deliveries covered by commercial insurance over the period 2005–2010 to derive their cost figures. Approximately 6,500 of these deliveries were IVF-related pregnancies. Reported costs were the sum of mother's medical expenses from 27 weeks prior to delivery to 30 days after, plus infant medical expenses up to the first birthday. Average health care costs were $5,510 (singleton) to $27,469 (triplet) higher for IVF pregnancies compared to other births.

[12] 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.

for IVF, requiring insurers to cover a certain number of IVF treatment cycles.[13] 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, Henne, and Becker (2008); and Buckles (2013) examine the impact of insurance mandates on population-level fertility rates, multiple birth rates, and child health. Hamilton and McManus (2012), Jain, Harlow, and Hornstein (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) 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.[14]

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 percent of IVF cycles in 2009 involved a transfer of one embryo), it is widely practiced in Europe. For example, 69 percent 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 Vélez 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.

C. *The Clinic We Study and Its Market*

In order to put our study of a single IVF clinic in context, we assess how the clinic and market we study compares to the full US 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 US clinic, with 370.2 cycles per year compared to the US 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 due to demographic patterns, or a response to insurance coverage, which may encourage potential patients to try

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

[14] Some studies exist in the medical literature that use patient-level IVF data (e.g., Malizia, Hacker, and Penzias 2009; Jungheim et al. 2017), but their focus and approach are naturally different from our own.

Table 1—This Paper's Clinic in Context

	This paper's clinic	All US clinics
Number of cycles per year	370.2	217.1
Patient median age	34	35–37 (see note)
Patients with 2+ infertility diagnoses (Y = 1)	0.11	0.12
Male partners with infertility diagnosis (Y = 1)	0.18	0.17
ICSI use (Y = 1)	0.50	0.59
Embryos per cycle	2.41	2.68
Birth rate per cycle (Y = 1)	0.335	0.299

Notes: 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 US as a whole. In the CDC's data, patient ages are reported in intervals.

IVF more quickly 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, is 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 US averages. These treatment patterns, along with a greater birthrate per cycle (33.5 percent at the clinic versus 29.9 percent 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 US 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 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 percent of St. Louis-area women lived in zip codes with median home values above $100,000; nationally the share was 56.7 percent. The market's share of African American women ages 25–44 was 19.9 percent in 2000, while the US share was 13.3 percent.

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, which has a mandate. 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 four IVF cycles if the woman has no children.[15] 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

[15] If the patient has already had a birth through IVF, the number of remaining covered cycles is set to two. 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.

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. Despite the absence of a mandate, some patients from Missouri have private insurance for IVF; in these cases the clinic has found that IVF coverage details are similar to those of Illinois patients. Some Missouri employers may choose to offer IVF coverage as a means to attract and retain employees. We calculate that potential patients from Illinois are three times more likely (30 percent versus 11 percent) to have IVF coverage than women from Missouri, but even in Illinois insurance penetration is fairly low.

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.[16]

II. Model Preliminaries

A. *Timing*

We consider two timing concepts in the model below. First, there are decision *periods* when active patients choose whether to start or delay an IVF cycle. Second, in an IVF treatment cycle there are four treatment *stages* during which patients make one choice per stage. The time unit of the model is a three-month period. We index the age of the patient (in three-month periods) with a and calendar time periods (also quarterly) with t, and we use j to index stages within a period during which an IVF cycle is started. At each age a, within stage j, the patient selects an action $y_{j,a}$ from the set Y_j. The set Y_j does not vary with a or t. Note that there will be two indexes, a and t, to keep track of time-varying objects in the model; the primary index is "a." Age is important because we specify a finite horizon model, and also because age itself enters various stage technologies.

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.[17] 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

[16] 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.

[17] Due to data limitations, we do not study the potential patient's decisions of how quickly to turn to IVF, and how this is affected by insurance coverage. We assume, instead, that all couples with fertility problems exhaust the same set of alternatives before arriving at the IVF decision.

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 assume that policy changes that affect the value of pursuing IVF do not change the value of this outside option.

Individual patients vary in the age at which they consider reproduction. We denote as a_0 the patient's age (in quarters) when she considers IVF. The patient's choice over IVF, if necessary, arrives four quarters after her exogenous decision to first attempt reproduction, i.e., at age a_{-4}. This timeline is consistent with a simplified setting in which all women believe themselves to be fertile (but are unaware of their true infertility status) until their first pregnancy attempt at age a_{-4}. A woman who is fertile conceives after three months of attempts and gives birth to her first baby nine months later. An infertile woman experiences a year of unsuccessful pregnancy attempts until, at age a_0, she considers whether to pursue IVF and therefore become a patient at the clinic. If at that point she opts to pursue IVF treatment, the patient may continue to make decisions until a terminal age, $a^{\max}$. At this age, the IVF clinic will no longer treat the patient and her birth probability (via IVF or naturally) is zero.[18] This allows a maximum of $4 \times \left(a^{\max} - a_0 + 1\right)$ periods for a patient whose reproductive decisions start at age a_0. In our data we observe patients with a_0 between their late twenties and early forties. We define $a^{\min}$ to be the youngest patient age considered in our model. We set $a^{\min}$ to be the first quarter of age 28, and $a^{\max}$ is the final quarter of age 44. Once a patient's total number of children reaches three (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. This index is also necessary to describe the prevailing embryo-transfer guidelines offered by the American Society for Reproductive Medicine (ASRM); these guidelines changed once during our sample period, in 2004. Let $t_{i,0}$ represent the period during which we first observe patient i. We see a patient for the last time in T_i, which might be equal to $a^{\max}$ or $\overline{T}$, the end of the sample period. We assume that all treatment stages that follow from a treatment starting in period t also occur in period t.

B. *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 $\mathbf{Z}$, 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 $\mathbf{Z}$ into two parts. We track a patient's age (a), a measure of her wealth (z_w), number of prior children $(\tilde{k})$, record of previous payments for

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

IVF (z_p), insurance status (ι), and race (z_r) in the state vector $\mathbf{Z}^D$.[19] Some of these state variables are time-invariant and others vary in how they evolve between periods. Age increases exogenously by three months every period. Race is a permanent patient characteristic, and we assume wealth is constant also. We construct z_w with zip code-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 as a controlled Markov process which depends on treatment outcomes.[20] Likewise, the patient's record of prior IVF payments (z_p) and remaining insured cycles evolve endogenously according to the patient's decisions within the model. We initialize the number of insured cycles (ι) to four (the Illinois mandate value) for all patients who ever use insurance, and this number falls by one whenever an insured patient advances to the second stage of treatment, when eggs are removed during surgery.[21] The forward-looking patient is aware that IVF's price increases substantially after her fourth insured cycle.

The second part of $\mathbf{Z}$ includes the patient's biological characteristics, $\mathbf{Z}^B$. We assume that the patient learns her own $\mathbf{Z}^B$ if she initiates treatment. The characteristics in $\mathbf{Z}^B$ include: the women's antral follicle count (AFC score, z_{afc}), an indicator of her egg-producing ability; whether she has one or more specific infertility diagnoses (z_{ff}), e.g., endometriosis; and whether her partner has male-factor infertility (z_{mf}). At the treatment initiation decision, the patient considers the possible values of $\mathbf{Z}^B$ she may have using the population frequency of these characteristics conditional on her initiating age, $f_{\mathbf{Z}^B}(\mathbf{Z}^B \mid a_0)$. We assume that the distribution of $\mathbf{Z}^B$ conditional on a_0 is independent of race.

Our assumptions on $\mathbf{Z}$ 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 $\mathbf{Z}^B$. This is a simplification that abstracts away from real-world opportunities for potential patients to learn about $\mathbf{Z}^B$ during less sophisticated infertility treatments or stand-alone screenings. We make this simplification in order to reduce the dimensions of potential patient heterogeneity prior to treatment, and because of data scarcity about this heterogeneity among the population at large. Second, we assume that patients (and their doctors) use no biological data other than $\mathbf{Z}^B$ in choosing a treatment path for patients. We assume that the patient's observed biological state variables $\mathbf{Z}^B$ and age fully capture her relevant fertility characteristics; there are no unobserved state variables that directly affect treatment outcomes, and previous cycles' choices and outcomes also have no direct impact on the current cycle's outcomes. 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 $\mathbf{Z}^B$ is realized, we consolidate notation and refer to the state vector $\mathbf{Z} = [\mathbf{Z}^D, \mathbf{Z}^B]$. In addition to acting as a state variable which influences

[19] Bitler and Schmidt (2006) document differences in infertility diagnoses and IVF usage across racial, ethnic, and socioeconomic groups. Due to data limitations we consider a simple binary specification for z_r, which records whether a woman is African American.

[20] All patients start with $\tilde{k} = 0$, and $\tilde{k}$ evolves stochastically in each following period. The evolution of $\tilde{k}$ depends on whether a cycle is started and what choices are made within that cycle. If no cycle is started or an ongoing cycle is canceled, then $\tilde{k}_{a+1} = \tilde{k}_a$. Note that $\tilde{k}$ remains constant as well if a cycle is unsuccessful. If a cycle at age a reaches the embryo transfer stage and at least one child is born as a result of that cycle ($k_a > 1$), then $\tilde{k}_{a+4} = \tilde{k}_a + k_a$.

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

treatment outcomes, patient age also functions as a time index, so we add an "a" subscript to $\mathbf{Z}$ where appropriate. During an arbitrary age, $\mathbf{Z}_a = [\mathbf{Z}_a^D, \mathbf{Z}^B]$, and at treatment initiation the state variables have the value $\mathbf{Z}_{a_0}$. We assume that the doctor knows how the variables in $\mathbf{Z}_a$ affect treatment outcome probabilities. Each patient receives this information from her doctor and also knows her own preferences over treatment outcomes.

C. *Patients' Preferences*

Patients have preferences over birth outcomes (k), and these preferences can depend on the patient's existing number of children ($\tilde{k}$) at the start of an IVF cycle and other personal characteristics. Possible values of k are in $\{0, 1, 2, 3\}$, and $\tilde{k}$ takes values in $\{0, 1, 2\}$. (These values for $\tilde{k}$ cover 98 percent of the patient population at the clinic.) We allow patients to have permanent unobservable preference heterogeneity, indexed by τ. Let $U(k \mid \tilde{k}, \tau)$ represent the lump-sum utility payoff from a treatment cycle that produces k children in that individual cycle, conditional on $\tilde{k}$ prior children and τ. As a normalization, we assume that treatment outcomes with $k = 0$ always result in $U(k \mid \tilde{k}, \tau) = 0$ for all patients. Beyond this normalization, we place no restrictions (e.g., concavity) on how $U(k \mid \tilde{k}, \tau)$ changes with k; the restrictions, if present, could have the effect of imposing risk preferences over values of k. While we treat $U(k \mid \tilde{k}, \tau)$ as a lump-sum benefit, we interpret it as the expected utility from having k additional children, which may include health risks from high-order births. The utility from children in $U(k \mid \tilde{k}, \tau)$ is policy-invariant. Other aspects of the model, introduced below, may change due to policies we consider, and these can impact the forward-looking patient's continuation value from different birth outcomes.

To maintain tractability, we combine $U(k \mid \tilde{k}, \tau)$ additively with other payoff-related terms in an indirect intertemporal expected utility function. Patients undergoing treatment experience disutility, scaled by α, from paying positive prices. When a patient pays p within treatment, she has the immediate utility loss of αp. We allow the value of α to depend on a patient's wealth, so we write $\alpha(z_w)$. This specification for the effect of p on utility is consistent with a model where consumption of other goods enters additively into the utility, and there is a static budget constraint that allocates income between consumption of other goods and IVF expenses.[22] Likewise, a patient's price depends on her insurance status, so we write $p(\iota)$. 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.[23] We assume that a patient's utility falls by η_0 if she makes a choice outside of the

[22] We have access only to a time-invariant, proxy measure of wealth, z_w. Therefore, we chose not to include an intertemporal budget constraint and we do not treat wealth (assets) as a time-varying state variable. High wealth operates in the model by changing α and thus altering the utility cost of reduced consumption of other goods.

[23] Prior to the September 2004 rule change, the ASRM recommended against single-embryo transfers for all patients, as well as four-embryo transfers for patients under age 35 and undergoing their first cycle. After the guidelines changed in September 2004, the recommendation against four-embryo transfers was extended to all patients under age 38 regardless of cycle number, plus patients between ages 38 and 40 who were taking their first cycle. The guideline revision also removed the recommendation against single-embryo transfers for patients under age 35 on their first cycle.

guidelines, with no penalty otherwise. We specify the function $\eta(x, \mathbf{Z}_a)$ to combine the penalty value and the conditions under which it is applied, with x as the number of embryos and $\mathbf{Z}_a$ capturing the contemporary ASRM guidelines and their relationship to the patient's current age. We assume that ASRM guidelines changes come as a surprise to decision makers, i.e., patients always believe the current policy environment is permanent. When the ASRM guidelines change, patients immediately reoptimize their treatment plans for the current rules, and then resume the belief that the guidelines will never change again.

The remaining parts of patients' preferences include the value from starting versus delaying an individual treatment and a terminal value. 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 during any period in which she begins IVF treatment. Note, u_s can include any physical or psychological stress from undergoing IVF. Patients' terminal payoffs are captured by the parameter vector $u_T(z_p)$, which depends on her prior payments. The patient receives u_T at age $a^{\max} + 1$ regardless of whether she remains active in the model up until $a^{\max}$ or if her decision process ends due to $k + \tilde{k} \geq 3$ at some earlier a.

At each treatment node, the patient's benefit from the available options includes an additional taste shock, ε, which represents heterogeneity in patient's circumstances and preferences. Following Rust (1987), for computational ease, we assume that ε is distributed i.i.d. type 1 extreme value across patients, time periods, treatment stages, and alternatives within each stage. We offset Euler's constant so that $E[\varepsilon] = 0$.

Finally, we assume that patients discount future decision periods by the factor β. We assume that all discounting occurs across periods, and not across treatment stages. Treatment options and outcomes that occur t periods into the future are discounted by β^t. We do not estimate β in this paper, so we set its value equal to $\beta = 0.97$.[24]

D. *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 estradiol (E2) to be revealed at the start of Stage 2. The score (e) is a signal of the patient's success in generating eggs. During the first stage, however, the patient knows only the distribution of possible e values rather than the signal's realization. Let $f_e(e \mid \mathbf{Z}_a)$ represent the probability that a patient with characteristics $\mathbf{Z}_a$ receives a score with value e, which takes positive integer values. Moving to the second stage, we denote as $f_r(r \mid e, \mathbf{Z}_a)$ the probability

[24] In principle, the nonstationary nature of our model could allow for identification β. In practice, however, this parameter is often very difficult to estimate so we chose to fix it at a conventional value.

of successfully retrieving r eggs for a patient with Peak E2 score e and personal characteristics $\mathbf{Z}_a$. A patient with a greater value of r is more likely to generate a large number of embryos in the next stage. 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 $f_X(X \mid r, m, \mathbf{Z}_a)$, and note that it may be shifted by r, m, and the patient's state variables. Finally, in Stage 4 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 $f_k(k \mid x, \mathbf{Z}_a)$.

We specify the prices that patients may pay at three treatment stages. The price of action y in stage j is $p_{y,j}(\iota)$. 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.

III. 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. The patient's current incentives are affected by her future treatment opportunities and prices. 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 $W_{y,j}(\mathbf{Z}_a, \varepsilon_{y,j,a})$ 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 $W_{y,j}(\mathbf{Z}_a, \varepsilon_{y,j,a})$ depend on τ, but we suppress this term for notational simplicity. Let $\overline{W}_{y,j}$ be the systematic component of $W_{y,j}(\mathbf{Z}_a, \varepsilon_{y,j,a})$, i.e., $W_{y,j}$ net of the additive preference shock $\varepsilon_{y,j,a}$. We then have

$$W_{y,j}(\mathbf{Z}_a, \varepsilon_{y,j,a}) = \overline{W}_{y,j}(\mathbf{Z}_a) + \varepsilon_{y,j,a}. \tag{1}$$

The patient observes the realization of the vector $\varepsilon_{j,a}$ before making her choice during stage j. The patient's value at the start of stage j is

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

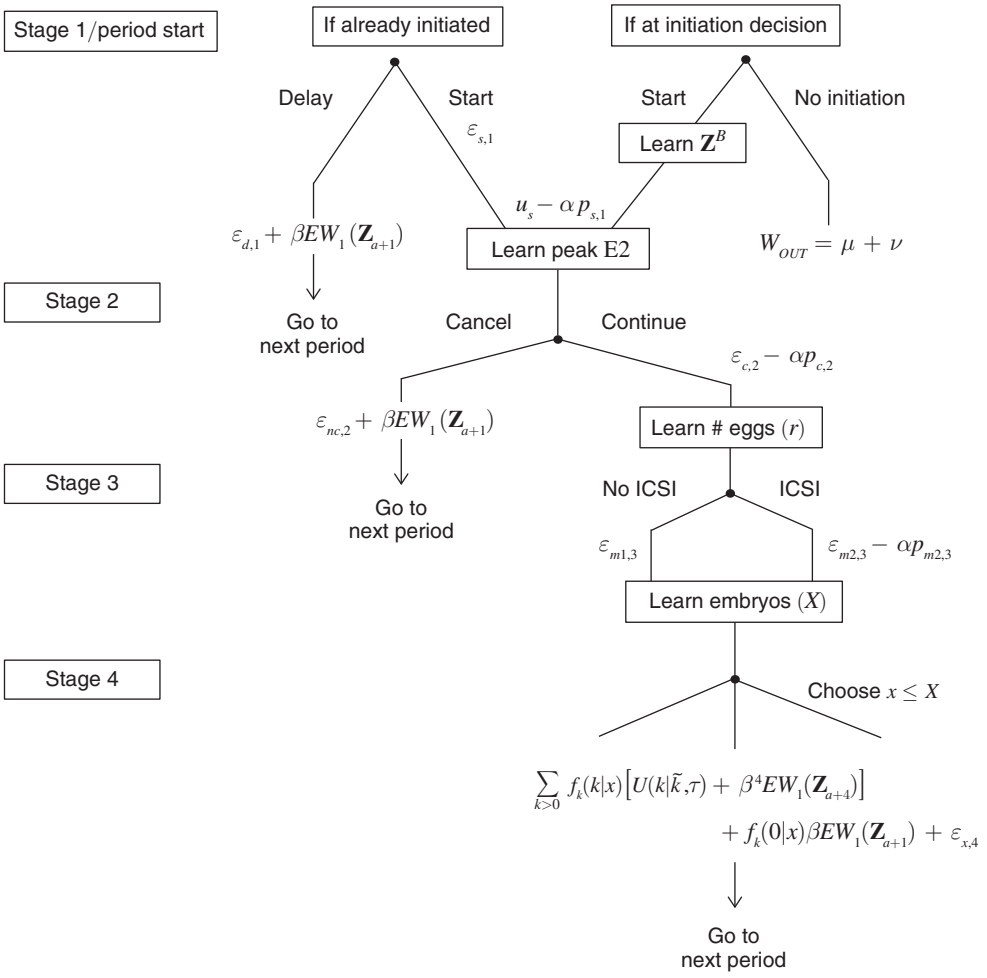

FIGURE 1. IVF TREATMENT STAGES

Notes: 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.

Here, $E\left[W_j(\mathbf{Z}_a, \varepsilon_{j,a})\right]$ represents the expected value from an optimal decision within treatment stage j, before observing the realization of $\varepsilon_{j,a}$. Due to the extreme value assumption for ε, we can write $E\left[W_j(\mathbf{Z}_a, \varepsilon_{j,a})\right]$ with the inclusive value expression:

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

We begin by focusing on the treatment stages that occur within IVF, after the patient has learned her value of $\mathbf{Z}^B$. Later in this section we return to the initiation decision.

A. *Stage 1: Start Treatment versus Delay*

In all periods after the initiation decision, patients who began IVF previously will return to Stage 1 and choose between the actions start (s) and delay (d). If the patient

starts treatment, she pays the price $p_{s,1}(\iota)$ out-of-pocket and begins a regimen of pharmaceuticals to promote egg production.

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

$$(4)\quad W_{s,1}(\mathbf{Z}_a, \varepsilon_{s,1,a}) = \overline{W}_{s,1}(\mathbf{Z}_a) + \varepsilon_{s,1,a}$$
$$= u_s - \alpha(z_w)p_s(\iota) + \varepsilon_{s,1,a} + \sum_e E[W_2(e, \mathbf{Z}_a, \varepsilon_{2,a})]f_e(e|\mathbf{Z}_a).$$

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

$$(5)\quad W_{d,1}(\mathbf{Z}_a, \varepsilon_{d,1,a}) = \overline{W}_{d,1}(\mathbf{Z}_a) + \varepsilon_{d,1,a}$$
$$= 0 + \beta E[W_1(\mathbf{Z}_{a+1}, \varepsilon_{1,a+1})] + \varepsilon_{d,1,a}.$$

Changes in $\mathbf{Z}$ 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 $\beta E[W_1(\mathbf{Z}_{a+1}, \varepsilon_{1,a+1})]$ accounts for the expectation of ε, the payoffs in $\overline{W}_1$ associated with starting or delaying IVF at age $a+1$, and the patient's option to choose the optimal action. If the patient is already at age $a^{\max}$, however, she receives the terminal value $W_T(\mathbf{Z}_{a^{\max}}) = u_T(z_p)$ 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.

B. *Stage 2: Continue versus 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 $(\mathbf{Z}_a)$ 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 $p_{c,2}(\iota)$ and undergoes a surgical process in which eggs are retrieved.

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

$$(6)\quad W_{nc,2}(e, \mathbf{Z}_a, \varepsilon_{nc,2,a}) = \overline{W}_{nc,2}(\mathbf{Z}_a) + \varepsilon_{nc,2,a}$$
$$= 0 + \beta E[W_1(\mathbf{Z}_{a+1}, \varepsilon_{1,a+1})] + \varepsilon_{nc,2,a}.$$

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

$$
\begin{aligned}
(7) \qquad W_{c,2}(e, \mathbf{Z}_a, \varepsilon_{c,2,a}) &= \overline{W}_{c,2}(e, \mathbf{Z}_a) + \varepsilon_{c,2,a} \\
&= -\alpha(z_w)\, p_c(\iota) + \varepsilon_{c,2,a} \\
&\quad + \sum_r E\big[W_3(r, \mathbf{Z}_a, \varepsilon_{3,a})\big] f_r(r|e, \mathbf{Z}_a).
\end{aligned}
$$

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

C. *Stage 3: Fertilization*

If treatment is not canceled, 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 $(\mathbf{Z}_a)$, 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 $\left(z_{mf} = 1\right)$ are likely to receive the greatest benefits from fertilizing via ICSI $\left(m = m_2\right)$. The price of option m is $p_{m,3}(\iota)$, which is positive for uninsured patients when $m = m_2$, and zero otherwise.

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, $f_X(X|r, m, \mathbf{Z}_a)$. We write the patient's choice-specific value from a third-stage action:

$$
\begin{aligned}
(8) \qquad W_{m,3}(r, \mathbf{Z}_a, \varepsilon_{m,3,a}) &= \overline{W}_{m,3}(r, \mathbf{Z}_a) + \varepsilon_{m,3,a} \\
&= -\alpha(z_w)\, p_{m,3}(\iota) + \varepsilon_{m,3,a} \\
&\quad + \sum_X E\big[W_4(X, \mathbf{Z}_a)\big] f_X(X|r, m, \mathbf{Z}_a).
\end{aligned}
$$

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

D. *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 $(\mathbf{Z})$. As a result of treatment, k children are born with probability $f_k(k|x, \mathbf{Z}_a)$. Under current policy, 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., nine months) before making her next reproductive decision.

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

$$(9)\quad W_{x,4}(X,\mathbf{Z}_a,\varepsilon_{x,4,a}) = \overline{W}_{x,4}(X,\mathbf{Z}_a)+\varepsilon_{x,4,a}$$
$$= \eta(x,\mathbf{Z}_a)+\varepsilon_{x,4,a}+f_k(0|x,\mathbf{Z}_a)\,\beta E\big[W_1(\mathbf{Z}_{a+1},\varepsilon_{1,a+1})\big]$$
$$+\sum_{k>0} f_k(k|x,\mathbf{Z}_a)\Big\{U(k|\tilde{k}_a,\tau)+\beta^4 E\big[W_1(\mathbf{Z}_{a+4},\varepsilon_{1,a+4})\big]\Big\}.$$

This expression includes the possibilities of failed treatment ($k=0$) and successful treatment ($k>0$). The future value of a patient's decision, $E\big[W_1(\cdot)\big]$, depends on the realization of the current treatment and the prices and policies that constrain the patient in future periods. If the treatment is successful, $\mathbf{Z}_a$ evolves to a value $\mathbf{Z}_{a+4}$ which reflects that the patient is one full year older and has k additional children. Moreover, this future value is discounted at β^4. If treatment fails, then the next decision's value is discounted by β, and $\mathbf{Z}_{a+1}$ reflects that the patient is just three months older. The value function captures the patient's benefit from building a family of a certain size during her fertile years, so statements about a patient's desire to avoid the risk of lifetime childlessness would be made in reference to W rather than $U(k|\tilde{k},\tau)$.

E. *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. We make the simplifying assumption that this potential patient does not yet know her values of $\mathbf{Z}^B$, but she knows the population distribution of $\mathbf{Z}^B$ values conditional on age, $f_{\mathbf{Z}^B}(\mathbf{Z}^B_{a_0}|\,a_0)$. The potential patient's expected value from starting treatment is

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

where we make the distinction between the state variables known prior to treatment ($\mathbf{Z}^D_{a_0}$) and those learned after treatment begins. The potential patient compares $W\big(\mathbf{Z}^D_{a_0}\big)$ to the utility from foregoing treatment, which we specify as $W^{OUT}(z_r) = \mu(z_r)+\nu$. The parameter $\mu(z_r)$ captures the mean value of the outside option for all potential patients of race z_r. (We use the notation $\mu(z_r)$ and μ to refer to the outside option's mean value for an individual patient of a given race; we write $\boldsymbol{\mu}$ to refer to the vector of individual μ values that we estimate.) There is evidence that African American women are less inclined to pursue infertility treatment, and in our model this would be reflected in a higher value of μ. The value of ν is specific to each potential patient and captures heterogeneity in the value of permanently foregoing IVF treatment; it explains why potential patients with the same $\mathbf{Z}^D_{a_0}$ and (unobserved) τ make different choices with respect to ever pursuing IVF.[25] One possible interpretation of ν 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. Notice,

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

τ and ν play fundamentally different roles in the model. Here, τ characterizes permanent differences in utility from children. Potential patients know both their ν and τ before embarking in IVF treatment. The ν term captures idiosyncratic preferences for taking the outside option relative to pursuing IVF but plays no further role once the potential patient becomes a patient.[26]

Under these assumptions, the potential patient enters the clinic to initiate her first IVF cycle if the expected value of pursuing IVF is greater than the value of foregoing treatment, and she exits the model otherwise. We let the indicator I equal 1 whenever a potential patient initiates IVF treatment, and equal 0 otherwise. Then

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

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 patient's first cycle, i.e., $\overline{W}_{s,1}\left(\mathbf{Z}_{a_0}, g\right)$. We let $g = g_E$ 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, the value of $\overline{W}_{s,1}$ (and all other $\overline{W}_{y,j}$ that follow) changes and therefore initiation decisions are affected.

IV. Data

A. *Clinic Data*

Our primary data cover individual patient histories at the clinic during 2001–2009. We observe all treatment cycles conducted during this period for patients who underwent their first-ever IVF cycle between 2001 and 2007, plus we observe cycles during 2001–2009 for some patients who first received IVF before 2001. While the data allow us to describe many patients' IVF histories from the start of their 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), received their first-ever IVF cycle at the clinic between 2001 and 2007, and have complete data on their personal characteristics and treatment details.[27] 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 2,167 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

[26] In a more general model, observationally similar patients, even with same τ and facing same ε, could make different decisions, perhaps with those with high values of ν being more reluctant to use IVF multiple times than those with low values of ν. For computational tractability, though, we work with a simpler model in which once a woman decides to use IVF, heterogeneity in ν is no longer relevant for her decision-making within the clinic.

[27] In practice, patients may choose to freeze excess embryos for potential later use, but we do not examine that decision. Frozen embryo cycles account for only 12 percent of the clinic's treatments during the sample period.

Table 2—Patient-Level Characteristics

	Main sample $N = 587$		First-stage sample $N = 1{,}106$	
	Mean	SD	Mean	SD
Panel A. Demographic state variables ($\mathbf{Z}^D$)				
Patient age at initiation	34.30	4.02	33.36	4.70
Insured at initiation? (Y = 1)	0.54	0.50	0.59	0.49
Wealthy zip code? (Y = 1)	0.82	0.39	0.79	0.41
Prior children at initiation	0.00	0.00	0.32	0.58
African American? (Y = 1)	0.05	0.21	0.05	0.21
Panel B. Biological state variables ($\mathbf{Z}^B$)				
AFC score	14.34	7.96	14.61	8.13
Female fertility problem? (Y = 1)	0.80	0.40	0.69	0.46
Male fertility problem? (Y = 1)	0.34	0.48	0.30	0.46
Panel C. Aggregate actions and outcomes				
Total cycles	1.75	1.02	1.96	1.21
Birth during sample period? (Y = 1)	0.53	0.50	0.54	0.50

Notes: We use the main sample in second-stage estimation of patients' choices. We use the first-stage sample to estimate treatment technologies.

health state outcomes from Stages 2–4 of IVF cycles. In the full first-stage sample we have complete data on 5,052 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 1,106 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 insured patients are from Illinois but not all; likewise, most Illinois patients are insured but not all of them. Most patients' homes are in a zip code with a median house price above $100,000, which we use as a proxy for patient wealth (z_w). We track race as a binary variable which indicates whether the patient is African American $(z_r = 1)$ or not $(z_r = 0)$; 95 percent of patients have $z_r = 0$. 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 $(\mathbf{Z}^B)$ 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 one-half 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 percent cancellation rate. Most patients (60 percent) fertilize their eggs with ICSI; this rate is closer to 90 percent when male-factor infertility is present.[28] 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

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

Table 3—Actions and Outcomes within Treatment

	Main sample			First-stage sample		
	Observations	Mean	SD	Observations	Mean	SD
Panel A. Stage 1–4 actions						
Cancel treatment? (Y = 1)	1,027	0.14	0.35	1,859	0.14	0.35
Fertilization method? (ICSI = 1)	879	0.60	0.49	1,597	0.59	0.49
Number of embryos transferred	879	2.28	0.82	1,596	2.32	0.83
Panel B. Stage 1–3 outcomes						
Peak E2 score	1,027	16.82	9.73	1,858	17.10	9.57
Eggs retrieved	879	10.60	5.46	1,597	10.87	5.51
Embryos generated	879	6.12	3.74	1,597	6.37	3.85
4+ embryos? (Y = 1)	879	0.74	0.44	1,597	0.77	0.42
Panel C. Stage 4 outcomes						
Children born	848	0.51	0.70	1,545	0.57	0.75
Singleton birth? (Y = 1)	848	0.27	0.45	1,545	0.28	0.45
Twin birth? (Y = 1)	848	0.12	0.32	1,545	0.13	0.34
Triplet birth? (Y = 1)	848	0.00	0.00	1,545	0.01	0.10

Notes: We use the main sample in second-stage estimation of patients' choices. We use the first-stage sample to estimate treatment technologies.

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 percent of cycles, and twins occur in an additional 12 percent. While we observe no triplet births in the main sample, they occur at a rate of about 1 percent in the larger first-stage sample; this allows us to account for triplet risk when estimating the structural model.[29]

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 percent of patients whose first cycle was unsuccessful returned for one or more additional cycles, while 16 percent 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 percent of all cases, while patients with scores in the twenty-fifth to seventy-fifth percentile cancel only 4 percent of cycles. Patients who are 35 or older take an average of 2.6 embryos in their first cycle, while younger patients take two embryos on average. Uninsured patients take more embryos (2.4) during their first cycle than insured patients (2.2), but this difference, which is statistically significant at the first cycle, shrinks in the second and third cycle as insurance coverage is drawn down.

B. *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,

[29] In 2009, 1.6 percent of all US IVF births from fresh non-donor cycles were triplets or more. Conditional on three or more embryos transferred, the triplet rate was 3.6 percent (Centers for Disease Control 2014).

its surrounding suburbs, and some rural towns outside of the metro area. This area captures almost all of the patients who ever visit the clinic; a small number come from greater distances.

We first describe the various data sources for the market data, and then we describe how they are assembled into an estimate of the "at risk" population. We use the Centers for Disease Control and Prevention's (CDC) Vital Statistics database to construct the market's distribution of maternal age at first birth. This distribution, along with estimates of infertility rates by age from Dunson, Baird, and Colombo (2004), allows us to construct an age distribution for women who may consider IVF. We supplement the Dunson et al. estimates with data from the National Survey of Family Growth (NSFG) to estimate the difference in infertility rates by race. 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 collect zip code-level data on race and the median home values from the 2000 Decennial Population Census. The home value data allow us to provide an estimate for the distribution of patients' wealth. We combine the various zip code-level data to 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 how we 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, in every quarter t there is a race-specific distribution of age at first (attempted) birth for these women $f_t(a \mid z_r)$. 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 $inf(a, z_r)$ be an age- and race-specific infertility rate which increases with age. Together, $\left(N^{stl}, f_t(a \mid z_r), inf(a, z_r)\right)$ provide the number of women, $\tilde{N}^{inf}_{a,z_r}$, of each age and race who realize that they are unable to conceive without IVF. These $\tilde{N}^{inf} = \sum_{z_r=0}^{1} \sum_{a=a^{\min}}^{a^{\max}} \tilde{N}^{inf}_{a,z_r}$ 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 by deflating $\tilde{N}^{inf}$ to match the clinic's market share as reported by the CDC. In total across all years of the sample period, we estimate that $N^{inf} = 2{,}781$ women consider treatment at the clinic we study. See Appendix B for additional details on the calculation of N^{inf} and its relationship to $f_{\mathbf{Z}^D}\left(\mathbf{Z}^D_{a_0}\right)$, the distribution of patient characteristics.

We compute race-specific empirical initiation shares, $s^{init}(z_r)$, for the clinic. We observe that $N^{clin} = 828$ new patients (including 35 African American) initiated treatment at the clinic during the period 2001–2007.[30] Using our estimates of $N^{inf}(z_r)$, we calculate the share

$$s^{init}(z_r) \approx \frac{N^{clin}(z_r)}{N^{inf}(z_r)}. \tag{10}$$

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

We find $s^{init}(0) = 0.072$ and $s^{init}(1) = 0.346$ which means that 7.2 percent of the clinic's African American potential patients and 34.6 percent of its remaining potential patients decided to pursue treatment. The remaining 1,953 potential patients could be induced to seek IVF treatment through large enough increases in $\overline{W}_{1,s}$.

V. Empirical Specification

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

A. *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 $\mathbf{Z}^B$, 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 $\mathbf{Z}^B$. 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 $f_e(e|\mathbf{Z}_a)$. In the data we observe e values between 0 and 10,196 pg/mL, with a mean and median around 1,600, and 99 percent 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, 500–1,000, 1,000–1,500, 1,500–2,000, 2,000–2,500, and over 2,500. 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 ages 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 $\mathbf{Z}_a$. In the data, r takes integer values from 0 to 38 with a mean of 10.6 and median of 10. The ninetieth percentile is at $r = 18$, and 99 percent 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 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 roles in determining outcomes is finished once the patient has her set of transferable embryos.

B. *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 $U(k|\tilde{k},\tau) = 0$ from zero-birth outcomes, we assume that outcomes with $k > 0$ provide utility according to

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

The parameter vector (u_1, u_2, u_3) captures the lump-sum payoff from a singleton, twin, and triplet birth to a patient with no prior children $(\tilde{k} = 0)$. Given the health risks and other challenges for triplets, we anticipate that $u_3 < u_2$ and $u_3 < u_1$, but these parameters are unrestricted in estimation. The parameter κ captures any difference in the marginal benefit of a birth to patients with prior children $(\tilde{k} > 0)$; diminishing marginal utility from children would imply that κ is negative. We account for the impact of ASRM guidelines on utility by specifying $\eta(x, \mathbf{Z}_a) = \eta_0 \times 1\{x, \mathbf{Z}_a\}$, where η_0 is a constant utility penalty and $1\{x, \mathbf{Z}_a\}$ is an indicator function that is equal to one for any choice of embryos (x) that is outside of the contemporary ASRM guidelines for a patient with state variables $\mathbf{Z}_a$.[31]

We assume a simple two-type structure for patients' permanent unobserved heterogeneity. A share of patients with type $\tau = 1$ has preferences for birth outcomes represented only by (u_1, u_2, u_3, κ), while the remaining patients (with $\tau = 2$) have, in addition, their utility payoff shifted by a scalar parameter ζ. A patient's probability of being of type $\tau = 2$ depends on her state values at the time she initiated treatment, $\mathbf{Z}^D_{a_0}$. Along with a_0, we allow the distribution of τ, conditional on initiation, to depend on a measure of her wealth level (z_w), her race (z_r), her initial number of insurance-covered cycles (ι_{a_0}), and a dummy (z_{asrm_0}) for the ASRM guideline regime when treatment started. The dependence of the distribution of τ on the state variables reflects the need to acknowledge that differences across patients' circumstances will result in a different selection of patients depending upon their intensity

[31] In practice, no patients in the main data choose to go outside of the guidelines by more than a single embryo, so we do not have the opportunity to estimate different penalties for different degrees of violation.

of taste for children. The value of initiation depends on various state variables and the unobserved type, and we expect a differential propensity to initiate by the two types. For example, conditional on being uninsured, one would expect a higher prevalence of patients with high taste for children (relative to the prevalence among potential patients). As a result, regardless of whether types are correlated with demographics in the pool of potential patients, the distribution of types within the clinic will surely depend on the state variables. We assume that the probability of a high type ($\tau = 2$) is

$$\Pr(\tau = 2 \,|\, \mathbf{Z}^D_{a_0}, I = 1, \boldsymbol{\rho}) = \frac{\exp(\rho_0 + \rho_1 a_0 + \rho_2 z_w + \rho_3 \iota_{a_0} + \rho_4 z_{asrm_0} + \rho_5 z_r)}{1 + \exp(\rho_0 + \rho_1 a_0 + \rho_2 z_w + \rho_3 \iota_0 + \rho_4 z_{asrm_0} + \rho_5 z_r)}.$$

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 $\boldsymbol{\rho}$ represent a column vector of $\{\rho_j\}_{j=0}^5$ values. In addition, we write $[1, \mathbf{Z}^D_{a_0}]$ as a vector containing 1 and an individual patient's row vector $\mathbf{Z}^D_{a_0}$, and we let Λ represent the logistic distribution function so that $\Lambda([1, \mathbf{Z}^D_{a_0}]\boldsymbol{\rho}) = \Pr(\tau = 2 \,|\, \mathbf{Z}^D_{a_0}, I = 1, \boldsymbol{\rho})$.

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 sacrifices) when she begins a treatment cycle. We assume that $u_s = \delta$, a scalar parameter. The value of u_s is 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 $\alpha(z_w)$, the disutility from paying a price p for some treatment component. We specify $\alpha(z_w)$ as an affine function of z_w: $\alpha(z_w) = \alpha_0 + \alpha_w z_w$. Since the effect of price is subtracted from within-stage value functions above, we expect α_0 to be positive for consistency with downward-sloping demand. If wealthier patients are less price sensitive, this will be captured through $\alpha_w < 0$.

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

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

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

VI. 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. Within this stage, we also estimate the distribution of $f_{\mathbf{Z}^B}(\mathbf{Z}^B|a_0)$ nonparametrically using frequencies of $\mathbf{Z}^B$ realizations from within the population of women who initiate treatment. In the second stage, we estimate the parameters in $\boldsymbol{\theta}$ using data exclusively from the population of 587 patients who are observed within the clinic. In the final stage, we estimate $\mu(z_r)$ using our estimates of $E\left[\overline{W}_{s,1}(\mathbf{Z}_{a_0}) \,|\, \mathbf{Z}^D_{a_0}\right]$ together with the market-level data.

A. *Within-Clinic Choices*

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

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

$$\Pr(d_{c,2,a,i} = 1 \,|\, \mathbf{Z}_a, \tau; \zeta, \boldsymbol{\varphi}) = \frac{\exp\left[\overline{W}_{c,2}(\mathbf{Z}_a, \tau; \zeta, \boldsymbol{\varphi})\right]}{\exp\left[\overline{W}_{c,2}(\mathbf{Z}_a, \tau; \zeta, \boldsymbol{\varphi}) + \overline{W}_{nc,2}(\mathbf{Z}_a, \tau; \zeta, \boldsymbol{\varphi})\right]}. \tag{11}$$

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

A patient's permanent unobserved type, τ, affects every period and stage of her decision problem. Let $\Pr(d_{y,j,a,i} = 1; \mathbf{Z}_a, \tau, \zeta, \boldsymbol{\varphi})$ represent the predicted probability that patient i takes the observed action $d_{y,j,a,i}$ if she has type τ. The patient is observed

starting in period $t_{i,0}$ and ending in T_i. Conditional on ζ and φ, the type-specific joint probability of observing patient i's sequence of choices is

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

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

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

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

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

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

In this second stage, the key model parameters recovered by the maximum likelihood procedure are the "price coefficient" (α_0) and the nonparametric utility from children, given by u_k for $k=1,2,3$. Identification of α_0 comes from various features of the clinic data. Specifically, it primarily draws on the following three sources of variation. First, since every patient that we observe at the clinic started at least one cycle, cross-sectional differences in insurance coverage during the first cycle provide identifying variation through different cancellation decisions. Greater price sensitivity induces patients facing higher prices (due to insurance variation) to be more selective in their decision to continue on to the second stage—the most expensive one—involving egg retrieval. Second, while all clinic patients started at least one cycle, uninsured patients are (all else equal) less likely to return for another attempt after an initial failure. Again, the greater the price sensitivity measured by α_0, the greater the observed differential return rate of uninsured versus insured patients. Third, additional variation comes from the observed behavior of the subsample of initially-insured patients as they approach the exhaustion of their covered cycles, relative to their behavior in earlier cycles. The greater the price sensitivity, the greater the tendency of insured patients to take actions that either preserve the last remaining insured cycles (by canceling more often) or make sure that the cycle succeeds (by increasing the number of embryos transferred). Identification of α_w comes from contrasting these arguments in subsamples of low versus high wealth patients.

The identification of the u_k values can be obtained with the embryo-transfer choice data under our assumption that patients know the probabilities of different birth outcomes for each potential embryo choice. Through her embryo-transfer choice, the patient essentially chooses among different potential lotteries over birth outcomes k. For example, if patients strongly dislike triplet outcomes (i.e. they have a very negative value for u_3), then they will avoid embryo choices that are more likely to produce triplets. While the identification of the u_ks does not require variation in ASRM guidelines, which changed in the middle of our sample period (2004), such variation is also helpful for identification. Comparing the transfer behavior of patients with similar characteristics facing different recommended choice sets helps

identify preferences for birth outcomes as well as the penalty parameter η_0. Of particular value are data on how similar patients behave when guidelines change with respect to single embryo transfers.

B. *Treatment Initiation*

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

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

where the variable I indicates whether a patient with characteristics $\mathbf{Z}^D_{a_0}$ started treatment at age a_0. We integrate over observed and unobserved potential patient characteristics to compute the model $\boldsymbol{\theta}$ predicted rates at which the clinic's potential patients of both races actually become patients, $\hat{s}^{init}\left(\mu(z_r); g_E, \hat{\boldsymbol{\theta}}\right)$. These race-specific distributions, $f_{\mathbf{Z}^D|z_r}(\mathbf{Z}^D_{a_0})$ for observed characteristics and $f_{\tau|z_r}(\tau|\,\mathbf{Z}^D_{a_0})$ for unobserved heterogeneity, will in general be different in the full potential patient pool versus among patients who choose to initiate treatment. We estimate the parameters $\mu(z_r)$ by solving

$$\hat{s}^{init}\left(\mu(z_r); g_E, \hat{\boldsymbol{\theta}}\right) = s^{init}(z_r),$$

for each race z_r under the empirical policy setting, g_E. Here, $s^{init}(z_r)$ is the race-specific, empirical initiation share that we construct from the count of actual patients given z_r and our estimate of the size of the potential patient pool with race z_r, and $\hat{s}^{init}$ is the initiation share predicted by the model for a given value of $\mu(z_r)$. We estimate μ for each race by finding the value that makes the model-predicted race-specific initiation share $\hat{s}^{init}$ match the observed share, s^{init}. Note that under alternative policies, $W\left(\mathbf{Z}^D_{a_0}, \tau, g, \boldsymbol{\varphi}, \zeta\right)$ will change but $\mu(z_r)$ remains fixed.

We approach the distributions $f_{\tau|z_r}(\tau|\,\mathbf{Z}^D_{a_0})$ and $f_{\mathbf{Z}^D|z_r}(\mathbf{Z}^D_{a_0})$ using different strategies. In Appendix A we show that our assumptions on: (i) the distribution of unobserved types conditional on treatment initiation, $f_\tau\left(\tau|\,\mathbf{Z}^D_{a_0}, I_i = 1, \boldsymbol{\rho}\right)$; and (ii) the initiation decision, are sufficient to back-out the unconditional (on initiation) race-specific distributions $f_{\tau|z_r}(\tau|\,\mathbf{Z}^D_{a_0})$. We write the incidence of $\tau = 2$ within-clinic as $\Lambda([1, \mathbf{Z}^D_{a_0}]\boldsymbol{\rho})$, and $\Lambda\left(W\left(\mathbf{Z}^D_{a_0}, \tau, g_E, \boldsymbol{\varphi}\right) - \mu(z_r)\right)$ provides the race-specific probability of initiation for $\tau = 1, 2$. In Appendix B we describe our approach to constructing $f_{\mathbf{Z}^D}(\mathbf{Z}^D_{a_0})$ using market data.

The estimates for $f_{\tau|z_r}\left(\tau|\,\mathbf{Z}^D_{a_0}\right)$ and $f_{\mathbf{Z}^D|z_r}\left(\mathbf{Z}^D_{a_0}\right)$ allow us to derive $s^{init}\left(\boldsymbol{\theta}, \mu(z_r), g_E\right)$, the model-predicted fraction of potential patients of race z_r who actually become patients. To obtain $s^{init}\left(\mu(z_r); g_E, \boldsymbol{\theta}\right)$, we integrate the initiation probability $\Pr(I = 1 \mid \mathbf{Z}^D_{a_0}, \boldsymbol{\theta}, \tau, \mu(z_r))$ over the race-specific distribution of $\mathbf{Z}^D_{a_0}$ and τ among potential patients,

$$\begin{aligned}\hat{s}^{init}\left(\mu(z_r); g_E, \hat{\boldsymbol{\theta}}\right) &= \Pr\left(I = 1 | \hat{\boldsymbol{\theta}}, \mu(z_r), g_E\right) \\ &= \sum_{\mathbf{Z}^D_{a_0}} \left[\sum_\tau \Lambda\left(W\left(\mathbf{Z}^D_{a_0}, \tau, g_E, \hat{\boldsymbol{\varphi}}, \hat{\zeta}\right) - \mu(z_r)\right) f_{\tau|z_r}\left(\tau|\mathbf{Z}^D_{a_0}, \boldsymbol{\mu}, \hat{\boldsymbol{\theta}}\right)\right] f_{\mathbf{Z}^D|z_r}\left(\mathbf{Z}^D_{a_0}\right).\end{aligned}$$

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

VII. Results

A. *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, $\mathbf{Z}$, which includes her AFC score, z_{afc}. At this point in the decision process, she considers her probable Peak E2 score (e), which will be revealed in Stage 2 if she starts treatment. In Figure 2 we display (estimated) 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 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 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 500–1,000 range) the mode of the distribution of eggs is 5–10 whereas if e is relatively high (2,000–2,500) the mode of the distribution of eggs is 11–20 and the probability of a low egg count (1–4) is almost 0. 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 four or more viable embryos and reducing the probability of 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 57 percent to 53 percent, 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 two or three

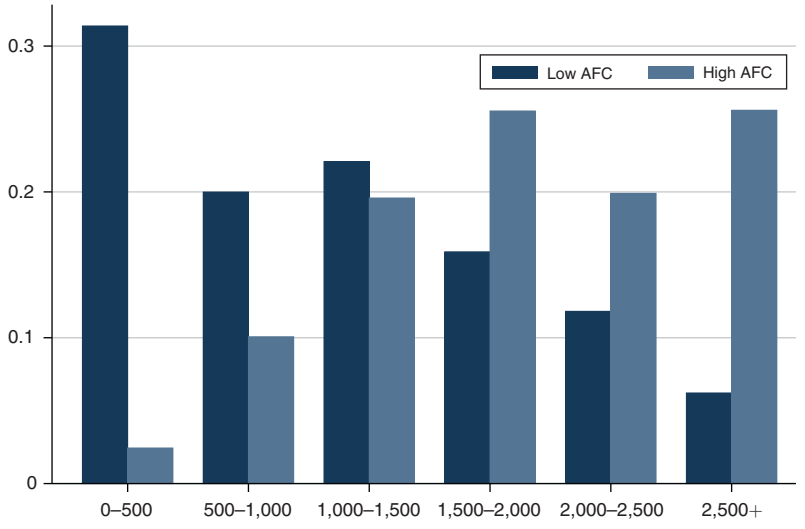

FIGURE 2. DISTRIBUTION OF PEAK E2 OUTCOMES BY AFC

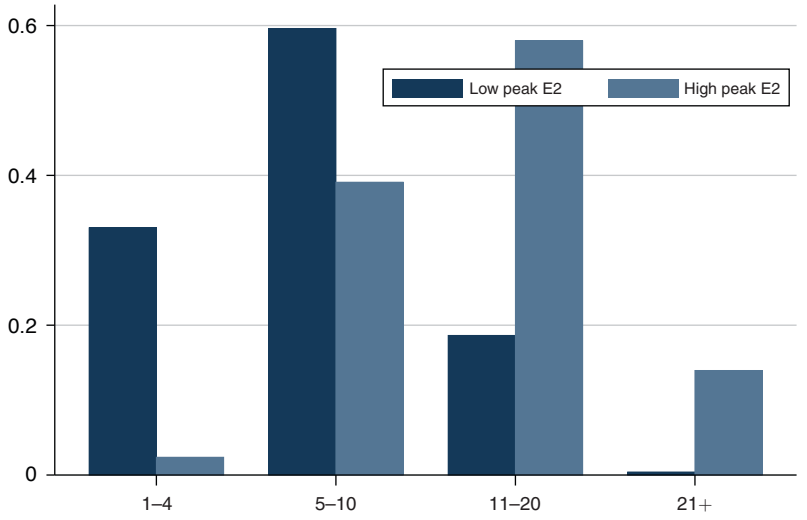

FIGURE 3. DISTRIBUTION OF RETRIEVED EGG COUNT BY PEAK E2

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 $(\text{age} > 35)$ women shifts to the left, noticeably increasing the odds of no live birth.

B. *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 $U(k \mid \tilde{k}, \tau)$,

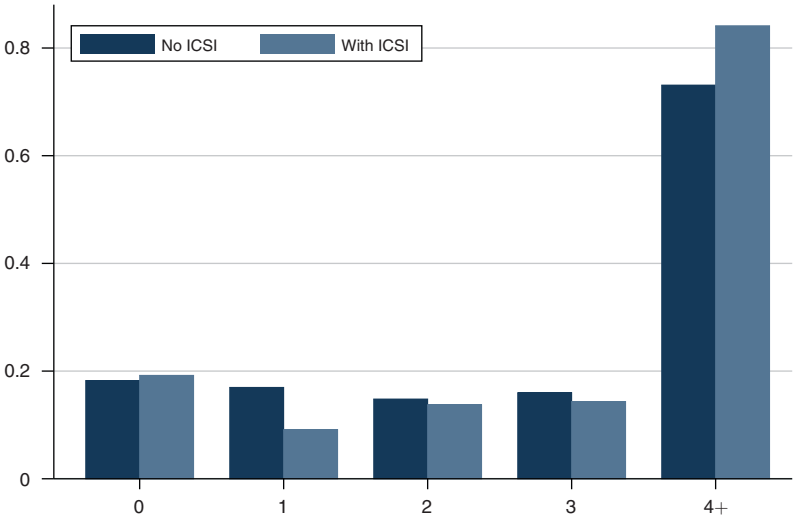

FIGURE 4. DISTRIBUTION OF EMBRYOS AVAILABLE BY ICSI USE, MALE FACTOR PATIENTS

α, δ, γ, and η. Our estimates of u_1, u_2, and u_3 represent payoffs from different birth outcomes to patients with $\tau = 1$ and no prior children. These estimates show that patients receive a positive payoff from a singleton or twin birth, with the latter valued slightly more. Triplet births, by contrast, have a negative utility payoff for patients. The estimate of κ indicates that patients with one or two prior children have their utility from births shifted downward substantially. For example, for a patient with $\tilde{k} > 0$ and $\tau = 1$, the estimated κ implies that the patient would prefer no additional children. The taste shifter ζ associated with type 2, however, is sufficient to increase the utility from additional births to be positive for patients with $\tilde{k} > 0$.

Table 4's results indicate that the baseline price disutility is significantly different from zero for all patients, and this disutility is smaller for high-wealth patients. We recover a significantly negative estimate for the start/delay parameter δ, which plays a large role in determining whether a patient returns for additional treatment cycles after her first. Our estimate of the parameter γ, for a patient's terminal payoff u_T, shows no significant difference between the utility of patients who have paid out-of-pocket for a treatment and those who have not. The final utility parameter on Table 4 is η_0, 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 τ within the treated population. We estimate that about half of the patient population has type $\tau = 2$ given their $\mathbf{Z}$ values. To interpret the individual ρ parameters, consider the case of patient wealth. The negative coefficient (ρ_2) on the wealth measure indicates that a high-wealth person selected from the treated population is less likely to have type $\tau = 2$ than a 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.

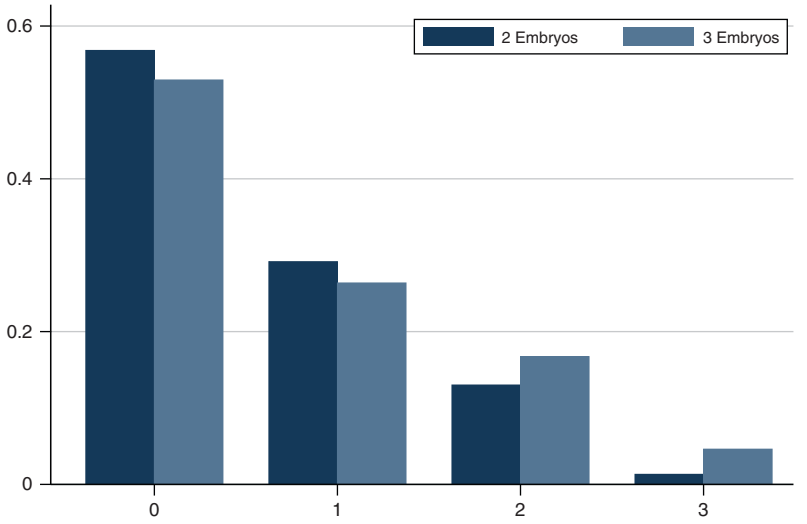

FIGURE 5. DISTRIBUTION OF BIRTHS BY EMBRYOS TRANSFERRED, PATIENT AGE 34–36

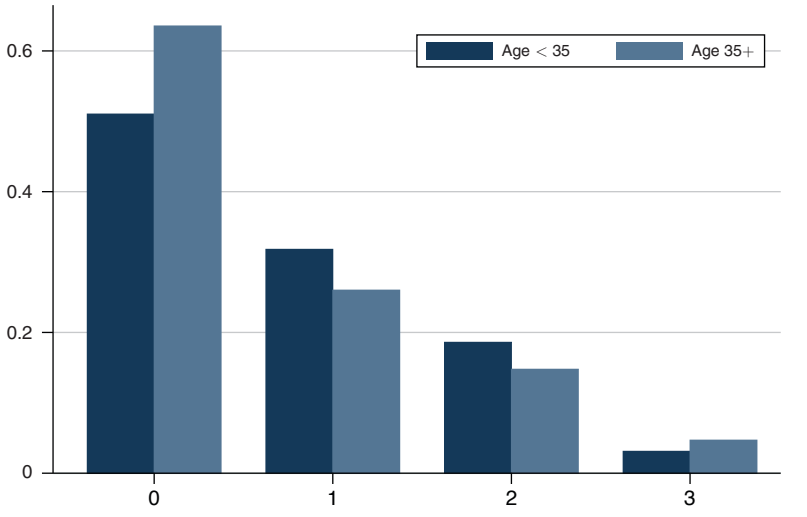

FIGURE 6. DISTRIBUTION OF BIRTHS BY PATIENT AGE, THREE EMBRYOS

Finally, in the third estimation step we recover $\hat{\mu}(1) = -0.34$ for African Americans and $\hat{\mu}(0) = -2.17$ for all other potential patients.[32] These values of μ ensure that the initiation model generates treatment initiation decisions such that, as estimated from our data, 30 percent of all potential clinic patients (7.2 percent

[32] The 95 percent confidence interval for $\hat{\mu}(1)$ is $[-0.85, 0.50]$ and for $\hat{\mu}(0)$ is $[-2.57, -1.57]$.

Table 4—Utility Parameter Estimates

Utility of 1 birth (u_1)	5.143
	(0.928)
Utility of 2 births (u_2)	5.866
	(1.682)
Utility of 3 births (u_3)	−14.095
	(4.411)
Utility shift when $\tilde{k} > 0$ (κ)	−11.868
	(0.959)
Preference shifter ζ	9.768
	(0.865)
Price sensitivity constant (α_0)	0.312
	(0.070)
Price sensitivity × wealth (α_w)	−0.127
	(0.066)
Terminal payoff × Prev. payment (γ)	0.228
	(0.631)
Utility shift from beginning treatment (δ)	−4.909
	(0.098)
Penalty for violating ASRM embryo guidelines (η_0)	−3.036
	(0.194)

Note: Standard errors are in parentheses.

Table 5—Utility-Type Distribution Parameter Estimates

Constant (ρ_0)	−1.269
	(0.552)
Age (ρ_1)	0.041
	(0.013)
Wealth (ρ_2)	−0.864
	(0.457)
Insurance (ρ_3)	−0.017
	(0.367)
ASRM regime (ρ_4)	0.469
	(0.340)
Race (ρ_5)	−1.032
	(0.879)

Note: Standard errors are in parentheses.

of African Americans and 34.6 percent of others) indeed choose to become clinic patients and undergo at least one IVF cycle.

C. *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; all predictions match the data fairly well. Start/delay decisions, which are observed most frequently in the data (and are assisted by the intercept term δ)

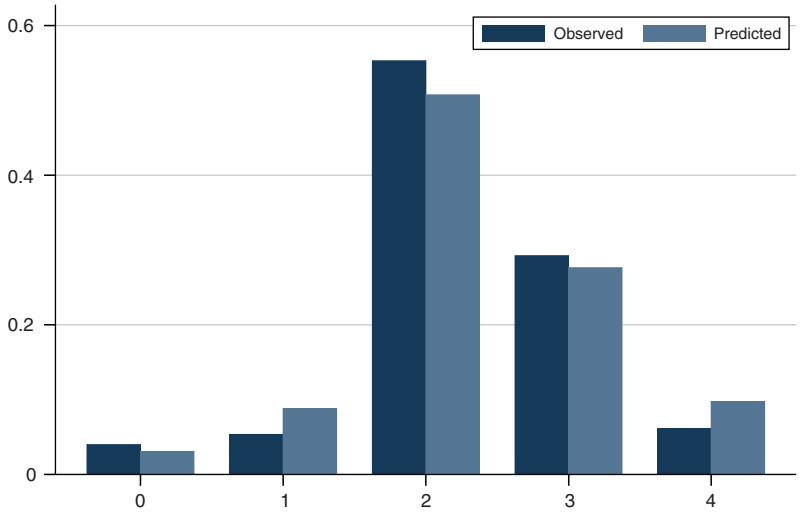

FIGURE 7. PREDICTED STAGE 4 DECISION ON EMBRYOS TRANSFERRED

have the tightest fit. In all periods after initiation, patients in the data choose "start" in 3.6 percent of all opportunities, while our estimated model predicts that patients choose "start" with frequency of 4.5 percent. Stage 2 and 3 predicted decisions also follow the data fairly closely. Some differences are to be expected, however, because these stages' fits depend on overall W_j values rather than individual parameters. We observe cancellation (Stage 2) and ICSI (Stage 3) frequencies of 14.4 percent and 39.7 percent, respectively, and our model provides choice probabilities of 13.5 percent and 54.1 percent. Figure 7 provides a comparison of distributions across Stage 4 choices. The estimated model succeeds in matching $x = 2$ as the most common choice in Stage 4, 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 $x = 4$).[33]

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 ($\mathbf{Z}$) as the patients in the data, but then random draws on medical outcomes and taste shocks determine choices and outcomes over time. For each patient we repeat the process ten times, allowing for the realization of different taste shocks and stochastic medical outcomes. We average over patients and their individual simulated histories in computing the statistics we report below.

We focus on two critical measures of effectiveness and efficiency of IVF treatment. First we ask, what proportion of patients eventually succeed in delivering at least one live birth through IVF, regardless of the number of attempted cycles required to do so? We find that 60 percent of our simulated patient histories include

[33] We have explored whether choices in a patient's second IVF cycle or beyond can be explained by outcomes during Stage 1–3 of her first IVF cycle. These variables have no significant impact on later-cycle choices. We view this result as evidence to support our assumption that the patient does not respond to health variables other than those in $\mathbf{Z}^B$ or revealed during the current cycle.

a birth, which is reasonably close to the empirical value of 53 percent reported on Table 2. Second, we investigate how many cycles an individual patient receives at the clinic. In our simulation, 49 percent of patients are observed taking a single cycle, 33 percent undergo two cycles, and 18 percent receive three or more cycles. These results compare well to the data, in which we see 47 percent, 28 percent, and 25 percent of patients receive one, two, or three or more cycles, respectively. Patients' responses to price variation, which come through differences in insurance status, also match well between the data and model. In the data, insured patients take an average of 1.97 cycles during their full treatment histories, while uninsured patients average 1.50 cycles. Our simulated patient histories contain an average of 1.99 cycles for insured patients and 1.56 for uninsured. Both the data and simulated histories contain only small differences in embryos transferred by insurance status; we return to the relationship between insurance and embryos below.

VIII. Counterfactual Experiments

We use the estimated model to consider a set of counterfactual policy experiments which analyze potential IVF patients' responses to changes in their decision environment (g). The policies vary in how they emphasize or highlight the key incentives and trade-offs in IVF. Three themes are central to all policies. First, with forward-looking patients, the full sequence of treatment choices can be affected by prices and restrictions at any point on the treatment path. This is especially important for prices or opportunities that might change several periods into the future, like the exhaustion of covered cycles under treatment-based insurance. Second, patients have a general incentive to be aggressive in treatment because they do not face the full costs of high-order births. Third, patients may have an incentive to take aggressive treatment in the current period in order to reduce the probability of paying for treatment in the future.

Extensive-margin choices are crucial for this analysis, so we employ the full at risk population of $N^{inf} = 2{,}781$ potential patients described above. While we abstract away from the impact of IVF policy on women's life-cycle choices over education, career, age at marriage, and age at first birth, these channels could affect the size and composition of the IVF patient pool.[34] The N^{inf} potential patients represent the portion of 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 percent of all cycles in US clinics.

For each potential patient, we draw age, wealth, insurance, race, 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 $W\left(\mathbf{Z}^{D}_{a_0}, \tau, g, \widehat{\varphi}, \widehat{\zeta}\right)$ for each

[34] Buckles (2005), Abramowitz (2014), and Gershoni and Low (2015) consider the impact of IVF policy on life-cycle choices. If these choices change with the counterfactual policies that we consider below, then our estimates of welfare impacts would change as well. We defer this issue to future research.

simulated woman. The values of $W\left(\mathbf{Z}^{D}_{a_0}, \tau, g, \widehat{\varphi}, \widehat{\zeta}\right)$ differ across policy experiments. We then allow potential patients to elect whether to begin treatment by comparing $W\left(\mathbf{Z}^{D}_{a_0}, \tau, g, \widehat{\varphi}, \widehat{\zeta}\right)$ to the population-wide utility parameters $\widehat{\mu}(z_r)$ and a simulated value for the potential patient's taste shock ν. For all potential patients, we simulate initiation choices and decision histories in the same way described above for evaluating model fit, including repeating the process ten times for each potential patient in N^{inf}. Potential patients who do not start treatment at a_0 exit the model forever.

We assume that the N^{inf} simulated potential patients arrive at the fertility decision uniformly over the 2001–2007 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 2001–2009, 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. 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 note that during the full sample period the clinic elected to keep its prices fixed at the same level. In fact, the typical IVF cycle price in the United States has remained approximately unchanged between the mid-1990s and the present, despite a substantial increase in the number of treated patients.

We report our main results in Tables 6–7 and Figures 8–9. The tables contain both point estimates of counterfactual outcomes and 95 percent confidence intervals calculated with the same bootstrap procedure described above.[35] The figures focus on embryo transfer choices and birth outcomes under selected policies in which these choices and outcomes are particularly interesting. 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; we refer to this as the "empirical baseline" and index it as g_E. We then consider a variety of insurance policies that provide patients with a specified number of treatments. We begin by simulating the market when no patients have treatment insurance (no insurance, g_N). We contrast this with a policy that provides Illinois-style insurance for four IVF cycles to all potential patients, without exceptions, in both states; this is "universal insurance for treatment" and is indexed as g_{IT}. We augment g_{IT} with three policies designed to reduce embryo transfers. An "IT + embryo cap" (g_C) policy limits all patients to transferring a single embryo. The policy "IT + actuarially fair top-up prices" (g_{AT}) allows patients to receive multiple embryos, but they bear the entire additional expected birth costs from multiple-embryo transfers. In the final experiment with treatment-focused insurance, we illustrate the implications of reduced but positive top-up prices ("IT + moderate top-up prices," g_{MT}). We contrast treatment-focused

[35] We draw 400 times from the sampling distribution of $\boldsymbol{\theta}$, and then estimate a new value of $\boldsymbol{\mu}$ for each draw. We use each pair $(\boldsymbol{\theta}, \boldsymbol{\mu})$ 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 $(\boldsymbol{\theta}, \boldsymbol{\mu})$ pairs) within a policy setting.

TABLE 6—INITIATIONS, CYCLES, AND OUTCOMES ACROSS POLICY SETTINGS

Policy setting (g)	Share initiating	*N* cycles if initiate	Share with birth	*N* births	*N* infants
Empirical baseline (E)	0.299	1.348	0.159	456.2	636.4
	(0.295, 0.305)	(1.280, 1.445)	(0.148, 0.173)	(422.1, 496.8)	(584.6, 691.1)
Panel A. Insured treatment policies					
No insurance (N)	0.242	1.277	0.127	360.7	505.2
	(0.227, 0.258)	(1.204, 1.396)	(0.113, 0.145)	(321.3, 413.3)	(446.0, 576.4)
Universal insurance for treatment (IT)	0.577	1.426	0.315	902	1,245.2
	(0.512, 0.625)	(1.365, 1.507)	(0.281, 0.347)	(808.0, 997.5)	(1,114.0, 1,373.0)
IT + embryo cap (C)	0.287	1.238	0.077	217.9	247.4
	(0.226, 0.342)	(1.127, 1.357)	(0.061, 0.093)	(172.7, 263.4)	(197.2, 300.6)
IT + actuarially fair top-up prices (AT)	0.323	1.271	0.113	320.5	397.7
	(0.280, 0.373)	(1.181, 1.396)	(0.100, 0.130)	(282.3, 371.6)	(349.0, 476.7)
IT + moderate top-up prices (MT)	0.449	1.356	0.206	586.7	778.8
	(0.407, 0.485)	(1.297, 1.461)	(0.183, 0.231)	(523.6, 656.1)	(688.6, 881.5)
Panel B. Insured outcome policies					
Universal insurance for outcomes (IO)	0.576	1.412	0.311	883.2	1,210.8
	(0.511, 0.625)	(1.353, 1.494)	(0.277, 0.343)	(788.9, 980.1)	(1081.9, 1,336.5)
Age-adjusted insurance for outcomes (IA)	0.484	1.4	0.278	787.6	1,087.9
	(0.442, 0.517)	(1.337, 1.486)	(0.251, 0.307)	(712.6, 869.7)	(979.2, 1,194.9)

Note: 95 percent confidence intervals are in parentheses.

TABLE 7—SURPLUS AND COSTS ACROSS POLICY SETTINGS

Policy setting (g)	Total surplus in risk population ($M)	Total IVF insurance cost ($M)	Total medical delivery cost ($M)	Medical cost per birth ($000)	Insurance + medical cost per birth ($000)
Empirical baseline (E)	10.3	3.6	31.4	68.9	76.9
	(8.3, 13.7)	(3.1, 4.0)	(28.6, 33.9)	(66.9, 69.5)	(74.4, 77.7)
Panel A. Insured treatment policies					
No insurance (N)	7.6	0	25.0	69.4	69.4
	(5.6, 10.8)	(0.0, 0.0)	(21.8, 28.5)	(67.0, 70.0)	(67.0, 70.0)
Universal insurance for treatment (IT)	18.4	16.4	60.5	67.1	85.2
	(16.0, 22.4)	(14.6, 18.0)	(54.4, 66.8)	(65.9, 68.2)	(84.4, 86.2)
IT + embryo cap (C)	5.6	7.3	8.5	38.9	72.4
	(4.4, 7.4)	(5.7, 8.8)	(6.8, 10.4)	(38.8, 39.7)	(71.8, 73.8)
IT + actuarially fair top-up prices (AT)	7.8	8.3	11.9	37.1	63.1
	(6.0, 11.4)	(7.3, 9.7)	(10.5, 13.8)	(36.0, 38.4)	(59.4, 66.2)
IT + moderate top-up prices (MT)	13.3	12.2	29.9	50.9	71.7
	(10.5, 17.9)	(11.1, 13.7)	(26.1, 34.3)	(48.4, 52.7)	(70.4, 72.7)
Panel B. Insured outcome policies					
Universal insurance for outcomes (IO)	18.5	16.2	58.3	66.1	84.4
	(16.0, 22.6)	(14.5, 18.0)	(52.4, 64.4)	(64.9, 67.2)	(83.6, 85.4)
Age-adjusted insurance for outcomes (IA)	15.7	12.0	52.7	66.9	82.1
	(13.3, 19.6)	(10.7, 13.3)	(46.8, 57.5)	(65.2, 67.7)	(80.5, 82.8)

Note: 95 percent confidence intervals are in parentheses.

insurance with outcome-focused insurance in two final policy simulations. First, we allow patients to receive as many insured treatments as they choose, provided they have one or fewer children when each cycle begins ("universal insurance for outcomes," g_{IO}). Second, we demonstrate the impact of dynamics and forward-looking behavior by simulating the impact of an outcome-based insurance policy in which

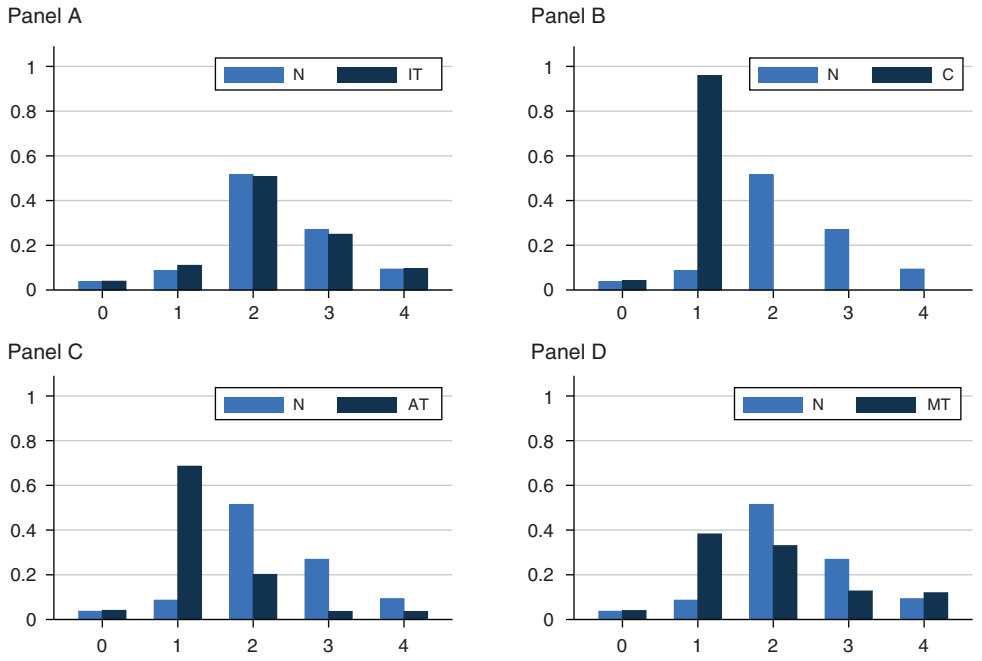

FIGURE 8. EMBRYO TRANSFER CHOICES IN COUNTERFACTUAL EXPERIMENTS

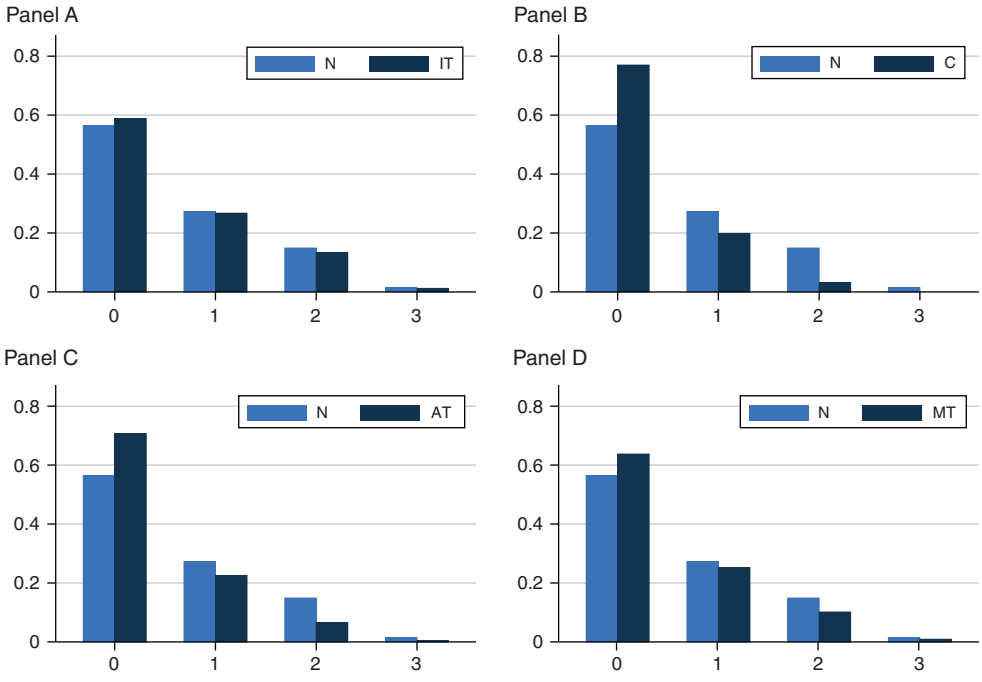

FIGURE 9. BIRTHS OUTCOMES IN COUNTERFACTUAL EXPERIMENTS

Notes: N = no insurance. IT = universal insurance for treatment. AT = insurance plus actuarially fair top-up prices. MT = insurance plus "moderate" top-up prices.

insurance benefits are reduced for older patients ("Age-adjusted insurance for outcomes," g_{IA}).

Before describing the individual policy experiments, we describe some of the outcomes that we calculate under the empirical distribution of insurance (g_E). Under the observed prices and constraints, we find that 29.9 percent of potential patients elect to begin treatment. This initiation rate, combined with success probabilities within the clinic, results in 15.9 percent of women in N^{inf} achieving at least one birth during the sample period. About 53 percent of simulated patients who begin IVF achieve a birth at some point in the treatment history; this matches the observed rate in the actual patient population. On average across 10 simulated histories under g_E, the baseline simulations average 456 births for the 2,781 patients in N^{inf}; these births deliver 636 infants to the population, implying an average number of infants per birth of 1.4.

For each patient who begins treatment we calculate $\Delta_i = W\left(\mathbf{Z}^D_{i,a_0}, \tau_i, g, \hat{\varphi}, \hat{\zeta}\right) - (\hat{\mu}(z_r) + \nu_i)$, which is a measure of the net utility gain from initiating IVF above the outside option. Patients who elect to forgo treatment receive $\Delta_i = 0$. We use $\hat{\alpha}_i$, our estimate of the disutility from payments for each patient i (which varies depending on i's wealth) to obtain a patient-specific dollar-valued surplus measure, $CS_i(g) = \Delta_i/\alpha_i$. Across all potential patients in N^{inf}, including those who do not initiate treatment, the total $CS(g_E) = \$10.3$ million, or about \$3,700 per person in the risk population (or \$12,400 conditional on initiation).[36] If insurers must pay the difference between insured patients' prices and the full price, the empirical baseline requires a total of \$3.6 million in payments from insurers to the clinic. Finally, we calculate the total medical costs of all pregnancies and births that occur under g_E, using the cost estimates from Lemos et al. (2013) discussed above.[37] Using these figures, the total pregnancy- and delivery-related medical cost of the baseline is \$31.4 million, or \$68,900 per birth. When insurance costs for treatment are included along with delivery costs, the average cost per birth is \$76,872 (or \$55,105 per child).

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 percent, we calculate that 6.4 percent fewer uninsured patients initiate treatment, implying an elasticity of -1.28 at the extensive margin. The same price increase has a slightly larger impact on the total number of uninsured cycles, which falls by 6.7 percent for an elasticity of -1.33. 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 percent

[36] Policies that have increased insurance costs (e.g., from universal insurance) will eventually pass these costs on to consumers in the form of higher premiums, and this could reduce welfare. Two factors, however, may moderate this welfare loss. First, the population of individuals purchasing health insurance policies is much larger that the risk population we focus on, so per-capita premium increases could be small. Second, insurance coverage of IVF would provide additional benefits to women who do not know whether they will join the risk pool. These issues are beyond the scope of our paper.

[37] We were unable to obtain actual cost data for individuals in our sample since deliveries may occur at any hospital chosen by the patient. Data from the National Inpatient Sample show that delivery costs in Missouri are similar to those nationwide, while Illinois is slightly above the national average.

increase in out-of-pocket expenses for insured patients results in 1 percent fewer insured patients initiating and a total reduction of 1.4 percent in insured cycles. While elasticities above -1 are inconsistent with profit maximization, the clinic may have different objectives than a traditional firm. These elasticities are comparable to others from the health care literature (e.g., Manning et al. 1987).

A. *Impact of Insured Treatment*

In our first set of counterfactual policies, we consider the impact of a substantial expansion of insurance in the market. To establish a benchmark, we begin by simulating treatment choices and outcomes when no patients have insurance (g_N). In this setting, we find that 24.2 percent of the at-risk population chooses to initiate IVF, and 12.7 percent of the at-risk population conclude treatment with one or more births. The availability of IVF generates \$7.6M in surplus for patients, who initiate treatment only if expected surplus is positive. While there are no insurance costs of treatment in this setting, patients who become pregnant and give birth generate \$25M in medical costs, which amount to \$69,414 per birth.

When Illinois-style insurance is extended to all potential patients (g_{IT}), the number of treated women and births increase substantially. We find that 57.7 percent of potential patients in the simulated risk pool initiate treatment, and 31.5 percent in this population experience a birth during the sample period.[38] The presence of insurance has an impact on both initiation and the choice whether to continue treatment after a failure. While patients without insurance (in g_N) take an average of 1.24 cycles over their treatment histories, insured patients receive 1.43 IVF cycles on average. Despite a reduction in the price of treatment, the distribution of embryos transferred is very similar under g_N and g_{IT} (Figure 8). Likewise, the distribution of births (Figure 9), shows little difference between g_N and g_{IT}. 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 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.[39]

The patient surplus benefits of Illinois-style insurance are substantial, with total $CS(g_{IT}) = \$18.4\text{M}$, which is \$10.8M greater than under g_N. To evaluate the full impact of g_{IT}, however, we must account for additional costs due to insurance payments and medical delivery costs. As reported in Table 7, the insurance costs of g_{IT} are substantial, at \$16.4M. The difference between the changes in consumer surplus and insurance costs is to be expected considering the traditional medical-demand

[38] If, contrary to our assumptions, potential patients have some information about $\mathbf{Z}^B$ prior to initiating treatment, we expect that this could affect the magnitudes of our policy predictions. For example, if prior information about $\mathbf{Z}^B$ implies that patients who select into treatment have better fertility characteristics than the average woman who experiences infertility, then policy changes which reduce IVF's price would have a smaller impact than we predict. On the other hand, if patients choose how quickly to move from lower-tech infertility treatments to IVF, then a price-reducing policy change might result in relatively more new patients with favorable fertility characteristics.

[39] By contrast, Hamilton and McManus (2012) 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.

"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 in supplemental analysis that increases in access and surplus are greater for patients from lower wealth areas ($z_w = 0$). Overall, insurance and delivery costs per birth are \$85,213 with policy g_{IT}, although the difference between this figure and the no-insurance value is almost entirely due to the introduction of insured treatment costs. Delivery costs per birth are largely unchanged because the distribution of birth outcomes is unchanged.

Embryo Caps.—We next explore the impact of restricting patients to transfer only a single embryo during treatment. We do this in the context of universal insurance so that our setting is close to the implementation of embryo caps as they have been introduced in practice, i.e., in countries with generous health insurance. To simulate the policy g_C, we solve the model again at the estimated parameters but we impose the restriction $x \leq 1$ instead of $x \leq X$ in Stage 4 (the embryo transfer stage). We also remove the utility penalty for single-embryo transfers for circumstances when these conflict with ASRM guidelines. We then use the new policy functions together with the same history of ε and medical technology shocks to simulate counterfactual patient histories under the one-embryo cap.

The restriction on embryo transfers reduces the expected value of IVF for patients considering treatment. The share of N^{inf} who initiate is 28.7 percent, or about half the initiation rate with g_{IT}. The cap has a large mechanical effect on the distribution of embryos transferred (see Figure 8), which in turn yields a substantial shift in the distribution of births (Figure 9). As a consequence, the share of patients with births (7.7 percent) declines more steeply than the initiation rate. The estimated Stage 4 function $f_k(k \mid x, \mathbf{Z})$ 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.

Despite an initiation rate that is greater than that of g_N, the embryo cap policy generates less consumer surplus (\$5.6M) than the no-insurance scenario. Total insurance and delivery costs are substantially reduced in magnitude, to \$7.3M and \$8.4M respectively, but important differences exist between measures of costs per birth. Medical costs per birth fall by almost \$30,000 relative to g_N and g_{IT}, down to \$38,900, due to the reduction in the numbers of twins and triplets. Patients require more IVF cycles to achieve a birth, however, so insurance expenses along with medical costs results in total per-birth costs equal to \$72,400, which is greater than the average in g_N.

Top-Up Prices.—Patients receive substantial value from the opportunity to transfer two or more embryos, which suggests that there may exist beneficial policies that allow patients to trade-off between the benefits and costs associated with transferring multiple embryos. The medical cost of multiple births 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_{IT}) in which too much risky behavior occurs in equilibrium.

We construct top-up prices that are paid when a patient transfers two or more embryos. Let c_k be the average medical cost of a delivering k infants, and let $\overline{\mathbf{Z}}$ denote the state variable values for a median-age treated patient with no additional fertility problems. We construct an (approximately) actuarially fair top-up price for transferring $x > 1$ embryos as

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

This expression acknowledges that there is some multiple birth risk for patients transferring a single embryo, $f_k(k|1, \overline{\mathbf{Z}})$, 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.[40] 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 the utility penalty (η_0) for any transfer outside of ASRM guidelines.

This top-up price policy (g_{AT}) generates greater total consumer surplus (\$7.8M) than g_N or g_C, which leads to an IVF initiation rate that is greater as well. In addition, despite positive insurance cost of \$8.3M under g_{AT}, the medical cost of births decreases by \$13.1M relative to g_N. Together, these results imply that a policy of insured treatment for all patients, but with actuarially fair top-up prices, is welfare-improving for both consumers and the insurance firms that pay for IVF and birth costs. Reduced total costs for insurers could lead to lower premiums for patients in a competitive insurance market. The policy g_{AT} has a slightly lower share of potential patients with a birth than g_N (11.3 percent versus 12.7 percent), but this is accompanied by a multiple-birth rate that declines substantially. The average number of infants per delivery is 1.24 under g_{AT}, which is about halfway between the rates of g_N and g_C. The total cost per birth falls to \$63,350 with g_N, which is the lowest value of all policies we consider in this paper. While both g_C and g_{AT} have per-birth medical costs of about \$37,000, the greater per-treatment success rate of g_{AT} drives total per-birth costs below those of g_C.

While g_{AT} has several attractive properties relative to g_N and g_C, the actuarially fair top-up prices can lead some patients to pay total prices that are substantially greater than under g_N. It is possible to scale-down g_{AT}'s top-up prices to achieve a variety of policy objectives. For example, the top-up prices could be set so that total consumer surplus is equal to the empirical status quo, or they could be set so that total IVF and medical costs with top-up prices are equal to their level under g_N. We conclude our consideration of top-up prices with an additional example: prices that

[40] For example, a patient with $\mathbf{Z} = \overline{\mathbf{Z}}$ who transfers two 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.

prevent any patient from paying a greater price for an IVF cycle than she would under g_N. We scale down the actuarially fair prices (while keeping their proportions fixed) so that patients pay \$5,000 to transfer a second embryo and \$8,000 for three or four embryos. These payments are on top of the insured treatment prices, which total \$3,000 for a full cycle with ICSI. Under g_N a patient who uses ICSI and receives four embryos also pays \$11,000. This policy of "moderate" top-up prices (g_{MT}) leads to 85 percent more patients initiating IVF than under g_N, and 62 percent more patients who ever experience a birth during the sample period. While total insurance plus medical costs are 68 percent greater with g_{MT} relative to g_N, the per-birth medical plus insurance cost is only 3.3 percent greater.

B. *Insuring Outcomes*

To highlight the scope of potential IVF policies and the impact of dynamics in this procedure, we study two outcome-focused insurance policies. In policy g_{IO}, we again expand insurance to the full population, but now with unlimited attempts at IVF for patients with one or fewer children when beginning a treatment cycle. In general, patients prefer an outcome-based policy that allows unlimited tries relative to treatment-based policies like g_{IT} that insure only a limited number of cycles. A dynamic model is needed to compare the welfare benefits of g_{IO} and g_{IT}, given their specific details. From the patient's perspective at the beginning of treatment, the entire surplus difference depends on how insurance coverage changes *after* the first cycle. From a cost perspective, the key trade-offs are between delivery costs (which are large when twins and triplets occur) and treatment costs (which are large when many attempts are needed until success). Insured-outcome policies might lead to more attempts with less chance of twins and triplets because forward-looking patients can be more conservative in their embryo transfers. Insured-treatment policies may drive up delivery costs as forward-looking patients try to succeed in the early (covered) attempts by being more aggressive in their treatments.

In the second outcome-focused policy we add age-based eligibility conditions, which are motivated by age-related fertility differences, as younger patients are typically more likely to succeed than older ones.[41] Patients under g_{IA} who begin a cycle younger than age 35 receive the full benefit of g_{IO}, while patients with ages between 35 and 40 pay prices halfway between the uninsured price and the insured one. (This amounts to a discount of about 35 percent.) Patients age 40 and above are completely excluded from insurance. This policy has an impact on dynamic decision-making because it becomes less generous when the patient crosses certain age thresholds, therefore encouraging patients to try to succeed before aging-out of the policy benefits. If patients accelerate their treatment attempts, the age-dependent policy may generate some cost savings as the attempts are conducted when the patient is younger and more fertile. Therefore, success is more likely to be achieved

[41] There are some similarities between our age-related policy experiment and the dynamic public finance literature's treatment of age-dependent tax policies (e.g., Weinzierl 2011). Just as productivity differences by age imply that age-related tax policies can be optimal, age-related fertility differences suggest opportunities for policy variation along this dimension.

in fewer attempts. On the other hand, incentivizing patients to succeed before they cross the age-based eligibility thresholds could increase costs if patients become more aggressive. In addition, patients who speed up treatment experience a utility penalty because they may deviate from their preferred treatment spacing across cycles.

While policies tied to cumulative patient outcomes and ages may place some record-keeping demands on insurers (especially if an individual changes providers), this is no more complicated than current Illinois policy, which requires tracking the cumulative number of cycles.

The treatment-focused policy g_{IT} and the outcome-focused policy g_{IO} have virtually identical effects on patient initiation, the probability of one or more births per patient, patient surplus, and insurance costs. The strong similarity between the two policies occurs because very few patients under g_{IO} choose to take five or more cycles. In the eyes of most patients, the two policies are similarly generous. This outcome was not guaranteed, and indeed if g_{IT} covers fewer than four cycles this treatment-based policy provides significantly less patient surplus than g_{IO}.

As expected, the policy with age-adjusted benefits, g_{IA}, has lower overall benefit to patients. Although older patients are most sharply affected by these limits, the differences in surplus and costs between g_{IO} and g_{IA} are not especially large. The older patients who would start treatment under g_{IO} but not g_{IA} are also the patients most likely to fail, so the share of patients with a birth falls by about 16 percent while insurance costs for treatment decline by 26 percent. Overall, treatment and birth costs fall by more than patient surplus.

Comparisons across patient ages offer the best opportunity to examine differences between g_{IO} and g_{IA}. In Figure 10 we graph the differences in expected value by age from initiating IVF for the first time. The largest difference in value is between ages 35 and 40, when insurance's price benefit is reduced but not zero. Although the total reduction in price benefit is greater for patients above age 40, the reduced likelihood of a birth at these ages dampens the value difference between the policies. It is interesting to note that patients younger than age 35 also have a lower expected value of initiating treatment under g_{IA}. This is due to cycles after age 35 being subject to the reduced price benefit. Note that if patients were myopic, the two policies would have the same value for patients initiating before age 35. The dynamic effects of g_{IA} are also apparent in simulated treatment histories for patients who initiate treatment between ages 33 and 35. Among patients who take more than one cycle in total, the average number of quarters between treatments is 6.0 under g_{IO}, but the patients under g_{IA} delay 5.3 quarters on average. A similar reduction in delay occurs for patients who initiate between ages 35 and 40.

IX. Conclusions

Health care 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 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

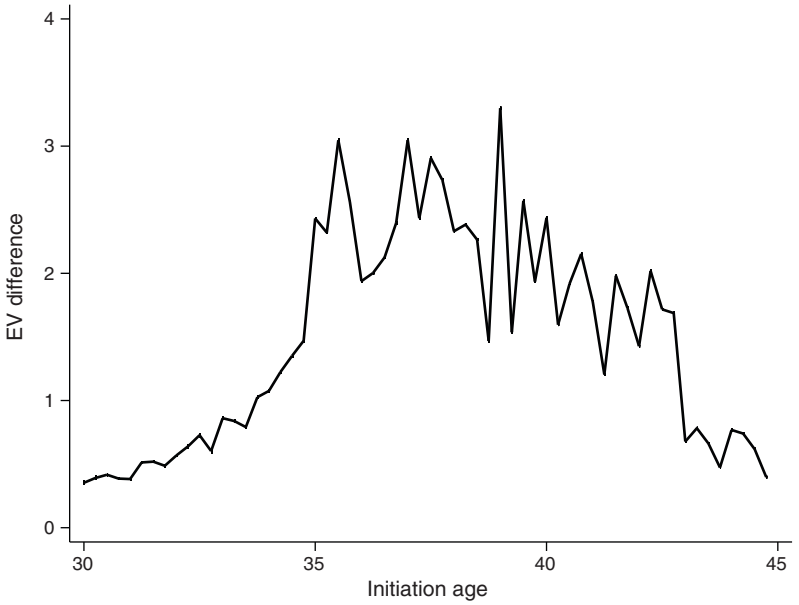

Figure 10. Difference in Patient Expected Value at Initiation between IO and IA

Note: The vertical axis plots the difference between expected value from policy IO (universal insured outcome) minus IA (insured outcome with age-adjusted benefits).

of embryo caps. To investigate the impact of these and other policies on patient welfare, outcomes, and health care costs, we estimate a structural 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 trade-off 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, MO, 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 IVF use. However, the mandate also increases both treatment and birth costs. We find similar results for an outcome-based policy that insures births as opposed to treatment. 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 (simultaneous with insured treatment) does indeed substantially lower aggregate costs, but this reduction is largely due to a reduction in patients who initiate treatment. Due to low

birth probabilities with an embryo cap, the total cost per birth is greater than when no patients have insurance. 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 "value-based" policy in the spirit of Baicker, Shepard, and Skinner (2013) and Einav, Finkelstein, and Williams (2016) 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. This policy generates more patient surplus than the embryo cap or the no-insurance benchmark, and insurance plus medical costs are lower than the no-insurance benchmark as well. 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.

Our study of IVF has several contributions for other areas of health economics. We emphasize the importance of how future treatment opportunities affect today's choices, which is also important in several diseases with long-term treatment strategies, such as heart disease and cancer. Our consideration of top-up prices is relevant for insurance policy in areas where patients may have discretion over treatment approach, which can be particularly important in areas where compromises must be struck between patients' utility and other medical concerns or costs. Finally, our framework illustrates the benefits of using a welfare approach to evaluate policies which may be different in structure (e.g., treatment- or outcome-based insurance) or different in goals (e.g., patients' surplus or number of births).

The favorable outcomes of the top-up price policies naturally leads to a question of why these contracts have not yet emerged in the real world. To understand this, it is important to distinguish between plans offered by insurers under a mandate, versus offerings by insurers not subject to a mandate. In a setting without mandates, adverse selection and unraveling could result in individual insurers being unwilling to offer unilaterally any plans with IVF coverage. Our results show that, in principle, insurers could be better off under a mandate which allows them to coordinate offering coverage for IVF access while encouraging less aggressive embryo transfer behavior through embryo caps or top-up prices.

Appendix A. Distribution of Types among Potential Patients

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

$$\Pr\left(\tau = 2 \middle| \mathbf{Z}_{a_0}^D, I = 1, \hat{\boldsymbol{\rho}}\right) = \Lambda([1, \mathbf{Z}_{a_0}^D]\hat{\boldsymbol{\rho}}). \tag{A1}$$

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

$$\Lambda\left(W\left(\mathbf{Z}_{a_0}^D, \tau, \hat{\boldsymbol{\varphi}}, \hat{\zeta}\right) - \mu(z_r)\right) \quad \text{for } \tau = 1, 2 \text{ and } z_r = 0, 1. \tag{A2}$$

For each $\mathbf{Z}^D_{a_0}$ we also know the total (i.e., unconditional on type) number of women of each race with other demographic characteristics $\mathbf{Z}^D_{a_0}$ who came into the clinic. Let this number be $N^{clin}_{\mathbf{Z}^D_{a_0}}(z_r)$. Together with $\Pr\big(\tau = 1|\mathbf{Z}^D_{a_0}, I_i = 1, \widehat{\boldsymbol{\rho}}\big)$ we then have an estimate of the race-specific number of patients of type 1 with characteristics $\mathbf{Z}^D_{a_0}$ who came into the clinic, say $N^{clin}_{\mathbf{Z}^D_{a_0},\,1,\,z_r}\big(\widehat{\boldsymbol{\rho}}\big)$, where

$$N^{clin}_{\mathbf{Z}^D_{a_0},\,1,\,z_r}\big(\hat{\boldsymbol{\rho}}\big) \;=\; N^{clin}_{\mathbf{Z}^D_{a_0}}(z_r) \times \Big[1 - f_\tau\big(\tau \;=\; 2\,|\,\mathbf{Z}^D_{a_0}, I = 1, \hat{\boldsymbol{\rho}}\big)\Big]. \tag{A3}$$

Similarly for type 2, we know

$$N^{clin}_{\mathbf{Z}^D_{a_0},\,2,\,z_r}\big(\hat{\boldsymbol{\rho}}\big) \;=\; N^{clin}_{\mathbf{Z}^D_{a_0}}(z_r) \;\times f_\tau\big(\tau = 2\,|\,\mathbf{Z}^D_{a_0}, I \;=\; 1, \hat{\boldsymbol{\rho}}\big). \tag{A4}$$

Note that while $N^{clin}_{\mathbf{Z}^D_{a_0}}(z_r)$ is data, $\Big[N^{clin}_{\mathbf{Z}^D_{a_0},\,1,\,z_r}\big(\widehat{\boldsymbol{\rho}}\big), N^{clin}_{\mathbf{Z}^D_{a_0},\,2,\,z_r}\big(\widehat{\boldsymbol{\rho}}\big)\Big]$ depend on $\boldsymbol{\rho}$, which is identified by the differential behavior of the two types in the (within-clinic) patient histories. Recall that $\boldsymbol{\rho}$ parameterizes the within-clinic distribution of types and is estimated in our second step with (φ, ζ).

Given $\mu(z_r)$, from the initiation model we know that $100 \times \Lambda\Big(W\big(\mathbf{Z}^D_{a_0}, \tau, \hat{\varphi}, \hat{\zeta}\big) - \mu(z_r)\Big)$ percent of potential patients of a given race, with additional initial nonbiological states $\mathbf{Z}^D_{a_0}$ and type τ will choose to initiate treatment. We also know that there we predict $N^{clin}_{\mathbf{Z}^D_{a_0},\,\tau,\,z_r}\big(\hat{\boldsymbol{\rho}}\big)$ patients of type τ. Then, it must be the case that the number of potential patients of each type, with that combination of demographic states $\mathbf{Z}^D_{a_0}$ and for that *race* is given by

$$\begin{aligned} N^{inf}_{\mathbf{Z}^D_{a_0},\,1,\,z_r} \;&=\; \frac{N^{clin}_{\mathbf{Z}^D_{a_0},\,1,\,z_r}\big(\widehat{\boldsymbol{\rho}}\big)}{\Lambda\Big(W\big(\mathbf{Z}^D_{a_0}, \tau \;=\; 1, \widehat{\varphi}\big) - \mu(z_r)\Big)} \\ &=\; \frac{N^{clin}_{\mathbf{Z}^D_{a_0}}(z_r)\left[1 - \Lambda\big([\,1, \mathbf{Z}^D_{a_0}]\hat{\boldsymbol{\rho}}\big)\right]}{\Lambda\Big(W\big(\mathbf{Z}^D_{a_0}, \tau \;=\; 1, \widehat{\varphi}\big) - \mu(z_r)\Big)}, \end{aligned} \tag{A5}$$

$$\begin{aligned} N^{inf}_{\mathbf{Z}^D_{a_0},\,2,\,z_r} \;&=\; \frac{N^{clin}_{\mathbf{Z}^D_{a_0},\,2,\,z_r}\big(\hat{\boldsymbol{\rho}}\big)}{\Lambda\Big(W\big(\mathbf{Z}^D_{a_0}, \tau = 2, \hat{\varphi}, \hat{\zeta}\big) - \mu(z_r)\Big)} \\ &=\; \frac{N^{clin}_{\mathbf{Z}^D_{a_0}}(z_r)\,\Lambda\big([\,1, \mathbf{Z}^D_{a_0}, z_r]\hat{\boldsymbol{\rho}}\big)}{\Lambda\Big(W\big(\mathbf{Z}^D_{a_0}, \tau \;=\; 2, \hat{\varphi}, \hat{\zeta}\big) - \mu(z_r)\Big)}. \end{aligned} \tag{A6}$$

Then we can estimate the unconditional (i.e., not conditional on $I = 1$) prevalence of type 2 among potential patients of that race with state $\mathbf{Z}^D_{a_0}$ as

$$f_{\tau|z_r}\big(\tau = 2|\mathbf{Z}^D_{a_0}\big) \approx \frac{N^{inf}_{\mathbf{Z}^D_{a_0},\,2,\,z_r}}{N^{inf}_{\mathbf{Z}^D_{a_0},\,1,\,z_r} + N^{inf}_{\mathbf{Z}^D_{a_0},\,2,\,z_r}} = \left(1 + \left[\frac{\dfrac{1 - \Lambda\big([\,1, \mathbf{Z}^D_{a_0}]\hat{\boldsymbol{\rho}}\big)}{\Lambda\Big(W\big(\mathbf{Z}^D_{a_0}, \tau = 1, \hat{\varphi}\big) - \mu(z_r)\Big)}}{\dfrac{\Lambda\big([\,1, \mathbf{Z}^D_{a_0}]\hat{\boldsymbol{\rho}}\big)}{\Lambda\Big(W\big(\mathbf{Z}^D_{a_0}, \tau = 2, \hat{\varphi}, \hat{\zeta}\big) - \mu(z_r)\Big)}}\right]\right)^{-1}. \tag{A7}$$

Note that for given $\mu(z_r)$, everything in the RHS is known, so $f_{\tau|z_r}\left(\tau = 2|\,\mathbf{Z}^D_{a_0}\right)$ is known and so is $f_{\tau|z_r}\left(\tau = 1|\mathbf{Z}^D_{a_0}\right) = 1 - f_{\tau|z_r}\left(\tau = 2|\,\mathbf{Z}^D_{a_0}, z_r\right)$. If, for a given race, both types were to select into the clinic at the same rate (i.e., they do not really have different preferences for children), we would have $W\left(\mathbf{Z}^D_{a_0}, \tau = 1, g_E, \boldsymbol{\theta}\right)$ $= W\left(\mathbf{Z}^D_{a_0}, \tau = 2, g_E, \boldsymbol{\theta}\right)$ so $\dfrac{\Lambda\left(W\left(\mathbf{Z}^D_{a_0}, \tau = 1, \boldsymbol{\theta}\right) - \mu(z_r)\right)}{\Lambda\left(W\left(\mathbf{Z}^D_{a_0}, \tau = 2, \boldsymbol{\theta}\right) - \mu(z_r)\right)} = 1$ and the distribution of types within the clinic and among potential patients would be the same, $f_{\tau|z_r}\left(\tau = 2|\,\mathbf{Z}^D_{a_0}, I = 1\right) = f_{\tau|z_r}\left(\tau = 2|\,\mathbf{Z}^D_{a_0}\right)$, which is not consistent with our estimates.

Appendix B. Approximating N^{inf} and $f_{\mathbf{Z}^D}\left(\mathbf{Z}^D_{a_0}\right)$

To obtain a model-predicted IVF initiation rate among potential patients of each race we must use an estimate of $f_{\mathbf{Z}^D|z_r}\left(\mathbf{Z}^D_{a_0}\right)$. Note that since the expected value of initiation depends on $\mathbf{Z}^D_{a_0}$, the distribution of $\mathbf{Z}^D_{a_0}$ among patients who initiate differs from the distribution among potential patients. In particular, we expect active patients at the clinic to be older, more likely to be covered by insurance, wealthier, and less likely to be African American. To approximate $f_{\mathbf{Z}^D}\left(\mathbf{Z}^D_{a_0}\right)$ among potential patients we use the following assumptions.

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

$$asrm \perp \left(\mathbf{Z}^B, \iota_{a_0}, z_w, a_0, z_r\right). \tag{A8}$$

ASSUMPTION 1 (Conditional Independence): *Conditional on age at initiation, the three biological state variables related to infertility, $\mathbf{Z}^B$, are independent of insurance, wealth, and race:*

$$\mathbf{Z}^B \perp \left(\iota_0, z_w, z_r\right)|\,a_0. \tag{A9}$$

Note that Assumption 1 and the fact that the value of $\mathbf{Z}^B$ only becomes observable after deciding to start a first cycle, imply that these variables will have the same conditional (on age) distribution in the risk set and in the clinic:

$$f\left(\mathbf{Z}^B|\,a_0\right) = f\left(\mathbf{Z}^B|\,a_0, I = 1\right). \tag{A10}$$

ASSUMPTION 2 (Surprise): *Individuals only become aware of their infertility at age a_0. Therefore, the joint distribution of wealth and insurance coverage among women at a_0 should be independent of whether they have any infertility problem (i.e., independent of whether they are among the potential patients). Therefore, $\Pr\left(\iota_{a_0}, z_w, z_r|\,a_0\right)$ is the same regardless of whether a woman is a potential patient. We further assume that $\Pr\left(\iota_{a_0}, z_w, z_r|\,a_0\right)$ $= \Pr\left(\iota_{a_0}, z_w, z_r\right)$ for all a_0.*

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

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

$$f_{\mathbf{Z}^D}(\mathbf{Z}^D_{a_0}) = f_{a_0, \iota_{a_0}, z_w, z_r}(a_0, \iota_{a_0}, z_w, z_r) f_{asrm}(z_{asrm_{a_0}}). \tag{A11}$$

To estimate $f_{a_0, \iota_{a_0}, z_w, z_r}(a_0, \iota_{a_0}, z_w, z_r)$ we note that

$$\begin{aligned} f_{a_0, \iota_{a_0}, z_w, z_r}(a_0, \iota_{a_0}, z_w, z_r) &= f_{a_0, \iota_{a_0}, z_w \mid z_r}(a_0, \iota_{a_0}, z_w \mid z_r) f_{z_r}(z_r) \\ &= f_{a_0 \mid z_r}(a_0 \mid z_r) f_{\iota_{a_0}, z_w \mid z_r}(\iota_{a_0}, z_w \mid z_r) f_{z_r}(z_r). \end{aligned}$$

Distribution of Age Among Potential Patients: We first estimate $f_{a_0 \mid z_r}(a_0 \mid z_r)$ using data from the St. Louis region on (first) births and the maternal age associated with those births, by race. Also because of Assumption 2, this gives us the distribution of age at first *attempted* birth (regardless of whether the attempt was successful) when combined with estimates of race-specific infertility rates by age. This provides the race-specific age distribution for our potential patients.

Joint Distribution of IVF Coverage and Wealth: Finally we collect data on the joint distribution of IVF coverage and wealth, (ι, z_w) conditional on race. To estimate $f_{\iota_{a_0}, z_w \mid z_r}(\iota_{a_0}, z_w \mid z_r)$ we consider

$$\begin{aligned} f_{\iota_{a_0}, z_w \mid z_r}(\iota_{a_0}, z_w \mid z_r) &= f_{\iota_{a_0} \mid z_w, z_r}(\iota_{a_0} \mid z_w, z_r) f_{z_w \mid z_r}(z_w \mid z_r) \\ &= f_{\iota_{a_0} \mid z_w}(\iota_{a_0} \mid z_w) f_{z_w \mid z_r}(z_w \mid z_r), \end{aligned}$$

and develop a strategy at the zip code level for estimating $f_{\iota}(\iota \mid z_w)$ and $f_{z_w \mid z_r}(z_w \mid z_r)$ using information from zip codes whose center is located within 75 miles from our clinic. To estimate $f_{z_w \mid z_r}(z_w \mid z_r)$ we assume patients from same zip code are homogeneous regarding (ι, z_w). In particular, we know whether each zip code in the St. Louis area is considered "wealthy" by construction: we defined $z_{w,l} = 1$ if zip code 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). We estimate $f_{z_w \mid z_r}(z_w \mid z_r)$ by

$$f_{z_w \mid z_r}(z_w = 1 \mid z_r) = \sum_{l \in STL} I\{z_{w,l} = 1\} \left\{ \frac{P_l^{z_r}}{\sum_{m \in STL} P_m^{z_r}} \right\}, \tag{A12}$$

where $P_l^{z_r}$ is the population of a given *race* (z_r) in zip code l within the St. Louis region. We have the population by race for each zip code, so we can construct $P_l^{z_r}$ easily.

To estimate $\Pr(\iota|w)$ we take the following steps. We have the percentage of the population who has 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) five-year estimate. A 2005 Mercer Survey of Employer Health Insurance Plans (Mercer Health Resources Consulting 2006) reported that 19 percent of large employers (500+ employees) and 11 percent of all firms (25+ employees) offered some form of IVF coverage as part of their health plans. We assume these same rates apply to Missouri zip codes. We then use data on the national size distribution of firms and employment in 2005 from the US Census Bureau's Statistics of US Business (SUSB) showing that among firms with 20+ employees, 60.7 percent of employees work for large firms, which account for only 2.87 percent of all US firms. Therefore, we use the adjustment factor $\psi^{MO} = 0.15$ to adjust the raw insurance coverage rates we obtain from the ACS. Then the IVF coverage for each Missouri zip code l is given by $f_{IVF}(\iota_l^{IVF}) = f_{priv}(priv_l) \times \psi^{MO}$.

Regarding Illinois counties, we know that there is a mandate. But small employers (<25 employees) and self-insured employers (regardless of size) are excluded.[42] According to a Kaiser Family Foundation (2007) report, 55 percent of workers nationally are covered by plans that are partially or fully self-insured.[43] 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 fewer 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 $f_{IVF}(\iota_l^{IVF}) = f_{priv}(priv_l) \times \psi^{IL}$.

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

$$\Pr(\iota = 4|\, z_w = 0) = \sum_{l:w_l=1} f_{IVF}(\iota_l^{IVF}) \left(\frac{\pi_l}{\sum_{l:w_l=1} \pi_l} \right). \tag{A13}$$

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

$$\hat{\Delta} = \left(\frac{\sum_{l:w_l=1,\, l \in IL} \pi_l}{\sum_{l:w_l=1} \pi_l} \right) \hat{\Delta}_{IL} + \left(\frac{\sum_{l:w_l=1,\, l \in MO} \pi_l}{\sum_{l:w_l=1} \pi_l} \right) \hat{\Delta}_{MO}. \tag{A14}$$

[42] Under this alternative definition of small employer, we interpolate the numbers in Moscarini and Postel-Vinay (2012) and find that 21.5 percent of employment is accounted for by firms with fewer than 25 employees.

[43] We assume that in these self-funded plans the same rate found in the Mercer survey (19 percent) 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.

Note, $\hat{\Delta}_s$ 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:

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

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

Size of Potential Patient Pool: In addition to the joint distribution of characteristics among potential patients, we need the size of the potential patient pool N^{inf}. We use $\tilde{N}^{inf}$ 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 race-specific 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 $\bar{N}_{a,z_r}$, which we obtain from Vital Statistics. The total number of women who *attempt* their first pregnancy at age a is N^{stl}_{a,z_r}. Of these, $\bar{N}_{a,z_r}$ succeed and have births recorded in Vital Statistics; the group that fails, $\tilde{N}^{inf}_{a,z_r}$, contributes to our set of potential patients. Therefore, $N^{stl}_{a,z_r} = \bar{N}_{a,z_r} + \tilde{N}^{inf}_{a,z_r}$. Then, using infertility rates by age and race among women who are attempting to get pregnant, $inf(a,z_r)$, we back out $\tilde{N}^{inf}_{a,z_r} = \dfrac{inf(a,z_r)}{\left[1 - inf(a,z_r)\right]}\bar{N}_a$. According to Vital Statistics, the larger counties in and around the St. Louis region have an average of 1,172 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 percent of births come from these counties.[44] So $\bar{N}^{75m}_{28-44} = 1{,}172 \times 1.104 = 1{,}294$. Using infertility rates by age and race and summing across ages, we can then determine that there are $\tilde{N}^{inf}_t = 257 \times 1.104 = 284$ new potential patients, on average, each quarter.[45] Since there are 28 quarters between 2001 and 2007, the size of the potential patient pool for our sample period is then $\tilde{N}^{inf}_{2001-07} = 28 \times \tilde{N}^{inf} = 28 \times 284 = 7{,}944$. 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 one-third, and we adjust $\tilde{N}^{inf}_{2001-07}$ in a proportional way. Ultimately the potential-patient population for our clinic is $N^{inf} = 2{,}781$. Of these, 486 are black and 2,295 are non-black.

[44] 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 zip codes 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 zip codes.

[45] To obtain age-specific infertility rates $inf(a,z_r)$, we interpolate between ages 28 to 39 and extrapolate for ages 40–44 the 12-month infertility estimates reported in Dunson, Baird, and Colombo (2004) and adjust these estimates using relative rates of infertility between black and white women from based on data from NSFG.

REFERENCES

Abbring, Jaap H., and James J. Heckman. 2007. "Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation." In *Handbook of Econometrics*, Vol. 6B, edited by James Heckman and Edward Leamer, 5145–303. Amsterdam: Elsevier.

Abramowitz, Joelle. 2014. "Turning Back the Ticking Clock: The Effect of Increased Affordability of Assisted Reproduction Technology on Women's Marriage Timing." *Journal of Population Economics* 27 (2): 603–33.

Aron-Dine, Aviva, Liran Einav, Amy Finkelstein, and Mark Cullen. 2015. "Moral Hazard in Health Insurance: Do Dynamic Incentives Matter?" *Review of Economics and Statistics* 97 (4): 725–41.

Baicker, Katherine, Mark Shepard, and Jonathan Skinner. 2013. "Public Financing of the Medicare Program Will Make Its Uniform Structure Increasingly Costly to Sustain." *Health Affairs* 32 (5): 882–90.

Bitler, Marianne P. 2008. "Effects of Increased Access to Infertility Treatment on Infant Health Outcomes: Evidence from Twin Births." Unpublished.

Bitler, Marianne P., and Lucie Schmidt. 2006. "Health Disparities and Infertility: Impacts of State-Level Insurance Mandates." *Fertility and Sterility* 85 (4): 858–65.

Bitler, Marianne P., and Lucie Schmidt. 2012. "Utilization of Infertility Treatments: The Effects of Insurance Mandates." *Demography* 49 (1): 125–49.

Buckles, Kasey S. 2005. "Stopping the Biological Clock: Infertility and the Career-Family Trade Off." Unpublished.

Buckles, Kasey S. 2013. "Infertility Insurance Mandates and Multiple Births." *Health Economics* 22 (7): 775–89.

Bundorf, Kate M., Melinda Henne, and Laurence Becker. 2008. "Mandated Health Insurance Benefits and the Utilization of Infertility Treatments." NBER Working Paper 12820.

Centers for Disease Control and Prevention. 2014. *Assisted Reproductive Technology National Summary Report Year 2002–2012.* Atlanta, GA: US Department of Health & Human Resources.

Chan, Tat Y., and Barton H. Hamilton. 2006. "Learning, Private Information, and the Economic Evaluation of Randomized Experiments." *Journal of Political Economy* 114 (6): 997–1040.

Chernew, Michael E., William E. Encinosa, and Richard A. Hirth. 2000. "Optimal Health Insurance: The Case of Observable, Severe Illness." *Journal of Health Economics* 19 (5): 585–609.

Chernew, Michael E., Allison B. Rosen, and A. Mark Fendrick. 2007. "Value-Based Insurance Design." *Health Affairs* 26 (2): w195–w203.

Csokmay, John M., Micah J. Hill, Rebecca J. Chason, Sasha Hennessy, Adita N. James, Jacques Cohen, Alan DeCherney, James H. Segars, and Mark D. Payson. 2011. "Experience with a Patient-Friendly Mandatory, Single-Blastocyst Transfer Policy: The Power of One." *Fertility and Sterility* 96 (3): 508–84.

Dunson, David B., Donna B. Baird, and Bernardo Colombo. 2004. "Increased Infertility with Age in Men and Women." *Obstetrics and Gynecology* 103 (1): 51–56.

Einav, Liran, Amy Finkelstein, and Heidi Williams. 2016. "Paying on the Margin for Medical Care: Evidence from Breast Cancer Treatments." *American Economic Journal: Economic Policy* 8 (1): 52–79.

European Society of Human Reproduction and Embryology. 2014. *ART Fact Sheet*. Brimbergen, Belgium: European Society of Human Reproduction and Embryology.

Fang, Hanming, and Alessandro Gavazza. 2011. "Dynamic Inefficiencies in an Employment-Based Health Insurance System: Theory and Evidence." *American Economic Review* 101 (7): 3047–77.

Feldstein, Martin, and Jonathan Gruber. 1995. "A Major Risk Approach to Health Insurance Reform." *Tax Policy and the Economy* 9: 103–30.

Gershoni, Naomi, and Corrine Low. 2015. "The Impact of Extended Reproductive Time Horizons: Evidence from Israel's Expansion of Access to IVF." Unpublished.

Hamilton, Barton, H., Emily Jungheim, Brian McManus, and Juan Pantano. 2018. "Healthcare Access, Costs, and Treatment Dynamics: Evidence from In Vitro Fertilization: Dataset." *American Economic Review.* https://doi.org/10.1257/aer.20161014.

Hamilton, Barton H., and Brian McManus. 2012. "The Effects on Insurance Mandates on Choices and Outcomes in Infertility Treatment Markets." *Health Economics* 21 (8): 944–1016.

Henne, Melina B., and M. Kate Bundorf. 2008. "Insurance Mandates and Trends in Infertility Treatments." *Fertility and Sterility* 89 (1): 66–73.

Højgaard, Astrid, Lars D. M. Ottosen, Ulrik Kesmodel, and Hans J. Ingerslev. 2007. "Patient Attitudes towards Twin Pregnancies and Single Embryo Transfer: A Questionnaire Study." *Human Reproduction* 22 (10): 2673–78.

Jain, Tarun, Bernard L. Harlow, and Mark D. Hornstein. 2002. "Insurance Coverage and Outcomes of an In Vitro Fertilization." *New England Journal of Medicine* 347 (9): 661–66.

Jungheim, Emily S., Man Yee Leung, George A. Macones, Lisa M. Pollack, and Barton H. Hamilton. 2017. "In Vitro Fertilization Insurance Coverage and Chances of a Live Birth." *Journal of American Medical Association* 317 (12): 1273–75.

Jungheim, Emily S., Ginny L. Ryan, Eric D. Levens, Alexandra F. Cunningham, George A. Macones, Kenneth R. Carson, Angeline N. Beltsos, and Randall R. Odem. 2010. "Embryo Transfer Practices in the United States: A Survey of Clinics Registered with the Society for Assisted Reproductive Technology." *Fertility and Sterility* 94 (4): 1432–36.

Kaiser Family Foundation. 2007. *Employer Health Benefits: 2007 Annual Survey.* Menlo Park, CA: Henry J. Kaiser Family Foundation.

Kowalski, Amanda E. 2015. "Estimating the Tradeoff between Risk Protection and Moral Hazard with a Nonlinear Budget Set Model of Health Insurance." *International Journal of Industrial Organization* 43: 122–35.

Lemos, Elkin V., Dongmu Zhang, Bradley J. Van Voorhis, and X. Henry Hu. 2013. "Healthcare Expenses Associated with Multiple vs. Singleton Pregnancies in the United States." *American Journal of Obstetrics and Gynecology* 209 (6): 586.e11–586.e11.

Malizia, Beth A., Michele R. Hacker, and Alan S. Penzias. 2009. "Cumulative Live-Birth Rates after In Vitro Fertilization." *New England Journal of Medicine* 360: 236–43.

Manning, Willard G., Joseph P. Newhouse, Naihua Duan, Emett B. Keeler, Arleen Leibowitz, and M. Susan Marquis. 1987. "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment." *American Economic Review* 77 (3): 251–77.

Mercer Health Resources Consulting. 2006. *2005 National Survey of Employer-Sponsored Health*. Washington, DC: Mercer Health and Benefits LLC.

Moscarini, Giuseppe, and Fabien Postel-Vinay. 2012. "The Contribution of Large and Small Employers to Job Creation in Times of High and Low Unemployment." *American Economic Review* 102 (6): 2509–39.

Murphy, Susan A. 2003. "Optimal Dynamic Treatment Regimes." *Journal of the Royal Statistical Society* 65 (2): 331–35.

Practice Committee of the Society for Assisted Reproductive Technology and Practice Committee for the American Society for Reproductive Health. 2012. "Elective Single-Embryo Transfer." *Fertility and Sterility* 97 (4): 835–42.

Robins, James M. 1997. "Casual Inference from Complex Longitudinal Data." In *Latent Variable Modeling and Applications to Causality: Lecture Notes in Statistics*, Vol. 120, edited by M. Berkane, 69–117. New York: Springer Verlag.

Rust, John. 1987. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica* 55 (5): 999–1033.

Ryan, Ginny L., Amy E. T Sparks, Christopher S. Sipe, Craig H. Syrop, Anuja Dokras, and Bradley J. Van Voorhis. 2007. "A Mandatory Singly Blastocyst Transfer Policy with Educational Campaign in United States IVF Program Reduces Multiple Gestation Rates without Sacrificing Pregnancy Rates." *Fertility and Sterility* 88 (2): 354–60.

Ryan, Ginny L., Sunny H. Zhang, Anuja Dorkas, Craig H. Syrop, and Bradley Jan Van Voorhis. 2004. "The Desire of Infertile Patients for Multiple Births." *Fertility and Sterility* 81 (3): 500–504.

Schmidt, Lucie. 2005. "Infertility Insurance Mandates and Fertility." *American Economic Review* 95 (2): 204–08.

Schmidt, Lucie. 2007. "Effects of Infertility Insurance Mandates on Fertility." *Journal of Health Economics* 26 (3): 431–46.

Simonstein, Frida, Michal Mashiach-Eizenberg, Ariel Revel, and Johnny S. Younis. 2014. "Assisted Reproduction Policies in Israel: A Retrospective Analysis of Vitro Fertilization-Embryo Transfer." *Fertility and Sterility* 102 (5): 1301–06.

Statistics of US Businesses (SUSB). 2005. "Data by Enterprise Employment Size." United States Census Bureau. https://www.census.gov/data/tables/2005/econ/susb/2005-susb-annual.html (accessed October 17, 2018).

Vélez, Maria P., Michelle P. Connolly, Isaac J. Kadoch, Sam W. Phillips, and Frederic Bissonnette. 2014. "Universal Coverage of IVF Pays Off." *Human Reproduction* 29 (6): 131–19.

Weinzierl, Matthew. 2011. "The Surprising Power of Age-Dependent Taxes." *Review of Economic Studies* 78 (4): 1490–1518.