

Wu, Bingxiao

Working Paper

Information presentation and consumer choice: Evidence from Assisted Reproductive Technology (ART) success rates reports

Working Paper, No. 2014-10

Provided in Cooperation with:

Department of Economics, Rutgers University

Suggested Citation: Wu, Bingxiao (2014) : Information presentation and consumer choice: Evidence from Assisted Reproductive Technology (ART) success rates reports, Working Paper, No. 2014-10, Rutgers University, Department of Economics, New Brunswick, NJ

This Version is available at:

<https://hdl.handle.net/10419/123820>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Information Presentation and Consumer Choice: Evidence from Assisted Reproductive Technology (ART) Success Rates Reports

Bingxiao Wu*

June 14, 2014

Abstract

Prior literature on quality disclosure focuses on whether information provision affects consumer choice. This paper extends this research and explores whether information presentation affects consumer responsiveness in the context of Assisted Reproductive Technology (ART) reports. I find that after CDC releases quality information on both “success rate” and “multiple-birth rate,” with the former highlighted, consumers only respond to “success rate;” after CDC changes the format by highlighting “multiple-birth rate,” consumers start to choose clinics with lower “multiple-birth rates.” It implies that proper design of information presentation is crucial in determining the effectiveness of public reporting. (JEL L15, I11, I18)

I. Introduction

In the marketplace of experience goods (Nelson 1970), consumers incur search costs to retrieve information on product price or quality. There are two types of

* Department of Economics, Rutgers University, 75 Hamilton St., New Brunswick, NJ 08901 (email: bwu@econ.rutgers.edu). This paper is an extension of the 2nd chapter of my Ph.D. dissertation. I thank my dissertation advisors David Dranove and Leemore Dafny for their guidance and support on this project. I also thank a number of seminar and conference participants for valuable suggestions. All remaining errors are my own.

search costs: the external costs and the internal costs (Smith, Venkatraman, and Dholakia 1999). The external costs refer to the opportunity costs of the time spent on searching and the monetary costs spent to acquire information.¹ The internal search costs refer to the fact that consumers have cognitive limitation to process the disclosed information and thus incur costs when comparing diverse choice alternatives (Shugan 1980).² Economists and policymakers have been largely focusing on efforts to reduce external search costs by making comparative quality/price information available.³ (Dranove and Jin 2010) and (Marshall et al. 2000) provide systematic reviews on the demand side response to information provision in public reporting. The internal search costs, however, have not received much attention in the economics literature. Although lab experiments by cognitive psychologists have revealed that information presentation significantly affects subjects' cognitive, affective and behavioral responses by changing the internal search costs (Zeithaml 1982), and researchers in the field of marketing also hold a consensus view that information context affects which information the decision maker attends to (Lurie and Mason 2007), there are surprisingly few empirical studies in the economics literature looking at whether information presentation in the context of "report cards"—public disclosure of comparative information on the performance of health care providers—affect consumer choices.

This article aims to contribute to the above basis by examining the effect of an exogenous change in the information context in ART (Assisted Reproductive Technology) Success Rate report cards on how consumers respond to the

¹ For example, consumers need to drive to another store to check whether the price is lower there; they need to spend time comparing the nutrition information on each brand of cereal; they need to buy Consumer Reports to learn about an SUV's reliability.

² The internal search costs are determined by consumer intelligence, education, prior knowledge and training (Smith, Venkatraman, and Dholakia 1999).

³ For example, third-party disclosure on quality information has been adopted in many industries: from Consumer Reports in the auto industry to Medicare's Hospital, Nursing Home and Home Health Compare.

multidimensional clinic quality. Specifically, it focuses on the "vividness" and "evaluability" of the information context. "Vividness" refers to the fact that people tend to construct an evaluative concept on the basis of the first information (the "primacy effect") and the most salient information (the "salience effect") they receive, and later use this concept as a basis for judgment (Hong and Wyer 1989). The differential "vividness" across multiple attributes might lead subjects to form cognitive bias towards paying more attention to the more vivid attributes while ignoring other attributes. In this case, if the more vivid attributes turn out to be the less important ones, "vividness" could decrease efficiency. "Evaluability" refers to the fact that consumers have a cognitive limitation in making inference when an attribute of a product is implicitly disclosed (e.g., when displaying the "raw price" and "volume" of a product, its "unit price" can be implicitly inferred by dividing the "raw price" by "volume") and thus unresponsive to that attribute.⁴

In this paper, I test whether "vividness" and "evaluability" affect consumer choice in the context of fertility-clinic report cards. In 1998, Centers of Disease and Control and Prevention (CDC) releases the 1st edition report cards. The report cards disclose 3-year lagged performance on the two most important quality dimensions of ART treatments: "success rate" (i.e., the percentage of cycles resulting in live births) and "multiple-birth rate" (i.e., the percentage of live births involving multiple infants) and highlight only the former measure—"success rate." In 2004, CDC changes the format of report cards and adds a new measure—"singleton success rate" (i.e., the percentage of transfers resulting in singleton live births). Exploring the relationship between market share and 3-year lagged quality measures before and after 1998/ 2004, I find that (a) after the release of the 1st edition report cards in 1998, market share increases for clinics with higher success rates; but is not responsive to multiple-birth rates; and (b) after the report-card

⁴ See Section 2.3 for a literature review on the efficiency loss induced by "vividness", and how "evaluability" affects consumer choice.

format change in 2004, market share starts to decrease for clinics with higher multiple-birth rates. These findings indicate that (a) the information context that makes some quality dimensions more vivid than others leads consumers to focus only on the most vivid ones and to ignore the other dimensions; and (b) after adding an explicit measure of an attribute (and thereby improving the "evaluability" of that attribute), consumers become more sensitive to that attribute.

This paper implies that making information available does not necessarily guarantee that consumers will respond; in addition, information presentation could lead consumers to form cognitive bias regarding the multiple quality measures. The remainder of the paper proceeds as follow. In Section II, I provide the institutional knowledge of the ART industry as well as the background information on the policy intervention. In Section III, I review previous studies and form the testable hypotheses. In Section IV, I describe the data and provide summary statistics. In section V, I show the empirical framework and explain the identification strategy. Section VI presents the estimation results and Section VII concludes.

II. Background

II.A. The fertility industry in the U.S.

Fertility clinics provide Assisted Reproductive Therapies (ART) for the treatment of infertility. In the U.S., there are currently over 500 fertility clinics. According to the latest CDC ART Success Rate Report, in year 2011, 163,039 ART cycles were performed at the 451 reporting clinics, which resulted in 47,818 live births (deliveries of one or more living infants) and 61,610 infants.⁵ This

⁵ Source: <http://www.cdc.gov/art/ART2011/index.htm>.

accounts to slightly more than 1% of total U.S. births. Figure 1 provided by the Centers for Disease Control and Prevention (CDC) shows the distribution of fertility clinics in the U.S. It is clear that the greatest number of clinics is in the eastern part of the country, and there is a great amount of variation in market concentration across cities.

[Insert Figure 1 Here]

In general, ART procedures involve surgically removing eggs from a woman's ovaries, combining them with sperm in the laboratory, and returning embryos to the woman's body or donating them to another woman. ART was first performed in the U.S. in 1981. Over the past 30 years, the development in drugs that stimulate egg production and the progress in embryo culturing and selecting technology have led to dramatic improvements in ART's effectiveness.⁶ A full cycle of treatment generally lasts 20 to 40 days, with approximately 15 visits in total. This high frequency of visiting normally prevents couples from choosing an out-of-town fertility clinic. In addition, since couples might experience more psychological pressure when travelling out of town to receive treatment, which could lead to further difficulty in pregnancy, patients are normally advised to choose a nearby clinic. The most common type of ART cycles transfers fresh non-donor embryos.⁷ I therefore focus on the market of fresh non-donor embryo transferring only in this study.

There are two primary quality measures of ART treatments that patients take into account when choosing a clinic: the success rate (i.e., the percentage of cycles resulting in live births) and the multiple-birth rate (i.e., the percentage of live births involving multiple infants). A higher success rate is without a doubt an

⁶ More information can be found at <http://www.cdc.gov/art/index.htm>.

⁷ There are also cycles transferring frozen non-donor eggs, fresh donor eggs and frozen donor eggs. In 2008, approximately 71% of all cycles transferred fresh non-donor embryos.

indicator of better quality and should be valued by consumers. The consumers' preference towards multiple-birth rate, however, can be ambiguous. On one hand, from a medical point of view, a multiple pregnancy has been recognized as the most frequent and most serious iatrogenic complication of ART (Evers 2002). In the U.S., the multiple-pregnancy rate is much higher with ART (30-40%) than with natural conception (3%). High-order pregnancies could lead to severe health risks for babies and mothers, and involve substantial medical and social costs.⁸ On the other hand, however, couples might not have enough medical knowledge to understand the full spectrum of the health consequences of a multiple pregnancy. Moreover, the skyscraping medical costs should mostly be covered by the insurers rather than the patients, and couples might enjoy having multiple kids at the same time for various reasons.

II.B. The quality report card mandate

In 1986, the Society for Assisted Reproductive Technology (SART), an organization of ART providers, started collecting data from its members on utilization and outcome. During that period, participation in reporting was voluntary and the collected information was not disclosed to the public. In 1989, Federal Trade Commission intervened in a case of false and misleading advertising by a fertility clinic. In 1992, Congress enacted the Fertility Clinic Success Rate and Certification Act of 1992, which requires CDC to collect and publish information on clinic quality for all fertility clinics in the U.S.

The first report card was published in Dec. 1997, disclosing the 1995 performance.⁹ Figure 2 shows the format of the 1st edition of report cards. As mentioned in Section II.A, the two most important quality measures of ART

⁸ For example, (Callahan et al. 1994) estimated that the maternal and neonatal hospital cost is estimated to be \$18,245 for a singleton birth, \$70,324 for twins and \$203,418 for triplets.

⁹ Because it takes a certain period of time for women to become pregnant and deliver, the published quality scores are the clinic's three-year lagged performance.

treatments are success rate and multiple-birth rate. The 1st edition reports both dimensions (among several other less important measures such as cancellation rate): percentage of cycles resulting in live births (i.e. success rate) and Percentage of live births having multiple infants (i.e. multiple-birth rate). However, the ways in which these two dimensions are presented are quite different: the success rate is highlighted, reported with confidence interval, and listed in the third row of the table (within 9 rows in total). In fact, it is the only measure that is highlighted in the report card. The multiple-birth rate, by contrast, is listed at the bottom of the table and is not highlighted. It is therefore easy for consumers to focus only on the success rate while neglecting multiple-birth rate given the format of the 1st edition report cards.

[Insert Figure 2 Here]

II.C. The report card format-changing mandate

It is controversial whether the public reporting system introduced in 1998 led to efficiency gains. On one hand, disclosing comparative quality information can facilitate consumer choice. In fact, (Bundorf et al. 2009) finds that patients do respond to “news” in the reported success rate—clinics with a higher success rate attracted more patients after public reporting. However, on the other hand, health professionals in fertility industry criticized that the public reporting system introduced in 1998 has had a negative impact on ART quality—it motivates clinics to transfer an excessive number of embryos to boost their success rate. Howard Jones, who performed the first successful IVF in the U.S. and is considered as the "father" of IVF in the U.S., advocated the abandonment of this clinic-specific quality reporting system (Jones Jr and A Schnorr 2001). He argued that clinic-specific success rates are misleading because patients lack adequate medical sophistication to interpret clinics-specific success rates, and these clinic-

specific success rates do not give consideration to the multiple variables involved in achieving a specific pregnancy rate. In 2002, The chairman of ESHRE (European Society of Human Reproduction and Embryology) suggested that the most appropriate outcome variable of ART procedures is the singleton live birth rate per cycle initiated (Evers 2002). (Min et al. 2004) also suggested that the success of an ART cycle is most appropriately measured as "birth emphasizing a successful singleton at term" (BESST).

To make things right, starting in December 2003, CDC added and highlighted a new entry in the report cards—the singleton success rate (i.e., the percentage of transfers resulting in singleton live births), which is approximately the success rate (i.e., percentage of cycles resulting in live births) multiplies the multiple-birth rate (i.e., percentage of live births having multiple infants).¹⁰ CDC declares the rationale for adding this singleton success-rate entry: (a) (on the demand side) to provide patients with a direct measure of "optimal" infant outcomes; and (b) (on the supply side) to make providers who limit the number of embryos transferred feel they are on an even playing field. Figure 3 shows the new version of report cards, which highlights the newly added singleton-success rate. Notice that singleton success rate could have been calculated using all the information on the old version—for a given clinic, the singleton success rate equals approximately success rate times multiple-birth rates; the new edition calculates the measure and shows it explicitly to consumers. This change makes a clinic with a very high multiple-birth rate look "worse" with a low singleton-success rate,¹¹ and presumably makes it easier for consumers to understand the harm of multiple-birth rate.

¹⁰ The number of cycles and the number of transfers are slightly different in that some of the cycles can be cancelled during the course of treatment thus do not result in transfers.

¹¹ A clinic with a 100% multiple-birth rate, i.e., all live births are multiple-order births, will have a 0% singleton success rate.

[Insert Figure 3 Here]

III. Analytical Framework

The notion that "revealed preference can be affected by the format of information presentation" traces back to the basic assumptions in "information processing theory," a building block of the field of cognitive psychology. According to the theory, information made available by the environment is processed by a series of processing systems in the human brain (e.g., attention, perception, short-term memory, etc.); these processing systems transform or alter the information in systematic ways (Eysenck and Keane 2000). Experimental studies have long been documenting the fact that information presentation can significantly affect subjects' cognitive, affective and behavioral responses (Zeithaml 1982). In the marketing-research literature, there is also a consensus view that the information context affects which information the decision maker attends to. Specifically, information context has three aspects: (a) the "vividness," or the salience of particular information; (b) the "evaluability," or the ease with which information can be compared; and (c) the "framing," or how a given representation changes the reference point or scale against which the information is evaluated (Lurie and Mason 2007).

Based on the descriptive models established by psychologists (e.g., (Bettman 1970) (Bettman and Kakkar 1977) and (Lussier and Olshavsky 1979)), normative models were developed to help explain those findings from lab experiments (e.g.,(Hagerty and Aaker 1984)). In these models, consumers are assumed to incur costs when processing information and perceive differential importance regarding the multiple attributes of an object. As a prediction of those models, both information processing costs and the perceived importance of the attributes would

affect information choice, which would ultimately affect the consumers' purchasing behavior.

In this section, I form hypotheses regarding how information presentation in the context of fertility-clinic report cards affects consumer choice. Specifically, I explain why we can expect to see that (a) the 1st edition of the report cards makes the "success rate" more vivid than the "multiple birth rate" and thus distorts the perceived importance of these two attributes (quality measures) of fertility clinics; and (b) the second version of the report cards, which adds a direct measure on the "singleton success rate," reduces consumers' information processing costs in making the calculation and enhances the evaluability of the "singleton success rate."

III.A. The effect of report-card publication

Early studies on impression formation by psychologists show that the format in which multiple attributes of an object are presented could affect how subjects form their impression of it. In particular, researchers find that when subjects receive several pieces of information about an object for the purpose of forming an impression, they construct an evaluative concept on the basis of the first information received (the "primacy effect") or the highlighted information (the "salience effect"), and later use this concept as their basis for forming a judgment (Hong and Wyer 1989).¹²

Findings in cognitive psychology were soon applied to exploring the determinants of consumers' purchasing behavior in marketing research. Early studies show that the "vividness" (defined in (Nisbett and Ross 1980) as the saliency or availability of the information) of a piece of presented information can

¹² Among the literature, the pioneer work is (Asch 1946) in which one group of participants is asked to form an impression of a person described as envious, stubborn, critical, impulsive, industrious and intelligent; the other group is asked to do the same about a person described as intelligent, industrious, impulsive, critical, stubborn and envious. The author finds that the second group rated the person more highly than the first group.

be easily enhanced by using pre-attentive graphic features, such as line orientation, width, length, and color, which are readily processed with little effort (Lurie and Mason 2007). Making some information more vivid than other might lead decision makers to ignore other relevant information (Glazer, Steckel, and Winer 1992), and overvalue the more salient information which sometimes turns out to be less important (Mandel and Johnson 2002). For example, (Soman and Shi 2003) show that providing consumers with map-based visual representations may lead them to overvalue the path characteristics of the trip and undervalue the overall trip time (even if the trip-time information is also presented). Similarly, using a colored format to present time-changing stock prices (Barber and Odean 2001) (Barber and Odean 2002) or market share information (Gilovich, Vallone, and Tversky 1985) might lead decision makers to act on "information" that is simply random noise. (Stone, Yates, and Parker 1997) show that the margin between the willingness to pay for the safer product relative to the riskier product is larger when the risk information is presented in a graphical format rather than a numerical format. Moreover, (Stone et al. 2003) further show that when using a graphical format to present risk information, highlighting the number of people harmed at the expense of the total number of people at risk for harm (which they define as "foreground:background salience") could lead to greater risk aversion.

In the above regards, information presentation in the case of ART report cards could affect how patients choose fertility clinics. As mentioned in Section II.A, the two primary quality measures of ART treatments are success rate and multiple-birth rate. Since success rate is an indicator of better quality, consumers should put a higher value on higher success rate. Since multiple births involve higher medical and social costs, as well as higher health risk for mothers and babies, consumers should have put a higher value on lower multiple-birth rate. However, it is not necessarily the case given the ways these two quality measures are presented in the 1st edition of report cards. In fact, the 1st edition can be

misleading in the sense that, although both measures are presented, “success rate” is listed in the third column of the report card and is highlighted—it is the only entry that is highlighted in the report card; by contrast, “multiple-birth rate” is placed at the bottom of the report card. Therefore, I hypothesize that the first-edition report card manipulated the consumers’ perceived importance of the multiple dimensions of quality and made them overvalue the “success rate” while undervaluing the “multiple-birth rate.” On the above basis, I make the following hypothesis.

Hypothesis 1: After CDC releases the 1st edition of report cards, consumers value a higher “success rate”, but not necessarily a lower “multiple-birth rate”; therefore, controlling for the trend before report cards, the market share of clinics with a higher “success rate” starts to increase, and the market share of clinics with a higher “multiple-birth rate” does not necessarily decrease.

III.B. The effect of report-card format change

Consumers incur costs processing information when they have a cognitive limitation (Shugan 1980) or when they need to spend money or time acquiring information (e.g., buying “report cards,” driving to another store to check prices or comparing the nutrition information for different brands on the shelf). This paper focuses on the first type of costs—the cognitive limitation.

The closest example is the unit-price legislation during the 1970s in the U.S.¹³ A traditional way of presenting product information is displaying the raw price in addition to the size of the product. Studies have found that although the unit price could have been calculated by dividing the raw price by size, providing a direct measure on unit price makes consumers more sensitive to the unit price. For

¹³ Starting with Massachusetts in 1971, several states have enacted a compulsory unit-pricing law, requiring the price of a product to be expressed in terms of the cost per unit of measure in addition to its raw price (Isakson and Maurizi 1973). For example, with unit pricing a consumer knows that a ten-ounce container of orange juice priced at \$1.5 costs 15 cents per ounce.

example(Granger and Billson 1972) show that the consumers' preferences in terms of package sizes (versus raw prices) change significantly when price-per-unit information is provided: consumers switch more heavily to larger sizes (whose unit price is lower but raw price is higher). The mechanism through which unit-pricing facilitates consumer purchasing decisions is revealed in (Russo, Krieser, and Miyashita 1975). According to this study, the posting of unit prices eliminates the most formidable step in making price comparisons, namely, the calculation of a price that is comparative across all sizes. In fact, when only the raw price is presented, shoppers make tremendous errors (43% to 53%) in figuring out the most economical product, even though they have unlimited time, are highly motivated, can make all the calculations they want, and are typically college educated (Russo, Krieser, and Miyashita 1975).

The idea that "consumers become more sensitive to a measure of a certain attribute of the product when that measure is explicitly calculated and provided" applies similarly to my setting. Under the quality reporting system, the 1st edition of the report cards does not provide a direct measure on the "singleton success rate," although the "singleton success rate" could have been calculated using the available information on the "success rate" and "multiple birth rate." The new version does the calculation for consumers and displays the number in highlighted form. Therefore, we expect consumers to become more sensitive to the "singleton birth rate" after the second version is introduced. Since the "multiple birth rate" is negatively correlated with the "singleton success rate," we expect that after the change in the report card format, a lower "multiple-birth rate" would increase the market share more. On the above basis, I form the following hypothesis.

Hypothesis 2: After CDC changes the format of report cards, consumes starts to value a lower "multiple birth rate"; therefore, controlling for the trend before the format change, the market share of clinics with a higher "multiple birth rate" starts to decrease.

The null hypothesis against the above information-processing hypothesis is that other things equal, clinics with differential "multiple-birth rates" do not see a difference in the change of market share after the change in the report card format. This null hypothesis holds if consumers are able to correctly calculate the "singleton success rate" score based on the "success rate" and the "multiple birth rate" or if they simply do not care about the "multiple birth rate," which means the only item that enters their utility is the "success rate."

IV. Data

The primary data set used in this paper is quality and market share information reported by fertility clinics to SART from 1993 to 2006. It covers around 90% clinics for each year. For each age group (i.e., younger than 35, 35-37, 38-40, older than 40), the data contains yearly clinic information on its total number of cycles performed, the average number of embryos transferred, several outcome measures (e.g., the birth success rate and multiple-birth rate) and clinic characteristics (e.g., years of tenure, SART membership, whether it accepts single women, whether it offers surrogate services, whether it offers donor-egg services). In addition, I obtain data containing the information about whether the clinic is affiliated with teaching hospitals and universities.¹⁴ The 13-year panel is long enough to cover the two mandatory changes on public reporting in the information context faced by consumers—(a) CDC starts reporting ART success rates in Dec. 1997, and (b) CDC changes the format of ART success rates reports in Dec. 2003. Figure 4 shows the timeline of these two events.

[Insert Figure 4 Here]

¹⁴ I appreciate the generous help from Kate Bundorf and Barton Hamilton in sharing the data.

Since the report cards disclose unadjusted success rate (i.e. number of births divided by number of cycles) for each age group, I generate an age-adjusted success rate in the following way. For each year t , I first regress the clinic-level unadjusted success $SR_{jt}^{\text{unadjusted}}$ on the share of patients in each age group. I then predict the success rate for each clinic: $SR_{jt}^{\text{predicted}}$. Second, I generate the age-adjusted success rate for each clinic using

$$SR_{jt}^{\text{adjusted}} = \frac{SR_{jt}^{\text{unadjusted}}}{SR_{jt}^{\text{predicted}}} \times SR_t^{\text{average}},$$

in which SR_t^{average} is the average unadjusted success rate for all clinics in year t .¹⁵ Table 1 presents the summary statistics of clinic-year level observations between 1996 and 2006. Notice that I use market share information from 1996 to 2006 and combine them with the 3-year lagged quality measures (i.e., quality measures from 1993 to 2003), as well as the 1-year lagged quality measures (i.e., quality measures from 1995 to 2005). To minimize the small-sample size bias, I exclude clinic-year observations of which the 3-year/1-year lagged number of cycles is less than 15¹⁶, or the 3-year/1-year lagged number of births is 0.

[Insert Table 1 Here]

V. Empirical Framework

Our basic empirical strategy is to compare the relationship between market share and the two quality measures (i.e. success rate and multiple-birth rate) before and after the change of report card format. Following (Bundorf et al. 2009) and (Hamilton and McManus 2005), I define markets as Metropolitan Statistical Areas (MSAs). As discussed in Section II.A, the high frequency of clinic visits

¹⁵ Alternatively, following specifications in (Bundorf et al. 2009), I replace adjusted success rate with unadjusted rate and age-mix. The results are robust to these changes.

¹⁶ The results are robust to alternative choices, such as 5, 10 or 20.

often prevents couples from traveling out of their own MSAs to seek treatment. Table 2 shows that in year 2001, for example, there are 118 markets and 68 of them contain more than one fertility clinics; for markets containing more than one clinics, there exists substantial difference between the highest and the lowest score in success rate and multiple-birth rate. This within-market variation in multiple-birth rate for the same year serves as our primary source of identification. The publication of the 1st edition report cards that highlights success rate, and the subsequent change in report-card format which makes a clinic with a higher multiple-birth rate looks “worse”, therefore effectively creates many "mini-experiments," which I explore to identify how consumers respond to report-card publication/format change.

[Insert Table 2 Here]

I estimate a discrete choice demand model by assuming that the consumers’ choice of clinic follows a nested logit model. I define the choice set facing the consumer as either no treatment or treatment from one of the clinics in the local market. As is well known, the standard assumption of i.i.d. errors in the traditional logit model produces stringent restrictions on substitution patterns across options. Therefore, I create a separate "nest" for choosing among clinics in the local market to permit the substitution among clinics to differ from the substitution between clinics and no treatment. This allows for consumers’ tastes to be correlated across clinics. As in (Bundorf et al. 2009), I define the potential population to be women of child-bearing age (25-49). Therefore, the utility of consumer i in year t from choosing clinic j (and thus nest g) in market c can be written as:

$$(1) \quad u_{ijct} = X_{jct}\beta + \xi_{jct} + \varphi_i\zeta_g + (1 - \sigma)\varepsilon_{ijct}$$

, X_{jct} are the observed clinic-market-year characteristics (described below), ξ_{jct} represents the mean utility of unobserved clinic-market-year characteristics, ζ_g is a dummy for choices in nest g , and ε_{ijt} is an i.i.d. extreme value error term. φ_i is an independent random term that reflects individual-specific preferences for nest g . Following (Berry 1994), I assume the distribution of φ_i is such that $\varphi_i + (1 - \sigma)\varepsilon_{ijct}$ is an extreme value random variable. Under this assumption, σ ranges between 0 and 1. As σ approaches 1, the within-nest correlation of utility levels is high, and as σ approaches 0 the substitution pattern does not differ across nests. I normalize the mean utility of the outside good "no treatment" to be zero. Consumer i makes choice j if and only if $\forall k \geq 0$ and $k \neq j$, $u_{ijct} \geq u_{ikct}$.

For my primary specification, I follow (Bundorf et al. 2009) and use a difference-in-difference approach to identify the demand-side response to format change. Since I observe quality levels of each clinic both before and after report-card publication/format change, I use the correlation between quality and consumer choice prior to the event as a control for the time-invariant correlation between quality and the unobserved sources of quality information that determine consumer choice. I define the observed clinic-market-year characteristics to be

$$(2) \quad X_{jct} = \sum_l (\text{score}_{jct-3}^l + \text{score}_{jct-3}^l \times \text{After}_t) + \text{Characteristics}_{jct} + \kappa_c + \psi_t$$

The specification includes the score for each quality measure l (i.e., success rate or multiple-birth rate) released at time t , score_{jct-3}^l to capture the consumers' market learning about quality, as well as interactions between each score score_{jct-3}^l and an after dummy After_t . For the effect of the report-card publication, After_t takes the value of 1 beginning in 1998; for the effect of the report-card format change, After_t takes the value of 1 beginning in 2004. Since I observe quality scores both before and after the mandatory report-card publication/format change, this allows me to use the correlation between quality

scores and consumer choices before the report-card publication/format change format change to control for time-invariant relationship between quality scores and unobserved sources of quality information, as well as correlation between quality scores and unobserved clinic characteristics. The coefficient before the interaction $score_{jct-3}^l \times After_t$ thus captures how consumers respond to the new information in report cards. In all specifications, I include the market fixed effect κ_c to control for time-invariant differences in unobserved consumer utility across markets, as well as the year fixed effect ψ_t to control for unobserved factors trending over time that affect all markets.

For robustness checks, I test two alternative specifications of the discrete choice model. In the first alternative specification (3), I add the one-year lagged quality score. Because the report card only reports three-year lagged performance, including the one-year lagged performance addresses the possibility that the demand change is in response to the clinics' contemporaneous changes in quality. It also controls for the time-varying correlation between unobserved sources of quality information and the reported quality measures.

$$(3) \quad X_{jct} = \sum_l (score_{jct-3}^l + score_{jct-3}^l \times After_t + score_{jct-1}^l + score_{jct-1}^l \times After_t) + Characteristics_{jct} + \kappa_c + \psi_t$$

In the second alternative specification (4), I aim to test whether the observed response to report-card publication/format change is merely an artifact of pre-intervention trend. I replace the dummy variable $After_t$ with a set of year dummy variables $Year_t$. If the coefficients for years after the intervention are significantly different from those before the intervention, it indicates there is discontinuity in response and suggests the after-intervention response is not merely an extension from pre-intervention trend.

$$(4) \quad X_{jct} = \sum_l (score_{jct-3}^l + score_{jct-3}^l \times Year_t) + Characteristics_{jct} + \kappa_c + \psi_t$$

In the third alternative specification (5), I aim to quantify the response by control for pre-intervention trend in the consumers' response to quality. As in (Dafny and Dranove 2008), I add a separate time trend for each quality measure to capture market learning about that dimension over time.

$$(5) \quad X_{jct} = \sum_l (score_{jct-3}^l \times f(Year_t) + score_{jct-3}^l \times After_t) + Characteristics_{jct} + \kappa_c + \psi_t$$

Specifically I choose the time trend to be of log functional form: $f(Year_t) = \ln(Year_t)$. The concave trend captures the fact that a decreasing percentage of the population learns over time. Since I need a long pre-intervention trend to capture the trend, I use data from 1998 to 2006 (which captures the full time period when the 1st edition is on release) to test only the effect of report-card format change.

Following (Berry, Levinsohn, and Pakes 1995), the utility function in (1) can be rewritten as

$$(2) \quad \ln(s_{jct}) - \ln(s_{0t}) = X_{jct}\beta + \xi_{jct} + \sigma \ln(s_{(j|g)ct}),$$

where s_{jct} denotes the absolute market share of clinic j in market c and year t , s_{0t} denotes the market share of the outside good (i.e. no treatment), and $s_{(j|g)ct}$ denotes the within-group market share—clinic j 's market share among patients who choose to receive treatment in market c . Since the within-group market share is correlated with the mean utility of unobserved clinic quality ξ_{jct} , I need some instruments uncorrelated with the unobserved clinic quality to predict the within-group share. As proposed in (Berry 1994), a proper choice of these instruments would be the average characteristics of competing products that are relatively fixed over time or at least which clinics would not choose in response to the (unobserved) quality of competitors. Following (Bundorf et al. 2009), my choice of competitor characteristics includes whether the clinics accept single women, whether the clinics accept gestational carriers, and whether the clinics accept egg donors.

VI. Results

I first test hypothesis 1 using data from 1996 to 2000. Table 3.1 presents the results from specification (2) and (3) in Column 1 and 2 respectively. Column 2 differs from column 1 in that I add one-year lagged performance to control for responses to unobserved contemporaneous changes in clinic quality and unobserved clinic characteristics. The results do not differ much. Comparing row 1 and row 3, I conclude that after the publication of report cards in 1998, the market share of clinics with higher success rates increases; while that for clinics with lower multiple-birth rates does not increase. This confirms hypothesis 1 in that consumers tend to be favor the highlighted quality measure—the success rate, while remain relatively unresponsive to the reported but not highlighted multiple-birth rate. I then replace the dummy variable for “after the intervention” with a set of year dummies as in specification (4). I omit the last year before the publication of report cards (i.e., year 1997); thus the coefficients capture deviations from responses in the omitted (baseline) year. The results are presented in Table 3.2. It is clear that after CDC releases the report cards in 1998, market share immediately responses to the success rate information. However, that for multiple-birth rate is non-existent. These results suggest the observed response to success rate is not an artifact of pre-report cards trend. This again confirms hypothesis 1.

[Insert Table 3.1 and 3.2 Here]

Similarly, I test hypothesis 2 using data from 2001 to 2006. I employ specification (2) and (3) and report results in Table 4.1. Column 2 includes one-year lagged performance to control for responses to unobserved contemporaneous changes in clinic quality and unobserved clinic characteristics. The results do not differ much. Row 2 shows that under the 1st edition of report cards (i.e., 2001-

2003), the effect of a clinic's reported success rate (i.e., 3-year lagged success rate) on utility is positive, suggesting that consumers are more likely to choose a clinic with a higher success rate. Row 4 shows that consumers also tend to prefer clinics with higher multiple-birth rates. Row 3 presents our main result of interest: after the format change—the effect of a clinic's multiple-birth rate on utility is negative—suggesting that consumers started to dislike higher multiple-birth rates. Row 1 serves as a falsification test: it shows that consumers do not appear to respond the same way to success rate after the format change.

[Insert Table 4.1 Here]

These results are limited because they do not really show a discontinuity in the revealed preference at the time of format changing. One concern is that infertile couples might gradually become more educated and realize the harm of multiple pregnancies. As discussed in Section V, I solve this issue by replacing the after time dummy with a set of year dummies as in specification (4). Table 4.2 shows the corresponding results without and with one-year lagged performance as controls for unobserved quality change or clinic characteristics. I omit the last year before the report-card format change. Row 6 to row 10 jointly show that there is no pre-change trend towards a lower multiple-birth rate clinic, suggesting the observed response to multiple-birth rate should be induced by the format change—the newly added and highlighted singleton success rate makes a clinic with lower multiple-birth rate look more attractive. Another robustness check I adopt is to add an interaction term between the 3-year lagged quality measures and a log-year term as in specification (5). I use data from year 1998 to capture a long enough pre-change period (i.e., 1998 to 2003). The result is shown in Table 4.3. Row 2 shows that even after controlling for a log pre-change trend in consumers' unobserved preference, the effect of multiple-birth rate on utility is

still significantly positive. Row 3 and 4 present the effect of market based learning. Overall, there is a trend of towards clinics with higher success rate and multiple-birth rates; nevertheless, the format change in 2004 leads consumers away from clinics with higher multiple-birth rates. Column 1 suggests that after controlling for pre-change trend, there is no significant change in how success rate affects utility before and after the format change.

[Insert Table 4.2 and 4.3 Here]

VII. Conclusions

This study examines how information presentation in public reporting systems influences consumer choice by analyzing the ART success rates reports for all fertility clinics in the U.S. The data set covers the time period before and after the release of the 1st edition report cards in 1998, and the enforcement of a mandatory change in the report card format in 2004. The 1st edition reports both success rates and multiple-birth rates while highlighting only success rates; while the 2nd edition adds and highlights a "singleton success rate" measure, which could have been calculated using all the information reported in the old edition. All else equal, a clinic with a higher multiple-birth rate will have a lower singleton success rate. I find that after the report card format change, consumers become more sensitive to multiple-birth-rate; they start to choose clinics with lower multiple-birth rates.

Previous studies on information disclosure focus on whether public disclosure of quality information affects consumer choice. This paper extends the literature by showing that information presentation is a key determinant of consumers' responsiveness to report cards. These results are robust to alternative specifications, including adding 1-year lagged quality measures which are not reported to capture unobserved quality change in response to public reporting or

report-card format change, as well as unobserved clinic characteristics, and controlling for the possibility of pre-intervention trend. One caveat is that I do not observe contemporaneous price change by fertility clinics. To the extent that clinics with higher multiple-birth rate reduced price in response to report-card format change and thus attract some price-sensitive consumers (e.g., those with no insurance coverage for infertility treatments), the estimated effects of the report-card publication/format change on consumer choice will be an underestimate of the actual response. Overall the evidence suggests that, when launching a public reporting system with the hope of facilitating consumer shopping decision, simply disclosing information on multiple product attributes is insufficient to guarantee there is a demand side response because consumers incur internal costs to understand or to analyze the disclosed information; furthermore, improper design of information presentation might lead consumers to overvalue/undervalue certain product attributes. Therefore, policymakers should always keep in mind the importance of information presentation when designing report cards.

There is one interesting extension of this paper—since information presentation affects consumer choice, and providers have an incentive to tailor its product in response to consumer responsiveness, there is a potential that information presentation may also evoke a supply side response in improving quality along the more vividly/explicitly displaced dimension. In this setting of ART success rate reports, clinics may have an incentive to transplant greater number of embryos to boost the success rate in response to the release of the 1st edition report cards. This incentive should be minimized after the report-card format change, as consumers start to move away from clinics with very high multiple-birth rates. Further research using detailed supply side information is needed to test whether this indeed happens.

REFERENCES

- Asch, Solomon E. 1946. "Forming impressions of personality." *The Journal of Abnormal and Social Psychology* 41 (3):258.
- Barber, Brad M, and Terrance Odean. 2001. "The internet and the investor." *Journal of Economic Perspectives*:41-54.
- Barber, Brad M, and Terrance Odean. 2002. "Online investors: do the slow die first?" *Review of Financial Studies* 15 (2):455-488.
- Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile prices in market equilibrium." *Econometrica: Journal of the Econometric Society*:841-890.
- Berry, Steven T. 1994. "Estimating discrete-choice models of product differentiation." *The RAND Journal of Economics*:242-262.
- Bettman, James R. 1970. "Information processing models of consumer behavior." *Journal of Marketing Research*:370-376.
- Bettman, James R, and Pradeep Kakkar. 1977. "Effects of information presentation format on consumer information acquisition strategies." *Journal of Consumer Research* 3 (4):233.
- Bundorf, M Kate, Natalie Chun, Gopi Shah Goda, and Daniel P Kessler. 2009. "Do markets respond to quality information? The case of fertility clinics." *Journal of health economics* 28 (3):718-727.
- Callahan, Tamara L, Janet E Hall, Susan L Ettner, Cindy L Christiansen, Michael F Greene, and William F Crowley Jr. 1994. "The economic impact of multiple-gestation pregnancies and the contribution of assisted-reproduction techniques to their incidence." *New England Journal of Medicine* 331 (4):244-249.
- Dafny, Leemore, and David Dranove. 2008. "Do report cards tell consumers anything they don't already know? The case of Medicare HMOs." *The RAND Journal of Economics* 39 (3):790-821.
- Dranove, David, and Ginger Zhe Jin. 2010. "Quality Disclosure and Certification: Theory and Practice." *Journal of Economic Literature* 48 (4):935-63.
- Evers, Johannes LH. 2002. "Female subfertility." *The Lancet* 360 (9327):151-159.
- Eysenck, Michael W, and Mark T Keane. 2000. *Cognitive psychology: A student's handbook*: Psychology Press.
- Gilovich, Thomas, Robert Vallone, and Amos Tversky. 1985. "The hot hand in basketball: On the misperception of random sequences." *Cognitive psychology* 17 (3):295-314.
- Glazer, Rashi, Joel H Steckel, and Russell S Winer. 1992. "Locally rational decision making: The distracting effect of information on managerial performance." *Management Science* 38 (2):212-226.

- Granger, Clive WJ, and Andrew Billson. 1972. "Consumers' Attitudes Toward Package Size and Price." *Journal of Marketing Research (JMR)* 9 (3).
- Hagerty, Michael R, and David A Aaker. 1984. "A normative model of consumer information processing." *Marketing Science* 3 (3):227-246.
- Hamilton, Barton H, and Brian McManus. 2005. "Infertility treatment markets: The effects of competition and policy." Washington University Manuscript.
- Hong, Sung-Tai, and Robert S Wyer. 1989. "Effects of country-of-origin and product-attribute information on product evaluation: An information processing perspective." *Journal of Consumer Research*.
- Isakson, Hans R, and Alex R Maurizi. 1973. "The Consumer Economics of Unit Pricing." *Journal of Marketing Research (JMR)* 10 (3).
- Jones Jr, Howard W, and John A Schnorr. 2001. "Multiple pregnancies: a call for action." *Fertility and sterility* 75 (1):11-13.
- Lurie, Nicholas H, and Charlotte H Mason. 2007. "Visual representation: Implications for decision making." *Journal of Marketing* 71 (1):160-177.
- Lussier, Denis A, and Richard W Olshavsky. 1979. "Task complexity and contingent processing in brand choice." *Journal of Consumer Research*:154-165.
- Mandel, Naomi, and Eric J Johnson. 2002. "When web pages influence choice: effects of visual primes on experts and novices." *Journal of Consumer Research* 29 (2):235-245.
- Marshall, MN, PG Shekelle, S Leatherman, and RH Brook. 2000. "The public release of performance data: what do we expect to gain? A review of the evidence." *JAMA: the journal of the American Medical Association* 283 (14):1866.
- Min, Jason K, Sue A Breheny, Vivien MacLachlan, and David L Healy. 2004. "What is the most relevant standard of success in assisted reproduction? The singleton, term gestation, live birth rate per cycle initiated: the BESST endpoint for assisted reproduction." *Human Reproduction* 19 (1):3-7.
- Nelson, Phillip. 1970. "Information and consumer behavior." *The Journal of Political Economy*:311-329.
- Nisbett, Richard E, and Lee Ross. 1980. *Human inference: Strategies and shortcomings of social judgment*: Prentice Hall.
- Russo, J Edward, Gene Krieser, and Sally Miyashita. 1975. "An effective display of unit price information." *Journal of Marketing* 39 (2):11-19.
- Shugan, Steven M. 1980. "The cost of thinking." *Journal of Consumer Research*.
- Smith, Gerald E, Meera P Venkatraman, and Ruby Roy Dholakia. 1999. "Diagnosing the search cost effect: Waiting time and the moderating impact of prior category knowledge." *Journal of Economic Psychology* 20 (3):285-314.

- Soman, Dilip, and Mengze Shi. 2003. "Virtual progress: The effect of path characteristics on perceptions of progress and choice." *Management Science* 49 (9):1229-1250.
- Stone, Eric R, Winston R Sieck, Benita E Bull, J Frank Yates, Stephanie C Parks, and Carolyn J Rush. 2003. "Foreground: background salience: Explaining the effects of graphical displays on risk avoidance." *Organizational Behavior and Human Decision Processes* 90 (1):19-36.
- Stone, Eric R, J Frank Yates, and Andrew M Parker. 1997. "Effects of numerical and graphical displays on professed risk-taking behavior." *Journal of Experimental Psychology: Applied* 3 (4):243.
- Zeithaml, Valarie A. 1982. "Consumer response to in-store price information environments." *Journal of Consumer Research*:357-369.

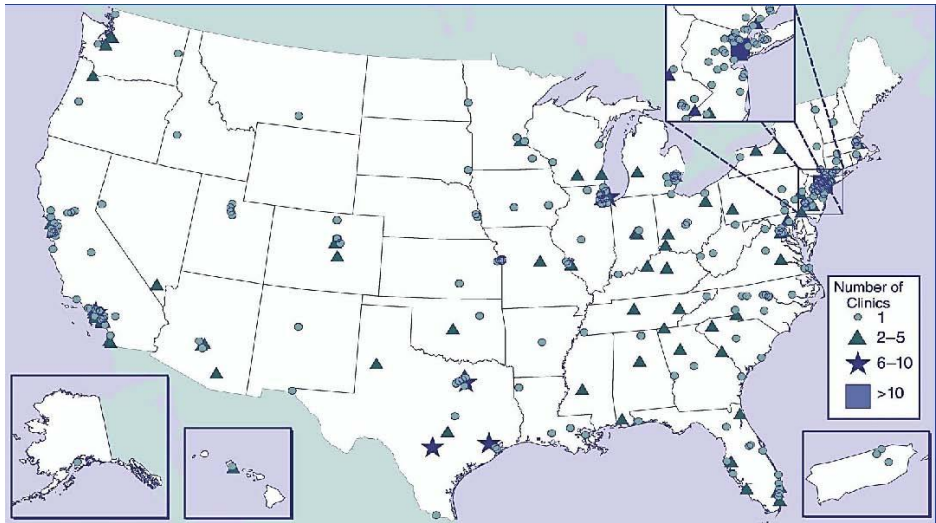


FIGURE 1. LOCATIONS OF ART CLINICS IN THE U.S. AND PUERTO RICO, 2011

Notes: Each point represents a city. The actual market defined in this paper, as in prior literature, is Metropolitan Statistical Areas, which might incorporate multiple cities. (Source: <http://www.cdc.gov/art/ART2011/section1.htm>)

1997 ART CYCLE PROFILE		Patient Diagnosis			
	Type of ART	Tubal factor	21%	Other factor	33%
IVF	99%	Ovulatory dysfunction	12%	Unknown factor	<1%
GIFT	0%	Diminished ovarian reserve		Multiple Factors:	
ZIFT	1%	Endometriosis	17%	Female factors	
Combination	0%	Uterine Factor	2%	Female & Male factors	
		Male factor	15%		
	Procedural Factors				
	With ICSI	31%			
	Unstimulated	0%			
	Used gestational carrier				
	Used PGD				
	With eSET				

1997 PREGNANCY SUCCESS RATES		Age of Women				
Type of Cycle		<35	35-37	38-40	41-42	43-44
Fresh Embryos From Nondonor Eggs						
Number of cycles		204	139	125	89	
Percentage of embryos transferred resulting in implantation ^a						
Percentage of cycles resulting in pregnancies ^b		68.6	51.8	42.4	30.3	
Percentage of cycles resulting in live births ^{b,c}		59.3	46.8	31.2	20.2	
(Confidence Interval)		(52.6 - 66.1)	(38.5 - 55.1)	(23.1 - 39.3)	(11.9 - 28.6)	
Percentage of retrievals resulting in live births ^{b,c}		63.7	59.1	41.9	33.3	
Percentage of transfers resulting in live births ^{b,c}		65.1	59.1	42.4	34.0	
Percentage of cancellations ^b		6.9	20.9	25.6	39.3	
Average number of embryos transferred		3.6	3.5	3.9	4.4	
Percentage of pregnancies with twins ^b		37.1	34.7	26.4	22.2	
Percentage of pregnancies with triplets or more ^b		17.9	9.7	7.5	3.7	
Percentage of live births having multiple infants ^{b,c}		52.1	41.5	33.3	5/18	
Frozen Embryos From Nondonor Eggs						
Number of transfers		31	11	13	11	
Percentage of transfers resulting in live births ^{b,c}		38.7	3/11	1/13	0/11	
Average number of embryos transferred		3.7	4.3	4.2	4.5	
		All Ages Combined^e				
		Fresh Embryos		Frozen Embryos		
		100		31		
Donor Eggs						
Number of transfers						
Percentage of transfers resulting in live births ^{b,c}						
Average number of embryos transferred						
Current Clinic Services and Profile						
Current Name: Sample Clinic Name						
Donor egg?	Yes	Gestational Carriers?	Yes	SART Member?	Yes	Verified lab accreditation?
Donor embryo?		Cryopreservation?				
Single women?	Yes					

FIGURE 2. A SAMPLE OF ART SUCCESS RATES REPORTS—1ST EDITION (VALID DEC. 1997 – DEC. 2002)

Notes: The reported quality measures are three-year lagged performance. That is, scores released in Dec. 1997 reports quality in 1995.

2001 ART CYCLE PROFILE				Patient Diagnosis			
Type of ART	Type of ART ^a			Patient Diagnosis			
	IVF	100%	Procedural Factors		Tubal factor	8%	Other factor
GIFT	0%	With ICSI	58%	Ovulatory dysfunction	2%	Unknown factor	9%
ZIFT	0%	Unstimulated	0%	Diminished ovarian reserve	18%	Multiple Factors:	
Combination	0%	Used gestational carrier	3%	Endometriosis	9%	Female factors	17%
		Used PGD		Uterine Factor	2%	Female & Male factors	11%
		With eSET		Male factor	11%		

2001 PREGNANCY SUCCESS RATES					
Type of Cycle	Age of Women				
	<35	35-37	38-40	41-42	43-44
Fresh Embryos From Nondonor Eggs					
Number of cycles	221	121	108	48	
Percentage of embryos transferred resulting in implantation ^b					
Percentage of cycles resulting in pregnancies ^b	66.1	64.5	41.7	45.8	
Percentage of cycles resulting in live births ^{b,c}	58.4	52.9	32.4	20.8	
(Confidence Interval)	(51.9 - 64.9)(44.0 - 61.8)(23.6 - 41.2)(9.3 - 32.3)				
Percentage of retrievals resulting in live births ^{b,c}	60.3	53.8	36.1	24.4	
Percentage of transfers resulting in live births ^{b,c}	60.6	54.2	36.8	24.4	
Percentage of transfers resulting in singleton live births ^b	27.7	29.7	23.2	19.5	
Percentage of cancellations ^b	3.2	1.7	10.2	14.6	
Average number of embryos transferred	3.2	3.4	4.1	4.5	
Percentage of pregnancies with twins ^b	39.7	42.3	24.4	18.2	
Percentage of pregnancies with triplets or more ^b	15.8	12.8	22.2	4.5	
Percentage of live births having multiple infants ^{b,c}	54.3	45.3	37.1	2 / 10	
Frozen Embryos From Nondonor Eggs					
Number of transfers	42	26	14	4	
Percentage of transfers resulting in live births ^{b,c}	52.4	38.5	6 / 14	0 / 4	
Average number of embryos transferred	3.2	2.8	3.9	2.8	
All Ages Combined^d					
Donor Eggs		Fresh Embryos		Frozen Embryos	
Number of transfers		197			56
Percentage of transfers resulting in live births ^{b,c}		70.6			33.9
Average number of embryos transferred		2.9			3.6
Current Clinic Services and Profile					
Current Name: Sample Clinic Name					
Donor egg?	Yes	Gestational Carriers?	Yes	SART Member?	Yes
Donor embryo?	Yes	Cryopreservation?	Yes	Verified lab accreditation?	Yes
Single women?	Yes				

FIGURE 3. A SAMPLE OF ART SUCCESS RATES REPORTS—2ND EDITION (VALID DEC. 2003 – PRESENT)

Notes: The reported quality measures are three-year lagged performance. That is, scores released in Dec. 2003 reports quality in 2001.

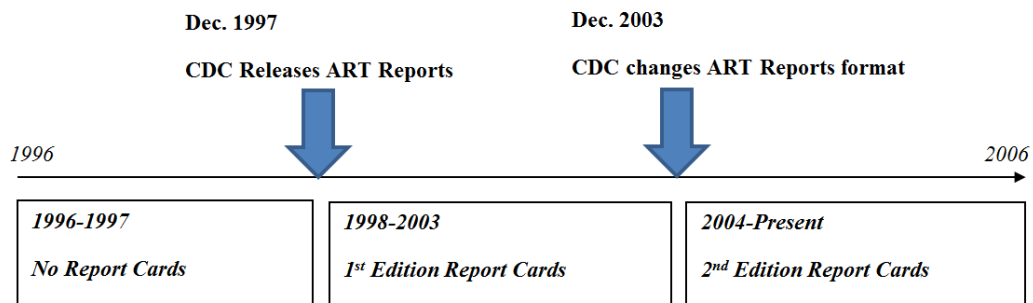


FIGURE 4. TIMELINE OF THE TWO MANDATORY CHANGES

Notes: The reported quality measures are three-year lagged performance. That is, scores released in Dec. 2003 reports quality in 2001.

TABLE 1—SUMMARY STATISTICS ON CLINIC CHARACTERISTICS (YEAR 1996-2006)

	Mean	SD	Min	Max	N
Number of Treatment Cycles	198	280	0	3073	4131
Adjusted Success Rate (3-year lagged)	.243	.087	0	.748	3073
Adjusted Success Rate (1-year lagged)	.258	.086	0	.802	3625
Multiple Birth Rate (3-year lagged)	.354	.140	0	1	3073
Multiple Birth Rate (1-year lagged)	.349	.130	0	1	3625
Tenure (number of years since opening)	6.87	3.62	1	14	4131
SART Membership	.922	.269	0	1	4131
Accept Single Women	.830	.376	0	1	4131
Accept Gestational Carriers	.600	.490	0	1	4131
Accept Egg Donors	.679	.467	0	1	4131
Teaching Hospital or University Affiliation	.186	.389	0	1	4131

Notes: Unit of observation is clinic-year. When reporting 3-year/1-year lagged quality measures, observations of which the 3-year/1-year lagged performance is missing, the 3-year/1-year lagged number of births is 0, or the 3-year/1-year lagged number of cycles is less than 15 are excluded.

TABLE 2—VARIATION IN QUALITY SCORES WITHIN THE SAME MARKET (YEAR 2001)

Number of Clinics in the Market	Number of Markets	Mean of (Max-Min)	
		Success Rate	Multiple-birth Rate
1	50	-	-
2	28	.09	.14
3+	40	.23	.31
Total	118	.17	.24

Notes: This table presents the average difference between the highest and the lower quality scores within a market, for markets which contain 2 or more clinics, for success rate and multiple-birth rate respectively.

TABLE 3.1—EFFECT OF SUCCESS RATE AND MULTIPLE-BIRTH RATE ON MEAN UTILITY, BEFORE AND AFTER THE RELEASE OF ART SUCCESS RATES REPORTS IN 1998. (PANEL 1996 TO 2000)

	(1)	(2)
$SuccessRate_{t-3} \times After97_t$.394 (.305)	.216 (.330)
$SuccessRate_{t-3}$.077 (.210)	.085 (.255)
$MultipleBirthRate_{t-3} \times After97_t$	-.005 (.134)	.084 (.169)
$MultipleBirthRate_{t-3}$.008 (.102)	-.026 (.122)
$SuccessRate_{t-1} \times After97_t$.315 (.255)
$SuccessRate_{t-1}$.203 (.303)
$MultipleBirthRate_{t-1} \times After97_t$.092 (.176)
$MultipleBirthRate_{t-1}$.077 (.106)
$\ln(\text{within MSA market share})$.809*** (.117)	.752*** (.145)

N

1138

1089

Notes: $SuccessRate_{t-3}$ and $SuccessRate_{t-1}$ are adjusted by the share of patients in each age group. $After97_t$ takes a value of 1 if after year 1997. All models control for additional clinic characteristics, including tenure, SART membership, whether the clinic accepts single women, whether the clinic accepts egg donors, whether the clinic accepts gestational carriers, whether the clinic is affiliated with a teaching hospital or university. All models also control for market- and time (year)-fixed effects. Standard errors in parentheses are clustered by clinic. Observations where the 3-year/1-year lagged performance is missing, the number of births is 0, or the number of cycles is less than 15 are dropped from the analysis.

Robust Standard Errors are in parenthesis.

*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

TABLE 3.2—ROBUSTNESS CHECK (EXAMINING PRE-CHANGE TREND): EFFECT OF SUCCESS RATE AND MULTIPLE-BIRTH RATE ON MEAN UTILITY, BEFORE AND AFTER THE RELEASE OF ART SUCCESS RATES REPORTS IN 1998. (PANEL 1996 TO 2000)

	(1)	(2)
$SuccessRate_{t-3} \times Year1996$ (Before)	.049 (.404)	.067 (.376)
$SuccessRate_{t-3} \times Year1998$ (After)	.297 (.358)	.244 (.378)
$SuccessRate_{t-3} \times Year1999$ (After)	.487 (.405)	.419 (.416)
$SuccessRate_{t-3} \times Year2000$ (After)	.523 (.403)	.412 (.424)
$SuccessRate_{t-3} \times Year1996$ (Before)	-.079 (.143)	-.074 (.162)
$SuccessRate_{t-3} \times Year1998$ (After)	.038 (.167)	.071 (.184)
$SuccessRate_{t-3} \times Year1999$ (After)	-.204 (.139)	-.115 (.172)
$SuccessRate_{t-3} \times Year2000$ (After)	.032 (.201)	.158 (.246)
$SuccessRate_{t-1}$.412 (.265)
$MultipleBirthRate_{t-1}$.122* (.074)
$\ln(\text{within MSA market share})$.801*** (.116)	.757*** (.139)
N	1138	1089

Notes: $SuccessRate_{t-3}$ and $SuccessRate_{t-1}$ are adjusted by the share of patients in each age group. All models control for additional clinic characteristics, including tenure, SART membership, whether the clinic accepts single women, whether the clinic accepts egg donors, whether the clinic accepts gestational carriers, whether the clinic is affiliated with a teaching hospital or university. All models also control for market- and time (year)-fixed effects. Standard errors in parentheses are clustered by clinic. Observations where the 3-year/1-year lagged performance is missing, the number of births is 0, or the number of cycles is less than 15 are dropped from the analysis.

Robust Standard Errors are in parenthesis.

*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

TABLE 4.1—EFFECT OF SUCCESS RATE AND MULTIPLE-BIRTH RATE ON MEAN UTILITY, BEFORE AND AFTER THE REPORT CARD FORMAT CHANGE IN 2004 (PANEL 2001 TO 2006)

	(1)	(2)
$SuccessRate_{t-3} \times After03_t$	-.050 (.151)	-.125 (.181)
$SuccessRate_{t-3}$.535** (.254)	.454* (.243)

$MultipleBirthRate_{t-3} \times After03_t$	-0.244** (.103)	-0.202** (.093)
$MultipleBirthRate_{t-3}$.186** (.088)	.112 (.069)
$SuccessRate_{t-1} \times After03_t$.054 (.172)
$SuccessRate_{t-1}$.203 (.170)
$MultipleBirthRate_{t-1} \times After03_t$		-.200** (.101)
$MultipleBirthRate_{t-1}$.227** (.097)
$\ln(\text{within MSA market share})$.774*** (.094)	.766*** (.103)
N	1933	1882

Notes: $SuccessRate_{t-3}$ and $SuccessRate_{t-1}$ are adjusted by the share of patients in each age group. $After03_t$ takes a value of 1 if after year 2003. All models control for additional clinic characteristics, including tenure, SART membership, whether the clinic accepts single women, whether the clinic accepts egg donors, whether the clinic accepts gestational carriers, whether the clinic is affiliated with a teaching hospital or university. All models also control for market- and time (year)-fixed effects. Standard errors in parentheses are clustered by clinic. Observations where the 3-year/1-year lagged performance is missing, the number of births is 0, or the number of cycles is less than 15 are dropped from the analysis.

Robust Standard Errors are in parenthesis.

*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

TABLE 4.2—ROBUSTNESS CHECK (EXAMINING PRE-CHANGE TREND): EFFECT OF SUCCESS RATE AND MULTIPLE-BIRTH RATE ON MEAN UTILITY, BEFORE AND AFTER THE REPORT CARD FORMAT CHANGE IN 2004 (PANEL 2001 TO 2006)

	(1)	(2)
$SuccessRate_{t-3} \times Year2001$ (Before)	-.021 (.253)	.088 (.232)
$SuccessRate_{t-3} \times Year2002$ (Before)	.096 (.239)	.178 (.237)
$SuccessRate_{t-3} \times Year2004$ (After)	-.039 (.206)	.028 (.182)
$SuccessRate_{t-3} \times Year2005$ (After)	-.022 (.202)	-.024 (.189)
$SuccessRate_{t-3} \times Year2006$ (After)	-.018 (.228)	-.071 (.228)
$MultipleBirthRate_{t-3} \times Year2001$ (Before)	.023 (.168)	-.0004 (.160)
$MultipleBirthRate_{t-3} \times Year2002$ (Before)	-.143 (.180)	-.095 (.176)
$MultipleBirthRate_{t-3} \times Year2004$ (After)	-.299* (.180)	-.287 (.180)
$MultipleBirthRate_{t-3} \times Year2005$ (After)	-.348* (.181)	-.299* (.176)
$MultipleBirthRate_{t-3} \times Year2006$ (After)	-.305* (.173)	-.285* (.163)
$SuccessRate_{t-1}$.177 (.124)
$MultipleBirthRate_{t-1}$.135* (.076)
$\ln(\text{within MSA market share})$.767*** (.092)	.761*** (.103)
N	1933	1882

Notes: $SuccessRate_{t-3}$ and $SuccessRate_{t-1}$ are adjusted by the share of patients in each age group. $After_t$ takes a value of 1 if after year 2003. All models control for additional clinic characteristics, including tenure, SART membership, whether the clinic accepts single women, whether the clinic accepts egg donors, whether the clinic accepts gestational carriers, whether the clinic is affiliated with a teaching hospital or university. All models also control for market- and time (year)-fixed effects. Standard errors in parentheses are clustered by clinic. Observations where the 3-year/1-year lagged performance is missing, the number of births is 0, or the number of cycles is less than 15 are dropped from the analysis.

Robust Standard Errors are in parenthesis.

*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

TABLE 4.3—ROBUSTNESS CHECK (CONTROLLING FOR PRE-CHANGE TREND): EFFECT OF SUCCESS RATE AND MULTIPLE-BIRTH RATE ON MEAN UTILITY, BEFORE AND AFTER THE REPORT CARD FORMAT CHANGE IN 2004 (PANEL 1998 TO 2006)

SuccessRate_{t-3} × After03_t	-106 (.176)
MultipleBirthRate_{t-3} × After03_t	-164* (.08)
SuccessRate_{t-3} × ln(Year_t)	.194* (.115)
MultipleBirthRate_{t-3} × ln(Year_t)	-.021 (.053)
ln(within MSA market share)	.774*** (.081)
N	2695

Notes: SuccessRate_{t-3} and SuccessRate_{t-1} are adjusted by the share of patients in each age group. After03_t takes a value of 1 if after year 2003. All models control for additional clinic characteristics, including tenure, SART membership, whether the clinic accepts single women, whether the clinic accepts egg donors, whether the clinic accepts gestational carriers, whether the clinic is affiliated with a teaching hospital or university. All models also control for market- and time (year)-fixed effects. Standard errors in parentheses are clustered by clinic. Observations where the 3-year/1-year lagged performance is missing, the number of births is 0, or the number of cycles is less than 15 are dropped from the analysis.

Robust Standard Errors are in parenthesis.

*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.