

# A Live Baby or Your Money Back: The Marketing of In Vitro Fertilization Procedures

David C. Schmittlein • Donald G. Morrison

*The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6371*

*Anderson Graduate School of Management, University of California at Los Angeles,*

*Los Angeles, California 90095-1481*

*schmittlein@wharton.upenn.edu • donald.morrison@anderson.ucla.edu*

---

Many clinics that offer in vitro fertilization (IVF) have begun to market the following options to couples: (1) an a la carte program where the couple pays \$7,500 per attempt regardless of the outcome; or (2) a money-back-guarantee program where the couple pays a \$15,000 fee that covers up to three attempts, however, if after three cycles there is no live-birth delivery, then the full \$15,000 is refunded.

We assess the a la carte versus the money-back-guarantee programs, and find the surprising result that the money-back-guarantee program appears (for the patients) to be “too good to be true.” That is, the money-back guarantee yields a substantial *negative* expected profit per couple for the clinics. More importantly from the patients’ perspective, the money-back guarantee is the better option for all couples with less than 0.5 success probability per cycle. Virtually all traditional IVF patients have had per-cycle success probabilities below 0.5.

A detailed analysis of the key variables—i.e., success rate per attempt, heterogeneity of couples’ rates of success, individual couples’ “learning” on successive attempts, and cost to the clinic per attempt—shows that these money-back guarantees are unprofitable for the clinics. Since presumably clinics are not in business to lose money, the standard analysis must be missing something major. We suggest that the marketing of money-back guarantees is inducing couples who would previously have used—successfully—other less invasive procedures with fewer side effects and less risk of multiple births to decide to proceed directly to IVF, and that this scenario makes the money-back guarantees profitable for the clinics.

The implications of earlier use of IVF are then considered from an overall public policy point of view. Just as mothers everywhere tell their children, “When something looks too good to be true, then it is too good to be true!”

*(Marketing; In Vitro; Assisted Reproduction; Health Care Marketing)*

---

## 1. Introduction

In the two decades since the first “test tube” baby was born, in vitro fertilization (IVF) has become the “last best hope” for a child for hundreds of thousands of infertile couples. Typically, such a couple has already attempted natural conception, the use

of fertility-enhancing drugs, and intrauterine insemination without success (see Figure 1). Because IVF is rarely covered by health insurance in the United States (Davis 1996, Freudenheim 1998), assisted reproductive technology (ART) clinics are typically chosen by the couple, which would expect to pay between

\$7,000 and \$10,000 for each in vitro attempt (Robertson and Scheyer 1997, Rubin and Zitner 2000). The chance of a live healthy birth from that procedure has been 20% to 25%. While the ART clinics may not use the language of marketing strategy, the evolution of this competitive service environment has led them to focus on two of the four major marketing decision variables: product performance and price.<sup>1</sup>

From the customer (patient) standpoint, the most important product performance measure is the live-birth probability per initiated IVF cycle. One of our goals in this paper is to show what such performance measures say (and do not say) about a couple's prospects for success with a particular clinic. To our knowledge, this kind of analysis has not been undertaken to date. We find that clinics do substantially differ from each other in the success rates that one can expect them to experience in the future. One-year success performance, however, provides a useful guide to that future for only about half the North American clinics. For the remainder, the average performance across all clinics would be a better guide to their future than the clinic's experience to date.

The second marketing variable—price—has received even more attention. Since 1997, many ART clinics have begun to offer as an option to patients "money-back guarantees"—i.e., "a live baby or your money back," in addition to the option of paying per cycle as above. For example, one Minnesota clinic has offered three IVF tries for approximately \$15,000, with a guarantee to refund the fee if a live birth does not occur via these three attempts. More than 60 clinics are believed to offer similar incentives to prospective customers (Trafford 1997).

Understandably, this pricing innovation has attracted significant attention from consumers. In light of the medical decision context, the complicated economic incentives for the ART clinics, the asymmetry

in information between these clinics and prospective patients, the amounts of money involved, and the emotional state of some infertile couples, this pricing "innovation" has also generated controversy in professional publications (Hyman and Silver 1998, Murray 1997, Robertson and Schneyer 1997) and in popular media (Freudenheim 1998, Norris 1997, Trafford 1997).

In this paper, we show how to estimate the economic consequences of these guarantees for any clinic. Doing so requires knowledge of this clinic's overall success rate *and* a stochastic model of IVF success for any given couple on successive IVF attempts. The latter model will incorporate both heterogeneity in success rate across couples and the learning effect observed to occur across IVF cycles. Models to date in the literature have not captured these two effects.

We also analyze factors that affect the economics of the new money-back policies for clinics. Some of these factors are compatible with the goals of patients and others are not. Prospective patients may, through specific questioning, infer which factors are driving the pricing policy of a certain clinic. For clinic managers, our analyses provide a mechanism for integrating cost and history information to set a sensible pricing program.

Using an analysis of clinic performance statistics, and of patients' success rates on successive IVF cycles, we show that these guarantees are "too good to be true." Our analysis suggests that guarantees of the sort described above are made economically viable for clinics by pursuit of *less*-infertile couples who are in the beginning stages of fertility assistance, rather than using IVF as a "last resort" as had been the case previously. Specifically, we conclude that

(1) These guarantees—which are typically offered to virtually all customers—are not economically viable for the average clinic and average couple.

(2) The guarantees, in fact, are not viable for most of the "better" clinics either and, in practice, are currently being offered by many "average" performing clinics.

(3) The guarantees are not made viable by the "economies of scale" in adding to the current customer base of a clinic.

<sup>1</sup>The other two marketing variables—promotion and place (distribution)—are important but have remained more stable and are managed similarly to many other medical procedures. That is, promotional activities include relationships with hospitals and physicians, and brochures, informational presentations, and so on. Distribution decisions trade-off customer convenience of multiple locations with cost efficiency of service provision from one or a few central locations.

(4) The guarantees *are* viable if new, relatively fertile couples are induced to proceed directly to IVF instead of trying natural conception or less invasive procedures. For these couples, on standard economic bases, these IVF “guarantees” are not a good deal.

The next section describes the practice of IVF. This is followed by a straightforward economic assessment of the “money-back guarantee” for a typical clinic. Next we examine two factors that might be imagined explain such offers: heterogeneity in clinic performance, and heterogeneity/learning effects for couples making successive IVF attempts. Finally, we consider the factor that *does* appear to be responsible for these offers: targeting a new customer base characterized by higher IVF success rates.

## 2. The Prevalence and Practice of IVF

### The Scope and Scale of IVF

During its 20 years of experience, IVF has gone from an expensive procedure available in a handful of clinics whose success rate was extremely low to one that is expensive, widely accepted, available in numerous clinics, with a modest success rate. This, of course, can be seen as a glass either “half empty” or “half full.” In 1996, when the money-back guarantees at issue in this paper were being initiated, the approximate 300 clinics in the United States and Canada that performed IVF accounted for about 45,500 “standard” IVF cycles, i.e., cycles using fresh, nondonor eggs (U.S. Centers for Disease Control and Prevention 2002b). The number of babies born as a result was about 10,000 in 1996. Thus, more than 35,000 of the 45,500 IVF cycles failed to produce a live birth; but approximately 10,000 couples achieved at least one child—an outcome highly unlikely in the absence of the IVF procedure. (The natural-conception birthrate per cycle for couples classified as infertile is about 1.6% (Gleicher et al. 1996).)

### IVF Decision Process for Patients: High Risk, High Return, and When to Stop?

The most meaningful measure of success for each in vitro cycle started is the live-birth probability.

This success rate has increased from 6% in the early 1980s to about 22% in 1996 when money-back guarantees were being planned by clinics, to 25% by 1999, the most recent data available (Society for Assisted Reproductive Technology (SART) and the American Society for Reproductive Medicine (ASRM) 1996a, b; U.S. Centers for Disease Control and Prevention 2002a, b). Taking into account the typical number of IVF cycles pursued by a couple, however, more than half of them complete their attempts at IVF without taking home a child.

Accompanying each IVF cycle’s uncertain outcome is a substantial cost, typically in the range of \$7,000–\$10,000. Couples often pursue multiple IVF cycles if needed, so a total cost in the range of \$10,000–\$30,000 is common. In some European countries, national health insurance pays for a certain number of IVF tries, but coverage of IVF by U.S. health insurers is usually limited (e.g., one cycle) or more often nonexistent (Friedler et al. 1992, Haan and Rutten 1989, *Strictly Business* 1997). As a result, most U.S. couples considering IVF are placing the largest-scale single economic gamble of their lives. That is, they may make a few investment decisions involving more money (e.g., a home purchase) but none whose “payoff” is so random and stark (i.e., about a 50-50 chance of a genetically related child versus losing the entire investment with no benefit).

In addition to the economic risk above, other factors also make IVF decisions difficult for patients. These include the risk of multiple conception (twins, triplets, and so on) with attendant health risks for the fetuses or infants, and risk to the female of hyperstimulation and other possible long-term health risks (Dawood 1996, ESHRE Capri Workshop 1996). They also include the difficulty in deciding when to stop IVF attempts, a decision faced after each (failed) IVF cycle. Couples typically want to feel that they did “all they could” to conceive (Golombok 1992, Stolberg 1997, *Strictly Business* 1997), but success rates for IVF do *not* appear to substantially drop after several failures for a particular couple. Clinics report this finding to those considering an additional cycle (Haan et al. 1991b), making it difficult to stop.

### IVF Decision Process for Clinics

IVF represents a large and growing service opportunity. In 1994, the amount spent on such procedures in the United States and Canada was approximately \$300 million. IVF procedures are likely to continue to grow in popularity, because it is estimated that 10%–15% of all married couples in the reproductive age group are infertile (Diczfalusy and Crosignani 1996). To the extent that couples are paying for the procedure themselves, opportunities to “lock in” patients through the emerging relationships in the health-care market are minimized, leaving at least the potential for regional competition across clinics. Indeed, metropolitan areas are generally now served by at least two such clinics.

### 3. Money-Back Guarantees: Economics for a Typical Patient and Clinic

The money-back guarantees for IVF are largely standard: The patient pays about \$15,000 for up to three IVF attempts as needed to produce a live-birth delivery. If after three attempts the couple has not succeeded, the \$15,000 is refunded. Alternatively, the patient is offered the choice of paying “a la carte,” about \$7,500 per attempt.

We initially will analyze the economic effects for clinics and patients as of the time that money-back guarantees were being created as marketing programs, i.e., 1996. (Later, we will note that the basic economic conclusions had not changed by 2002.) A clinic’s average cost per IVF cycle was estimated to be \$5,000 in 1988 (Wagner and St. Clair 1989). By 1996, a 2% annual nominal cost increase would lead to an IVF cost per cycle of \$6,000. This cost figure, combined with the recent past’s a la carte price of \$7,500, leads to a return on investment (ROI) (actually, return on cost) of 25%. This is a high return in light of the competition between clinics and, accordingly, we conclude that cost per cycle in 1996 has risen to at least \$6,000 per cycle.

With notation  $C$  = clinic’s cost per cycle,  $p$  = success probability on any IVF attempt,  $\pi$  = clinic’s profit (or

loss), and  $G$  = the patient’s payment for the guarantee, the clinic’s expected profit resulting from the guarantee is

$$E[\pi] = G - C - C(1 - p) - C(1 - p)^2 - G(1 - p)^3. \quad (1)$$

With  $C = 6,000$ ,  $G = 15,000$ ,  $p = 0.22$  (as noted in §2), expected profit for the clinic is minus \$6,448 per patient. Via this calculation, the money-back guarantee is not close to break even. To put this in perspective, recall that ROI was 25% for patients paying a la carte (with  $C = 6,000$ ). Under the guarantee, the expected number of IVF cycles that a patient will undergo is 2.4, so the average incurred cost is 2.4 times \$6,000 or \$14,400. Accordingly, the return on cost under the guarantee is  $-6,448/14,400$  or minus 45%.

Our simple Equation (1) assesses financial outcomes only for the IVF cycles themselves. A particular new patient at an IVF clinic may also undergo some testing, whose cost is separate from the guarantee (and nonrefundable). Discussion with clinic operators suggests that such testing could approach \$4,000, and a generous estimate of the profit on such testing is \$3,000. So to be conservative in our assessment of a clinic’s loss on the money-back guarantee, we will reduce the \$6,448 loss from the previous paragraph by a \$3,000 gain on testing; leaving an expected loss per IVF patient of \$3,448.

To see a clinic’s difficulty another way, note that setting  $C = 6,000$  and  $G = 15,000$ , and assuming a \$3,000 profit from testing as above, the success percentage per attempt  $p$  would have to be 31% (rather than 22%) to break even. It would need to be 40% to generate a 20% return on expected cost.

While altruism on a grand scale or gross mismanagement could explain this conundrum, we are interested in examining other explanations for the guarantee policy. We deal with four here, and then turn to the ones that require a more serious examination of clinic performance statistics and patient success dynamics.

#### Marginal Cost vs. Average Cost

One might posit that the clinic’s marginal cost of conducting one more IVF cycle is more relevant in calculating profitability, and that marginal cost

would be much lower than average cost. In this case, however, several factors undermine such a notion, specifically:

(1) Only 10% of the clinic cost is equipment or location (Haan and van Steen 1992) and the trend is to individualize equipment (e.g., incubator) to the specific patient.

(2) Of the total IVF clinic costs, 40% is material/lab tests/drugs (Haan and van Steen 1992), which are not subject to many economies of scale.

(3) Of total clinic costs, 50% is personnel (Haan and van Steen 1992), and clinics were already operating at an efficient economic scale, doing on average over 100 IVF procedures a year, each cycle requiring 4–6 office visits.

(4) Clinics offering the guarantee are finding that a substantial fraction of patients take it, making average cost more relevant than marginal cost.

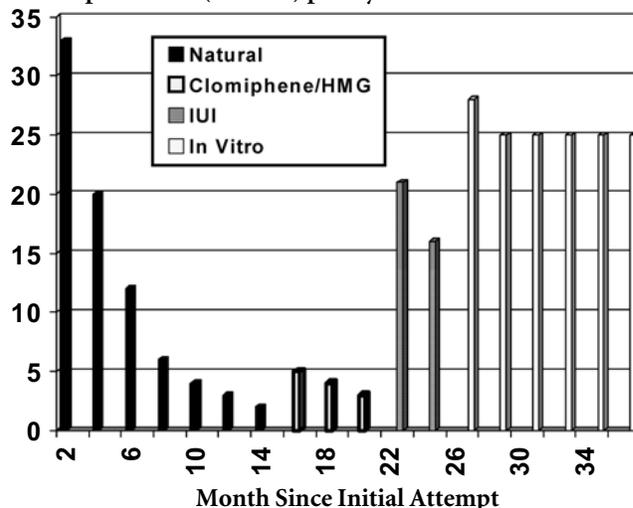
#### Patient Selection: Aggressive Screening

Another possible enhancement to profit is limiting the guarantee program to patients whose likelihood of IVF success is “high.” Clinics generally make such restrictions public, and many do limit the age of the female to less than 40 years, and some also eliminate couples with a male infertility factor. Based on clinic performance data (SART 1996), these two restrictions raise the IVF success probability by three percentage points. Beyond this, clinics claim that they do not “discriminate” and, indeed, the available studies have not succeeded in identifying additional observable factors that predict IVF success (Haan et al. 1991a, Stolwijk et al. 1996, Zhou et al. 1996). We will, as a result, assume that patient screening adds only 3 percentage points to the clinic’s success rate.

#### Breakthrough in IVF Success Rates

A third possible explanation for the guarantee’s financial loss is a breakthrough in IVF success rates since 1996 that was well anticipated by clinics as they launched money-back guarantees in 1997 and 1998. This is unlikely for two reasons. First, to go from 22% to 40% in just a year or two would mean a significant breakthrough and the medical literature points to no such advance during this period. Second, IVF success, summarized in Figure 2, shows a slow and steady

Figure 1 Typical Progression of Assisted Reproduction Treatment  
 Conception Rate (Percent) per Cycle\*



Source. Hull, M. G. R. 1992. Infertility treatment: Relative effectiveness of conventional and assisted conception methods. *Human Reproduction* 7 785–796.

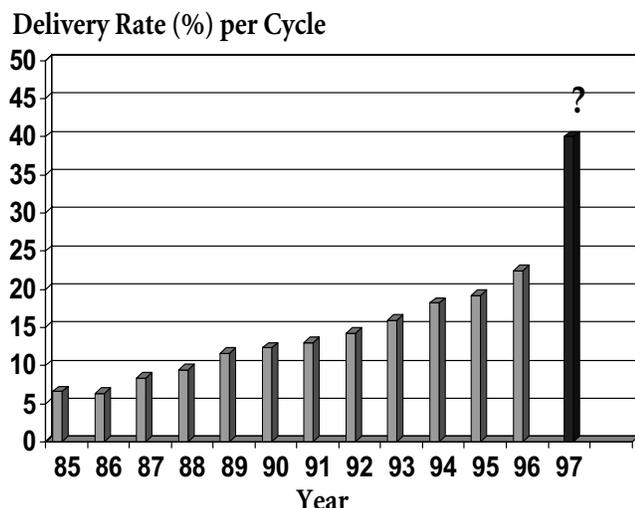
Note. \*Conception rate does not equal live-delivery rate. In later cycles live-delivery rate is approximately 20% lower than conception rate.

increase through 1996. Indeed, that slow steady pace is now known to have persisted through 1999. The rate of success, in fact, increased only from 22% to 25% between 1996 and 1999, quite consistent with the historical trend (U.S. Centers for Disease Control and Prevention 2002a, b).

#### Risk Transfer

Another approach to improve the financials for the guarantee is to shift the balance between minimizing risk of the IVF procedure and maximizing the probability that a live birth occurs. This means transferring additional risk to the patient, and could be accomplished in multiple ways. For instance, the drug regimen can be increased to stimulate production of more eggs. This also increases the risk of hyperstimulation, a potentially serious health consequence for the female (ESHRE Capri Workshop 1996). The success rate can also be increased by using more viable embryos (ESHRE Capri Workshop 1996). This also increases the risk of multiple conception, and higher-order multiples carry both health risks for the infants (prenatal death, cerebral palsy, low birth weight) and

**Figure 2** Historical Trend in IVF Success Rate, North America



Sources. Dawood 1996; SART 1996; US Centers for Disease Control and Prevention 1999

for the mother (preeclampsia, hydramnios varicosities, anaemia) (Dawood 1996, ESHRE Capri Workshop 1996). Clinics claim that they do not do these things (*Strictly Business* 1997) and, indeed, these kinds of activities and their negative outcomes are reasonably observable. In short, risk transfer to patients would likely become public and we assume that this does not occur on a significant scale.

In summary, the four factors considered here are estimated, collectively, to increase the probability of success by a total of 3 percentage points from the baseline level of 22% in the 1996 data. That is, those clinics that *do* screen on age (<40 years) and male factor infertility can achieve a 25% chance of a live birth per IVF cycle in 1997—well under the 40% chance required for break even on the guarantee.

### Heterogeneity

Staying within the framework of our profitability Equation (1), there are really only two additional factors to consider, and they both involve heterogeneity. First, Equation (1) implicitly assumes that all clinics have the same success rate, and second, it assumes that this success rate does not differ from cycle to cycle for an individual customer.

Regarding the first of these factors, if clinics differ reliably in their success rates, the “better” ones may

be able to afford the guarantee and, indeed, such a guarantee might signal a strong clinic to a prospective patient. In the next section, we analyze this across-clinic heterogeneity.

Our second source of heterogeneity concerns the repeated experiences of a single patient. Heterogeneity across patients in seriousness of infertility will tend to drive down success probability across repeated cycles. That is, the “healthier” patients will tend to have success on the early IVF cycles, leaving them out of the sample for later cycles. On the other hand, useful information is acquired *during* the IVF cycles, which can counteract this negative effect of heterogeneity. Specifically, the drug regimen used may be optimized to a particular patient on later IVF cycles, or some patients may be counseled out of additional IVF attempts when previous cycles suggest that IVF will not work for them (Haan et al. 1991b). To examine both heterogeneity across patients and learning across repeated attempts will require some new probability modeling. We will turn to such a model after the next section’s summary of heterogeneity across clinics.

## 4. Variation in Performance Across IVF Clinics

We consider the extent to which clinics differ in prospects for IVF success by examining the 1994 clinic performance data reported to SART for the U.S. eastern region (SART and ASRM 1996b, U.S. Centers for Disease Control and Prevention 2002b). (Clinic-specific reports for 1995–1999—the most recent data available—collapse IVF with other ART procedures such as zygote intrafallopian transfer (ZIFT) and gamete intrafallopian transfer (GIFT), and so are not as valuable for considering guarantees that apply only to IVF.) The leading infertility support organization in the United States (RESOLVE) promotes these statistics to couples considering IVF (RESOLVE 1997, p. 14). This information includes the number of live deliveries per IVF cycle started, for each of the 101 reporting clinics. In total, these clinics began 14,322 IVF cycles and had 2,646 deliveries, for a delivery rate of 0.185. During 1994, the average number of IVF cycles was 142 per clinic. We noted earlier

**Figure 3** Assessing Heterogeneity in Clinic Performance

Original "Top 5"			Original "Bottom 5"		
Rank	Original Observation.	Updated Probability	Rank	Original Observation.	Updated Probability
1	2/4=0.500	0.235	5	1/28=0.036	0.144
2	11/27=0.407	0.285	4	0/1=0.000	0.207
3	8/20=0.400	0.270	3	0/4=0.000	0.194
4	15/39=0.385	0.292	2	0/5=0.000	0.190
5	69/182=0.379	0.346	1	0/24=0.000	0.138

Updated "Top 5"			Updated "Bottom 5"		
Rank	Original Observation.	Updated Probability	Rank	Original Observation.	Updated Probability
1	69/182=0.379	0.346	5	3/39=0.077	0.149
2	169/480=0.352	0.340	4	1/28=0.036	0.144
3	55/166=0.331	0.306	3	0/24=0.000	0.138
4	15/39=0.385	0.292	2	23/213=0.108	0.126
5	26/77=0.338	0.291	1	23/222=0.104	0.122

that age and presence of a male infertility factor are sometimes used to exclude patients from the guarantee programs. Accordingly, we focus our analysis on the 8,714 IVF cycles for which the female was under age 40 and there was no male infertility factor.

In any one year, the success rate does greatly vary from clinic to clinic. The top two panels of Figure 3 highlight this—showing the “top 5” and “bottom 5” clinics. From the standpoint of both patients and clinic managers, of course, the relevant quantity is not the observed success rate in some previous year. Rather, it is the best estimate, based on that history, of future success for that clinic. For this purpose, the historical success rates are deficient estimators, due to the influence of sampling variation. Indeed, clinics that are observed to be outliers will tend to be those that happened to do few cycles—irrespective of long-run success rates. This is evident in Figure 3—all five of the “worst” clinics and four of the five “best” clinics did very few IVF cycles.

Let  $p$  denote a particular clinic’s long-run success probability, and  $x$  denote the number of live-birth deliveries arising from  $n$  IVF cycles started in some time period. We are interested in estimating  $p$  from  $x$  and  $n$ —or specifically from the observed success rate  $x/n$ . In doing so, we adopt the empirical Bayes approach of estimating both the sampling variance for successes  $x$ , and the real variance in success probability  $p$  across clinics (Maritz 1970). The estimate for  $p$  is then a combination of this clinic’s historical success

rate ( $x/n$ ) and the observed average success across clinics ( $E[p]$ ).

Specifically, we assume that the clinic-specific probability of success  $p$  is distributed beta across clinics and, accordingly, births  $x$  at a specific clinic follow the well-known beta-binomial (BB) model. This model has been highly effective for representing heterogeneous Bernoulli processes in marketing and in other social science applications, including biomedical research (Crowder 1978, Greene 1982, Griffiths 1973). Based on its two beta distribution parameters  $a$  and  $b$ , the expected future success rate for a clinic that was observed to experience  $x$  successes in  $n$  IVF attempts is

$$E[p|x, n, a, b] = \frac{n}{a+b+n} \frac{x}{n} + \frac{a+b}{a+b+n} \frac{a}{a+b}, \quad (2)$$

which represents a weighted average of the clinic’s observed success rate ( $x/n$ ) and the average success rate for all clinics ( $a/(a+b)$ ).

Maximum likelihood estimates of the model parameters for this set of 101 clinics are ( $a = 9.44$ ;  $b = 35.15$ ), which corresponds to a mean and standard deviation of real long-run success rates across clinics of ( $E[p] = 0.212$ ;  $\sigma_p = 0.0605$ ). Accordingly, using the one- and two-sigma heuristics, two-thirds of clinics have a true delivery rate between 15% and 27% (one sigma) and 95% of clinics have a true delivery rate between 9% and 33%. These results rule out heterogeneity in clinics as a contributor to the viability of IVF money-back guarantees. Virtually none of the clinics have a stable success rate that is even at break even (31%). We focused here on eastern region clinics to help remove geographic differences. However, the same beta-binomial model estimated with all 244 North American clinics resulted in MLEs  $a = 8.41$  and  $b = 32.04$ , which produce the same kind of confidence interval results as above.

As Equation (2) makes clear, expected future success is driven not only by observed success rate ( $x/n$ ), but by the number of IVF cycles performed ( $n$ ) relative to the sum of the two beta distribution parameters  $a + b$ . This is highlighted in Figure 3. Note how our “top 5” and “bottom 5” clinics’ expectations are changed when an empirical Bayes updated probability is calculated for each clinic. In the bottom half

of Figure 3, we provide the updated “top 5” and “bottom 5” clinics, i.e., clinics ranked on the probability of success on future IVF attempts. Only three have expected probability greater than 0.3 and none exceed 0.35. Because about 60 of 300 North American clinics are offering the IVF guarantee, again, we see that differential clinic performance is not the explanation.

Of course, there *are* real differences in clinic performance that should persist and matter to patients. Clinics in the top third of real performance ( $p$ ) are about twice as likely to experience a live-birth delivery as those in the bottom third. The challenge for a patient is to see a large enough base of experience to “tell” the strong performers from the weak.

## 5. Patient Outcomes Across Successive IVF Attempts

Our Equation (1) for economics of the money-back guarantee assumed that all of a clinic’s patients have the same success probability  $p$ , and that this success probability for a single couple does not vary across successive IVF attempts. Neither assumption is likely to well represent the IVF process, and each can greatly affect the expected financial outcome. First we will discuss the effect of patient heterogeneity, under the simplifying assumption that  $p$ -values remain stationary for each patient across successive attempts. Then, we will incorporate nonstationarity in the patient probabilities, and estimate a formal model that captures both phenomena.

A couple that succeeds on IVF cycle  $t$  naturally does not progress to cycle  $t+1$ , i.e., the only couples that enter attempt  $t+1$  are those unsuccessful thus far. When success probabilities differ from couple to couple, those entering cycle  $t+1$  will tend to be the ones who started out with relatively low  $p$ -values. In short, the success rates observed for a random set of patients for successive cycles should decline from cycle to cycle, as a result of this adverse selection effect. This effect of heterogeneity was not reflected in Equation (1) and is, in fact, detrimental to the economics of the IVF guarantee. Recall that the assumption of homogeneity across patients in Equation (1) produced a net loss of \$3,448 per patient, including a profit from testing. Imagine the extreme

case of patient heterogeneity with the same average success rate  $p = 0.22$  as assumed earlier. That is, imagine that 22% of patients have an IVF success rate on any cycle of 100%, and the remaining 78% of patients have a success rate of 0. Using the same arithmetic as in Equation (1), the clinic will net \$15,000–\$6,000 from 22% of the patients (who succeed on the first try), and will “net”  $\$15,000 - 3 \times \$6,000 = -\$15,000$  from the remaining 78%, who do not succeed even after three tries. The expected loss per patient under this complete heterogeneity scenario is minus \$12,060, 3.5 times greater than the loss of \$3,448 if patients are homogeneous in success rates. Heterogeneity scenarios between these two extremes examined here will produce losses between \$3,448 and \$12,060. In short, heterogeneity makes the guarantee programs, which “push” all patients through three tries if they need them, even more economically disadvantageous for the clinics.

There is, however, a potentially countervailing effect related to successive IVF attempts. While our heterogeneity across patients scenario above assumes that the success rate for any individual patient does not change from attempt to attempt, anecdotal reports suggest that the prospects for success can be increased for a particular couple from attempt to attempt. This “learning” effect stems from two kinds of sources. First, information on the body’s response to the drug regimen, sperm, egg, and embryo quality can be used to help improve the chances for success on later IVF attempts, at least for a while. So, the actual “ $p$ -value” for a patient is not constant from attempt to attempt, but may instead increase. Second, the information gleaned from initial cycles will sometimes reveal that the prospects for IVF success are remote, and the couple can be counseled to move on to another therapy (donor egg, intracytoplasmic sperm injection (ICSI)) or to adoption. This second learning effect helps remove low  $p$ -value patients from later cycles and, thus, runs directly counter to heterogeneity’s adverse selection effect discussed above. Of course, only the former of these two learning effects would actually benefit a clinic offering a money-back guarantee. We will not be able to differentiate between these two possible learning effects, but we will be able to estimate the combined effect from historical data, placing

an upper limit on the learning effect's benefit for clinics offering the guarantee.

Which of these two countervailing effects, heterogeneity depressing success rates across cycles or learning increasing them, is larger? Many studies have provided data concerning success prospects on successive attempts, though none have incorporated both learning and heterogeneity effects (Alsaili et al. 1995, Check et al. 1994, Haan et al. 1991b, Hershlag et al. 1991, Stolwijk et al. 1996, Tan et al. 1994a, Zhou et al. 1996).

Let  $p_1$  denote a particular couple's success probability on the *first* IVF attempt. To incorporate learning in this model, we rewrite this cycle 1 success probability  $p_1$  as

$$p_1 = \frac{x}{x+y}, \quad x > 0, y > 0, \quad (3)$$

which, of course, can be done without loss of generality. It is useful to think of  $x$  as representing the impact of success factors and  $y$  as representing the factors making the IVF cycle likely to fail. The probability  $p$  is then the result of the relative magnitude of these two sets of factors as in Equation (3). Now, if learning occurs, this couple's success probability should be higher on later attempts, which can be accomplished by replacing  $x$  in Equation (3) with  $x + \lambda$  ( $\lambda \geq 0$ ). This approach would, however, restrict learning to a "one-shot" effect, i.e., no incremental learning after IVF cycle 1. Our proposed model for learning assumes more generally that there may, indeed, be learning on *each* IVF cycle, but that the magnitude of this learning may decline across cycles. As a result, the formula for the couple's success probability on attempt  $t$  ( $t \geq 2$ ) is

$$p_t = \frac{x + \sum_{i=2}^t \lambda_i}{x + y + \sum_{i=2}^t \lambda_i}, \quad \lambda_i \geq 0, \quad (4)$$

where  $\lambda_i$  represents the incremental learning effect for attempt  $i$  based on the learning on attempt  $i - 1$ . On the first IVF attempt, Equation (4) simply reduces to Equation (3). On the second attempt, the numerator "success" factor increases from  $x$  to  $x + \lambda_2$ , so  $\lambda_2$  represents the contribution of learning from the first cycle. On the third IVF attempt, the numerator is  $x + \lambda_2 + \lambda_3$ .

To incorporate heterogeneity across patients, we allow the baseline success factor  $x$ , baseline failure factor  $y$ , and learning effects  $\lambda_t$  to vary across

patients. Specifically, each of these three factors is assumed to follow an (independent) gamma distribution, with idiosyncratic shape parameters and a common scale parameter

$$f(x | r_x, \alpha) = \frac{\alpha^{r_x}}{\Gamma(r_x)} x^{r_x-1} e^{-\alpha x},$$

$$f(y | r_y, \alpha) = \frac{\alpha^{r_y}}{\Gamma(r_y)} y^{r_y-1} e^{-\alpha y},$$

$$f(\lambda_t | r_{\lambda_t}, \alpha) = \frac{\alpha^{r_{\lambda_t}}}{\Gamma(r_{\lambda_t})} \lambda_t^{r_{\lambda_t}-1} e^{-\alpha \lambda_t}.$$

For each of these gamma distributions, the mean is its shape parameter divided by the scale parameter, i.e.,  $r/\alpha$ . So, the average magnitude of the positive effects  $x$  and  $\lambda_t$ , and of the negative effect  $y$ , is proportional to the respective distribution's shape parameter. We will assume that the average size of the incremental learning effect decreases geometrically with repeated IVF attempts, with rate  $1 - K$  ( $0 \leq K \leq 1$ ). That is, because  $K \leq 1$ , the learning effect due to the information from the second IVF cycle will not generally be as great as learning from the first cycle. To capture this phenomenon, the shape parameter for each learning effect's gamma distribution is specified as

$$r_{\lambda_t} = r_{\lambda} K^{t-2} \quad \text{for } t = 2, 3, \dots \quad (5)$$

If  $K$  is close to 1, the incremental learning dies out only slowly across cycles. For example, imagine that  $r_{\lambda} = 1$  and  $K = 1/2$ . Then, on the second IVF attempt, in addition to the positive effect  $x$  (whose mean is  $r_x/\alpha$ ), we have a learning effect in the numerator (and denominator) of Equation (4) whose average size is  $r_{\lambda}/\alpha$ . On the third IVF attempt, we add an *additional* learning effect to the numerator and denominator of Equation (4), but now the average size of the incremental effect is only  $Kr_{\lambda}/\alpha = r_{\lambda}/2\alpha$ , i.e., half the size of the learning effect for the second cycle. With  $K = 1/2$ , the incremental learning effect for the fourth cycle would again drop by half, to only 1/4 what it was on the second cycle, and so on.

These assumptions mean that for any IVF attempt, the success probability is distributed beta across patients (Johnson and Kotz 1970, p. 38). For the first attempt (no learning) success probabilities are simply distributed beta( $r_x, r_y$ ), with mean  $E[p] = r_x/(r_x + r_y)$ .

For the second attempt, we have to reflect both the effects of heterogeneity and of learning. The latter is already specified above. The effect of heterogeneity is easy to assess: Given that the first attempt failed, we must update the beta distribution for success probability as in any BB model. That is, the parameter  $r_y$  is updated by one unit for each observed IVF failure (Greene 1982, Schmittlein 1989). So, on the second IVF attempt we replace  $r_y$  with  $r_y + 1$ . As a result, for patients entering the second IVF attempt, the probability of success is distributed beta( $r_x + r_\lambda, r_y + 1$ ), with mean  $E[p] = (r_x + r_\lambda) / (r_x + r_\lambda + r_y + 1)$ . In the same way, for the  $t$ th attempt, given that the first  $t - 1$  attempts failed, the expected success probability is

$$E[p_t | r_x, r_y, r_\lambda, K] = \frac{r_x + r_\lambda \sum_{i=2}^t K^{i-2}}{r_x + r_y + (t-1) + r_\lambda \sum_{i=2}^t K^{i-2}}. \quad (6)$$

In the absence of any learning ( $r_\lambda = 0$ ), Equation (6) is simply the success probability for attempt  $t$  (conditional on failure through attempt  $t - 1$ ) for the beta-geometric (BG) model, i.e., the waiting time version of the BB model. With the learning effect as specified above, we will call the formulation (6) the beta-geometric with learning (BGL) model. Equation (6) provides the basis for estimating the parameters ( $r_x, r_y, r_\lambda, K$ ) via maximum likelihood, from a histogram of success rates on successive IVF attempts.

The most promising such histogram appears to have been published by Tan et al. (1994a). In this case, all 3,824 patients who began an initial IVF cycle at one particular clinic were tracked through successive IVF attempts. The number of patients entering each successive attempt was recorded, along with the number that achieved a live-birth delivery. The Tan et al. (1994a) data are particularly valuable because of the large number of patients tracked, and because the outcome measure is live births. Many of the studies examining successive IVF attempts (Alsalili et al. 1995, Hershlag et al. 1991, Stolwijk et al. 1996) look instead at "ongoing pregnancies" as a "success," yet often about 20% of those pregnancies do not result in deliveries, and this percentage can vary substantially.

For the Tan et al. (1994a) data, the maximum likelihood estimates of the BGL model are

$$r_x = 0.263, \quad r_y = 2.247, \quad r_\lambda = 0.130, \quad K = 0.510.$$

These parameters tell an interesting story about the effects of learning and heterogeneity and about the trade-off between them. First, consider just the learning effect. Because  $r_\lambda$  is about half the size of  $r_x$ , on the second IVF attempt, the learning effect increases the positive factors for IVF success by about half. A further increase accrues for the third attempt, but because  $K$  is approximately 0.5, the contribution from learning on the second IVF cycle is only half as valuable as the learning from the first cycle. The incremental learning from the third cycle is only 1/4 ( $=0.5^2$ ) what it was from the first cycle and so on.

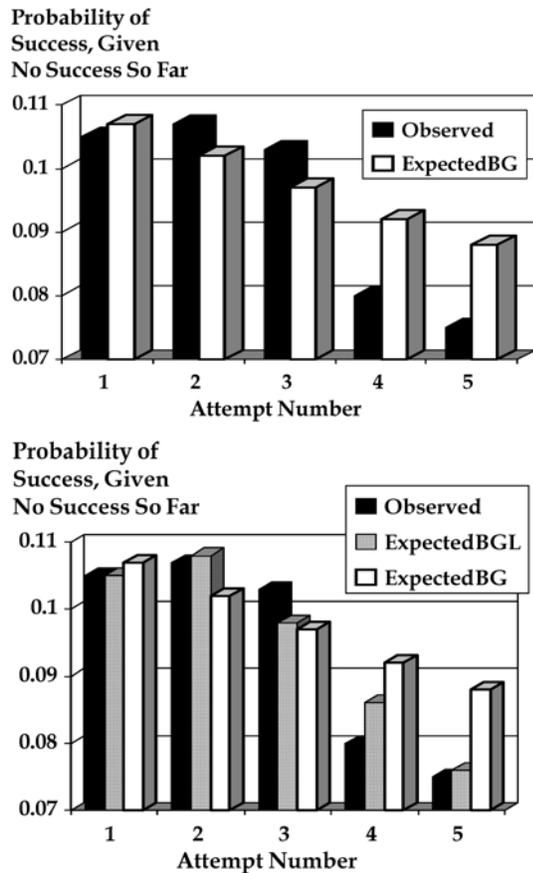
Now, we add in the effect of heterogeneity. Recall that failure on an IVF cycle increases de facto the  $r_y$  parameter by 1 for succeeding cycles. Neither heterogeneity nor learning operate on the first cycle, for which the average probability of success was  $r_x / (r_x + r_y) = 0.105$ . For patients who do not succeed on the first attempt, what happens on the second try? Via Equation (6), the net positive factors ( $r_x + r_\lambda$ ) increase by 50% relative to attempt 1 (i.e.,  $0.263 + 0.130$ ) due to learning. But due to heterogeneity, our updated sense of this patient's success rate (having failed at attempt 1), adds 1 to the negative factors parameter  $r_y$ , thus, increasing it from 2.247 to 3.247, i.e., also an increase of essentially 50%. In short, then, for the second IVF cycle, the effects of learning and heterogeneity offset, and the overall probability of success remains unchanged from what it was on cycle 1. Learning has succeeded in counteracting adverse selection.

On later cycles, this balance between learning and adverse selection is, of course, not maintained. The incremental learning effect decreases (decaying by approximately half at each cycle). The negative effect of adverse selection, by comparison, remains constant from cycle to cycle. So, by IVF cycle 4, 5, and so on, the probability of success falls.

Figure 4 illustrates this set of results.<sup>2</sup> The top half shows the fit of the BG (i.e., no learning) model to

<sup>2</sup>The Figure 4 data cover patients who began treatment between 1984 and 1990. This is why the success rate is less than the 22% figure that was relevant as guarantee programs were launched. The pattern on successive cycles, which is of primary interest, is not affected by this downward shift in success rates.

Figure 4 Modeling the Likelihood of Success on Repeated IVF Attempts



patients' delivery rates on successive IVF cycles. It clearly cannot represent the slight uptick in success observed on the second IVF cycle. Further, when the learning effect diminishes and the effect of adverse selection takes over on cycles 4 and 5, the BG cannot capture this phenomenon either. By contrast, the BGL is able to represent both phenomena.

This general pattern of interaction between learning and heterogeneity effects is not a fluke in the Tan et al. (1994a, b) data. While the specifics of the histograms do differ across studies, we also estimated the BGL model on four other published repeat cycle data sets that covered different clinics, time periods, and outcome measures (pregnancy versus live birth) (Alsaili et al. 1995, Haan et al. 1991b, Hershlag et al. 1991, Stolwijk et al. 1996). The average BGL parameter

values across these four data sets are

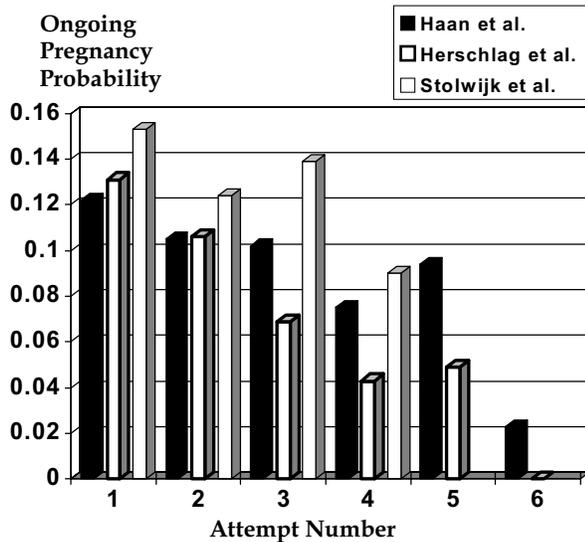
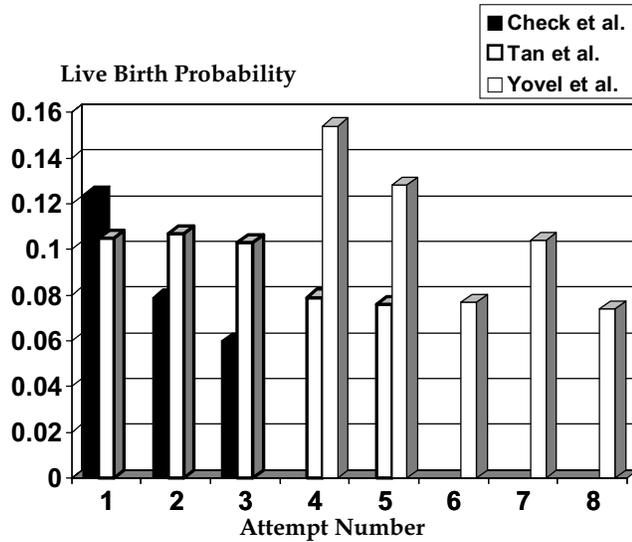
$$r_x = 0.057, \quad r_y = 0.399, \quad r_\lambda = 0.125, \quad K = 0.646.$$

Note that the average size of the learning effect ( $r_\lambda$ ) and the decay rate for this effect (indicated by  $K$ ) are similar to our results for the Tan et al. (1994a, b) data. (The parameters  $r_x$  and  $r_y$  differ more from the values earlier due to the different time periods and outcome measures in these studies.) One other qualitative conclusion here matches that observed earlier. Namely, in moving to the second IVF cycle, the impact of learning and the impact of heterogeneity essentially cancel, leaving the probability of success essentially the same as it was on attempt 1. Beyond cycle 3, however, the effect of heterogeneity (adverse selection) takes over and success rates drop. Figure 5 shows the success rate on repeated IVF attempts for the expanded set of studies reporting such histograms.

This analysis of IVF success dynamics allows us to draw several conclusions. First, learning and adverse selection (heterogeneity) are each substantial influences on the success prospects for patients who continue IVF attempts. Second, across the initial three IVF cycles, which are most relevant for our money-back-guarantee scenario, these two influences cancel, and success prospects can be viewed as constant across these cycles, as our BB analysis of clinic heterogeneity assumed in the previous section. Further, the constant success probability across cycles 1–3 represents both “good news” and “bad news” for clinics that offer a money-back guarantee. The good news is that adverse selection does not represent an additional financial drawback to the guarantee, over and above our calculation in Equation (1). The bad news is that while learning does counter adverse selection, it does not manage to add anything to our analysis in Equation (1). In short, the financial viability of the money-back guarantee remains unexplained.

A third conclusion from our analyses here concerns success prospects *beyond* the third IVF attempt. Across a variety of studies, our analyses suggest that success prospects begin to decline for such later attempts. In the next section, we leave the framework of Equation (1) and investigate in greater detail these later IVF cycles: What may happen *after* the money is refunded for a third (failed) IVF attempt?

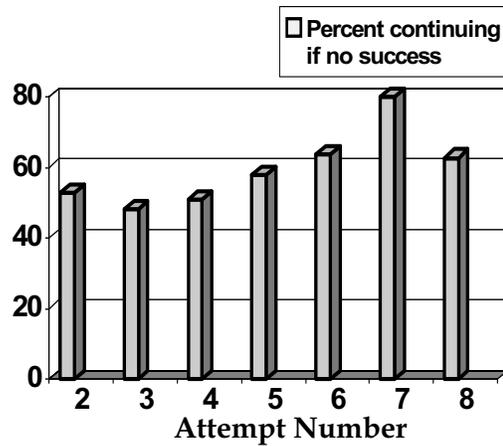
Figure 5 Outcome Probability on Successive IVF Attempts in Various Empirical Studies



## 6. Perseverance and the House Money Effect

For a typical IVF clinic offering the standard money-back guarantee, our analysis so far has been able to reduce the expected loss per patient from \$3,448 (shown in §3; based on Equation (1) and the clinic's potential profit from testing) to \$2,203. The latter figure arises simply by substituting  $p = 0.25$  in Equation (1), i.e., the increase of 3 percentage points in the

Figure 6 The Propensity to Continue IVF After Failed Attempts



Sources (pooled): Check (1994), Tar et al. (1994), Yovel (1994).

per cycle IVF success rate that stemmed from patient screening.

In this section, and in the next, we explore ways to close the remaining gap by examining factors that lie outside the framework of Equation (1). In that formula (and the ensuing sections), we have examined the guarantee's outcomes across the three possible IVF attempts that it covered. In this section, we look ahead at "what comes after" the refund of a couple's \$15,000. It is easy to describe the options. A couple can conclude their attempts at assisted reproduction, or the couple can continue with IVF, paying a la carte for additional cycles. Of course, the refund can also provide cash for such attempts.

One might think it highly unlikely that a couple, after three failures, would elect to continue pursuing IVF cycles (and paying for them), but the empirical evidence to date suggests otherwise. Looking at the pattern of behavior prior to the offering of money-back guarantees—i.e., when all couples not covered by insurance were paying for each attempt—the inclination to continue IVF cycle after cycle is striking. Pooling results from three empirical studies, Figure 6 shows that approximately 60% of those who fail on an IVF cycle go on to the next cycle, and this statistic remains virtually constant across the first eight IVF cycles.

Under the money-back guarantee, what will happen after a third failure/refund? Certainly, clinic man-

agers did not know upon instituting the guarantee. But they did have access to the empirical pattern of Figure 6. Indeed, for several reasons, the 60% perseverance rate seems a good working assumption. These reasons stem from the decision process a patient is likely to use in considering whether to continue on to the next IVF attempt. First, the 60% rate has been robust to the actual decision process the patient is likely to be using. One might imagine that the first attempt after refund "feels" like the decision an a la carte patient faces after failure on the "first for pay" attempt. That is, this is the first time a couple really had to make a decision about going further. This would suggest that the perseverance rate for refund receivers would be similar to the rate historically observed for a la carte patients after their first IVF failure. On the other hand, it might act more like a failure on the "third IVF attempt overall," because this is the number of failures the refund receivers actually experienced. But based on the empirical evidence for a la carte patients in Figure 6, such distinctions have not mattered, i.e., the 60% perseverance has been a constant across cycles.

Second, the refund receivers have just obtained a check for roughly \$15,000, which can be expected to lead to a "house money" effect (Thaler and Johnson 1990). That is, the decision-process literature predicts a heightened proclivity to gamble with the "found money" just obtained. Note that it will take at least six months for a couple to exhaust their initial three IVF attempts, so the \$15,000 refund is likely to have been viewed as "new" (found) money. Acting counter to the house money effect is a possible framing effect of the guarantee itself. That is, setting the refund at three IVF failures may suggest to the couple that three is the "right" number of attempts to represent "all that I can do." Certainly, there is anecdotal evidence that couples find it difficult to establish such a benchmark without some external criterion (Stolberg 1997, *Strictly Business* 1997). On the other hand, clinics provide data indicating a roughly constant success rate on repeated attempts (Haan et al. 1991b). This encourages couples to go on regardless of this kind of benchmark, i.e., the odds of pregnancy are presumably no worse on the next try than on the last. In fact, this kind of logic, and the powerful desire for a genetically related

child, pushes some patients to pursue more than 20 IVF attempts (Stolberg 1997).

So, we assume that 60% of patients who experience a failure will go on to the next IVF cycle, for cycles 4–8 following the refund. There may, of course, be additional cycles beyond the eighth, but even with the 60% perseverance, the number of such cycles is so small that it can be ignored. To calculate the consequences of cycles 4–8, we need to know the IVF success rate on these cycles, because success also removes the patient from later cycles. Based on the BGL model parameters reported in the last section (using the Tan et al. 1994a, b data), the success probability is essentially constant across the first three IVF attempts. We denote this probability  $p_{1-3}$ , and set it equal to 0.25, as discussed earlier. Again, using the BGL parameters from the Tan et al. (1994a, b) data in Equation (6), we calculate the proportionate drop in the success probability for each of attempts 4–7, compared to the average success probability on attempts 1–3. Applying this proportionate drop to the value  $p_{1-3} = 0.25$ , the success probabilities on cycles 4–7 become 0.204, 0.180, 0.159, and 0.142, respectively.

With this parameter set, 60% of those receiving a refund will go on to pay a la carte for IVF cycle 4. Of this 60%, approximately 20.4% are expected to succeed on cycle 4, so 79.6% of the 60% (i.e., 48%) of the refund receivers will go on to fail on cycle 4. And of these, again, 60% (i.e., 29% of all the refund receivers) will go on to attempt IVF cycle 5, and so on. The number of patients going on from cycle to cycle after the refund, therefore, decays with retention rate  $(0.6)(1 - p_t)$  for each cycle  $t$ .

The economic consequence of these later a la carte cycles for a clinic is then easy to calculate, because the typical contribution margin per cycle is (price cost) = \$7,500 – \$6,000 = \$1,500. Considering only IVF cycles 4–8, with the failure/perseverance pattern described above, the average number of additional a la carte cycles pursued by a refund receiver is 1.13 cycles. Multiplying this figure by the \$1,500 contribution per cycle means an additional \$1,695 economic contribution per refund receiver.

This last figure must be adjusted before tallying it against the current \$2,203 loss per patient described at the beginning of this section. The only patients

who may proceed to a la carte are those who receive a refund, so the \$1,695 figure must be reduced by that proportion. With  $p = 0.25$ , approximately 42.2% of couples will fail through three tries and receive the refund. So, the economic contribution of a la carte cycles 4–8 per patient who begins the guarantee program is  $\$1,695 \times 0.422 = \$715$ . This positive amount reduces the per-patient loss from the guarantee to  $\$2,203 - \$715 = \$1,488$ .

The money-back guarantee still does not make money without one last factor.

## 7. Abandoning the “No First Use” Policy

We noted at the outset of the paper that IVF has become the “last best hope” for many infertile couples. Figure 1 highlighted the treatments that typically have preceded IVF: 6–12 months attempting natural conception, drug therapy, and intrauterine insemination (IUI) being common. Doctors would tend to recommend IVF only after these options for two reasons. First, IVF is more invasive and carries some risks (including greater risk of multiples). Second, IVF is expensive and not often covered by insurance in the United States. While this sequencing of treatment may minimize health risks and economic cost, it has not minimized emotional costs for the patient. While Figure 1 shows that across the entire sequence preceding IVF, many couples can achieve a child without needing in vitro, in fact, incurring “failure” month after month can be emotionally debilitating (Golombok 1992). Further, during the time that patients pursue these treatments, they may be concerned about their own aging—e.g., a 38-year-old couple sees background pregnancy rates that virtually fall off the table for 40 to 44-year-old couples relative to ages 35 to 39.

Enter into this process the IVF money-back guarantee, with mass media advertising directed toward patients rather than doctors. For those willing to spend \$15,000, it emphasizes the complete lack of (economic) risk. It holds the promise of a baby now, rather than (maybe) a few years from now. It minimizes the emotional cost in repeated failure. It is being marketed no longer as the “last hope” but,

instead, with the positioning: Why wait? For infertility clinics that offer the IVF guarantee, accessing the large pool of patients in earlier stages of treatment—who are more fertile than IVF’s traditional patients—can, in fact, make the money-back guarantee programs earn money. The assumptions we require for this calculation may not apply to all clinics, but they are reasonable and provide an explanation for the guarantees that have, until now, been absent.

Consider the following scenario for earlier use of IVF in infertility treatment:

(1) Couples wishing a child pursue natural conception for six months.

(2) If no natural conception in six months, half the couples pursue drug therapy (the common progression, see Figure 1) and half take the IVF guarantee.

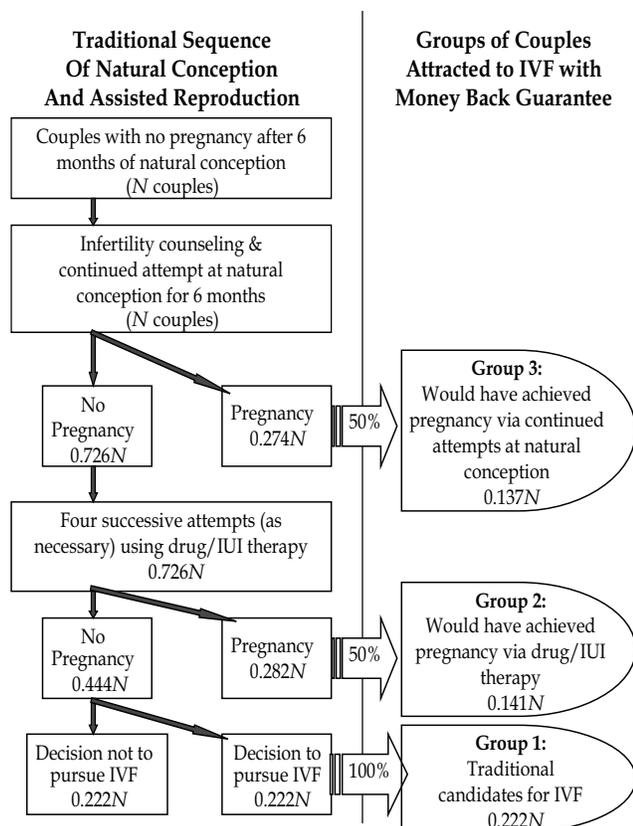
(3) For those pursuing drug therapy, if no success in two attempts, half pursue IUI (the common historical progression) and half take the IVF guarantee.

(4) For those pursuing IUI, if no success in two attempts, half pursue IVF. The other half pursue some other option (e.g., adoption).

This sequence is illustrated in Figure 7. Of course, other assumptions are possible. IVF clinic guarantees may attract couples even earlier in the natural conception process. The IVF guarantee may take more than half the drug/IUI patients—after all, the latter treatments may cost some patients, offer a lower probability of success and, accordingly, a high likelihood of emotional distress, and delay the arrival of a child. Finally, less than 50% may choose to pursue IVF, either due to counseling regarding the source of infertility, or because of the \$15,000 cost under the guarantee. *Each* of these alternative assumptions will *increase* the clinic’s economic reward for pursuing infertility patients earlier. So in this sense, the calculation below is a conservative estimate of the payoff.

We need one more assessment to calculate the financial implications here: the success rate for IVF per cycle in the more fertile populations. In this assessment, it is useful to consider the steps required for an ongoing pregnancy, i.e., production of viable egg and sperm, egg fertilization, embryo implantation, and sustained ongoing pregnancy. In couples that could conceive via natural conception, drug/IUI therapy, by far, the weakest link in this chain of events

Figure 7 Event Tree for Sequential Use of Assisted Reproductive Procedures

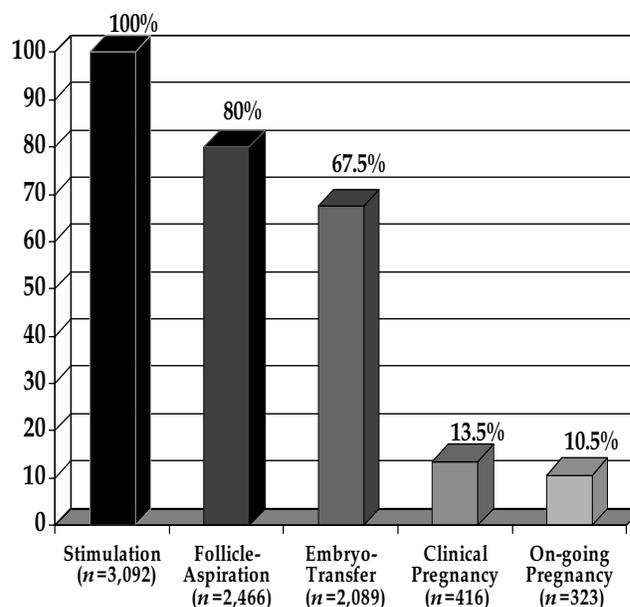


is fertilization. On the other hand, fertilization is the strong point of IVF. As Figure 8 clarifies, two-thirds of couples using IVF produce at least one fertilized, growing embryo. For these relatively infertile couples, the *weak* point of IVF is implantation—this is the stage where more than half the IVF cycle failures arise. Implantation for such patients is often difficult due to the source of infertility, e.g., endometrial dysfunction. Yet, this is a relatively successful stage for more fertile couples, i.e., those who could conceive via natural conception or IUI.

Specifically, we assume that couples who would have pursued months 6–12 of natural delivery (after six months without a conception) would have an IVF success rate of 50% per cycle. For couples that do not conceive naturally and would have pursued drug/IUI therapy, we assume a 40% per cycle IVF success rate.

To calculate the effect of pursuing infertility patients earlier, let  $N$  represent the number of couples that

Figure 8 Patient Progress Through Successive Stages Within an IVF Cycle



Source. Adapted from Haan et al. (1991a).

complete six months of natural conception without a pregnancy. The sequence of reproductive procedures chosen by the couples, and outcomes experienced, are illustrated in Figure 7. Prior to the IVF guarantee programs, the statistics in Figure 1 indicate that about 0.274N would achieve a delivery during an additional six months of natural conception attempts. We imagine that the IVF guarantee would capture 50% of these patients, or 0.137N (“Group 3” in Figure 7).

This leaves 0.726N patients who complete 12 months without a delivery and go on to drug/IUI therapy. Again, via the statistics in Figure 1, after four attempts, an additional 0.282N would have had a baby—and we assume that the IVF guarantee captures 50% of them or 0.141N (“Group 2” in Figure 7). Finally,  $0.726N - 0.282N = 0.444N$  would not achieve a birth after the drug/IUI sequence, and we assumed that half these—or 0.222N—go on to IVF (“Group 1” in Figure 7).

Under the guarantee, with the impact of attracting earlier infertility patients, the clinic’s population of IVF-guarantee patients is, therefore, composed of three groups:

(1) Group 1, of size  $0.222N$ , is the “usual” IVF patients who have repeatedly failed to conceive naturally or with drug/IUI therapy, and who, accordingly, have success probability on IVF cycles 1–3 of 0.25 (with success on later cycles 4–8 as described in the previous section).

(2) Group 2, of size  $0.141N$ , is composed of patients who have repeatedly failed to conceive naturally, and who *would* have conceived via drug/IUI therapy, but elect to go to IVF instead, and who, accordingly, have success probability = 0.4 on each of IVF cycles 1–3.

(3) Group 3, of size  $0.137N$ , is composed of patients who did not conceive in six months of natural conception, but who *would* have conceived in another six months of such attempts, but who elect to go with the IVF guarantee instead, and who, accordingly, have success probability = 0.5 on each of IVF cycles 1–3.

For Groups 2 and 3, the success probability on each of cycles 4–7 is envisioned to fall off relative to cycles 1–3 by the same proportionate decline as for Group 1, using the BGL model and parameters as in the last section. Then, for each group, the per-patient profit (or loss) from the IVF guarantee (including the possibility of IVF cycles 4–8) can be calculated as in the last section (para 6 of §6 above), together with Equation (1). Doing so, the average return per patient for each group is

$$\begin{array}{ll} \text{Group 1: } -\$1,488, & \text{Group 2: } +\$3,328, \\ & \text{Group 3: } +\$5,802. \end{array}$$

The economic consequence of the IVF guarantee is then just the weighted average of the within-group returns, the weights being the relative sizes of the three groups listed above. (Group 1 accounts for 44% of IVF patients, Groups 2 and 3 account for 28% and 27%, respectively.) With our scenario for attracting patients to IVF earlier in their fertility treatment, this overall return is positive at last, at \$1,844 per patient (approximately a 13% return on average cost).<sup>3</sup>

<sup>3</sup> The key to a positive economic return is obviously the high IVF success rate for both Group 2 and Group 3 couples. While we believe that the success rates assumed (Group 3 = 50%; Group 2 = 40%) are reasonable and, indeed, conservative, we note that the money-back guarantee is profitable even if both groups' success rates are reduced to as low as 35%, which is a 10% “premium” over the 25% success rate for Group 1.

The financial effect of pursuing these more-fertile couples is obviously dramatic and positive for clinics. From the patients' standpoint, most of the members of Groups 2 and 3 will conceive via IVF, and “soon,” indeed, on the first or second try. They are already reconciled to spending \$15,000 for a child, so have essentially no regret in pursuing IVF “early.” They as a group are more likely to experience multiple births, but at least in the case of twins, many couples will view this as a “bonus.” In expected value terms they paid collectively more than they needed to, but are unlikely to complain.

What about Group 1? These are the couples who historically made it to IVF. The money-back guarantee lets them pursue this course without extreme economic hardship. About half of Group 1 will not, in fact, succeed in three IVF attempts, but they will get their money back. Of course, whether these patients keep that money for other uses or “let it ride” on more IVF cycles remains to be seen.

Finally, we note that our analysis has focused on explaining the economic viability of money-back guarantees for IVF as they were becoming widely available, i.e., in 1997 and 1998 (based on published success rate data from 1996). This is not, however, simply a historical exercise. In 2002, the most recent IVF success rates published for North American clinics cover the year 2000, during which the overall success rate was 25% (U.S. Centers for Disease Control and Prevention 2002b), a modest 3 percentage point increase over the 22% figure from 1996 used in our analysis. The additional 3 percentage points do not, of course, explain the continued viability of money-back guarantees for these clinics, and the clinics have persisted in offering the guarantees during this period. In short, our analysis and conclusions regarding the guarantees is as applicable now as it was in 1997–1998 when they were being launched.

## 8. Conclusion

We conclude by examining the implications of our analysis for patients, clinics, and regulatory agencies or public policy advocates.

**Patients.** For the Group 1 patients (typical IVF patients pre-money-back guarantees), the implications are clear. If they pursue IVF, they should take the money-back guarantee. For these patients who have already tried less invasive procedures, their expected cost will be dramatically less than the a la carte option. To see this, we start with the same model (1) that calculated the clinic's expected profit. Under the a la carte method of  $\$A$  per cycle, the patient pays  $A$  if the first cycle is successful,  $2A$  if success occurs on the second cycle, and  $3A$  otherwise. (We are assuming that the couple deciding between the a la carte and money-back guarantee commits to three cycles.) Thus, the expected cost to the patient under a la carte is

$$Ap + 2Ap(1 - p) + 3A[1 - p - p(1 - p)].$$

Under the money-back guarantee with an upfront payment of  $G$ , the couple pays  $G$  only if there are not three consecutive failures, so the expected cost to the patient is

$$G[1 - (1 - p)^3].$$

Equating these two expected costs and rearranging terms gives the cubic equation (in  $p$ )

$$(Gp - A)(p^2 - 3p + 3) = 0. \quad (7)$$

The only relevant (real) solution to Equation (7) is the root from the first term:  $p = A/G$ .

With the monetary values used today of  $A = \$7,500$  and  $G = \$15,000$  the break even value is  $p = 0.5$ . Any couple with a  $p$ -value less than 0.5 should opt for the money-back guarantee. For the risk-averse couple who would be very unhappy going a la carte for three cycles, paying  $\$22,500$ , and getting no baby, the break even  $p$ -value will be even higher than 0.5. In any event, the typical Group 1 traditional patient for IVF should most definitely take the money-back guarantee. Their best guess  $p$ -values will be far below the nonrisk averse break even value of 0.5.

The typically younger and less infertile Group 2 and Group 3 patients need to assess their personal  $p$ -value to some appropriate level above the Group 1 value. (There is no single obviously correct way to do this, but §7 gives some suggestions.) These patients should then decide if the increase in the chance of having a baby *now* is worth the potential side effects

of IVF and the increased likelihood of having multiple births. This is a subjective trade-off, but couples should at least explicitly consider these pros and cons of doing IVF "early." Assuming a Group 2 or Group 3 couple decides on IVF now, they should use their adjusted  $p$ -value to decide the expected costs of the a la carte versus money-back guarantee option. Those patients who opt for the money-back guarantee and receive a refund should explicitly consider the economic, psychological, and side-effect costs yet again before treating that refund as "house money."

Of course, all of the above is predicated on patients gaining as much information as possible on the success rates for the IVF clinics that are geographically feasible for them. They should be especially wary of high success rates based on small sample sizes or high success rates where the prevalence of multiple births is above average. The couple should also find out as much as possible on how aggressively each clinic screens its patients, i.e., what kind of patient pool produced the clinic's historical success rate.

**Clinic Managers.** The clinic manager needs to estimate the overall success rate per cycle for patients the clinic is currently attracting, and this is especially true if the clinic offers the money-back guarantee. The clear-cut economic questions should be addressed via Equation (1). The key ethical considerations for a clinic follow. Are patients given data that are appropriate for their particular situations? Are the downsides of IVF vis-à-vis less invasive procedures adequately addressed? Are the patients who receive their money back being pressured to "let it ride" on additional a la carte IVF attempts?

**Public Policy Advocates.** Organizations such as SART and RESOLVE that provide information to infertility patients play an important role, that is increasingly challenging with new payment options and expanded competition. The data, model, projection, and scenario analyses in this paper can help in patient education and, indeed, a menu-driven interactive computer version of this work would be valuable for patients. It is also important to expand monitoring of clinics. Some of the key variables—e.g., success rate by age, incidence of multiple births—have been comprehensively collected. Other key variables—e.g.,

success rates by previous infertility treatment history, pricing policy, and number of IVF attempt for the patient—are not available clinic by clinic. This information would add incentive for the clinics to “do the right thing.”

**The Simple Model Works.** Finally, we stress the need for formal probabilistic models to analyze the existing IVF data. These models give insights not available with the usual summary statistics. But perhaps the best news from our efforts is that the simplest possible model, Equation (1), “works.” That is, assuming that each patient has the same baseline success probability (wrong) and that this probability stays unchanged on successive cycles (wrong) leads to a good economic analysis of the money-back guarantee. The heterogeneity of these base rates across patients (which drives down the aggregate probability of success on repeated attempts) is counterbalanced by the individual patient “learning” (which drives up the aggregate probability of success on repeated attempts), at least for the first three cycles. Clinics know their cost per cycle,  $C$ , and both the clinics and the patients know the up-front money-back guarantee fee,  $G$ , and the a la carte price,  $A$ . Because the homogeneous (Bernoulli) model in Equation (1) provides a good approximation to reality, then the clinics need to know only their overall success rate  $p$  to see if the money-back guarantee will be profitable.

For those deciding between the a la carte and money-back guarantee, only  $A$ ,  $G$ , and their personal  $p$ -value matter. The break even  $p$ -value is just  $A/G$ . With coaching the couple can estimate their personal  $p$ -value. Even just telling couples the  $A/G$  break even point and a  $p$ -value for “couples like them” would improve on current patient education.

### Acknowledgments

The authors appreciate helpful comments contributed by Eric Bradlow, Barbara Kahn, Mary Frances Luce, and the participants of the 1998 UCLA Winter Marketing Camp.

### References

Alsallil, M., A. Yuzpe, I. Tummon, J. Parker, J. Martin, S. Daniel, M. Rebel, J. Nisker. 1995. Cumulative pregnancy rates and pregnancy outcomes after in-vitro fertilization: >5000 cycles at one centre. *Human Reproduction* **10** 470–474.

- Check, J. H., A. Baker, D. Lurie, K. Benfer, C. Callan. 1994. Comparison of the cumulative probability of pregnancy after in vitro fertilization—Embryo transfer by infertility factor and age. *Fertility Sterility* **61** 257–261.
- Crowder, M. J. 1978. Beta-binomial ANOVA for proportions. *Appl. Statist.* **27**(1) 34–37.
- Davis, K. 1996. The agonizing price of infertility. *Kiplinger Online* (May) www.kiplinger.com, 1–7.
- Dawood, M. Y. 1996. In vitro fertilization, gamete intrafallopian transfer, and superovulation with intrauterine insemination: Efficacy and potential health hazards on babies delivered. *Amer. J. Obstetrics Gynecology* **174** 1208–1217.
- Diczfalusy, E., P. G. Crosignani. 1996. Introduction: From reproductive endocrinology to reproductive health. *Human Reproduction* **11** 1776–1777.
- European Society for Human Reproduction and Embryology (ESHRE) Capri Workshop. 1996. Infertility revisited: The state of the art today and tomorrow. *Human Reproduction* **11** 1779–1807.
- Freudenheim, M. 1998. Aetna is reducing fertility benefits: Advanced treatment coverage is eliminated by a pioneer. *The New York Times* (January 10) A1–A11.
- Friedler, S., S. Mashiach, N. Laufer. 1992. Births in Israel resulting from in-vitro fertilization/embryo transfer, 1982–1989: National Registry of the Israeli Association for Fertility Research. *Human Reproduction* **7** 1159–1163.
- Gleicher, N., B. VanderLaan, D. Pratt, V. Karande. 1996. Background pregnancy rates in an infertile population. *Human Reproduction* **11** 1011–1012.
- Golombok, S. 1992. Psychological functioning in infertility patients. *Human Reproduction* **7** 208–212.
- Greene, J. D. 1982. *Consumer Behavior Models for Non-Statisticians*. Praeger, New York.
- Griffiths, D. A. 1973. Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of a disease. *Biometrics* **29** 637–648.
- Haan, G., F. Rutten. 1989. No cure, no pay: An acceptable way of financing fertility treatments? *Health Policy* **13** 239–249.
- , R. van Steen. 1992. Costs in relation to effects of in-vitro fertilization. *Human Reproduction* **7** 982–986.
- , R. E. Bernardus, J. M. G. Hollanders, B. A. Leerentveld, F. M. Prak, N. Naaktgeboren. 1991a. Results of IVF from a prospective multicentre study. *Human Reproduction* **6** 805–810.
- , ———, ———, ———, ———. 1991b. Selective drop-out in successive in-vitro fertilization attempts: The pendulum danger. *Human Reproduction* **6** 939–943.
- Hershlag, A., A. H. DeCherney, E. H. Kaplan, G. Lavy, R. A. Loy. 1991. Heterogeneity in patient populations explains differences in in vitro fertilization programs. *Fertility Sterility* **56** 913–917.
- Hull, M. G. R. 1992. Infertility treatment: Relative effectiveness of conventional and assisted conception methods. *Human Reproduction* **7** 785–796.

- Hyman, D. A., C. Silver. 1998. Letter to the editor. *J. Law, Medicine Ethics* **26**(1) 94–95.
- Johnson, N. L., S. Kotz. 1970. *Continuous Univariate Distributions*, Vol. 2. John Wiley and Sons, New York.
- Maritz, J. S. 1970. *Empirical Bayes Methods*. Methuen, London, U.K.
- Murray, T. H. 1997. Money-back guarantees for IVF: An ethical critique. *J. Law, Medicine Ethics* **25** 292–294.
- Norris, F. 1997. The many lives made miserable by a low-down high-roller. *The New York Times* (March 9) E7.
- RESOLVE. 1997. *RESOLVE National Newsletter* **22**(1).
- Robertson, J. A., T. J. Schneyer. 1997. Professional self-regulation and shared-risk programs for in vitro fertilization. *J. Law, Medicine Ethics* **25** 283–291.
- Rubin, A. J., A. Zitner. 2000. Infertility cases spur an illicit drug market. *Los Angeles Times* (September 10) A20, A22.
- Schmittlein, D. C. 1989. Surprising inferences from unsurprising observations: Do conditional expectations really regress to the mean? *Amer. Statist.* **43**(3) 176–183.
- Society for Assisted Reproductive Technology and The American Society for Reproductive Medicine. 1996a. Assisted reproductive technology in the United States and Canada: 1994 results generated from the American Society for Reproductive Medicine/Society for Assisted Reproductive Technology Registry. *Fertility Sterility* **66** 697–705.
- . 1996b. Clinic specific report for 1994. Society for Assisted Reproductive Technology and the American Society for Reproductive Medicine, Birmingham, AL.
- Stolberg, S. G. 1997. For the infertile, a high-tech treadmill. *The New York Times* (December 14) 1, 36.
- Stolwijk, A. M., C. J. C. M. Hamilton, J. M. G. Hollanders, L. A. Bastiaans, G. A. Zielhuis. 1996. A more realistic approach to the cumulative pregnancy rate after in-vitro fertilization. *Human Reproduction* **11** 660–663.
- Strictly Business. 1997. *The Newshour with Jim Lehrer* (January 27). Overland Park, KS.
- Tan, S.-L., P. Doyle, N. Maconochie, R. G. Edwards, A. Baken, J. Bekir, P. Brinsden, S. Campbell. 1994a. Pregnancy and birth rates of live infants after in vitro fertilization in women with and without previous in vitro fertilization pregnancies: A study of eight thousand cycles at one center. *Amer. J. Obstetrics Gynecology* **170** 34–40.
- , N. Maconochie, P. Doyle, S. Campbell, A. Balen, J. Bekir, P. Brinsden, R. G. Edwards, H. S. Jacobs. 1994b. Cumulative conception and live-birth rates after in vitro fertilization with and without the use of long, short, and ultrashort regimens of the gonadotropin-releasing hormone agonist buserelin. *Amer. J. Obstetrics Gynecology* **171** 513–520.
- Thaler, R. H., E. J. Johnson. 1990. Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice. *Management Sci.* **36** 643–660.
- Trafford, A. 1997. Medicine's money back warranty. *The Washington Post* (August 5) Z06.
- U.S. Centers for Disease Control and Prevention. 2002a. 1996 assisted reproductive technology success rates national summary and fertility clinic reports. [www.cdc.gov/nccdphp/drh/art96/](http://www.cdc.gov/nccdphp/drh/art96/).
- . 2002b. 1999 assisted reproductive technology success rates national summary and fertility clinic reports. [www.cdc.gov/nccdphp/drh/art99/](http://www.cdc.gov/nccdphp/drh/art99/).
- Wagner, M., P. St. Clair. 1989. Are in-vitro fertilization and embryo transfer of benefit to all? *Lancet* **ii** 1027–1030.
- Yovel, I., E. Geva, J. B. Lessing, Y. Yaron, A. Botchan, A. Amit. 1994. Analysis of the fourth to eighth in-vitro fertilization treatments after three previously failed attempts. *Human Reproduction* **9** 738–741.
- Zhou, H., C. R. Weinberg, A. J. Wilcox, D. D. Baird. 1996. A random-effects model for cycle viability in fertility studies. *J. Amer. Statist. Assoc.* **91** 1413–1422.

*Accepted by Jon Caulkins, former department editor; received March 1999. This paper was with the authors 2 years for 1 revision.*

Copyright 2003, by INFORMS, all rights reserved. Copyright of Management Science is the property of INFORMS: Institute for Operations Research and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.