Predicting personalized multiple birth risks after in vitro fertilization–double embryo transfer

Benjamin M. Lannon, M.D.,[a,b,c] Bokyung Choi, M.Sc.,[d,e] Michele R. Hacker, Sc.D.,[b,c] Laura E. Dodge, M.P.H.,[b] Beth A. Malizia, M.D.,[f] C. Brent Barrett, Ph.D.,[a,b,c] Wing H. Wong, Ph.D.,[d,g] Mylene W. M. Yao, M.D.,[d] and Alan S. Penzias, M.D.[a,b,c]

[a] Boston IVF, Waltham, Massachusetts; [b] Department of Obstetrics and Gynecology, Beth Israel Deaconess Medical Center, Boston, Massachusetts; [c] Department of Obstetrics, Gynecology, and Reproductive Biology, Harvard Medical School, Boston, Massachusetts; [d] Univfy Inc., Los Altos, California; [e] Department of Applied Physics and [g] Department of Statistics, School of Humanities and Sciences, Stanford University, Stanford, California; and [f] Alabama Fertility Specialists, Birmingham, Alabama

Objective: To report and evaluate the performance and utility of an approach to predicting IVF–double embryo transfer (DET) multiple birth risks that is evidence-based, clinic-specific, and considers each patient's clinical profile.

Design: Retrospective prediction modeling.

Setting: An outpatient university-affiliated IVF clinic.

Patient(s): We used boosted tree methods to analyze 2,413 independent IVF-DET treatment cycles that resulted in live births. The IVF cycles were retrieved from a database that comprised more than 33,000 IVF cycles.

Intervention(s): None.

Main Outcome Measure(s): The performance of this prediction model, MBP-BIVF, was validated by an independent data set, to evaluate predictive power, discrimination, dynamic range, and reclassification.

Result(s): Multiple birth probabilities ranging from 11.8% to 54.8% were predicted by the model and were significantly different from control predictions in more than half of the patients. The prediction model showed an improvement of 146% in predictive power and 16.0% in discrimination over control. The population standard error was 1.8%.

Conclusion(s): We showed that IVF patients have inherently different risks of multiple birth, even when DET is specified, and this risk can be predicted before ET. The use of clinic-specific prediction models provides an evidence-based and personalized method to counsel patients. (Fertil Steril® 2012;98:69–76. ©2012 by American Society for Reproductive Medicine.)

Key Words: IVF prediction, prediction model, personalized prognosis, multiple birth, elective single embryo transfer

The greatest risk of advanced reproductive technology today is multiple gestation. The issue of how many embryos to transfer during an IVF treatment cycle is a challenging one for patients and their physicians. Historically, the transfer of multiple embryos led to a greater chance of pregnancy but also an increased risk of multiple birth and its associated costs, including higher risks of neonatal and obstetric complications. Elective single embryo transfer (eSET) eliminates all risk of multiple pregnancy except spontaneous identical twinning but may compromise a patient's chance of having a live birth altogether (1–4).

Although guidelines in the United States have succeeded in reducing the incidence of high-order multiple gestations (triplets or more), the incidence of twin gestations has remained relatively unchanged (5, 6). Among the more than 46,000 live births that resulted from approximately 148,000 IVF cycles initiated in the United States in 2008, 30% were twin gestations, whereas less than 2% were high-order multiples (6).

Under recommendations proposed by the American Society of Reproductive Medicine, patients are placed into categories defined primarily by age and embryo quality to determine whether eSET should be strongly recommended (7). Physicians are asked to adapt these guidelines according

to clinic-specific protocols (8, 9). However, how each clinic should adapt them is unclear (10). One recent study reported the implementation of mandatory SET for women younger than 38 years without a prior IVF failure who produced at least seven zygotes and one good-quality blastocyst (11). Although a laudable effort, its bar is set so high as to address only a fraction of all patients being treated. Deviations from guidelines are inevitable as healthcare providers consider an individual patient's reproductive history and perception of the risks of multiple birth (7, 12–15). Finally, despite reports advocating eSET, especially by European groups, large-scale standardization of criteria and implementation have been difficult to achieve (16). Reasons include patients' skepticism in accepting an option that may decrease the odds of pregnancy and result in additional IVF treatment cycles, as well as conflicting reports on the success of eSET protocols to reduce multiple birth rates or maintain live birth rates (10, 17–19). There is a need to move beyond eSET and to develop an evidence-based method to assess a patient's individual risk of multiple birth, and support counseling and decision making (15).

Clinical decision making aimed specifically at reducing the odds of multiple pregnancy without compromising the chance of conception may be improved by validated prediction models that use patient-specific reproductive health history, response to current treatment, and embryo developmental parameters. A number of predictive models have been developed, but most have focused on assessing an individual's chance of achieving a pregnancy without addressing the specific risk of a multiple birth (18–23). The utility of these models is also limited by methodology, particularly the lack of validation (24). Recent work by Banerjee et al. (25) showed that data from a previous failed IVF treatment cycle could be used to provide a valid and personalized probability of live birth in a future cycle.

Here we address the issue of number of embryos to transfer after IVF by establishing a model to predict a patient's individual risk of multiple births. Although this risk is thought to increase with the number of transferred embryos, the magnitude of the influence of other clinical factors is unclear. We hypothesize that patients have unique probabilities of multiple birth and that these probabilities are influenced by their reproductive health data and by characteristics of their embryos.

Using a large, rigorously maintained data source and strict eligibility criteria, we constructed a training dataset drawn from a cohort of cycles in which at least one live birth resulted after the transfer of two fresh, nondonor embryos. We used this dataset to establish a model for multiple birth prediction at our center (MBP-BIVF) to provide patient-, treatment-, and clinic-specific risk of multiple birth. We tested the performance of this model in an independent set of eligible cycles by quantifying and comparing its performance against those of an age-based control model (24, 25). By limiting analysis to cycles in which live births—single or multiple—resulted after double embryo transfer (DET) in IVF treatments, we showed that patients have very different risks of multiple births even when only two embryos are transferred. Further, we established an approach that may

be applied to improve clinical counseling, with an aim to decrease the incidence of multiple births after IVF.

MATERIALS AND METHODS
Patients, IVF Treatments, and Clinical Outcomes

The retrospective cohort included 33,741 IVF treatment cycles performed at Boston IVF, an academically affiliated large private practice located in Waltham, Massachusetts, from January 1, 2000 to December 31, 2009. Cycles were included if they used transfer of fresh embryos and the patients' own eggs. Patients underwent ovarian stimulation protocols according to physician preference, as described previously (26). Embryos were cultured using standard methods. Ultrasound-guided ET was performed at 3–5 days after oocyte retrieval according to clinic protocol, which for most of the 10-year period preferred day-3 transfer. The number of embryos transferred was based on national and clinic guidelines, as well as individual patient needs. Multiple gestations were confirmed by the presence of more than one fetal heartbeat on ultrasound. Patients were followed for at least 1 year from the start of their IVF cycles to confirm pregnancy outcome.

Data Collection and Exclusion Criteria

Baseline demographic, clinical, and laboratory data were collected according to standard clinic practices, as described previously (27). Medical record review was used to confirm pregnancy outcome. Retrospective data collection, aggregation, deidentification, and analysis for this research project were approved by the Committee on Clinical Investigations at Beth Israel Deaconess Medical Center, Boston, Massachusetts.

We excluded all treatment cycles performed for a patient if any one of her cycles met the following initial exclusion criteria: the first IVF treatment cycle at this clinic did not fall within the study period, the IVF cycles were performed after a patient already had a live birth resulting from IVF, the IVF cycle was cancelled before oocyte retrieval, clinical outcome was not known, or the patient's age was 43.0 years or greater at her first cycle. Subsequently, we excluded IVF treatment cycles that met the second set of exclusion criteria: the cycle resulted in no live birth, occurred after three fresh cycles, or the number of transferred embryo(s) did not equal two.

Variables

Of the 43 variables analyzed, 9 were baseline clinical factors, 13 pertained to IVF treatment cycle response, protocol, or sperm parameters, and 21 were variables that described oocytes, fertilization, embryo development, transferred embryos, or manipulation such as intracytoplasmic sperm injection, assisted hatching, and ET day.

Statistical Analysis

Eligible cycles were assigned to a training set if the start date was between 2000 and 2007, or to the test set if they were initiated between 2008 and 2009. Briefly, we computed the

log-likelihood based on the Bernoulli distribution and applied generalized boosted models (GBM), a free software implementation of a stochastic gradient-boosting algorithm, to build a boosted tree model (MBP-BIVF) using a maximum of 70,000 trees and a 10-fold cross-validation (28). The MBP-BIVF model was compared with an age-based control model (Age-BIVF) that was generated by applying GBM to patient's age alone, according to age categories (<35, 35–37, 38–40, 41–42, and ≥43 years) that are used by the Society for Assisted Reproductive Technologies and Centers for Disease Control and Prevention (6).

Models were generated using the training set and validated using the test set. All the results reported are validated test results, not merely a description of the training set. We used treatment cycles with successful live births, to determine the posterior probability of having multiple births conditioned upon having live birth(s) and the collective phenotype profile of the patient, sperm count of her male partner or donor sperm, and her embryos. Predictive power is described as the improvement in the log-likelihood of predicting the probability of multiple births with MBP-BIVF relative to Age-BIVF prediction, in the context of Baseline-BIVF. Log-likelihoods were computed using GBM. Baseline-BIVF refers to the performance of a prediction model if no predictors were used—the overall multiple birth rate of those 2,413 treatments.

Additional cases were excluded from the analysis according to the second set of exclusion criteria—no live births (n = 11,611), number of transferred embryos were one or more than two (n = 2,132), or the patient already had three unsuccessful treatments (n = 70). After all exclusion criteria were applied, a total of 2,413 cycles were eligible for the development and validation of MBP-BIVF.

Of these 2,413 cycles, 1,789 cycles from 2000 to 2007 (8 years) were assigned to the training set and 624 cycles from 2008 to 2009 (2 years) to the test set (Fig. 1). Some variables had significantly different mean values in the training and the test sets, which demonstrated the independent nature of the training and test sets (Table 1).

The prediction model assigned relative importance to each prognostic factor, with the total relative importance set arbitrarily at 100. The top 10 prognostic factors accounted for approximately 70% of the total relative importance (Fig. 2). Variables with an individual influence of <2.5% were placed under "Others," which included the following variables in order of decreasing relative importance: percentage of oocytes with normal maturation, total motile sperm before wash, total amount of gonadotropins used, average grade of transferred embryos, percentage of "high implantation potential" (HIP) embryos that were cryopreserved, serum day 3 FSH level, percentage of transferred embryos at the eight-cell stage, percentage of eight-cell embryos, number of oocytes, percentage

$$\% \text{ Improvement} = \frac{(\text{Log-likelihood}_{\text{MBP-BIVF}} - \text{Log-likelihood}_{\text{Baseline}}) - (\text{Log-likelihood}_{\text{Age-BIVF}} - \text{Log-likelihood}_{\text{Baseline}})}{(\text{Log-likelihood}_{\text{Age-BIVF}} - \text{Log-likelihood}_{\text{Baseline}})} \times 100\%$$

We calculated the standard error for predicted multiple birth probabilities for the entire population comprising the test set for the MBP-BIVF and Age-BIVF models. In addition, we calculated the standard error of the mean multiple birth probability for each age group for Age-BIVF, as well as the bootstrap estimation of standard error for each of five groups representing quintiles of predicted probabilities of multiple birth for MBP-BIVF. Receiver operating characteristic analysis was used to test the ability of MBP-BIVF to discriminate patients with different probabilities of multiple births. Dynamic range describes the probabilities of live birth that can be predicted using MBP-BIVF, compared with Age-BIVF.

RESULTS
Training and Test Sets

Of 33,741 IVF treatment cycles performed during the 10-year period, 25,595 cycles used fresh IVF and patients' own eggs and involved 11,720 unique patients. For reference, 5,940 of those 25,595 fresh, nondonor eggs, IVF treatments resulted in live births, of which 1,682 were multiple births. Thus, the overall live and multiple birth rates were 22.8% and 28.3%, respectively.

Of those 25,595 treatment cycles, a total of 16,226 cycles were included according to the first set of exclusion criteria.

of oocytes that fertilized normally (e.g., precisely two pronuclei were observed), number of eight-cell embryos, number of embryos, number of transferred embryos with four or fewer cells each, gravidity, number of days of gonadotropin stimulation, number of transferred embryos that were HIP, number of cryopreserved embryos, history of ectopic pregnancy, history of pregnancy loss ≤20 weeks, parity, month, season, number of transferred embryos that had eight cells, donor sperm, number of embryos with eight or more cells each, use of intracytoplasmic sperm injection, cycle number (e.g., whether it is the patient's first, second, or third IVF treatment at this clinic), year, use of assisted hatching, male factor as a cause of infertility, method of sperm collection, number of transferred embryos that had eight or more cells each, and day of ET. (HIP refers to the center's internal designation of embryos that are grade 3, have more than six cells each, and have normal rates of division.) Interestingly, two variables—"number of transferred embryos with more than eight cells each" and day of ET—each had a relative importance of zero after all other variables have been accounted for.

Predictive Power

The log-likelihoods were computed to be −381.86 for Baseline-BIVF, −376.60 for Age-BIVF, and −368.69 for

Source of clinical data used in training and test sets. We show the source of IVF cycles that were available for training and validating the MBP-BIVF model after two rounds of exclusion.

Lannon. Personalized risk of multiple birth. Fertil Steril 2012.
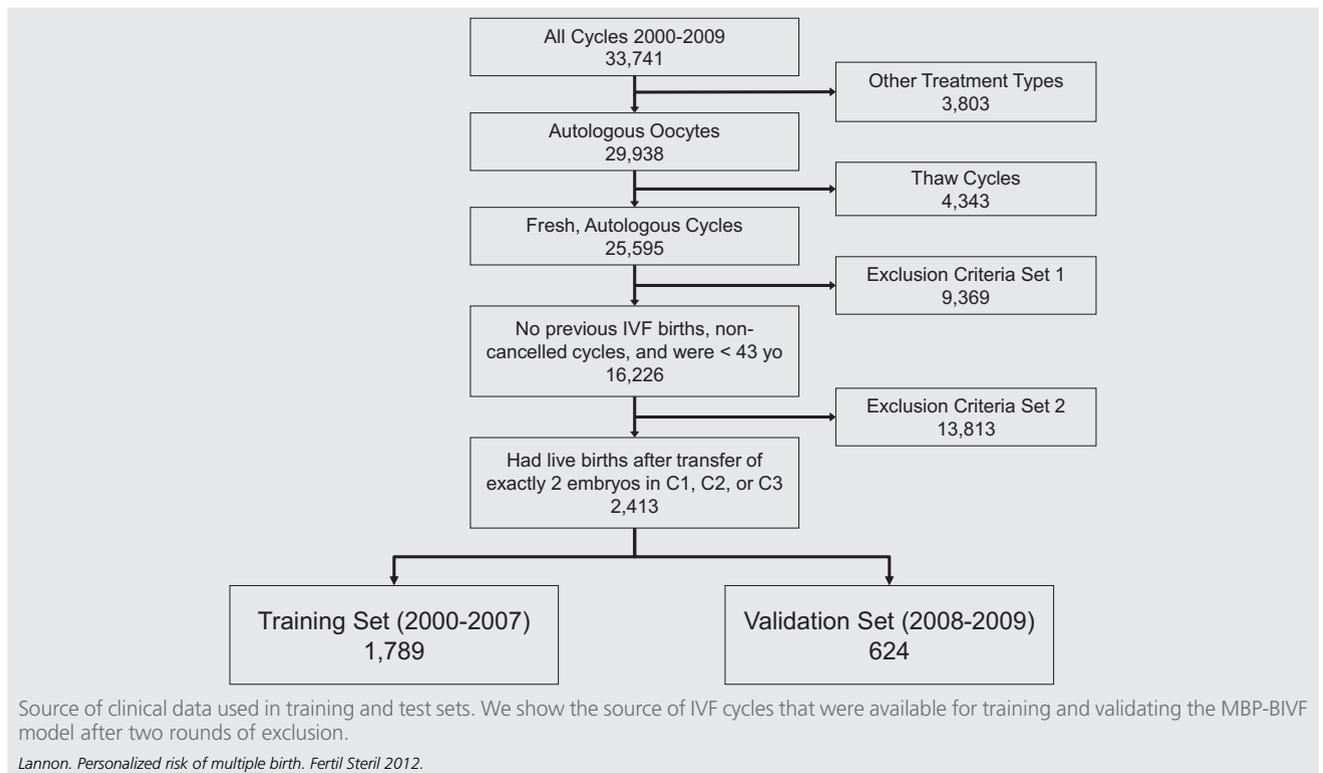
MBP-BIVF. MBP-BIVF showed a 146% improvement over Age-BIVF in its ability to predict the probability of multiple births. The prediction error for the population is 1.8% for both MBP-BIVF and Age-BIVF.

Discrimination

Receiver operating characteristic analysis revealed the area under the curve for MBP-BIVF and Age-BIVF to be 0.632 and 0.544, respectively. Thus, the ability of MBP-BIVF to

TABLE 1

Variables that are significantly different in the training (2000–2007) and test (2008–2009) data sets.

Variable	2000–2007 training set (n = 1,789)		2008–2009 external validation set (n = 624)		
	Mean[a]	SD	Mean	SD	*P* value
Year	2003.80	2.14	2008.52	0.50	0
Serum peak E_2 (IU/L)	1,800.60	1,042.22	2,353.78	1,287.26	.00
No. of sperm motile after wash ($\times 10^6$/mL)	12.56	11.45	16.34	12.25	.00
Rate of embryo cryopreservation	0.21	0.24	0.15	0.19	.00
No. embryos ≥ 8 cells	1.31	0.95	1.05	0.95	.00
No. cryopreserved embryos	2.04	2.71	1.48	2.19	.00
No. transferred embryos ≥ 8 cells	1.23	0.83	1.02	0.90	.00
Cycle length (d)	12.45	2.28	12.97	2.29	.00
Total amount of gonadotropin (IU/mL)	2,195.75	1,126.65	2,438.71	1,401.84	.00
No. of sperm motile before wash ($\times 10^6$/mL)	33.32	29.11	28.06	29.05	.00
No. 8-cell embryos	1.09	0.89	0.95	0.91	.00
Donor sperm	0.04	0.19	0.07	0.26	.00
Mean no. cells per embryo	7.02	1.34	6.79	1.57	.01
No. transferred embryos ≥ 8 cells	1.03	0.82	0.92	0.87	.01
No. oocytes	11.85	6.43	12.71	6.97	.02
Age	33.00	3.54	32.55	3.70	.02
Gravida	0.79	1.14	0.67	1.07	.04
Para	0.27	0.58	0.21	0.51	.04
Male factor	0.28	0.45	0.33	0.47	.04
Transferred embryos at the 8-cell stage (%)	57.41	39.07	62.10	39.26	.05

[a] For continuous variables, the mean indicates the mean value of each variable. For categorical variables, the mean indicates the average number of positive occurrences.

Lannon. Personalized risk of multiple birth. Fertil Steril 2012.

FIGURE 2
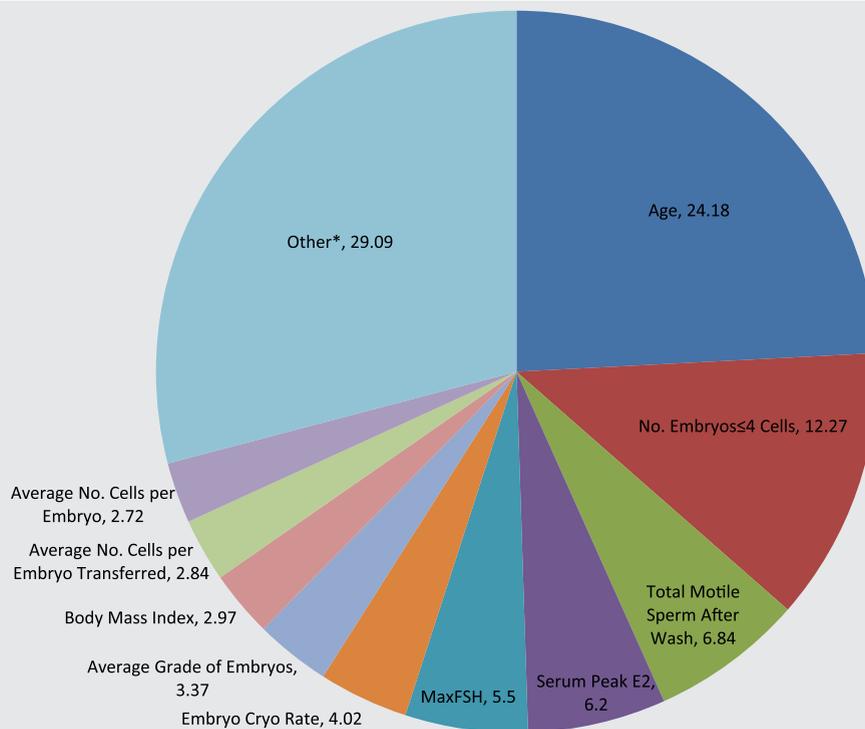
Prognostic factors and their relative influence on multiple birth risk in the MBP-BIVF model. Of 43 variables analyzed, the top 10 nonredundant variables and their "relative influence" are shown.

Lannon. Personalized risk of multiple birth. Fertil Steril 2012.

discriminate patients with differential probabilities of multiple births improved by 16.0%, compared with Age-BIVF (data not shown). Practically, the ability to discriminate can be expressed in terms of percentile ranking of patients based on their risk of multiple births. For example, multiple birth risks of >40% and >35% corresponded to the top 6.6 and 19.2 percentiles in the population that was analyzed.

Dynamic Range and Reclassification

Compared with Age-BIVF, which can only provide one of four probabilities of live birth (<35 years, 33.3%; 35–37 years, 27.6%; 38–40 years, 11.6%; 41–42 years, 11.1%), MBP-BIVF showed a significantly improved dynamic range of multiple live birth probabilities that can be predicted, ranging from 11.8% to 54.8% (Fig. 3). Overall, approximately 55% of patients had predicted probabilities that were significantly different from those predicted by the Age-BIVF control model ($P < .05$).

DISCUSSION

The major findings of this study are that patients, even when stratified by age, have inherently different risks of multiple birth, and these risks can be predicted. Knowing which patients are at risk of twin pregnancy before placing embryos should improve patient counseling substantially regarding the number of embryos to transfer. Previous reports of the risk of multiple births typically focused on the risk merely as a function of patient age and the number of transferred embryos (9, 10, 13). That understanding contributed to the drastic decrease in the rate of high-order multiple births but was not sufficient to establish criteria for eSET to reduce the rate of twin births. Here we show that even when only two embryos were transferred, patients' risks of twins ranged from 12% to 55%. The precise risk and error estimate for a particular patient can be computed by using readily available clinical data pertaining to the patients, their male partners, and their embryos. By restricting our analysis to cycles in which only two embryos were transferred, we have removed the potential paradoxical increase in multiple birth risk in patients who may be receiving more than two embryos on the basis of their poor prognosis.

We were able to generate these findings because of the unique strengths of our research design and methods and our large data set. Most prediction models previously proposed in reproductive medicine have primarily suffered from failed external validation (24). Logistic regression, as conventionally applied to generate previously published prediction models, may not readily extract uniquely predictive components from variables pertaining to the female patient, her male partner, and their embryos (24). Further, the conventional method of selecting a few prognostic factors a priori may also limit the power of the prediction model. Among many advantages of the boosted tree, it allowed us to eliminate the need to limit analysis to a few variables a priori, and it maximizes information that is contributed

FIGURE 3
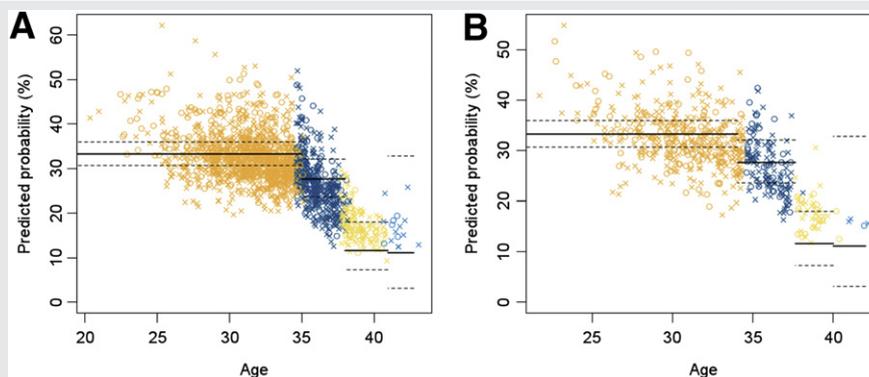
Patient-specific predictions and reclassification. This figure is a visual summary of the results pertaining to dynamic range (i.e., multiple birth probabilities that were predicted by MBP-BIVF and the control, Age-BIVF); reclassification, an important criteria in determining utility (i.e., the portion of cases with MBP-BIVF–predicted multiple birth probabilities that fall outside of the 95% confidence interval (CI) of Age-BIVF–predicted multiple birth probabilities; and prediction results from which a clinic-specific percentile ranking corresponding to each stratification of multiple birth risks can be developed (see Discussion). The multiple birth probability predicted by MBP-BIVF for each of 1,789 patients in the training set (A) and 624 patients from test set (B) was plotted against age, which is conventionally the primary predictor and also the predictor used by the control model. Further, if the observed outcome was multiple birth, the cycle was represented by "O," and if the observed outcome was singleton birth (e.g., not multiple birth), then the cycle was represented by "X." The bold lines indicate probabilities of multiple birth predicted by the Age-BIVF control model, with their CIs indicated by dotted lines above and below the bold line. For example, the Age-BIVF model predicted multiple birth probabilities of 33.3% (95% CI 31.0%-35.6%) for patients with age <35 years; 27.6% (95% CI 23.7%-31.5%) for age 35–37 years; 11.6% (95% CI 7.2%-16%) for age 38–40 years; and 11.1% (95% CI 0-32.5%) for age 41–42 years. Note that MBP-BIVF provides a wide range of predicted probabilities specific to the patients' clinical parameters, which contrasts with the four discrete probabilities provided by Age-BIVF. The estimated standard errors of the mean multiple birth probabilities for patients grouped by quintiles ranged from 4.0% to 8.4%, whereas the standard errors for the age groups in Age-BIVF ranged from 2.3% to 21.4%.

Lannon. Personalized risk of multiple birth. Fertil Steril 2012.

by each variable, thus it allows us to extract good prediction from our dataset (28–31).

Additional strengths of our research design include a highly stringent set of exclusion criteria to ensure that each patient is uniquely represented, as well as the allocation of data into mutually exclusive training and test sets. Our use of a phenotype-rich data source, an extensive and rigorously maintained database comprising more than 33,000 IVF treatment cycles over a span of 10 years, made it possible to apply stringent criteria and implement training and testing sets that comprise unrelated treatment cycles and patients.

We did not know whether treatment protocols, embryo culture techniques, and patient populations vary sufficiently among clinics to enable validation of a prediction model in multiple clinics from a single dataset. Hence, we took an unconventional approach of generating a clinic-specific prediction model using data from 2000–2007, and tested whether it could predict outcomes from that clinic's 2008–2009 data set.

This research study accomplished three goals. First, it provided a prediction model that is validated for patients attending this clinic. Although there was a multitude of variables that were not analyzed (e.g., socioeconomic factors, zip code), and it is certainly possible that their inclusion might have improved the model's performance, their absence did not compromise our objective to develop an approach to building multiple birth prediction models that perform well and are practical and easily applied to other clinics. Second, the use of the most recently available data in the test set

makes this model relevant for current patients. Third, we demonstrate that our model is clinically meaningful at two levels—the patient as an individual and the clinic. It is clinically meaningful at the patient-level because the prediction model allows us, for example, to input data from two patients of the same age on the day of ET and know that by transferring two embryos, one of them has a 55% chance of a twin pregnancy and the other has only a 12% chance. Each patient's predicted probability of multiple birth is further considered in terms of multiple birth risk as a percentile of the clinic's population. For example, the patient with 55% chance of a twin pregnancy falls within the top 5 percentile, whereas the patient with 12% chance belongs to the lowest 4 percentile. This knowledge before embryo placement could have an enormous impact on the utilization of eSET.

At the clinic level, the ability to objectively correlate a patient's multiple birth probability with her risk as a percentile among the clinic's entire population allows physicians to consider eSET utilization criteria without compromising each physician's personal style of care giving.

There are many ways to use the patient- and clinic-level information. For example, one way is to establish a risk percentile to serve as a guide or threshold above which eSET would be very strongly urged. Then after a period of time, outcomes would be analyzed and the clinic could reassess whether the criteria should be revised. Therefore, the first round of prediction modeling results as presented here should not be considered the end goal. Rather, it establishes the beginning of an advanced, evidence-based process of clinical protocol implementation that is informed by personalized

prognosis. This approach establishes the critical connection between personalized medicine and an institution's quality assurance objectives, thus rendering debates about the merits of personalized vs. protocol-based care obsolete.

Until further research reveals whether outcomes prediction models in reproductive medicine can be generalized while remaining sufficiently discriminative to be clinically meaningful, the applicability of this model and the relative importance of its prognostic factors are limited to our center and patient population. The relative importance of individual prognostic factors in our model should be taken only as a set of mathematical relationships that pertain to this clinic and not over-interpreted as having causal roles in success. New hypotheses nonetheless could be generated and further tested in appropriate clinical trials or animal models in the laboratory.

In contrast to the clinic-specific applicability of this prediction model, the overall approach to developing and validating clinic-specific prediction models has now been performed in two unique clinical settings. Specifically, the center used in this analysis is in a state with mandated infertility insurance coverage, whereas Banerjee et al. (25) reported a prediction model that is based on clinical data from a state that does not have mandated coverage. The Banerjee model predicted the probability of live birth for women who had previously failed at least one IVF treatment, and an earlier prediction model reported by Jun et al. (32) predicted the probability of pregnancy for a current IVF treatment. In contrast, we report a prediction model that computes the probability of multiple births for women who had a live birth after IVF. The validation of the clinic-specific approach and use of boosted tree in these widely different clinical scenarios and healthcare environments suggests that these methods may be applied to other clinics to establish an evidence-based, clinic-specific, and personalized approach toward reducing the rate of multiple births after IVF.

Selecting the appropriate number of embryos to transfer in an IVF cycle is an important decision requiring clear understanding of the risks of multiple pregnancy. We show that age alone is limited in guiding counseling and decision making. Our prediction model computes the risk of multiple birth with higher predictive power than age alone and clearly identifies patients with different risks before placing their embryos. By using our own data and extracting a maximal amount of predictive information from many clinical variables in a validated model, we are able to provide patient-specific estimates of risks of multiple births. Our approach represents a powerful way to comply with national guidelines calling for clinics to derive clinic-specific protocols to improve counseling of patients on their risks of multiple births and reduce the rate of multiple births.

REFERENCES

1. van Heesch MM, Bonsel GJ, Dumoulin JC, Evers JL, van der Hoeven MA, Severens JL, et al. Long term costs and effects of reducing the number of twin pregnancies in IVF by single embryo transfer: the TwinSing study. BMC Pediatr 2010;10:75.

2. Lukassen HG, Schonbeck Y, Adang EM, Braat DD, Zielhuis GA, Kremer JA. Cost analysis of singleton versus twin pregnancies after in vitro fertilization. Fertil Steril 2004;81:1240–6.

3. Bernasko J, Lynch L, Lapinski R, Berkowitz RL. Twin pregnancies conceived by assisted reproductive techniques: maternal and neonatal outcomes. Obstet Gynecol 1997;89:368–72.

4. ESHRE Capri Workshop Group. Multiple gestation pregnancy. Hum Reprod 2000;15:1856–64.

5. Luke B, Brown MB, Grainger DA, Cedars M, Klein N, Stern JE, et al. Practice patterns and outcomes with the use of single embryo transfer in the United States. Fertil Steril 2010;93:490–8.

6. Centers for Disease Control and Prevention, American Society for Reproductive Medicine, Society for Assisted Reproductive Technology. 2008 assisted reproductive technology success rates: national summary and fertility clinic reports. Atlanta, GA: US Department of Health and Human Services; 2010.

7. Practice Committee of the American Society for Reproductive Medicine, Practice Committee of the Society for Assisted Reproductive Technology. Guidelines on number of embryos transferred. Fertil Steril 2009;92:1518–9.

8. Stern JE, Goldman MB, Hatasaka H, MacKenzie TA, Racowsky C, Surrey ES, et al. Optimizing the number of blastocyst stage embryos to transfer on day 5 or 6 in women 38 years of age and older: a Society for Assisted Reproductive Technology database study. Fertil Steril 2009;91:157–66.

9. Stern JE, Goldman MB, Hatasaka H, MacKenzie TA, Surrey ES, Racowsky C, et al. Optimizing the number of cleavage stage embryos to transfer on day 3 in women 38 years of age and older: a Society for Assisted Reproductive Technology database study. Fertil Steril 2009;91:767–76.

10. Stern JE, Cedars MI, Jain T, Klein NA, Beaird CM, Grainger DA, et al. Assisted reproductive technology practice patterns and the impact of embryo transfer guidelines in the United States. Fertil Steril 2007;88:275–82.

11. Kresowik JD, Stegmann BJ, Sparks AE, Ryan GL, van Voorhis BJ. Five-years of a mandatory single-embryo transfer (mSET) policy dramatically reduces twinning rate without lowering pregnancy rates. Fertil Steril 2011;96:1367–9.

12. Jungheim ES, Ryan GL, Levens ED, Cunningham AF, Macones GA, Carson KR, et al. Embryo transfer practices in the United States: a survey of clinics registered with the Society for Assisted Reproductive Technology. Fertil Steril 2010;94:1432–6.

13. Gleicher N, Barad D. Twin pregnancy, contrary to consensus, is a desirable outcome in infertility. Fertil Steril 2009;91:2426–31.

14. Gleicher N, Oleske DM, Tur-Kaspa I, Vidali A, Karande V. Reducing the risk of high-order multiple pregnancy after ovarian stimulation with gonadotropins. N Engl J Med 2000;343:2–7.

15. Chervenak FA, McCullough LB, Rosenwaks Z. Ethical dimensions of the number of embryos to be transferred in in vitro fertilization. J Assist Reprod Genet 2001;18:583–7.

16. Voelker R. Researchers in Canada call for policy to mandate single-embryo transfer in IVF. JAMA 2011;305:1848.

17. Twisk M, van der Veen F, Repping S, Heineman MJ, Korevaar JC, Bossuyt PM. Preferences of subfertile women regarding elective single embryo transfer: additional in vitro fertilization cycles are acceptable, lower pregnancy rates are not. Fertil Steril 2007;88:1006–9.

18. van Loendersloot LL, van Wely M, Limpens J, Bossuyt PM, Repping S, van der Veen F. Predictive factors in in vitro fertilization (IVF): a systematic review and meta-analysis. Hum Reprod Update 2010;16:577–89.

19. Hunault CC, Eijkemans MJ, Pieters MH, te Velde ER, Habbema JD, Fauser BC, et al. A prediction model for selecting patients undergoing in vitro fertilization for elective single embryo transfer. Fertil Steril 2002;77:725–32.

20. Baker VL, Luke B, Brown MB, Alvero R, Frattarelli JL, Usadi R, et al. Multivariate analysis of factors affecting probability of pregnancy and live birth with in vitro fertilization: an analysis of the Society for Assisted Reproductive Technology Clinic Outcomes Reporting System. Fertil Steril 2010;94:1410–6.

21. Nelson SM, Lawlor DA. Predicting live birth, preterm delivery, and low birth weight in infants born from in vitro fertilisation: a prospective study of 144,018 treatment cycles. PLoS Med 2011;8:e1000386.

22. Roberts SA, McGowan L, Mark Hirst W, Vail A, Rutherford A, Lieberman BA, et al. Reducing the incidence of twins from IVF treatments: predictive modelling from a retrospective cohort. Hum Reprod 2011;26:569–75.

23. Tur R, Barri PN, Coroleu B, Buxaderas R, Martinez F, Balasch J. Risk factors for high-order multiple implantation after ovarian stimulation with gonadotrophins: evidence from a large series of 1878 consecutive pregnancies in a single centre. Hum Reprod 2001;16:2124–9.

24. Leushuis E, van der Steeg JW, Steures P, Bossuyt PM, Eijkemans MJ, van der Veen F, et al. Prediction models in reproductive medicine: a critical appraisal. Hum Reprod Update 2009;15:537–52.

25. Banerjee P, Choi B, Shahine LK, Jun SH, O'Leary K, Lathi RB, et al. Deep phenotyping to predict live birth outcomes in in vitro fertilization. Proc Natl Acad Sci U S A 2010;107:13570–5.

26. Eaton JL, Hacker MR, Harris D, Thornton KL, Penzias AS. Assessment of day-3 morphology and euploidy for individual chromosomes in embryos that develop to the blastocyst stage. Fertil Steril 2009;91:2432–6.

27. Malizia BA, Hacker MR, Penzias AS. Cumulative live-birth rates after in vitro fertilization. N Engl J Med 2009;360:236–43.

28. Ridgeway G. Generalized boosted regression models. R package 1.6.3. 2007.

29. Breiman L, Friedman J, Olshen R, Stone C. Classification and regression trees. Belmont, CA: Wadsworth; 1984.

30. Friedman J. Stochastic gradient boosting. Technical report. Stanford, CA: Stanford University, Department of Statistics; 1999.

31. Friedman J. Greedy function approximation: a stochastic boosting machine. Stanford, CA: Stanford University, Department of Statistics; 1999.

32. Jun SH, Choi B, Shahine L, Westphal LM, Behr B, Reijo Pera RA, et al. Defining human embryo phenotypes by cohort-specific prognostic factors. PLoS One 2008;3:e2562.