A Live Baby or Your Money Back:
The Marketing of *in vitro* Fertilization Procedures

David C. Schmittlein¹
Donald G. Morrison

May 4, 1999

¹ David C. Schmittlein is the Ira A. Lipman Professor and Professor of Marketing at The Wharton School, University of Pennsylvania, Philadelphia, PA 19104-6371.
Donald G. Morrison is the William E. Leonhard Professor at the Anderson Graduate School of Management, University of California at Los Angeles, Los Angeles, CA 90095-1481.

The authors appreciate helpful comments contributed by Eric Bradlow, Barbara Kahn, Mary Frances Luce, and the participants of the 1998 UCLA Winter Marketing Camp.
A Live Baby or Your Money Back:  
The Marketing of in vitro Fertilization Procedures

Abstract

A large number of clinics that offer in vitro fertilization (IVF) have begun to aggressively market the following options to couples seeking to have a genetically related baby: (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 up-front 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.

If the couple contemplating these two choices knew the success probability for each attempt, then a simple analysis would show whether the a la carte or money-back-guarantee program is better. Unfortunately, it is difficult for couples to gather the relevant data and even more daunting to adjust the aggregate data to their own situation.

In this article we assemble the most recent available data and develop a model that allows patients, clinics, and public policy advocates to assess the a la carte vs. money-back-guarantee programs. The most surprising result of our analysis is that the money-back-guarantee program appears (for the patients) to be “too good to be true.” That is, with reasonable projections from the most recent data, 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 turns out to be the better option for all couples with less than 0.5 success probability per cycle. (The breakeven probability is even higher if risk aversion is considered.) Virtually all traditional IVF patients can be considered to have per-cycle success probabilities well below 0.5. Can it be that clinics are offering money-back-guarantees that both lose money for the clinics and give the patients a deal that is far better for them than the traditional a la carte payment method?

After a detailed analysis of the key variables – i.e., success rate per attempt, heterogeneity of couples’ base rates of success, individual couples’ “learning” on successive attempts, and cost to the clinic per attempt – nothing makes these money-back-guarantees profitable for the clinics. Since presumably clinics are not in business to lose money, the analysis must be missing something major. Based on the kind of aggressive marketing (e.g., mass media) for the money-back guarantees, we believe it is bringing in younger and less infertile patients than those who were in IVF clinics prior to 1996. In other words, the marketing of money-back guarantees may be inducing couples who would previously have used – successfully – other less invasive procedures with fewer potential side effects and less risk of multiple births to decide, “Why wait: let’s jump to IVF now and we’ll get our money back if it doesn’t work.” We show that under this scenario the money-back-guarantees can be profitable for the clinics.

The implications of earlier use of IVF are then considered for patients, clinic managers, and from an overall public policy point of view. Although the clinics that make profits and the couples who either receive a baby or their money-back are unlikely to complain, there are some significant downsides to the marketing of IVF money-back-guarantees that need to be understood. Just as mothers everywhere tell their children, “When something looks too good to be true, then it is too good to be true!”
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. *In vitro* procedures are often the last chance for a child genetically related to the parents, and so represent a final step leading to one of the following: a live birth, pursuit of adoption, or accepting the prospect of remaining childless.

Since *in vitro* is rarely covered by health insurance in the U.S. (Davis 1996, Freudenheim 1998), Assisted Reproductive Technology (ART) clinics are typically chosen by a couple, rather than directed to a clinic by a managed health care provider (e.g. HMO). Until recently, the couple would expect to pay approximately $7,000 for each *in vitro* attempt, and the chance of a live healthy birth from that procedure would be 10-20 percent. 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: namely product performance and price.²

From the customer (patient) standpoint the most important product performance measure is the live birth probability per initiated IVF cycle. North American clinics have reported this information for publication, and use it in informational sessions to “educate” prospective patients about the clinic. One of our goals in this article is to show what such
performance measures say (and do not say) about a couple’s prospects for success with a particular clinic. Such an inference is not only important to prospective consumers (patients), but also to the clinic’s managers in determining pricing policy. To our knowledge this kind of analysis has not been undertaken to date. We find that clinics do differ substantially from each other, not only in historical success rates, but in the 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 the beginning of 1997 many ART clinics have begun to offer as an option to patients “volume discounts” and “money-back guarantees”; i.e. “a live baby or your money back”, in addition to the option of paying separately 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. Over 60 clinics are believed to offer similar incentives to prospective customers (Trafford 1997). Considering the probabilistic prospects for success, these pricing innovations have attracted significant attention from consumers.

In light of the medical decision context, the complicated economic incentives for the ART clinics, the asymmetry in diagnostic information between these clinics and

---

2 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, etc. Distribution decisions trade off customer convenience of multiple locations with cost efficiency of service provision from one or a few central locations.
prospective patients, the amounts of money involved, and the emotional state of some infertile couples, these pricing “innovations” have generated great controversy in both professional/academic publications (Hyman and Silver 1998; Murray 1997; Robertson and Schneyer 1997) and in popular media (Freudenheim 1998; Norris 1997; Trafford 1997). For instance *The Newshour with Jim Lehrer* on U.S. public television devoted a segment in January 1997 to such pricing practices.

In this article we show how to estimate the economic consequences of guarantees and volume discounts for any given clinic. Doing so requires both an appropriate stochastic model for clinic success rate (i.e. relative to other clinics) and a stochastic model of IVF success for any given couple on successive IVF attempts. In particular 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 will also analyze the main factors that affect the economics of the new money-back/volume-discount policies for clinics. Some of the factors are incentive compatible with the goals of patients - and others are not. Prospective patients can, through specific questioning, get a sense of which factors are driving the pricing policy of any particular clinic. For clinic managers our analyses provide a concrete mechanism for integrating cost and history information to set a sensible pricing program going forward.

Using an analysis of clinic performance statistics, and of patients’ success rates on successive IVF cycles, we will show that – in one sense at least – these guarantees are “too good to be true.” Our analysis suggests that volume discounts and guarantees of the sort described above will only be economically viable for clinics that pursue *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, for the average current patient couple.

2. The guarantees are in fact 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.

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.

5. Accordingly, a large economic transfer from more fertile couples to less fertile couples is beginning.

6. Given couples’ desire for a baby, and the risk aversion of such patients, it seems unlikely that they will complain about a “guarantee” program that could be viewed as a price discrimination device or an expensive insurance policy.

The next section describes the prevalence and practice of in vitro fertilization procedures. This is followed by a straightforward economic assessment of the “quantity discount plus money-back guarantee” policy for a typical clinic, which shows how simple economies of scale do not appear to make these offers viable. We next examine empirically the two key factors that might be imagined to make such offers viable:
heterogeneity in clinic performance and heterogeneity/learning effects for couples making successive attempts at IVF. Finally, we consider the factor that does appear to be responsible for these offers: namely, targeting a new customer/patient base that is characterized by much higher IVF success rates.

2. THE PREVALENCE AND PRACTICE OF IN VITRO FERTILIZATION

The Scope and Scale of IVF

Over its twenty years of experience IVF has developed from an exotic, complex, expensive procedure available in a handful of clinics worldwide whose success rate was very low, to one that is widely accepted, complex, expensive, available in a dispersed set of clinics, and having a modest success rate. This can of course be seen as a glass either “half empty” or “half full”. That is, in 1996 the approximately 300 clinics in the U.S. and Canada that performed IVF accounted for over 45,500 “standard” IVF cycles, i.e. cycles using fresh, non-donor eggs (U.S. Centers for Disease Control and Prevention 1999). The number of babies born as a result of these IVF cycles was about 10,000 in 1996. This represents both the “bad news” and the “good news”: i.e. over 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 for them in the absence of the IVF procedure. (The natural-conception birthrate per cycle for couples classified as infertile is only about 1.6 percent (Gleicher 1996).) In the U.S. alone in vitro fertilization has accounted for over 40,000 babies since it began to be performed (Strictly Business 1997; U.S. Centers for Disease Control and Prevention 1999). Worldwide, over 130,000 IVF cycles were
conducted during 1993, leading to over 17,000 deliveries (ESHRE Capri Workshop 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 delivery (live birth) probability. This success rate has increased from about 6 percent in the early 1980’s to about 22 percent in 1996 (Society for Assisted Reproductive Technology (SART) and The American Society for Reproductive Medicine (ASRM) 1996a; 1996b; U.S. Centers for Disease Control and Prevention 1999). On one hand this threefold increase is impressive and clearly accounts for part of the growth in IVF. On the other hand, a substantial majority of couples undergoing a particular IVF cycle will not see a live birth delivery result. Further, taking into account the typical number of IVF cycles pursued by a couple, over half of them complete their attempts at IVF without taking home a child.

Accompanying each IVF cycle’s very uncertain outcome is a very substantial cost, typically in the range of $6,000 to $8,000. Couples often commit themselves to multiple IVF cycles should they be needed, so a total cost in the range of $10,000-20,000 is common. In some European countries and in Israel national health insurance pays for a certain number of IVF tries, but coverage of IVF by U.S. private health insurers is usually either very 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 at one time (e.g. a home purchase; possibly an automobile or a college degree) 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 prospect of remaining childless and the economic risk above,
other factors also make IVF decisions difficult for patients. These include mood side-
effects due to drugs taken during the IVF process, the risk of multiple conception (twins,
triplets etc.) 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 successfully (Golombok 1992; Stolberg 1997; Strictly
Business 1997) but success rates published for IVF do not appear to drop substantially
after several failures for a particular couple. Clinics routinely report this finding to those
considering an additional cycle (Haan et al. 1991b), making it difficult to decide to stop at
any point. Furthermore, after one or more unsuccessful IVF tries, an additional attempt
offers the hope to make all the sunk costs to date (from previous attempts) worthwhile
(loss aversion).

**IVF Decision Process For Clinics**

From the standpoint of assisted reproduction clinics IVF and a few related
procedures (GIFT, ZIFT, ICSI) represent a large and growing service opportunity. In
1994 the amount spent on such procedures in the U.S. and Canada was approximately
$300 million. IVF procedures are an important source of revenue for such clinics. They
have been growing in popularity and are likely to continue to, since it is estimated that 10-
15 percent of all married couples in the reproductive age group are infertile (Tan et. al.
This means a market, for all assisted reproductive procedures, of 6 million couples in the U.S. and 60-80 million globally (ESHRE Capri Workshop 1996, p.1779; 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 healthcare delivery 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.

To assess the economics of pricing programs for IVF clinics two kinds of data are required, i.e., clinic-by-clinic overall performance rates, and the prospects for IVF success on repeated attempts for a single couple. Individual U.S. and Canadian clinics report performance information (i.e. live birth rates) annually in a standardized format. We will base our analysis on the 1994 data (collected in 1995) and the 1996 data (SART and ASRM 1996b; U.S. Centers for Disease Control and Prevention 1999). The latter contain the most recent statistics published to date. Unfortunately, clinic-specific reports for 1995 and 1996 collapse IVF with other ART procedures (ZIFT, GIFT) and so are not as valuable for considering the economics of guarantees that apply only to IVF. Accordingly some of our individual-clinic analyses will use the 1994 data. The leading infertility support organization in the United States (RESOLVE) promotes access to these statistics in its member newsletter, i.e. to couples considering IVF (RESOLVE 1997, p.14). To the best of our knowledge our statistical analysis of clinic performance is the first such use of these data.

For the second set of information, i.e. prospects for a live birth on successive IVF attempts for a single couple, we will rely on several published studies that report this
information for a cohort of patients (Check et al. 1994; Tan et al. 1994a; Yovel et al. 1994). We also examine a similar outcome – the prospect for an ongoing pregnancy – for successive IVF attempts (Haan et al. 1991b; Herschlag et al. 1991; Stolwijk et al. 1996). While none of these studies individually offer a very reliable base of experience for IVF attempts 3, 4, 5, etc. for a single couple, taken together they will enable us to draw some fundamental conclusions about a patient’s success prospects over time.

3. MONEY-BACK GUARANTEES:
ECONOMICS FOR A TYPICAL PATIENT AND CLINIC

We have already presented most of the information needed for a simple analysis of the expected financial consequences for standard money-back guarantees on IVF. Based on both popular press descriptions of these guarantees and on clinic advertising copy, they have essentially the same features for all clinics. Specifically, the patient is offered the choice of paying “a la carte”, about $7,500 per attempt, or accepting the “guarantee.” The latter has the patient pay about $15,000, which covers 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.

A clinic’s average cost per IVF cycle was estimated to be $5,000 in 1988 (Wagner and St.Clair 1989). Since that time a 2 percent annual nominal cost increase (or equivalently a 2 percent annual real cost decrease) would lead to an IVF cost per cycle of $6,000 by 1998. This cost figure, combined with the recent past’s a la carte price of $7,500, leads to a return on investment (actually, return on cost) of 25 percent. This is a
high return in light of the substantial competition between clinics, and accordingly it seems safe to conclude that cost per cycle in 1998 has risen to at least $6,000 per cycle.

With notation \( C = \text{clinic’s cost per cycle} \), \( p = \text{success probability on any IVF attempt} \), \( \pi = \text{profit (or loss)} \), and \( G = \text{the patient’s payment for the guarantee} \), the clinic’s expected profit resulting from the guarantee, the expected profit is:

\[
\]

With \( C = 6,000 \), \( G = 15,000 \), \( p = .22 \) (as discussed in the previous section), expected profit for the clinic is minus $6,448 per patient. Via this calculation the money-back guarantee is not even close to breakeven, let alone profitability. To put this in perspective, recall that ROI was 25 percent for patients paying a la carte (with \( C = 6,000 \)).

Under the guarantee (with \( p = .22 \)), 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 incurred cost under the guarantee is \(-6,448/14,400\), or minus 45 percent.

Our simple formula (1) assesses financial outcomes only for the IVF cycles themselves. A particular new patient at an IVF clinic may also be required to undergo some testing, whose cost is separate from the guarantee (and nonrefundable). Discussion with clinic operators suggests that the cost of such testing could approach $4,000, and a
generous estimate of the profit on such testing is $3,000 per patient. 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 potential $3,000 gain on testing; leaving a net expected loss per new IVF patient of $3,448.

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

While altruism on a grand scale or gross mismanagement of many clinics could explain this conundrum, we are interested in examining other potential explanations for the guarantee policy, and in their consequences for patients and clinics. Many of the “natural” possible explanations can be ruled out readily. 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 versus 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. $C = 6,000$ is a conservative assessment of average cost,

2. Only about 10 percent of the clinic cost is equipment or location (Haan and van Steen 1992) and in fact the trend is to individualize equipment (e.g. embryo incubator) to the specific patient,
3. Of the total IVF clinic costs 40 percent is material/lab tests/drugs (Haan and van Steen 1992), which are not subject to many economies of scale,

4. Of total clinic costs, 50 percent 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. These clinics also offer the broad array of infertility treatments (clomiphene, HCG, IUI, GIFT, ZIFT, ICSI, donor egg) in addition to IVF. So additional IVF cycles of any magnitude will require hiring additional personnel, and clinics are finding that many patients are opting for the guarantee.

Accordingly, relevant marginal costs, even in the face of increased patient volume created by the guarantee, are not likely to be meaningfully below the $6,000 figure used in equation (1).

**Patient Selection: Aggressive Screening**

Another possible enhancement to profit from the guarantees is limiting the guarantee program only to patients whose likelihood of IVF success is “high”. In fact, clinics generally make any 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 the clinic performance data (SART 1996), these two restrictions raise the IVF success probability by three percentage points – a notable increase but hardly approaching breakeven. Beyond this, clinics claim that they do not “discriminate,” and indeed the available empirical studies have not succeeded in identifying additional observable factors that are predictive of IVF success (Haan et al. 1991a; Stolwijk et al. 1996; Zhou et al. 1996). We will, as a result, assume that patient screening adds no more than the equivalent of 3 percentage points to the clinic’s success rate.
Breakthrough In IVF Success Rates

A third possible explanation for the guarantee’s apparent financial loss is a breakthrough in IVF success rates since 1996. (The IVF money back guarantee became a significant offering of clinics during 1997.) This is highly unlikely for two reasons. First, to go from 22 percent to 40 percent in just a year or two would indeed mean a significant breakthrough, employed by nearly all clinics, and the medical literature points to no such advance in technology, method, etc., during this period. Second, the history of IVF success, summarized in Figure 2, shows a remarkably slow and steady increase through 1996.

Risk Transfer

Another way that clinics can – at least in principle – improve the financials for the guarantee program is by shifting the traditional balance between minimizing risk of the IVF procedure itself and maximizing the probability that a live birth occurs. In short, this means transferring additional risk to the patient. This could be accomplished in multiple ways. For instance the drug regimen can be increased early in the treatment cycle, 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 probability of a live birth can also be increased by replacing more viable embryos into the female (ESHRE Capri Workshop 1996). This also greatly increases the risk of multiple conception (twins, triplets, etc.), and higher order multiples carry increased risk of both health risks for the infants (including prenatal death, cerebral palsy, low birth weight) and for the mother (pre-clampsia, hydramnios varicosities, anaemia) (Dawood 1996; ESHRE Capri Workshop 1996).
Clinics, naturally, claim that they do not do these things (Strictly Business 1997) and, indeed, these kinds of activities and the negative outcomes they can cause are reasonably observable. In short, a pattern of 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 three percentage points from the baseline level of 22 percent in the 1996 data. That is, those clinics that do screen on age (<40 years) and male factor infertility “gain” 3 percentage points. This amounts to a 25 percent chance of a live birth per IVF cycle in 1997 – as opposed to the 40 percent chance required to make the money-back guarantee viable long-term.

**Heterogeneity**

Staying within the framework of our profitability formula (1) above, there are really only two additional factors to consider, and they both involve heterogeneity. That is, our formula above implicitly assumes that all clinics have the same success rate, and that this success rate does not differ from cycle to cycle for an individual customer. If clinics differ reliably from each other 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 general, our analysis of the predictive value of historical success experience for individual clinics will be relevant for both clinic managers and patients.

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.
Those making it to the third, fourth etc. cycle will tend to be those with lower success rates. On the other hand, doctors do acquire useful information during the IVF cycles, which can be used to counteract this negative effect of heterogeneity. Specifically, the drug regimen used may be optimized to a particular patient on later IVF cycles, or certain patients may be counseled out of additional IVF attempts when interim outcomes of previous cycles suggest that IVF will not work for them (Haan et al. 1991b).

In the next two sections we offer an empirical analysis of these two key kinds of heterogeneity, and assess their consequences for the kinds of IVF pricing programs at issue.

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. This information includes the number of live birth deliveries per IVF cycle started, for each of the 101 reporting clinics. It is estimated that approximately ninety percent of the IVF clinics reported their data to SART (SART 1996). In total these clinics began 14,322 IVF cycles and had 2,646 deliveries, for a delivery rate of .185. During 1994 the average number of IVF cycles was 142 per clinic. We noted earlier that age and presence of a male infertility factor measurably affect IVF success, and are sometimes used to exclude patients from the guarantee programs. Accordingly, we focus our analysis on the 8714 IVF cycles for which the female was under age 40 and there was no male infertility factor. This will make the sample most relevant to our guarantee scenario, and will also help
ensure that across-clinic differences observed are not really manifestations of different patient populations.

In any one year the success rate does indeed vary greatly from clinic to clinic. The top two panels of Figure 3 highlight this fact – showing the “top 5” and “bottom 5” clinics. Three of our 101 clinics had success rates over 40 percent, the breakeven point for the money-back guarantee. On the other end of the scale, four clinics had no successes at all during 1994. 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.

For instance, if 10 clinics each had a “true” success probability of .20 for each patient who entered the clinic, the observed success rates for the clinics would vary naturally according to the binomial distribution. This binomial proportion variance decreases with the number of cycles performed, so the clinics that are observed to be “outliers” (on each side – good and bad) will tend to be those that happened to do relatively few cycles – irrespective of long-run success rates. This is of course evident in Figure 3 – all five of the “worst” clinics and four of the five “best” clinics did very few IVF cycles. Indeed our “best” clinic with a success rate of 50 percent can hardly be counted on to perform well in the future, since its success was based on only 4 IVF cycles performed during all of 1994 (with 2 deliveries resulting).

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 usual 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 “best” (minimum error variance; shrinkage) 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, with p.d.f.

\[
f(p \mid a, b) = \frac{p^{a-1}(1-p)^{b-1}}{B(a, b)}, \quad a>0, b>0,
\]

with mean \( E[p]=a/(a+b) \) and variance \( \text{Var}[p]=ab/[(a+b)^2(a+b+1)] \). Then the number of successes \( X \) in \( n \) IVF attempts for a randomly chosen clinic follows the beta binomial distribution:

\[
P[X \mid n,a,b] = \binom{n}{X} \frac{B(a + X, b + n - X)}{B(a,b)}, \quad X=0,1,2,\ldots,n,
\]

and the key formula of interest, i.e., the expected future success rate for a clinic that was observed to experience \( x \) successes in \( n \) IVF attempts, is:

\[
E[p \mid x,n,a,b] = \left( \frac{n}{a+b+n} \right) \left( \frac{x}{n} \right) + \left( \frac{a+b}{a+b+n} \right) \left( \frac{a}{a+b} \right),
\]

which represents a weighted average of the clinic’s observed success rate \( (x/n) \) and the average success rate for all clinics \( (a/(a+b)) \). The beta binomial has been a highly effective and robust model for heterogeneous Bernoulli processes in marketing and in other social science applications, including biomedical research (Crowder 1978; Greene 1982; Griffiths 1973).
Maximum likelihood estimates of the two parameters a and b for this set of 101 clinics are:

\[ a = 9.44; \quad b = 35.15, \]

which corresponds to a mean and standard deviation of real long-run success rates across clinics of

\[ E[p]=.212; \quad \sigma_p=.0605. \]

Accordingly, using the simple one-sigma and two-sigma heuristics, two-thirds of clinics have a true delivery rate between 15 percent and 27 percent (one sigma); and 95 percent of clinics have a true delivery rate between 9 percent and 33 percent.

These results essentially rule out heterogeneity in clinics as any significant contributor to the viability of IVG money-back guarantees. Based on the experience during 1994, virtually none of the clinics have a stable success rate going forward that is even at breakeven (31%), let alone sufficient to generate a reasonable return (40%). We focussed here on eastern region clinics, to help remove possible geographic differences as a confound. We note in passing, however, that 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 on performance across clinics as described in the last paragraph.

As equation (4) makes clear, expected success for any single clinic 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 \(a+b\). That is, when \(n<a+b\) (and so the term \(n/(a+b+n)\) in (4) is less than \(\frac{1}{2}\)) the expected future success probability is more determined by the average success across clinics than by this clinic’s particular history.
Since $a+b=44.59$ in our data, both managers and patients should regard success rates from clinics that do less than, say, 50 cycles per year with the most extreme caution.

This kind of conclusion is highlighted in Figure 3. Note how our “top 5” and “bottom 5” clinics’ expectations are changed when an empirical Bayes updated probability (via equation (4)) is calculated for each clinic. Four of the “top 5” clinics did fewer than 40 cycles, and should all be treated as about average in performance. In the bottom half of Figure 3 we provide the “updated” top 5 and bottom 5 clinics – i.e. clinics ranked on the updated probability of success on future IVF attempts. Only three of our 101 clinics have expected future success probability greater than .3 – and none are greater than .35. Since approximately 60 of 300 North American clinics are offering the IVF guarantee, again we see that differential clinic skill/performance is not the explanation. Consider, as one piece of anecdotal evidence, the Minnesota clinic highlighted in the popular press as an innovator in offering this guarantee (Strictly Business 1997). Its 1994 experience (among age<40, no-male-factor patients) according to the SART data was 10 live deliveries in a total of just 32 IVF cycles started. The observed success rate was thus .31 – and after applying the shrinkage formula (4) this clinic’s expected future success rate drops to just .254.

Our discussion here has focussed on implications for the pricing policy of clinics. It is also appropriate to offer here some interim conclusions on the “product P” element of the marketing mix: i.e. on the real variation in clinic performance. Our cautions above notwithstanding, there are real differences in clinic performance that should persist through time. The homogeneous Bernoulli model (constant $p$ for all clinics) was firmly rejected by the usual likelihood ratio test ($\chi^2_{(1)} = 131$) in favor of the heterogeneous beta
binomial. Further, differences across clinics are of a magnitude to 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. Finally here, we note the usual caveat regarding medical outcome statistics – namely that the patients most in need of treatment may well tend to find their way to the strongest medical providers. Our analysis here obviously does not control for any such effect, nor for clinics’ policies in accepting patients or managing the health/outcome risks noted earlier. For example, consider our “best” eastern clinic after updating the past history, i.e., the clinic ranked #1 in our updated “Top 5” in Figure 3. Without taking anything away from the clinic’s abilities, we note that

1. the patients in the “under-40” age range in this geographic area may well be “more” under 40 than at some other clinics, and

2. the SART data show that among the IVF deliveries at this clinic during 1994, more than half (47 of 90) were multiples (twins, triplets, etc.). This compares with a multiples rate of only 32 percent across all North American clinics during 1991-1993 (Dawood 1996).

The prospective IVF patient obviously needs to consider both the statistical reliability of reported clinic performance statistics and also the kinds of patient profiles and treatment aggressiveness that may account for those statistics.

5. PATIENT OUTCOMES ACROSS SUCCESSIVE IVF ATTEMPTS

Our formula (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 represent very well the IVF process, and each can affect greatly the expected financial outcome. We will first 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 who 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 equation (1)’s assumption of homogeneity across patients produced a net loss of $3448 per patient, including a profit from testing. Imagine the extreme case of patient heterogeneity, with the same average success rate p=.22 as assumed earlier. That is, imagine that 22 percent of patients have an IVF success rate on any cycle of 100 percent, and the remaining 78 percent of patients have success rate of 0. Using the same arithmetic as in equation (1), the clinic will net $15,000-$6,000 from 22 percent of the patients (who succeed on the first try), and will “net” $15,000-3×$6,000-$15,000 from the remaining 78 percent, 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 $3448 if patients are homogeneous in success rates. Heterogeneity scenarios between these two
extremes examined here will produce losses between $3448 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 in the past 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 somewhat. Second, for some patients the information gleaned from initial cycles will reveal that the prospects for success via IVF are remote, and the couple can be counseled to move on to either another assisted reproduction therapy (e.g., donor egg, ICSI) or to other options such as 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. This will allow us to place an upper limit on the learning effect’s benefit for IVF 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 (Alsalili 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).

Since the SART data do not connect IVF success/failure with the specific cycle for a couple, we will rely on the histograms in these published studies to calibrate a formal probability model.

Let $p_1$ denote a particular couple’s success probability on the first IVF attempt.

To incorporate learning in this model it will be useful to re-write this cycle-1 success probability $p_1$ as

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

which of course we can do 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 (6). Now, if learning occurs, this couple’s success probability should be higher on later attempts, which can be accomplished by replacing $x$ in (6) 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:
\[ p_i = \frac{x + \sum_{i=2}^{i} \lambda_i}{x + y + \sum_{i=2}^{i} \lambda_i}, \quad \lambda_i \geq 0, \quad (7) \]

where \( \lambda_i \) represents the incremental learning effect for attempt \( i \) based on the learning on attempt \( i-1 \). On the first IVF attempt (7) simply reduces to (6). On the second attempt the numerator “success” factor increases from \( x \) to \( x + \lambda_2 \), so \( \lambda_2 \) essentially represents the contribution of learning from the first IVF cycle. On the third IVF attempt the numerator becomes \( x + \lambda_2 + \lambda_3 \).

To incorporate heterogeneity across patients we allow the baseline success factor \( x \), baseline failure factor \( y \) and learning effects \( \lambda_i \) 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:

\[
f(x \mid r_x, \alpha) = \frac{\alpha^{r_x}}{\Gamma(r_x)} x^{r_x-1} e^{-\alpha x};
\]

\[
f(y \mid r_y, \alpha) = \frac{\alpha^{r_y}}{\Gamma(r_y)} y^{r_y-1} e^{-\alpha y}; \quad \text{and}
\]

\[
f(\lambda_i \mid r_{\lambda_i}, \alpha) = \frac{\alpha^{r_{\lambda_i}}}{\Gamma(r_{\lambda_i})} \lambda_i^{r_{\lambda_i}-1} e^{-\alpha \lambda_i}.
\]

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 \( \lambda_i \), 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 ≤ \( K \) ≤ 1). That is, since \( K \leq 1 \) 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:

\[ r_{\lambda_t} = r_{\lambda} K^{-t-2}, \quad \text{for } t=2,3,\ldots \]  

(9)

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 (7) whose average size is \( r_{\lambda}/\alpha \). On the third IVF attempt we add an additional learning effect to the numerator and denominator of (7), 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 ¼ what it was on the second cycle, etc.

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 beta-binomial 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_y)/(r_x+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:

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

(10)

In the absence of any learning (r_\lambda=0) this formula (10) 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 Beta Binomial. With the learning effect as specified above we will call the formulation (10) the Beta-Geometric with Learning (BGL).

Equation (10) 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 3824 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. data are particularly valuable because of the large number of patients tracked, and also 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 percent of those pregnancies do not result in deliveries, and this percentage can vary substantially.

For the Tan et al. data the maximum likelihood estimates of the BGL model are:

\[ r_x = .263; \quad r_y = 2.247; \quad r_\lambda = .130; \quad K = .510. \]
These particular parameter values tell a very interesting story about the effects of learning and heterogeneity, and about the tradeoff between the two. First, consider just the learning effect. Since $r_\lambda$ 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 IVF attempt, but since $K$ is approximately .5, the contribution from learning on the second IVF cycle is only about half as valuable as the learning from the first cycle. The incremental learning from the third cycle is only $\frac{1}{4} = (.5)^2$ 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 (in these data) was $r_x/(r_x+r_y) = .105$. For patients who do not succeed on the first attempt, what happens on the second try? Via equation (10), the net positive factors ($r_x+r_\lambda$) increase by 50 percent relative to attempt #1 (i.e. .263+.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 percent. 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 the adverse selection effect.

On later cycles this balance between learning and adverse selection is of course not maintained. This is due to the decreasing size of the incremental learning effect (decaying by approximately half at each cycle). The negative effect of adverse selection, by
comparison, remains constant from cycle to cycle. So for IVF cycles 3, 4 etc. the probability of success begins to fall.

Figure 4 illustrates this set of results. The top half of the figure shows the fit of the Beta-Geometric (i.e., no-learning) model to 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 Beta-Geometric cannot capture this phenomenon either. By contrast the Beta-Geometric with Learning (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. data. While the specifics of the histograms do differ across studies, we also estimated the BGL model on four other published repeat-cycle datasets that covered different clinics, time periods, and outcome measures (pregnancy vs. live birth) (Alsalili et al. 1995; Haan et al. 1991b; Hershlag et al. 1991; Stolwijk et al. 1996). The average BGL parameter values across these four datasets 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 very similar to our results for the Tan et al. 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

---

3 The Figure 4 data cover patients who began treatment between 1984 and 1990. This is why the success rate per cycle is less than the 18 percent figure used in the current study. The pattern on successive cycles, which is of primary interest here, not affected by this downward shift in success rates.
of heterogeneity essentially cancel, leaving the probability of success essentially the same as it was on attempt #1. On cycles 3 and up, as above however, the effect of heterogeneity (adverse selection) takes over and success rates drop, albeit slowly at first. 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 essentially cancel. As a result, success prospects can be viewed as constant across these cycles, as our beta-binomial 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 is real and does counter adverse selection, it does not manage to add anything to our analysis in equation (1). In short, the financial viability of the in vitro 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. There is a substantial consequence here for the money-back guarantee policies. In the next section we leave the framework of equation
(1) and investigate in greater detail these later IVF cycles: i.e., what may happen after the money is refunded for a third (failed) IVF attempt?

6. PERSEVERENCE 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 section 3’s $3448 (based on equation (1) and the clinic’s potential profit from testing), to $2203. The latter figure arises simply by substituting p=.25 in equation (1) – i.e. the increase of three percentage points in the per-cycle IVF success rate that stemmed from patient screening. So it still appears that clinics could not afford to offer the guarantees on a large scale.

In this section, and in the next, we look 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, either to pursue further natural conception and/or to attempt adoption, possibly using the refund to cover the lion’s share of the latter’s costs. Alternatively, the couple can continue with IVF, paying a la carte for additional cycles. Of course, the refund can also provide the 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 1997 – 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 the probability that a couple “goes on” after each IVF failure. Approximately 60 percent of those who fail on an IVF cycle elect to go on to the next cycle, and this statistic remains virtually constant over the first eight IVF cycles.

Under the money-back guarantee what will happen after a third failure/refund? No one knows at this point. Certainly clinic managers did not know upon instituting the guarantee. But they did have access to the empirical pattern of Figure 6. Indeed, for a couple of reasons the 60-percent-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 percent 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, for both patients 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 observed historically 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”, since this is the number of failures the refund receivers actually experienced. But as Figure 6 makes clear, based on the empirical evidence for a la carte patients, such distinctions have really not mattered – i.e. the 60 percent 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 (relative to a la carte patients) 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 very 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 patients historical 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 over twenty IVF attempts (Stolberg 1997).

So we assume that 60 percent 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 percent 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, since success also removes the patient from consideration for later cycles. Based on the BGL model parameters reported in the last section (using the Tan et al. data), the success probability is essentially constant across
the first three IVF attempts. We denote this probability $p_{1-3}$. Using the BGL model, the success probability drops across attempts 4-7, in relation to $p_{1-3}$, as follows:

$$p_4 = 0.818 \times p_{1-3},$$

$$p_5 = 0.719 \times p_{1-3},$$

$$p_6 = 0.636 \times p_{1-3},$$

and

$$p_7 = 0.569 \times p_{1-3}.$$  

Substituting our success rate across the first three cycles $p_{1-3} = 0.25$ as above, the success probabilities on cycles 4 through 7 become 0.204, 0.180, 0.159 and 0.142, respectively.

With this parameter set, 60 percent of those receiving a refund will go on to pay *a la carte* for IVF cycle number 4. Of this 60 percent, approximately 20.4 percent are expected to succeed on cycle 4, so 79.6 percent of the 60 percent (i.e. 48 percent) of the refund receivers will go on to fail on cycle 4. And of these, again 60 percent (i.e. 29 percent 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, since the typical contribution margin per cycle is (price-cost) = $7,500 - $6,000 = $1,500. Considering only IVF cycles 4 through 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 $1695 economic contribution per refund-receiver.

This last figure must be adjusted before tallying it against the current $2203 loss per patient described at the beginning of this section. The only patients that may proceed
to *a la carte* are those who receive a refund, so the $1695 figure must be reduced by that proportion. With \( p = 0.25 \), approximately 42.2 percent of couples will fail through three tries and receive the refund. So the economic contribution of *a la carte* cycles 4-8 per patient that begins the guarantee program is \( 1695 \times 0.422 = 715 \). This positive amount reduces the per-patient loss from the guarantee to \( 2203 - 715 = 1488 \).

Even allowing for some potential error in our assessments, the money-back guarantee still does not make economic sense. 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 (or more) months attempting natural conception, drug therapy, and IUI being common. Doctors would tend to recommend IVF only after these options had been exhausted, 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 U.S.. 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, over the year or more that patients pursue these earlier 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 40-44 year old couples relative to those aged 35-39.
Enter into this process the IVF money-back guarantee, by late 1997 offered by many IVF clinics. The guarantee is advertised in mass media (magazines, newspapers) and directed by the clinics toward patients rather than their doctors. For patients who are willing to spend $15,000 for a child, it emphasizes the complete lack of (economic) risk. It holds the promise of a baby now, rather than (maybe) a couple years from now. It minimizes the emotional cost in repeated failures and waiting. It is, in short, being marketed no longer as the “last hope” – but instead with the positioning “Why wait”?

For infertility clinics that offer the IVF guarantee, tapping into the large pool of patients in these earlier stages of treatment – who are of course substantially more fertile than IVF’s traditional patients – can result in a large positive financial contribution. We will show in this section that such a contribution can readily make the money-back guarantee programs do more than break even for the clinic. The set of assumptions we require for this calculation may not apply to all clinics, but they are reasonable, and do provide an explanation for the guarantees that has been absent up until now.

Consider the following scenario for increased use of IVF earlier in the infertility treatment process.

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 historical 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) or accept the probability of remaining childless.

Of course, other assumptions are possible. IVF clinics may attract some couples even earlier in the natural conception process, with their guarantee. The IVF guarantee may take more than half the drug therapy or IUI patients – after all, the latter treatments may cost some patients, offer a much lower probability of success and accordingly a high likelihood of emotional distress (including marital stress), and delay the arrival of a child. Finally, less than 50 percent may choose to pursue IVF, either due to counseling regarding their specific source of infertility, or because of the $15,000 cost under the guarantee. Of course if any of these alternative assumptions hold they 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 in order to calculate the financial implications here; namely, the success rate that IVF is expected to achieve per cycle in the more fertile populations. In this assessment it is useful to consider the sequence of 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 therapy or IUI, by far the weakest link in this chain of events is fertilization. Attrition at the other stages is relatively low. On the other hand, fertilization is the strong point of IVF. As Figure 7 makes clear, two-thirds of couples using IVF make it to a (usually several) fertilized, growing embryo. For these relatively infertile couples, the weak point of IVF is implantation – this is the stage where well over 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 – specifically 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 6 months without a conception) would have an IVF success rate of 60 percent per cycle. For couples who do not conceive naturally and would have pursued drug therapy and/or IUI, we assume a 50 percent per-cycle IVF success rate.

To calculate the financial effect of pursuing infertility patients earlier, let N represent the number of couples who complete 6 months of natural conception without a pregnancy. Historically, prior to the IVF guarantee programs, Figure 1’s statistics indicate that about .274N would achieve a delivery during an additional 6 months of natural conception attempts. We imagine that the IVF guarantee would capture 50 percent of these patients, or .137N.

This leaves .726N patients who complete 12 months without a delivery, and go on to drug and/or IUI therapy. Again via the statistics in Figure 1 after four attempts an additional .282N will historically have had a live birth delivery – and we assume that now the IVF guarantee will capture 50 percent of them, or .141N. Finally, .726N-.282N = .444N do not achieve a live birth after the drug/IUI sequence, and we assumed that half these patients – or .222N – go on to IVF.

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

1. Group 1, of size .222N, is the “usual” IVF patients who have failed repeatedly to conceive naturally or with drug or IUI therapy, and who accordingly have
success probability on IVF cycles 1-3 of .25 (with success on later cycles 4-8 as described in the previous section),

2. Group 2, of size .141N, is composed of patients who have failed repeatedly 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=.5 on each of IVF cycles 1-3.

3. Group 3, of size .137N, is composed of patients who did not conceive in 6 months of natural conception, but who would have conceived in another 6 months of such attempts, but who elect to go with the IVF guarantee instead, and who accordingly have success probability=.6 on each of IVF cycles 1-3.

For Groups 2 and 3 the success probability on each of cycles 4-7 can be calculated using the BGL model 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, together with equation (1). Doing so, the average return per patient for each group is:

Group 1: -$1488

Group 2: +$2784

Group 3: +$4765

The overall 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, which are listed above. (Group 1 accounts for 44 percent of IVF patients, and Groups 2 and 3 account for 28 percent and 27 percent, respectively.) With our assumed scenario for attracting patients to IVF earlier in their fertility treatment, this overall return

40
is positive at last, at $1430 per patient. This figure represents approximately a ten percent return on average cost for the clinic. As noted above, the prospects for attracting more-fertile patients to IVF may be better than we have assumed, which would increase the return calculated here.

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 (twins, triplets), but at least in the case of twins many apparently-infertile couples will view this as a “bonus” rather than a drawback. In expected value terms they paid collectively more than they needed to, but are extraordinarily 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.

Relative to this latter decision for Group 1, one final consequence of pursuing patients earlier for IVF is worth stressing. Reported overall success rates for IVF are going to rise, in the absence of any substantial published breakthrough in procedures. This rise is of course irrelevant to the Group 1 patients who have exhausted other options. Their prospects have not gone up dramatically. What has happened instead is a change in the total IVF patient pool. It would be unfortunate in the extreme if Group 1 patients
were misled to believe that, following three IVF failures and return of their $15,000, the “new and improved” published IVF success rates actually applied to them, in considering paying *a la carte* for additional cycles.

8. CONCLUSION

We conclude by examining the implications of our analysis for the key stakeholders: 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 crystal clear. If they decide to go the IVF route, they should take the money-back guarantee option. 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 – this keeps the probability of getting a live birth the same for both options.) Thus the expected cost to the patient under *a la carte* is:


Under the money-back-guarantee with an upfront payment of G, the couple pays G only if there are not three consecutive failures. The expected cost to the patient under the money-back-guarantee 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. (11)

The second term is always positive for p between 0 and 1, so the only relevant (real) solution to (11) is the root from the first term:

\[ p = \frac{A}{G}. \]

With the monetary values used today of A=$7,500 and G=$15,000 the breakeven value is p=0.5. Thus any couple with a p-value less than 0.5 should opt for the money-back guarantee. For the typical risk-averse couple who would be very unhappy going \textit{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. For those who are highly risk averse (in this monetary sense) it could easily exceed 0.7. 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 breakeven value of 0.5.

For the typically younger and less infertile Group 2 and Group 3 patients, a deeper analysis is required. They need to adjust their baseline p value to some appropriate level above the Group 1 value. (There is no single obviously correct way to do this, but section 7 gives some suggestions.) These patients should then decide if the increase in the chance of having a baby \textit{now} is worth the potential adverse side effects of IVF and the increased likelihood of having multiple births. This is a highly 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 \textit{a la carte} versus money-back-guarantee option. Those patients who opt for the money-back-guarantee and receive their refund should explicitly consider the economic, psychological, and side-effect costs yet again before treating their...
refund as “house money.” They may well decide to use this money for a fourth (and fifth) attempt, but they should put the same degree of thought into this decision as they did in initially deciding to enter an IVF program. (Group 1 patients should also do this analysis.)

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 (less than about fifty) sample sizes or high success rates where the prevalence of multiple births is well 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 manager should be especially wary of using outdated p-values if the “carpet-bombing” mass media advertising of a money-back guarantee (undertaken by either this clinic or competitors) is bringing in patients to IVF “earlier”.

The key ethical considerations for a clinic are the following. Are patients being given data that are appropriate for their particular situations, as illustrated above? Are the downsides of IVF _vis-à-vis_ less invasive procedures adequately addressed? Are the patients who receive their money back being unduly 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 between clinics. The data, model, projection, and scenario analyses in this paper can be of help in patient education; and indeed a menu-driven interactive computer version of this work would be valuable for some potential IVF patients. It is also important to continue and expand monitoring of individual clinics. Some of the key variables—e.g. success rate by age group, incidence of multiple births—have been collected comprehensively. Other key variables—e.g. success rates by previous infertility treatment history, number of embryos placed, pricing policy, and number of IVF attempt for the patient—are not available clinic by clinic. This information would provide added incentive for the clinics to “do the right thing.”

It is especially important for public policy groups to be active in this situation, since patients themselves, individually or collectively, are unlikely to pressure the clinics. For those less infertile couples who are recruited to IVF “too soon,” they are likely to obtain a baby—or two or three—and not complain. The couples who are not successful, but who get their money back, are also unlikely to be very vocal—even if they wind up spending it all on additional unsuccessful attempts.

Under the current pricing policies and cost structures the clinics will lose significant money offering money-back-guarantees to their traditional patient population. What these clinics may be doing to make these guarantees viable should be examined very carefully. Aggressive advertising for more fertile patients, highly selective screening, and aggressive treatment with more side effects and higher probability of multiple births are ways that the guarantee could be made profitable.
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. They also allow for the sensitivity analysis and the implications of different future scenarios. But perhaps the best news from our efforts is that the simplest possible model, (1), “works.” That is, assuming that each patient has the same base-line success probability (wrong) and that this probability stays unchanged on successive cycles (wrong) leads to a very 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. Since the homogeneous (Bernoulli) model in (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 the couple deciding between the a la carte and money-back-guarantee options, only A, G and their personal p value enter the equation. The monetary risk neutral breakeven p value is simply A/G. With a little coaching the couple can get a reasonable estimate for their p-value as well as an appropriate risk-averse upward adjustment in the breakeven p. Of course, just telling couples the A/G breakeven point and the approximate p-value for “couples like them” would be a major improvement on the current state of patient education – especially if the money-back-guarantee option continues to expand.
REFERENCES


**Figure 1**

Typical Progression of Assisted Reproduction Treatment

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

Figure 2
Historical Trend in IVF Success Rate, North America

Sources:
Figure 3
Assessing Heterogeneity in Clinic Performance

<table>
<thead>
<tr>
<th>Original “Top 5”</th>
<th></th>
<th>Updated “Top 5”</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank</td>
<td>Original Observation</td>
<td>Updated Probability</td>
<td>Rank</td>
</tr>
<tr>
<td>1</td>
<td>2/4=.500</td>
<td>.235</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>8/20=.400</td>
<td>.270</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>15/39=.385</td>
<td>.292</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>69/182=.379</td>
<td>.346</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Original “Bottom 5”</th>
<th></th>
<th>Updated “Bottom 5”</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank</td>
<td>Original Observation</td>
<td>Updated Probability</td>
<td>Rank</td>
</tr>
<tr>
<td>1</td>
<td>69/182=.379</td>
<td>.346</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>169/480=.352</td>
<td>.340</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>55/166=.331</td>
<td>.306</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>15/39=.385</td>
<td>.292</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>26/77=.338</td>
<td>.291</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 4
Modeling The Likelihood Of Success
On Repeated IVF Attempts

Probability of Success, Given No Success So Far

Attempts Number

Probability of Success, Given No Success So Far

Attempts Number
Figure 5
Outcome Probability On Successive IVF Attempts
In Various Empirical Studies

Live Birth Probability

Attempt Number

Ongoing Pregnancy Probability

Attempt Number
Figure 6
The Propensity To Continue IVF After Failed Attempts

Sources (pooled):
Check (1994); Tan et al (1994); Yovel (1994)
Figure 7
Patient Progress Through Successive Stages Within An IVF Cycle
(adapted from Haan et al. 1991a)