

**Supplementary Tables and Figures for
“Machine learning center-specific models show improved IVF live birth predictions over US
national registry-based model”**

Authors:

Mylene W. M. Yao MD^{1*‡}, Elizabeth T. Nguyen PhD^{1*}, Matthew G. Retzliff MD², L. April Gago MD³, John E. Nichols MD⁴, John F. Payne MD⁴, Barry A. Ripps MD⁵, Michael Opsahl MD⁶, Jeremy Groll MD⁷, Ronald Beesley MD⁶, Gregory Neal MD², Jaye Adams MD², Lorie Nowak PhD⁷, Trevor Swanson PhD¹, Xiaocong Chen MSc¹.

*These authors contributed equally.

¹R&D Department, Univfy, Los Altos, CA, US.

²Fertility Center of San Antonio, San Antonio, TX, US.

³Gago Center for Fertility, Brighton, MI, US.

⁴Piedmont Reproductive Endocrinology Group, Greenville, SC, US.

⁵NewLIFE Fertility, Pensacola, FL, US.

⁶Poma Fertility, Kirkland, WA, US.

⁷SpringCreek Fertility, Dayton, OH, US.

‡Corresponding Author:

Author Name: Mylene W. M. Yao, MD

Affiliation: Univfy Inc.

Mailing Address: 171 Main Street, #139, Los Altos, CA 94022

Phone: 1-650-799-8003

Email: mylene.yao@univfy.com

Supplementary Table 1. This table shows the weighted ROC-AUC and unweighted ROC-AUC with 95% confidence interval (95% CI) from the cross validation metrics for MLCS1 and MLCS2 models for each center. Two-sided paired DeLong's test was performed for each MLCS model ROC-AUC compared to its corresponding Age model ROC-AUC with the resulting paired DeLong's test statistical significance shown. Statistical significance is indicated by * next to the p-value if $p < 0.05$. The Posterior Log of Odds Ratio compared to Age model (PLORA) was positive for each MLCS model, indicating improved predictive power in MLCS over Age model.

Center	Model	Weighted ROC-AUC		Unweighted ROC-AUC (95% CI)		Test-Statistic	Two-sided, Paired DeLong's Test Statistical Significance	PLORA
		MLCS model	Age model	MLCS model	Age model			
916	MLCS1	0.68	0.52	0.69 (0.64, 0.73)	0.53 (0.48, 0.58)	Z = 8.33	$p < 0.0001^*$	12.2
	MLCS2	0.70	0.52	0.71 (0.69, 0.73)	0.55 (0.52, 0.57)	Z = 14.32	$p < 0.0001^*$	45.8
522	MLCS1	0.67	0.51	0.66 (0.58, 0.74)	0.51 (0.42, 0.60)	Z = 4.84	$p < 0.0001^*$	0.8
	MLCS2	0.67	0.44	0.65 (0.60, 0.71)	0.45 (0.39, 0.51)	Z = 7.99	$p < 0.0001^*$	9.1
869	MLCS1	0.59	0.36	0.60 (0.54, 0.65)	0.64 (0.58, 0.69)	Z = -0.81	$p = 0.4172$	0.1
	MLCS2	0.57	0.40	0.57 (0.53, 0.61)	0.59 (0.55, 0.63)	Z = -0.44	$p = 0.6628$	7.9
395	MLCS1	0.66	0.48	0.65 (0.61, 0.69)	0.47 (0.43, 0.51)	Z = 9.34	$p < 0.0001^*$	10.6
	MLCS2	0.66	0.48	0.65 (0.64, 0.66)	0.59 (0.57, 0.60)	Z = 9.92	$p < 0.0001^*$	34.2
635	MLCS1	0.69	0.47	0.69 (0.65, 0.72)	0.49 (0.45, 0.53)	Z = 11.08	$p < 0.0001^*$	16.0
	MLCS2	0.68	0.47	0.78 (0.66, 0.70)	0.47 (0.44, 0.50)	Z = 17.16	$p < 0.0001^*$	41.1
189	MLCS1	0.60	0.36	0.62 (0.56, 0.68)	0.41 (0.56, 0.68)	Z = 6.68	$p < 0.0001^*$	3.7
	MLCS2	0.67	0.43	0.67 (0.63, 0.71)	0.55 (0.50, 0.60)	Z = 2.94	$p = 0.003231^*$	13.6

Supplementary Table 2. This table shows the unweighted ROC-AUC and 95% confidence interval (95% CI) from the MLCS1 model's external live model validation (LMV) and MLCS1 model's internal cross validation for each center. A two-sided unpaired DeLong's test was performed to compare MLCS1 model LMV ROC-AUC versus MLCS1 model internal cross validation ROC-AUC and produced p-values based on D-statistics. Statistical significance is indicated by * next to the p-value if $p < 0.05$.

Center	MLCS1 model from external, live model validation (LMV)	MLCS1 model from internal validation	Test-Statistic	Statistical Significance
916	0.66 (0.62, 0.70)	0.69 (0.64, 0.73)	D(1320.5) = 0.83	p = 0.4093
522	0.63 (0.55, 0.72)	0.66 (0.58, 0.74)	D(356) = 0.41	p = 0.6803
869	0.50 (0.42, 0.58)	0.60 (0.54, 0.65)	D(407.62) = 1.97	p = 0.04994*
395	0.64 (0.59, 0.69)	0.65 (0.61, 0.69)	D(1068) = 0.23	p = 0.6866
635	0.66 (0.62, 0.69)	0.69 (0.65, 0.72)	D(1741.1) = 1.22	p = 0.2232
189	0.66 (0.58, 0.74)	0.62 (0.56, 0.68)	D(479.72) = -0.82	p = 0.4140

Supplementary Table 3A-F. These 6 tables show the key clinical variables expressed in mean (standard deviation) or incidence and clinical diagnostic frequencies (shown in percentage, %) observed in the internal cross-validation (CV) of MLCS1, the MLCS1 data set limited to first cycles (C1), and the out-of-time test set used for the external validation of MLCS1 for each of the 6 pseudo-named centers 916, 552, 869, 395, 635 and 189, respectively. First cycles in MLCS1 internal and external data sets were compared to test whether the distribution of these key clinical variables were different between these two data sets. The distribution of numeric clinical variables were compared using the two-sided, two-sample Kolmogorov-Smirnov test (with D-statistic shown). Clinical diagnosis frequencies and sperm variable incidences were compared using the Chi-square test (with χ^2 statistic shown) for homogeneity. Statistical significance is indicated by * next to the p-value if $p < 0.05$; "n/a" indicated that statistical comparison was not performed/appropriate.

Table 3A. Center 916					
Clinical variables or diagnoses	MLCS1 CV training and testing data set	MLCS1 CV training and testing data set (limited to C1)	MLCS1 external test set (C1 only)	Test-Statistic	Comparison of MLCS1 internal C1 vs. MLCS1 external data sets
Female Age	34.2 (4.1)	33.8 (4.1)	33.9 (4.1)	D(1217) = 0.04	p = 0.7344
AMH (ng/mL)	3.5 (3.7)	3.7 (3.8)	3.9 (4.5)	D(1207) = 0.03	p = 0.8821
BMI	26.9 (6.0)	27.1 (6.0)	28.3 (6.4)	D(822) = 0.09	p = 0.1699
Day 3 FSH (mIU/mL)	7.6 (4.4)	7.3 (3.4)	7.0 (3.2)	D(1160) = 0.06	p = 0.2029
No. prior term births	0.9 (1.0)	1.0 (1.1)	1.0 (1.1)	D(619) = 0.03	p = 0.994
No. prior pregnancies	1.2 (1.6)	1.2 (1.6)	1.1 (1.6)	D(1217) = 0.04	p = 0.7652
No. prior clinical pregnancy losses	0.9 (1.2)	1.0 (1.2)	0.7 (1.0)	D(619) = 0.10	p = 0.1234
Use of sperm ejaculate	98%	98%	96%	$\chi^2(1, N=1217) = 2.06$	p = 0.1508
Use of male partner's sperm	96%	96%	95%	$\chi^2(1, N=1217) = 0.46$	p = 0.4983
Clinical diagnoses:					
Diminished ovarian reserve	23%	19%	22%	$\chi^2(1, N=1217) = 0.94$	p = 0.3321
Male factor	35%	34%	32%	$\chi^2(1, N=1217) = 0.47$	p = 0.4948
Endometriosis	10%	11%	8%	$\chi^2(1, N=1217) = 3.88$	p = 0.04899*
Ovulation disorders	35%	32%	41%	$\chi^2(1, N=1217) = 8.39$	p = 0.003768*
Tubal factor	28%	27%	23%	$\chi^2(1, N=1217) = 2.94$	p = 0.08645
Uterine factor	6%	7%	7%	$\chi^2(1, N=1217) = 0.13$	p = 0.7205
Unexplained infertility	12%	13%	15%	$\chi^2(1, N=1217) = 1.55$	p = 0.2124
Recurrent pregnancy loss	0.3%	n/a	2%	n/a	n/a

Abbreviations: AMH = anti-mullerian hormone; BMI = body mass index; FSH = follicle stimulating hormone; AFC = antral follicle count.
 Explanations: 1) Sperm ejaculate was not used in cases of testicular biopsy, testicular sperm extraction (TESE), testicular sperm aspiration (TESA), microscopic epididymal sperm aspiration (MESA). 2) Male partner's sperm is not used in cases of donor sperm usage. 3) Clinical diagnoses were taken as-recorded and did not assume methods of clinical diagnoses were the same across centers. 4) Ovulation disorders included any finding of anovulation or irregular timing of ovulation regardless of underlying pathology including but not limited to polycystic ovarian syndrome (PCOS). If available, specific ovulatory disorder diagnoses would be tested for applicability by the model at the center level.
 5) Here, recurrent pregnancy loss is based on documentation at the center level.

Table 3B. Center 552

Clinical variables or diagnoses	MLCS1 CV training and testing data set	MLCS1 CV training and testing data set (limited to C1)	MLCS1 external test set (C1 only)	Test-Statistic	Comparison of MLCS1 internal C1 vs. MLCS1 external data sets
Female Age	33.8 (4.3)	33.3 (4.2)	33.2 (4.0)	D(327) = 0.05	p = 0.9863
AMH (ng/mL)	3.5 (4.0)	3.7 (4.2)	3.8 (3.7)	D(306) = 0.09	p = 0.5150
BMI	25.8 (5.3)	26.1 (5.3)	26.7 (5.5)	D(321) = 0.11	p = 0.3208
Day 3 FSH (mIU/mL)	8.4 (3.1)	8.2 (2.9)	8.2 (3.6)	D(306) = 0.16	p = 0.04915*
No. prior term births	0.6 (0.8)	0.7 (0.8)	0.7 (1.1)	D(128) = 0.09	p = 0.9688
No. prior pregnancies	0.8 (1.3)	0.8 (1.3)	0.9 (1.5)	D(327) = 0.06	p = 0.9537
No. prior clinical pregnancy losses	0.9 (1.5)	0.8 (1.4)	1.2 (1.4)	D(128) = 0.12	p = 0.7305
Use of sperm ejaculate	94%	92%	93%	X ² (1, N=327) = 0.00	p = 0.9984
Use of male partner's sperm	93%	93%	94%	X ² (1, N=327) = 0.00	p = 1.000
Clinical diagnoses:					
Diminished ovarian reserve	37%	31%	17%	X ² (1, N=327) = 8.84	p = 0.002948*
Male factor	40%	43%	41%	X ² (1, N=327) = 0.02	p = 0.8942
Endometriosis	19%	20%	13%	X ² (1, N=327) = 2.55	p = 0.1103
Ovulation disorders	40%	37%	16%	X ² (1, N=327) = 18.32	p < 0.0001*
Tubal factor	14%	15%	6%	X ² (1, N=327) = 6.84	p = 0.008893*
Uterine factor	7%	6%	5%	X ² (1, N=327) = 0.09	p = 0.7587
Unexplained infertility	6%	6%	4%	X ² (1, N=327) = 0.25	p = 0.6163
Recurrent pregnancy loss	2%	1%	11%	X ² (1, N=327) = 10.86	p = 0.0009842*

Table 3C. Center 869

Clinical variables or diagnoses	MLCS1 CV training and testing data set	MLCS1 CV training and testing data set (limited to C1)	MLCS1 external test set (C1 only)	Test-Statistic	Comparison of MLCS1 internal C1 vs. MLCS1 external data sets
Female Age	33.1 (4.0)	32.8 (4.0)	32.5 (4.0)	D(518) = 0.04	p = 0.9971
AMH (ng/mL)	4.5 (4.1)	4.6 (4.2)	2.9 (2.5)	D(515) = 0.23	p < 0.0001*
BMI	26.0 (5.5)	25.9 (5.3)	26.8 (5.5)	D(518) = 0.15	p = 0.005151*
Day 3 FSH (mIU/mL)	6.7 (2.6)	6.8 (2.8)	6.8 (2.5)	D(500) = 0.05	p = 0.8969
No. prior term births	0.6 (0.9)	0.7 (1.0)	0.8 (1.0)	D(213) = 0.11	p = 0.6199
No. prior pregnancies	0.9 (1.3)	0.8 (1.2)	0.7 (1.2)	D(518) = 0.02	p = 1.000
No. prior clinical pregnancy losses	1.0 (1.0)	1.1 (1.0)	0.6 (0.7)	D(213) = 0.20	p = 0.0317*
Use of sperm ejaculate	88%	90%	92%	X ² (1, N=518) = 0.49	p = 0.4818
Use of male partner's sperm	96%	96%	95%	X ² (1, N=518) = 0.24	p = 0.6233
Clinical diagnoses:					
Diminished ovarian reserve	19%	18%	36%	X ² (1, N=518) = 18.91	p < 0.000*
Male factor	41%	41%	45%	X ² (1, N=518) = 0.78	p = 0.3767
Endometriosis	15%	16%	17%	X ² (1, N=518) = 0.00	p = 0.9451
Ovulation disorders	29%	28%	45%	X ² (1, N=518) = 15.37	p < 0.000*
Tubal factor	15%	14%	28%	X ² (1, N=518) = 13.74	p = 0.0002099*
Uterine factor	0.5%	1%	0.5%	n/a	n/a
Unexplained infertility	9%	10%	8%	X ² (1, N=518) = 0.49	p = 0.4818
Recurrent pregnancy loss	2%	2%	2%	n/a	n/a

Table 3D. Center 395					
Clinical variables or diagnoses	MLCS1 CV training and testing data set	MLCS1 CV training and testing data set (limited to C1)	MLCS1 external test set (C1 only)	Test-Statistic	Comparison of MLCS1 internal C1 vs. MLCS1 external data sets
Female Age	34.4 (3.9)	34.2 (3.9)	34.1 (3.7)	D(1182) = 0.05	p = 0.3419
AMH (ng/mL)	2.8 (2.5)	2.9 (2.5)	3.3 (2.7)	D(1150) = 0.08	p = 0.05143
BMI	25.0 (5.5)	24.9 (5.6)	25.9 (5.0)	D(1171) = 0.15	p < 0.0001*
Day 3 FSH (mIU/mL)	8.1 (3.2)	8.0 (3.0)	6.9 (2.7)	D(1087) = 0.21	p < 0.0001*
No. prior term births	0.2 (0.4)	0.2 (0.4)	0.2 (0.4)	D(1171) = 0.02	p = 0.9943
No. prior pregnancies	n/a	n/a	n/a	n/a	n/a
No. prior clinical pregnancy losses	n/a	n/a	n/a	n/a	n/a
Use of sperm ejaculate	95%	95%	95%	$\chi^2(1, N=1182) = 0.01$	p = 0.9330
Use of male partner's sperm	95%	95%	96%	$\chi^2(1, N=1182) = 0.35$	p = 0.5553
Clinical diagnoses:					
Diminished ovarian reserve	19%	18%	10%	$\chi^2(1, N=1182) = 13.92$	p = 0.0001905*
Male factor	31%	30%	27%	$\chi^2(1, N=1182) = 1.05$	p = 0.3054
Endometriosis	7%	7%	5%	$\chi^2(1, N=1182) = 2.27$	p = 0.1315
Ovulation disorders	11%	11%	10%	$\chi^2(1, N=1182) = 0.29$	p = 0.5913
Tubal factor	12%	12%	11%	$\chi^2(1, N=1182) = 0.01$	p = 0.9321
Uterine factor	3%	3%	1%	$\chi^2(1, N=1182) = 6.51$	p = 0.0107*
Unexplained infertility	24%	25%	32%	$\chi^2(1, N=1182) = 7.69$	p = 0.005556*
Recurrent pregnancy loss	n/a	n/a	n/a	n/a	n/a

Table 3E. Center 635

Clinical variables or diagnoses	MLCS1 CV training and testing data set	MLCS1 CV training and testing data set (limited to C1)	MLCS1 external test set (C1 only)	Test-Statistic	Comparison of MLCS1 internal C1 vs. MLCS1 external data sets
Female Age	33.5 (4.1)	33.2 (4.0)	33.0 (4.3)	D(1691) = 0.04	p = 0.5658
AMH (ng/mL)	3.4 (3.4)	3.5 (3.4)	3.9 (3.8)	D(1454) = 0.07	p = 0.07228
BMI	27.2 (6.7)	27.2 (6.6)	28.2 (6.7)	D(1533) = 0.10	p = 0.001019*
Day 3 FSH (mIU/mL)	8.6 (4.1)	8.6 (4.2)	8.4 (4.0)	D(1245) = 0.08	p = 0.05689
No. prior term births	0.9 (1.1)	0.9 (1.1)	0.9 (1.2)	D(653) = 0.04	p = 0.9684
No. prior pregnancies	1.0 (1.4)	1.0 (1.4)	1.0 (1.4)	D(1518) = 0.01	p = 1.000
No. prior clinical pregnancy losses	1.0 (1.1)	1.0 (1.1)	0.8 (1.1)	D(652) = 0.11	p = 0.09396
Use of sperm ejaculate	93%	93%	94%	$\chi^2(1, N=1691) = 0.58$	p = 0.4471
Use of male partner's sperm	94%	95%	91%	$\chi^2(1, N=1691) = 7.06$	p = 0.007887*
Clinical diagnoses:					
Diminished ovarian reserve	7%	7%	17%	$\chi^2(1, N=1691) = 36.40$	p < 0.0001*
Male factor	14%	14%	29%	$\chi^2(1, N=1691) = 52.88$	p < 0.0001*
Endometriosis	2%	2%	7%	$\chi^2(1, N=1691) = 20.45$	p < 0.0001*
Ovulation disorders	15%	15%	39%	$\chi^2(1, N=1691) = 114.49$	p < 0.0001*
Tubal factor	4%	4%	15%	$\chi^2(1, N=1691) = 48.33$	p < 0.0001*
Uterine factor	1%	0.4%	3%	n/a	n/a
Unexplained infertility	4%	4%	20%	$\chi^2(1, N=1691) = 98.07$	p < 0.0001*
Recurrent pregnancy loss	n/a	n/a	4%	n/a	n/a

Table 3F. Center 189					
Clinical variables or diagnoses	MLCS1 CV training and testing data set	MLCS1 CV training and testing data set (limited to C1)	MLCS1 external test set (C1 only)	Test-Statistic	Comparison of MLCS1 internal C1 vs. MLCS1 external data sets
Female Age	33.1 (4.2)	32.5 (4.0)	32.5 (4.2)	D(525) = 0.03	p = 0.9984
AMH (ng/mL)	3.6 (3.8)	3.8 (4.0)	4.1 (3.5)	D(490) = 0.11	p = 0.1219
BMI	28.6 (6.8)	28.6 (6.7)	28.5 (6.0)	D(409) = 0.07	p = 0.7416
Day 3 FSH (mIU/mL)	6.4 (3.3)	6.4 (3.2)	6.1 (2.9)	D(427) = 0.13	p = 0.05131
No. prior term births	n/a	n/a	0.3 (0.7)	n/a	n/a
No. prior pregnancies	n/a	n/a	0.9 (1.4)	n/a	n/a
No. prior clinical pregnancy losses	n/a	n/a	0.5 (0.9)	n/a	n/a
Use of sperm ejaculate	96%	96%	95%	$\chi^2(1, N=525) = 0.10$	p = 0.7504
Use of male partner's sperm	92%	93%	93%	$\chi^2(1, N=525) = 0.00$	p = 1.000
Clinical diagnoses:					
Diminished ovarian reserve	22%	20%	21%	$\chi^2(1, N=525) = 0.02$	p = 0.8952
Male factor	38%	40%	50%	$\chi^2(1, N=525) = 4.31$	p = 0.03788*
Endometriosis	15%	13%	9%	$\chi^2(1, N=525) = 1.56$	p = 0.2116
Ovulation disorders	10%	11%	17%	$\chi^2(1, N=525) = 3.46$	p = 0.06281
Tubal factor	12%	12%	13%	$\chi^2(1, N=525) = 0.06$	p = 0.8000
Uterine factor	n/a	n/a	1%	$\chi^2(1, N=525) = 0.02$	n/a
Unexplained infertility	4%	6%	6%	$\chi^2(1, N=525) = 0.00$	p = 1.000
Recurrent pregnancy loss	2%	2%	4%	$\chi^2(1, N=525) = 0.89$	p = 0.3468

Supplementary Table 4. Model calibration was evaluated by computing the center-specific Brier score for MLCS2 and SART model outputs using the de novo model validation test sets (DNMV1; N=4645). Across clinics, Brier scores are lower for MLCS2 compared to SART models. See Table 4 for statistical comparison.

Center	Brier Score	
	MLCS2 models	SART model
916	0.22	0.23
552	0.21	0.22
869	0.23	0.24
395	0.21	0.23
635	0.22	0.23
189	0.20	0.22

Supplementary Figure 1. MLCS2 model predictors and their relative importance, cross-referenced with SART pretreatment model predictors. (1-3) The model predictors used by each center's MLCS2 model and their relative importance vary across centers, despite drawing from a similar set of clinical variables. The SART pretreatment model predictors were either also used by MLCS2 models or were determined to have no non-redundant predictive impact based on other predictors used by the MLCS2 models. The MLCS2 models used additional predictors not used by the SART model even though they were recorded in SART-CORS. See Yao et al, 2024 for more in-depth discussion of model predictors and relative importance (1).

References.

1. Yao MWM, J Jenkins, Nguyen ET, Swanson T, Menabrito M. Patient-centric IVF prognostics counseling using machine learning for the pragmatist. Semin Repro Med 2024, *accepted*.
2. McLernon DJ, Raja EA, Toner JP, Baker VL, Doody KJ, Seifer DB, Sparks AE, Wantman E, Lin PC, Bhattacharya S, Van Voorhis BJ. Predicting personalized cumulative live birth following in vitro fertilization. Fertil Steril. 2022 Feb;117(2):326-338. doi: 10.1016/j.fertnstert.2021.09.015. Epub 2021 Oct 19. PMID: 34674824.
3. Curchoe CL, Tarafdar O, Aquilina MC, Seifer DB. SART CORS IVF registry: looking to the past to shape future perspectives. J Assist Reprod Genet. 2022 Nov;39(11):2607-2616. doi: 10.1007/s10815-022-02634-6. Epub 2022 Oct 21. PMID: 36269502; PMCID: PMC9722991.

Abbreviations.

MLCS2 = machine learning, center-specific model version 2

SART = Society of Assisted Reproductive Technologies

AMH = serum anti-mullerian hormone level

BMI = body mass index

Day 3 FSH = serum day 3 follicular stimulating hormone level

PCOS = polycystic ovarian syndrome

PCO = polycystic ovaries

IVF = in vitro fertilization

ICSi = intracytoplasmic sperm injection

[illegible]

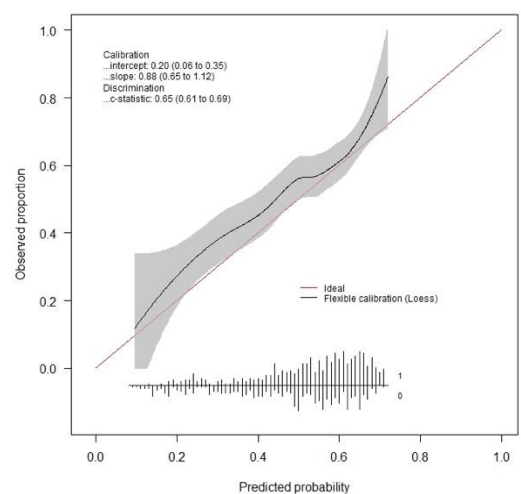
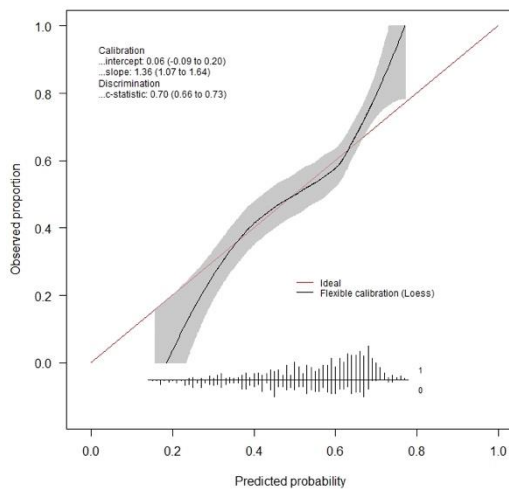
Supplementary Figure 2. Calibration and frequency distribution plots of live birth prediction (LBP) comparing MLCS2 Production Model and SART Model using each center's de novo model validation 1 (DNMV1) test sets. The flexible calibration curve and its 95% confidence limits are represented by the plotted dark line and grey areas, respectively, using loess, a non-parametric smoother.

Note: The calibration plot in the particular context of in vitro fertilization (IVF) live birth predictions for patients under 40 years of age (which is the upper age limit of this study) is limited by the sparsity of data points with lower LBPs. This was the case for both MLCS2 and SART model responses. With few subjects with lower LBPs, calibration performance may not be adequately evaluated. However, the Brier score incorporates both discrimination and calibration to provide an overall measure of performance, and the Brier score analysis between MLCS2 and SART model indicated that MLCS2 had a better predictive performance than the SART model.

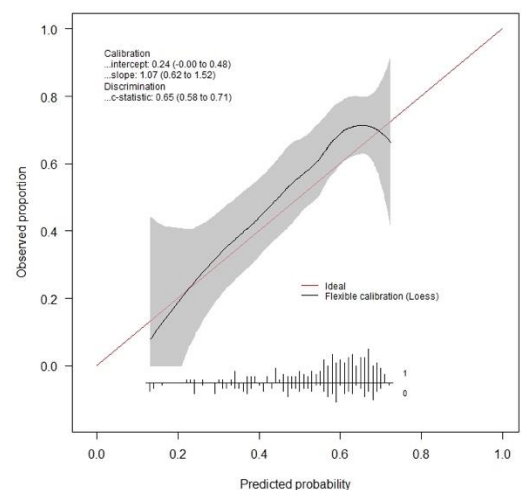
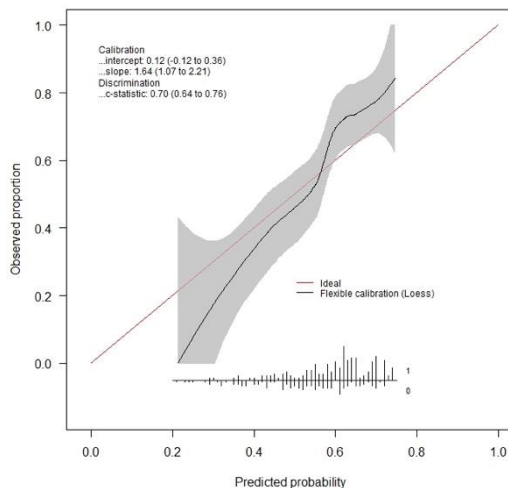
MLCS2 Model

SART Model

Clinic 916



Clinic 552

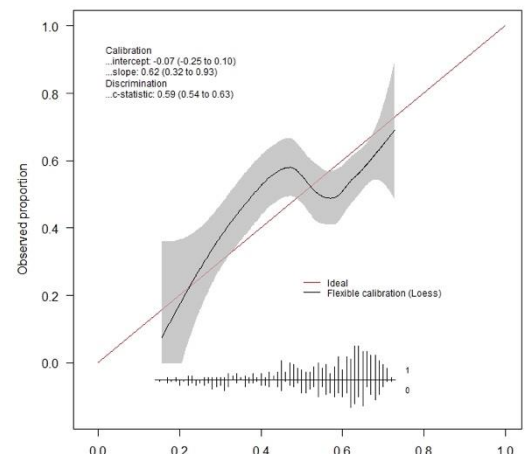
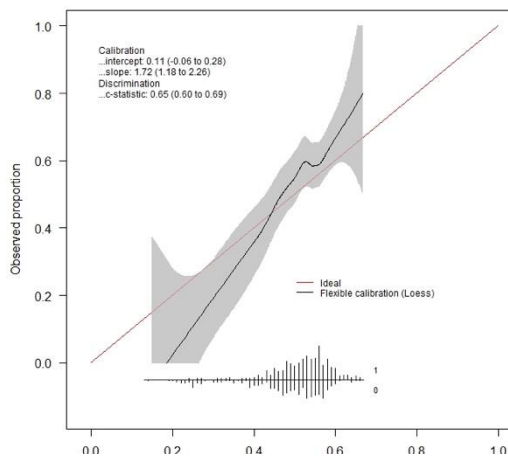


Supplementary Figure 2 (Continued).

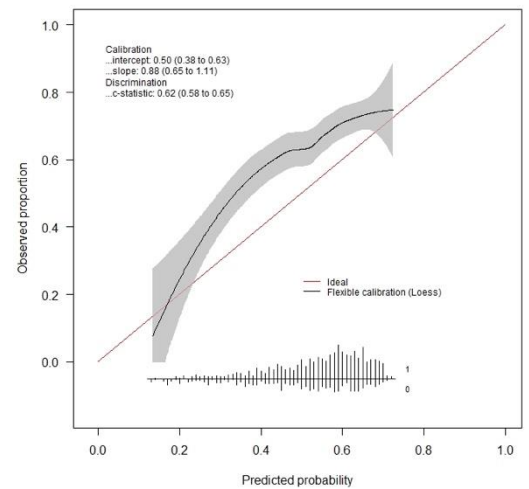
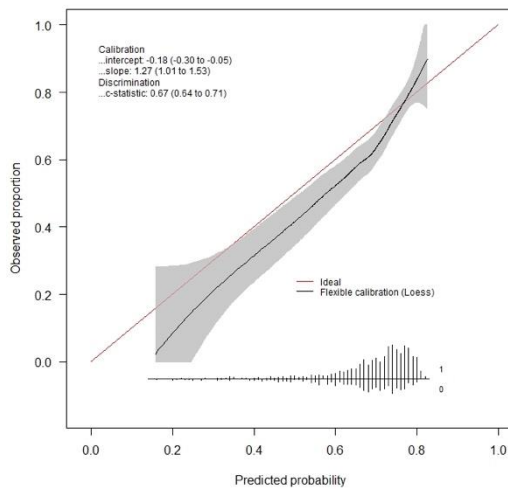
MLCS2 Model

SART Model

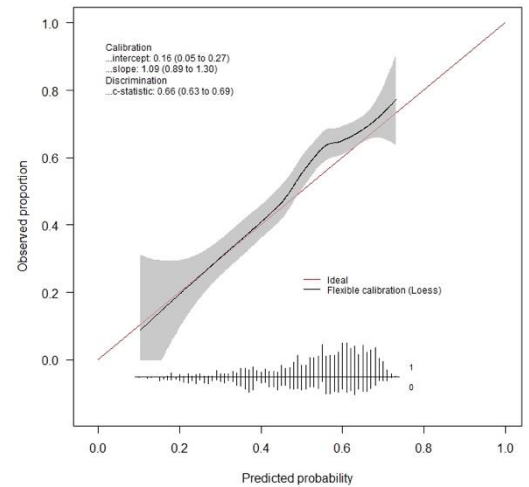
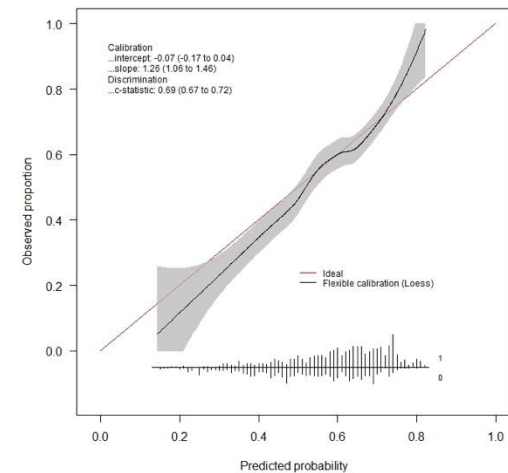
Clinic 869



Clinic 395

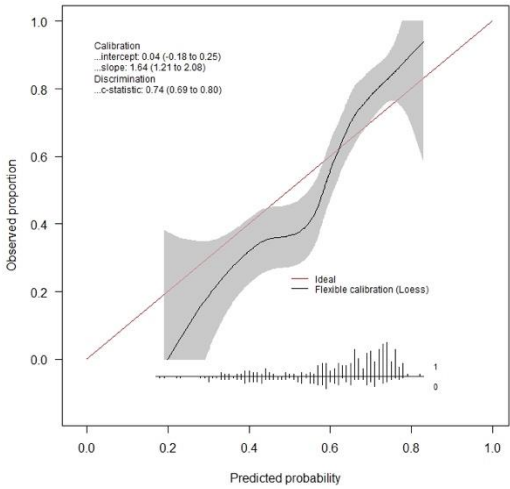


Clinic 635

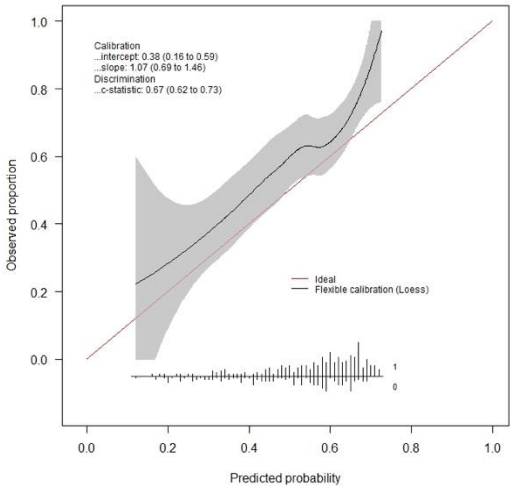


Clinic 189

MLCS2 Model



SART Model



Supplementary Figure 3A-F. Reclassification tables comparing in vitro fertilization (IVF) live birth prediction (LBP) for MLCS2 production model and SART model DNMV1 responses for each of the 6 centers: Centers 916, 552, 869, 395, 635, 189. The reclassification table shows the number of IVF cycles receiving predicted IVF live birth prediction (LBP) in each range by the SART vs. MLCS2 model, as well as the live birth rate (LBR) for each cell.

Abbreviations: MLCS2 = machine learning, center-specific model version 2; DNMV1 = de novo model validation test set 1, adapted to enable testing of MLCS2 and SART models; n = the number of patients being placed into each of the 16 groups. LBR = live birth rate of each group. Green = patient subgroups that have MLCS2-LBP and SART-LBP in the same prognostic category; Blue = patient subgroups with MLCS2 models placing patients in higher prognostic categories compared to SART model; Peach = patient subgroups with MLCS2 models placing patients in lower prognostic categories compared to SART model.

Supplementary Figure 3A. Clinic 916.

		MLCS2 Model Prognostic Categories (LBP range)			
		Low LBP < 25%	Medium LBP 25% to 49.9%	High LBP 50% to 74.9%	Very High ≥ 75%
SART Model Prognostic Categories (LBP range)	Low LBP < 25%	n = 21 LBR = 5%	n = 55 LBR = 40%	n = 0	n = 0
	Medium LBP 25% to 49.9%	n = 3 LBR = 0%	n = 164 LBR = 38%	n = 66 LBR = 68%	n = 0
	High LBP 50% to 74.9%	n = 0	n = 27 LBR = 48%	n = 469 LBR = 64%	n = 6 LBR = 100%
	Very High ≥ 75%	n = 0	n = 0	n = 0	n = 0
	Subtotal	n = 24 LBR = 4%	n = 246 LBR = 39%	n = 535 LBR = 65%	n = 6 LBR = 100%
		Subtotal			
		n = 811 LBR = 55%			

Supplementary Figure 3B. Clinic 552.

		MLCS2 Model Prognostic Categories (LBP range)			
		Low LBP < 25%	Medium LBP 25% to 49.9%	High LBP 50% to 74.9%	Very High LBP ≥ 75%
SART Model Prognostic Categories (LBP range)	Low LBP < 25%	n = 3 LBR = 0%	n = 10 LBR = 30%	n = 0	n = 0
	Medium LBP 25% to 49.9%	n = 0	n = 49 LBR = 41%	n = 23 LBR = 48%	n = 0
	High LBP 50% to 74.9%	n = 0	n = 8 LBR = 38%	n = 200 LBR = 70%	n = 0
	Very High LBP ≥ 75%	n = 0	n = 0	n = 0	n = 0
	Subtotal	n = 3 LBR = 0%	n = 67 LBR = 39%	n = 223 LBR = 67%	n = 0
		Subtotal			
		n = 293 LBR = 60%			

Supplementary Figure 3C. Clinic 869.

MLCS2 Model Prognostic Categories (LBP range)					
	Low LBP < 25%	Medium LBP 25% to 49.9%	High LBP 50% to 74.9%	Very High LBP ≥ 75%	Subtotal
SART Model Prognostic Categories (LBP range)	Low LBP < 25%	n = 10 LBR = 10%	n = 13 LBR = 38%	n = 0	n = 23 LBR = 26%
	Medium LBP 25% to 49.9%	n = 4 LBR = 0%	n = 114 LBR = 47%	n = 25 LBR = 76%	n = 143 LBR = 51%
	High LBP 50% to 74.9%	n = 0	n = 74 LBR = 35%	n = 307 LBR = 60%	n = 381 LBR = 55%
	Very High LBP ≥ 75%	n = 0	n = 0	n = 0	n = 0
	Subtotal	n = 14 LBR = 7%	n = 201 LBR = 42%	n = 332 LBR = 61%	n = 547 LBR = 53%

Supplementary Figure 3D. Clinic 395.

MLCS2 Model Prognostic Categories (LBP range)					
	Low LBP < 25%	Medium LBP 25% to 49.9%	High LBP 50% to 74.9%	Very High ≥ 75%	Subtotal
SART Model Prognostic Categories (LBP range)	Low LBP < 25%	n = 15 LBR = 7%	n = 31 LBR = 35%	n = 2 LBR = 50%	n = 48 LBR = 27%
	Medium LBP 25% to 49.9%	n = 0	n = 54 LBR = 33%	n = 270 LBR = 61%	n = 330 LBR = 57%
	High LBP 50% to 74.9%	n = 0	n = 1 LBR = 0%	n = 432 LBR = 63%	n = 786 LBR = 70%
	Very High ≥ 75%	n = 0	n = 0	n = 0	n = 0
	Subtotal	n = 15 LBR = 7%	n = 86 LBR = 34%	n = 704 LBR = 62%	n = 359 LBR = 79%

Supplementary Figure 3E. Clinic 635.

MLCS2 Model Prognostic Categories (LBP range)					
	Low LBP < 25%	Medium LBP 25% to 49.9%	High LBP 50% to 74.9%	Very High LBP ≥ 75%	Subtotal
SART Model Prognostic Categories (LBP range)	Low LBP < 25%	n = 27 LBR = 15%	n = 44 LBR = 23%	n = 0	n = 71 LBR = 20%
	Medium LBP 25% to 49.9%	n = 4 LBR = 0%	n = 303 LBR = 36%	n = 136 LBR = 58%	n = 443 LBR = 42%
	High LBP 50% to 74.9%	n = 0	n = 27 LBR = 30%	n = 820 LBR = 65%	n = 847 LBR = 66%
	Very High LBP ≥ 75%	n = 0	n = 0	n = 0	n = 0
	Subtotal	n = 31 LBR = 13%	n = 374 LBR = 34%	n = 956 LBR = 64%	n = 1361 LBR = 56%

Supplementary Figure 3F. Clinic 189.

MLCS2 Model Prognostic Categories (LBP range)					
	Low LBP < 25%	Medium LBP 25% to 49.9%	High LBP 50% to 74.9%	Very High LBP ≥ 75%	Subtotal
SART Model Prognostic Categories (LBP range)	Low LBP < 25%	n = 3 LBR = 0%	n = 15 LBR = 33%	n = 0	n = 18 LBR = 28%
	Medium LBP 25% to 49.9%	n = 1 LBR = 0%	n = 51 LBR = 35%	n = 46 LBR = 65%	n = 98 LBR = 49%
	High LBP 50% to 74.9%	n = 0	n = 7 LBR = 43%	n = 217 LBR = 68%	n = 224 LBR = 70%
	Very High LBP ≥ 75%	n = 0	n = 0	n = 0	n = 0
	Subtotal	n = 4 LBR = 0%	n = 73 LBR = 36%	n = 263 LBR = 67%	n = 340 LBR = 63%