Diminished ovarian reserve in the United States assisted reproductive technology population: diagnostic trends among 181,536 cycles from the Society for Assisted Reproductive Technology Clinic Outcomes Reporting System

Kate Devine, M.D.,[a] Sunni L. Mumford, Ph.D.,[b] Mae Wu, D.O.,[a] Alan H. DeCherney, M.D.,[a] Micah J. Hill, D.O.,[c] and Anthony Propst, M.D.[d]

[a] National Institutes of Health, Program in Reproductive and Adult Endocrinology and [b] National Institutes of Health, Epidemiology Branch, Division of Intramural Population Health Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development; and [c] Department of Obstetrics and Gynecology, Uniformed Services University of the Health Sciences, Bethesda, Maryland; and [d] Texas Fertility Center, Austin, Texas

Objective: To evaluate trends in diminished ovarian reserve (DOR) assignment in the Society for Assisted Reproductive Technology (SART) Clinic Outcomes Reporting System database and to evaluate its accuracy in predicting poor ovarian response (POR) as defined in European Society of Human Reproduction and Embryology's Bologna criteria (2011).

Design: Retrospective cohort study.

Setting: Not applicable.

Patient(s): A total of 181,536 fresh, autologous ART cycles reported to SART by U.S. clinics in 2004 and 2011 (earliest and most recent available reporting years).

Intervention(s): None.

Main Outcome Measure(s): DOR assignment was the primary exposure. POR, defined as cycle cancellation for poor response or less than 4 oocytes retrieved after conventional gonadotropin stimulation ($>$149 IU FSH daily), was the primary outcome. Secondary outcomes were live birth and number of oocytes retrieved. DOR prevalence, power of DOR and FSH ($</\geq$ 12 mIU/mL) to predict POR, and live birth in POR cycles were also calculated.

Result(s): DOR prevalence increased from 19% to 26% from 2004 to 2011. Among cycles clinically assigned as DOR, incidence of POR decreased from 32% to 30%, and live birth improved from 15% to 17%. Comparing basal FSH $\geq$ 12 versus clinical assignment of DOR, basal FSH had a higher specificity (92.2% vs. 81.6%) and positive predictive value (38.3% vs. 30.9%) for predicting POR. Live birth among POR cycles was 4%.

Conclusion(s): DOR diagnosis is increasing, and accuracy remains poor, despite the availability of additional diagnostic parameters such as antral follicle count and antimüllerian hormone. POR entailed poor outcomes, but the majority of patients clinically assigned as DOR did not experience POR. Development and use of more accurate predictors of POR are needed to minimize patient distress resulting from overdiagnosis. (Fertil Steril® 2015;104:612–9. ©2015 by American Society for Reproductive Medicine.)
Key Words: Diminished ovarian reserve, poor ovarian response, live birth, diagnosis, FSH

Use your smartphone to scan this QR code and connect to the discussion forum for this article now.*

* Download a free QR code scanner by searching for "QR scanner" in your smartphone's app store or app marketplace.

The diagnosis of diminished ovarian reserve (DOR) is clinically applied to patients with infertility who are thought likely to experience a poor ovarian response (POR) to ovarian stimulation. Assignment of DOR in the assisted reproductive technology (ART) population is increasing, although diagnostic criteria remain poorly defined [1, 2]. Per the American Society for Reproductive Medicine (ASRM), DOR has been defined in various ways, to include a reduced fecundability and/or POR to gonadotropin stimulation [1]. Diagnosis of DOR may cause understandable distress among patients [3]; therefore, it is important to minimize mis- and overdiagnosis. Unfortunately, recent data suggest that DOR is overdiagnosed. Butts et al. noted a significant upward trend in the prevalence of DOR [4]. Among their sample of younger than 40 ART patients diagnosed with DOR, a >21% live-birth rate was noted. This relatively high proportion of successful cycles does not accord with the expected poor prognosis of this diagnostic category. A standardized definition of DOR is needed to reduce overdiagnosis.

In 2011, the European Society of Human Reproduction and Embryology (ESHRE) outlined its Bologna criteria. This consensus recommended the minimum criteria needed to predict POR, namely, that at least two of the three following conditions be met: [1] age greater than or equal to 40 years or any conditions linked to decreased resting follicles; [2] low antral follicle count (AFC) or antimüllerian hormone (AMH); [3] prior poor ovarian response, that is, a history of cycle cancellation for poor response or fewer than four oocytes at retrieval after conventional gonadotropin stimulation (>149 IU FSH daily) [5]. These criteria have not been fully validated among a large population. However, Chai et al. recently found that subfertile patients who were deemed poor responders by these criteria had a lower chance of live birth after ART [6]. Similarly, Polyzos et al. have found an association between POR and poor treatment outcomes from both natural cycle IVF [7] and traditional IVF/intracytoplasmic sperm injection [8].

Given the upward trend in clinical DOR assignment and the lack of validated diagnostic criteria, we sought to quantify the increase in proportion of cycles clinically assigned as DOR (DOR prevalence), using data from the Society for Assisted Reproductive Technologies Clinic Online Reporting System (SART CORS). We also sought to determine whether the increase in DOR among ART patients represents improved detection versus overapplication. Therefore, we conducted a retrospective analysis of all autologous SART CORS cycles in 2004 and 2011, the earliest and most recent available reporting years, to evaluate the trends in clinical DOR assignment and the accuracy of DOR in predicting POR, as defined by the Bologna criteria.

MATERIALS AND METHODS
Data Source and Permissions

The SART Research Committee approved this study and provided data from SART CORS. SART CORS includes information from more than 90% of clinics performing ART in the United States. Deidentified data from ART cycles are entered by individual clinics, verified by SART, and reported to the Centers for Disease Control. The study was approved by the Uniformed Services University of the Health Sciences Institutional Review Board. The extracted data set included all fresh, autologous ART cycles from 2004 and 2011 (the earliest and latest available reporting years) for a total of 181,536 cycles.

Inclusion and Exclusion Criteria

All fresh, autologous ART cycles reported to SART CORS in 2004 and 2011 were included in the initial data set. Donor-recipient cycles were excluded. To minimize the impact of differences in practice patterns between 2004 and 2011, cycles with transfer of any cryopreserved embryos or embryos from cryopreserved oocytes (i.e., combined fresh/frozen transfers) were excluded. In addition, cycles initiated without the intent to transfer embryos (i.e., "batching cycles") were excluded. All of the above excluded cycles were removed by SART before we received the initial data set, which included 183,555 cycles. We then additionally excluded a total of 2,019 preimplantation genetic diagnosis/screening cycles (cycles with transfer of biopsied embryos), leaving 181,536 cycles for analysis. All of the 2,019 preimplantation genetic diagnosis/screening cycles were reported in 2011.

Definitions and Statistics

Prevalence of DOR diagnosis as clinically assigned and entered by the treating clinic was calculated among all included cycles, both by year and overall. DOR prevalence in 2004 was compared with 2011 using χ^2 analysis.

Unstimulated cycles (i.e., cycles with FSH dose of 0 or with no information available on FSH dose) were excluded in subsequent analyses, in which clinical, precycle assignment of DOR was the primary exposure and POR to gonadotropins during that cycle was the primary outcome. POR was identified using the Bologna criteria among cycles that were

cancelled for poor response or where fewer than four oocytes were obtained at retrieval performed after conventional gonadotropin stimulation (>149 IU FSH daily).

Secondary outcomes included the number of oocytes retrieved and live birth per cycle start. Subanalysis of live birth per cycle start among cycles meeting the Bologna criteria for POR was performed to evaluate the validity of this definition for POR. Comparisons were performed via analysis of variance for number of oocytes retrieved and using χ^2 analysis for live birth.

Regression Analyses

To evaluate patient and cycle characteristics potentially associated with clinical assignment of DOR diagnosis, univariate and multivariate logistic regression were performed. Factors assessed included age, number of prior fresh cycles, number of prior gonadotropin cycles, gravidity, and elevated basal FSH. Relative risk of DOR was first calculated for each factor separately. Adjusted (multivariate) models included all factors together. Using the same method, we tested these factors, as well as assignment of DOR diagnosis before cycle start, for association with poor response in the current cycle, as defined by the Bologna consensus. Statistical analysis was performed using SAS, version 9.3, software (SAS Institute, Inc.).

Diagnostic Parameters

The power of DOR to predict POR was also assessed via multiple parameters. Sensitivity, specificity, and positive and negative predictive values were calculated using standard definitions. The diagnostic accuracy of clinical DOR assignment was assessed over time and was compared with that of basal FSH ≥ 12 alone and age ≥ 40 years alone. FSH cutoff was selected a priori as it constituted the 90th centile of the data set and has been shown to reliably predict poor pregnancy rates (9). Given that the Bologna criteria and others (10–15) suggest that at least two criteria are needed to classify a patient as a poor responder, we also assessed the predictive power of clinical DOR diagnosis relative to the combination of basal FSH ≥ 12 and age ≥ 40 years. Cycles where basal FSH was not entered or was entered as 0 or as >50 mIU/mL were excluded from these analyses.

RESULTS

Overall clinical assignment of DOR (DOR prevalence) increased from 19% to 26% from 2004 to 2011 (P<.0001). There was a small but significant increase of 3.6 months in patient age at cycle start (P<.0001) as well as a 3.6% increase in percentage of women who were >40 years of age at cycle start (P<.0001). The prevalence of DOR was also higher in 2011 in patients younger than 40 years old (P<.0001). The DOR prevalence increased 37% in the overall cohort in 2011 and increased 42% among patients younger than 40 years (Table 1, Fig. 1).

Among cycles clinically assigned as DOR before cycle start, the incidence of POR (cycle cancellation for poor response or fewer than four oocytes obtained at retrieval) decreased from 32% to 30% from 2004 to 2011 (P=.0013), suggesting no improvement in the predictive power of the diagnosis over this time period. Overall, 69% of stimulated cycles assigned as DOR failed to meet the Bologna definition for poor response. The mean number of oocytes obtained during stimulated DOR cycles remained constant from 2004 to 2011 at approximately eight, and live birth per DOR cycle start increased slightly from 15% to 17% (P<.0001).

When the DOR cohort was limited to women with age younger than 40 years, the mean oocyte yield was 8.8 and 8.8 and the live-birth rate was 21% and 24% in 2004 and 2011, respectively. For all cycles in women younger than 40 years (with or without DOR diagnosis), the mean oocyte yield was 13.3 and the live-birth rate was 34% (Table 2, Fig. 1).

Despite the significant increase in DOR prevalence from 2004 to 2011, the overall incidence of POR among all stimulated ART cycles decreased from 16% to 15% over this time period (P=.0009). Live birth increased from 28% to 30% (P<.0001), and the mean number of oocytes retrieved decreased slightly from 12.4 to 12.3 (Table 2, Fig. 1). Live birth per cycle start among all cycles meeting the Bologna definition of POR was extremely low at 4% and improved from only 3% to 4% from 2004 to 2011 (P<.0001; Supplemental Table 1).

The models detailed in Supplemental Table 2 calculate the relative risk (RR) of DOR assignment based on several patient characteristics. Factors associated with clinical DOR assignment on univariate analysis were age at cycle start (<40 vs. ≥40 years): RR, 3.58; 95% confidence interval [CI], 3.52–3.64; and elevated basal FSH (<12 vs. ≥12 mIU/mL): RR, 3.15; 95% CI, 3.09–3.20. Factors protective against a

TABLE 1

Prevalence of diminished ovarian reserve by reporting year.

|	2004		2011		Combined	
Variable	All cycles (n = 87,352)	Age <40 (n = 69,403)	All cycles (n = 94,184)	Age <40 (n = 71,487)	All cycles (n = 181,536)	Age <40 (n = 140,890)
Age (y), mean ± SD	35.3 ± 4.6[a]	33.7 ± 3.7[b]	35.6 ± 4.8[a]	33.6 ± 3.8[b]	35.5 ± 4.7	33.7 ± 3.7
DOR prevalence, n (%)	16,392[a] (19)	8,262[a] (12)	24,909[a] (26)	12,052[a] (17)	41,301 (23)	20,314 (14)

Note: DOR = diminished ovarian reserve.
[a] P<.0001, 2004 vs. 2011.
[b] P=.0004, 2004 vs. 2011.

Devine. Overdiagnosis of DOR in SART ART cycles. Fertil Steril 2015.

FIGURE 1
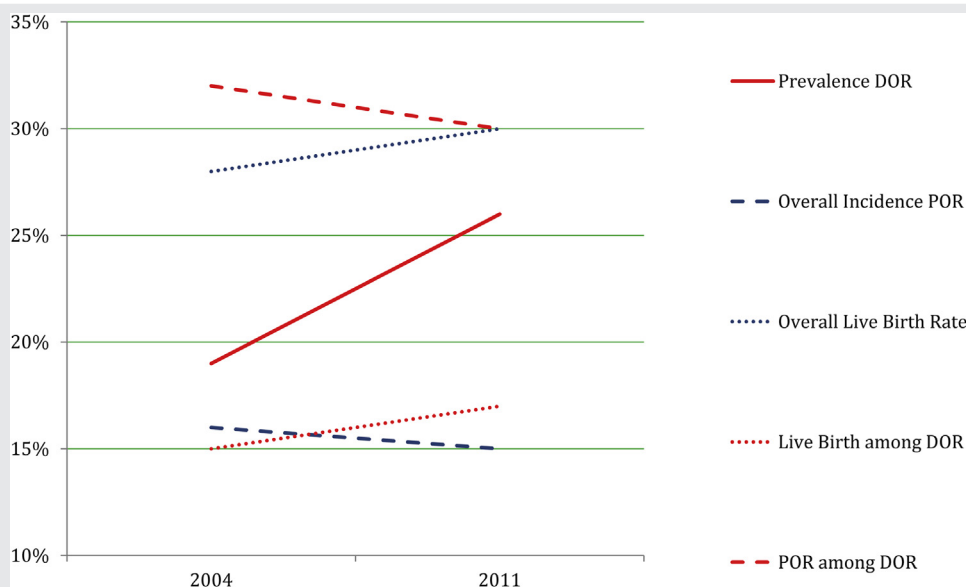
Trends in DOR prevalence, POR incidence, and live birth.

Devine. Overdiagnosis of DOR in SART ART cycles. Fertil Steril 2015.

diagnosis of DOR included low gravidity, having had fewer than two prior IVF cycles, and having had no prior gonadotropin ovulation induction cycles. However, having had no prior gonadotropin ovulation induction cycles was not associated with DOR in the adjusted analysis, and the RR of being assigned DOR in the setting of low gravidity reversed from being protective in the unadjusted analysis to a slight increase in RR of the diagnosis in the adjusted model. This was likely because both of these factors are associated with young age and thus were not independently protective after adjusting for patient age.

We performed similar analyses assessing the RR of POR in the current cycle with the same factors, this time adding DOR diagnosis as a covariate. Univariate analyses indicated strong associations with patient age (≥ 40 years), basal FSH (≥ 12 mIU/mL), and the assignment of DOR diagnosis with POR in the current cycle (RR, 2.37; 95% CI, 2.32–2.42; RR, 2.91; 95% CI, 2.84–2.99; and RR, 2.88; 95% CI, 2.82–2.94, respectively). Univariate analysis again indicated low gravidity and fewer than two prior ART cycles to be protective for POR. However, the association was not significant once age was controlled for in the adjusted analysis. Elevated basal FSH, DOR, and age remained highly associated with POR on multivariate analysis (Supplemental Table 3).

Finally, diagnostic power of DOR assignment was assessed over time and relative to other predictors of POR (Table 3). Sensitivity and negative predictive value (NPV) of DOR to predict POR were increased among 2011 cycles relative to 2004, concurrent with the observed 7% increase in clinical DOR assignment. However, specificity and positive predictive value (PPV) of the diagnosis decreased. The PPV in 2011 of DOR assignment for predicting POR was only

30.2%. Relative to DOR, basal FSH had a higher specificity (92.2% vs. 81.6%) and PPV (38.3% vs. 30.9%) for predicting POR but a lower sensitivity (25.7% vs. 45.7%) and NPV (86.9% vs. 89.3%). Age ≥ 40 years performed more poorly as a predictor of POR than clinical DOR diagnosis for all parameters assessed; however, sensitivity and NPV were higher for age alone than for elevated FSH alone (age vs. FSH: sensitivity, 40.4% vs. 25.7%; specificity, 81.0% vs. 92.2%). The combination of age ≥ 40 years and elevated basal FSH had the greatest specificity and PPV at 97.6% and 48.0%, respectively, although sensitivity was very poor at only 12.1%, and these combined criteria had the lowest NPV at 85.5%. The live-birth rate among those stimulated cycles in women 40 years or older with a basal FSH ≥ 12 mIU/mL was only 6.7%.

DISCUSSION

Our results demonstrate that DOR diagnosis is increasing in the U.S. ART population and that the upward trend largely represents overdiagnosis rather than improved detection. The prevalence of DOR among SART CORS cycles increased 7% from 2004 to 2011, and the ability of clinical DOR assignment to predict POR in the concurrent cycle worsened from 2004 to 2011. Furthermore, the finding of an extremely low live birth rate (4%) among cycles meeting the Bologna criteria for POR supports the definition selected for these criteria as one consistent with very poor cycle outcomes and is consistent with the findings of smaller retrospective analyses (8).

Our finding of increasing DOR prevalence in the U.S. ART population without a corresponding increase in the accuracy of this label as a predictor of poor ART outcomes accords with

TABLE 2

Outcomes among all stimulated[a] cycles versus those clinically assigned as diminished ovarian reserve.

All stimulated cycles	2004 Cycles		2011 Cycles		Combined 2004 and 2011	
	All (n = 84,099)	Age <40 (n = 66,779)	All (n = 90,102)	Age <40 (n = 68,642)	All (n = 174,201)	Age <40 (n = 135,421)
POR, n (%)	13,088[b] (16)	8,299[c] (12)	13,508[b] (15)	7,549[c] (11)	26,596 (15)	15,848 (12)
No. of oocytes retrieved, mean ± SD	12.4 ± 7.8[d]	13.2 ± 7.9[e]	12.3 ± 7.8[d]	13.3 ± 7.9[e]	12.4 ± 7.8	13.3 ± 7.9
Live birth, n (%)	23,231[c] (28)	21,375[c] (32)	26,920[c] (30)	24,375[c] (35)	50,151 (29)	45,750 (34)
DOR subsets	All DOR <40 (n = 15,752)	DOR Age <40 (n = 7,952)	All DOR <40 (n = 23,559)	DOR Age <40 (n = 11,457)	All DOR (n = 39,311)	All DOR <40 (n = 19,409)
POR, n (%)	5,004[f] (32)	2,331[g] (29)	7,123[f] (30)	3,095[g] (27)	12,127 (31)	5,403 (28)
No. of oocytes retrieved, mean ± SD	8.3 ± 5.9	8.8 ± 6.1	8.3 ± 5.9	8.8 ± 6.1	8.3 ± 5.9	8.8 ± 6.1
Live birth, n (%)	2,357[c] (15)	1,693[c] (21)	4,003[c] (17)	2,811[c] (24)	6,360 (16)	4,504 (23)

Note: DOR = diminished ovarian reserve; POR = poor ovarian response.
[a] Cycles with a daily FSH dose = 0 or missing were excluded.
[b] P=.0009, 2004 vs. 2011.
[c] P<.0001, 2004 vs. 2011.
[d] P=.0318, 2004 vs. 2011.
[e] P=.0023, 2004 vs. 2011.
[f] P=.0013, 2004 vs. 2011.
[g] P=.0001, 2004 vs. 2011.

Devine. Overdiagnosis of DOR in SART ART cycles. Fertil Steril 2015.

those of Butts et al. (4). Possible explanations for increased DOR prevalence include older age of women undergoing ART, the addition of diagnostic testing modalities, and clinic motivations to attract patients carrying the diagnosis and/or to explain suboptimal success rates.

The inverse relationship between age and the ovarian follicular pool has been well described (16–19) and corresponds with lower ART and controlled ovarian hyperstimulation outcomes among women diagnosed with DOR (20, 21). However, the difference in mean age at cycle start increased by only 3.6 months from 2004 to 2011, which did not account for the 37% relative increase in DOR prevalence. While it is important to note that the proportion of cycles starting at age 40 or greater increased by 3.6% in 2011 versus 2004, DOR assignment increased even more sharply when analysis was limited to patients with age younger than 40 years.

Increased use of AFC and AMH may have contributed to the observed upward trend in DOR assignment. We were unable to directly evaluate the contribution of AFC and AMH, as these data are not as yet available in the SART CORS database. However, our finding of a sharper rate of rise among younger patients, in whom age alone is unlikely to be used for DOR assignment, further suggests reliance on these tests. In recent years, AFC and AMH have been reported to have higher diagnostic accuracy than FSH and clinical history alone (22–26). Therefore, their increased use would be expected to result not only in a higher prevalence of DOR but also perhaps in an increase in poor ovarian response. Nevertheless, diagnostic accuracy has not improved. Among DOR cycles in 2011 versus 2004, incidence of poor response decreased by 2%, mean oocyte yield remained constant, and live birth improved slightly (17% vs. 15%). Improvements in embryo culture, especially increased use of extended culture and blastocyst transfer over this time period (27–31), likely contributed to the upward trend in live birth but not to the decrease in POR. Overall in 2011, 70% of cycles assigned a diagnosis of DOR did not experience POR as defined by the Bologna consensus. Furthermore, in patients younger than 40 years, 21% and 24% of DOR cycles resulted in live birth, in 2004 and 2011, respectively. This accords with the relatively high live-birth rate among younger DOR patients found by Butts et al. and strongly suggests overdiagnosis in this group (4).

The limited diagnostic utility of clinical DOR assignment was demonstrated most clearly by its low PPV for POR. Overall, if a patient received a diagnosis of DOR before a stimulated ART cycle, there was only a 30.8% chance that she would meet the Bologna criteria for poor response. Basal FSH ≥ 12 mIU/mL was more strongly associated with POR on multivariate regression analysis. Both basal FSH alone and FSH combined with age had higher PPV and specificity for predicting POR than did clinical DOR assignment. If, as we believe to be the case, the primary purpose of DOR assignment is to allow for better patient counseling that autologous ART is relatively less likely to result in success (1, 32, 33) and that alternative therapies should be considered, PPV and sensitivity of DOR are its most relevant predictive parameters.

TABLE 3

Sensitivity and specificity of diminished ovarian reserve and elevated follicle-stimulating hormone in predicting poor ovarian response in stimulated[a] ART cycles.

	POR		
Variable	Yes	No	% (95% CI)
DOR overall			
Yes	12,171	27,184	Sensitivity 45.69 (45.09, 46.29)
No	14,469	120,421	Specificity 81.58 (81.38, 81.78)
			PPV 30.93 (30.47, 31.39)
			NPV 89.27 (89.11, 89.44)
DOR 2004			
Yes	5,004	10,748	Sensitivity 38.23 (37.40, 39.07)
No	8,084	60,263	Specificity 84.86 (84.60, 85.13)
			PPV 31.77 (31.04, 32.50)
			NPV 88.17 (87.93, 88.41)
DOR 2011			
Yes	7,123	16,436	Sensitivity 52.73 (51.89, 53.58)
No	6,385	60,158	Specificity 78.54 (78.25, 78.83)
			PPV 30.23 (29.65, 30.83)
			NPV 90.40 (90.18, 90.63)
Elevated FSH (≥ 12)			
Yes	5,285	8,529	Sensitivity 25.72 (25.13, 26.33)
No	15,260	100,934	Specificity 92.21 (92.05, 92.37)
			PPV 38.26 (37.45, 39.07)
			NPV 86.87 (86.67, 87.06)
Age ≥ 40			
Yes	10,748	28,032	Sensitivity 40.41 (39.82, 41.00)
No	15,848	119,573	Specificity 81.01 (80.81, 81.21)
			PPV 27.72 (27.27, 28.16)
			NPV 88.30 (88.12, 88.47)
Elevated FSH (≥ 12) and age ≥ 40			
Yes	2,480	2,685	Sensitivity 12.07 (11.63, 12.52)
No	18,065	106,778	Specificity 97.55 (97.45, 97.64)
			PPV 48.02 (46.64, 49.39)
			NPV 85.53 (85.33, 85.72)

Note: DOR = diminished ovarian reserve; FSH = follicle stimulating hormone; NPV = negative predictive value; POR = poor ovarian response; PPV = positive predictive value.
[a] Cycles with a daily FSH dose = 0 or missing were excluded.

Devine. Overdiagnosis of DOR in SART ART cycles. Fertil Steril 2015.

The poor predictive performance of clinical DOR assignment in its current use underscores the importance of establishing standardized, validated diagnostic criteria. The Bologna criteria for POR suggest a uniform framework for predicting poor response to ovarian stimulation, which is one of the purposes of clinical DOR assignment pretreatment. Although several investigators have recently expressed concern regarding the heterogeneous population that the Bologna criteria represent (34, 35), many have already begun implementing these standards as inclusion criteria for clinical studies of the effectiveness of various interventions for poor responders (7, 36–38). However, to date, the Bologna criteria for POR have not been fully validated. In its current form, the SART CORS database does not have the capacity to do so, since it does not contain previous cycle outcomes, AFC, or AMH (the latter criterion will be available in future reporting years) (39). Although AMH has great potential as a diagnostic and prognostic tool for patients undergoing ART, standardization is needed, to decrease variability and improve accuracy among AMH assays in current use (40). Reliable collection and entry of AMH data by clinics into SART CORS may ultimately enable validation of the assay, further improving its utility.

A strength of the current study lies in its analysis of outcomes among cycles meeting the Bologna definition for poor response. Although we could not assess the association of previous poor response by the Bologna definition with current cycle outcomes, our finding of a live-birth rate of 4% among cycles meeting this definition substantiates in a large cohort the criterion as a dismal indicator. Furthermore, it accords with the finding of 6.0% live birth among 823 cycles meeting the Bologna criteria, as retrospectively analyzed by Polyzos et al. (8). The finding that the combination of basal FSH ≥ 12 mIU/mL and age ≥ 40 years resulted in the highest specificity and PPV for poor response in the current cycle supports the consensus's conclusion that at least two criteria should be necessary for diagnosis. Patients meeting these two criteria should be counseled about the 6.7% likelihood of live birth from autologous ART and should be advised to consider the use of donor oocytes.

Strengths of the current study include its large cohort, its evaluation of the trend in DOR diagnosis and diagnostic accuracy via analysis of reporting years separated by a 7-year interval, and its corroboration of the Bologna definition for poor response. The very poor PPV of DOR in U.S. ART cycles has clinical utility. Although this has previously been reported

for FSH (5, 9, 41), clinical DOR assignment, which takes multiple objective factors into consideration, would be expected to perform better. The information that 70% of all patients diagnosed with DOR had at least four oocytes retrieved and that 16% achieved live birth may provide some limited reassurance to patients informed by clinicians of this diagnosis.

Our study has several limitations, the most significant of which derive from its retrospective nature and the unavailability of relevant cycle characteristics in the data set. AFC, AMH, and previous cycle outcomes would further refine our understanding of patterns in DOR diagnosis and assist in validating the Bologna criteria for POR more completely. Once available in a large database, such as SART CORS, these, along with age, FSH, and other cycle characteristics, should be used to determine an optimal set of predictors of POR. Importantly, the fully deidentified nature of the provided data set made it impossible to link cycles. This precluded an analysis of whether and when a current DOR diagnosis might manifest as POR in the future. Furthermore, we could not conduct an analysis to account for repeated cycle bias; however, we performed sensitivity analyses of first cycles only, which revealed lower overall DOR prevalence but a similar trend in DOR diagnosis over time. In patients with DOR, the number of oocytes retrieved and live-birth rates among first cycles were similar to those reported among all cycles. Another limitation is the inability to objectively measure distress or anxiety felt by the patient. Furthermore, it is difficult to establish whether patients who have been diagnosed with DOR have been counseled on the definition. This particular aspect could not be investigated with the current data set. Future studies similar to Cizmeli et al. are needed to evaluate the psychological impact of this assignment in fertility treatment cycles (3).

Our findings indicate that, despite attempts to improve diagnostic standards, there exists a mismatch between patients clinically assigned as DOR and their ovarian response during ART cycles. This mismatch has not improved over the past decade. There certainly exists a group of patients who, despite aggressive stimulation, are unlikely to experience live birth after fresh, autologous ART (20, 42, 43). However, the present study suggests that when and if patients are informed that DOR is their SART CORS diagnosis, they should also be made aware that this designation in its current clinical use does not necessarily portend a poor response to ovarian stimulation. The Bologna consensus, absent full validation to date, provides clinicians well-reasoned guidelines for defining POR. The current analysis supports this definition as one that entails a very poor prognosis for live birth after ART.

It is important to acknowledge that DOR and POR are not interchangeable. If a patient is diagnosed with DOR on the basis of ovarian reserve testing, it does not signify that she will have POR during her stimulation. However, DOR when properly assigned should signify that a patient is at increased risk of POR and may indicate a higher starting dose of gonadotropins than for similar aged patients.

Overdiagnosis of DOR among SART cycles likely resulted from indiscriminate use of the category by individual clinics, in the absence of standards. Further studies are needed to pro-spectively validate POR diagnostic criteria (such as those laid out by ESHRE) to predict poor cycle outcomes in addition to standardized DOR definitions to guide treatment choices for infertile couples.

REFERENCES

1. ASRM Practice Committee. Testing and interpreting measures of ovarian reserve: a committee opinion. Fertil Steril 2012;98:1407–15.
2. Polyzos NP, Devroey P. A systematic review of randomized trials for the treatment of poor ovarian responders: is there any light at the end of the tunnel? Fertil Steril 2011;96:1058–61.
3. Cizmeli C, Lobel M, Franasiak J, Pastore L. Levels and associations among self-esteem, fertility distress, coping, and reaction to potentially being a genetic carrier in women with diminished ovarian reserve. Fertil Steril 2013;99: 2037–44.
4. Butts S, Ratcliffe S, Dokras A, Seifer D. Diagnosis and treatment of diminished ovarian reserve in assisted reproductive technology cycles of women up to age 40 years: the role of insurance mandates. Fertil Steril 2013;99: 382–8.
5. Ferraretti AP, Marca A, Fauser B, Tarlatzis B, Nargund G, Gianaroli L. ESHRE consensus on the definition of "poor response" to ovarian stimulation for in vitro fertilization: the Bologna criteria. Hum Reprod 2011;26:1616–24.
6. Chai J, Lee VC, Yeung TW, Li RW, Ho PC, Ng EH. Live birth and cumulative live birth rates in expected poor ovarian responders defined by the Bologna criteria following IVF/ICSI treatment. PLoS One 2015;10:e0119149.
7. Polyzos NP, Blockeel C, Verpoest W, De Vos M, Stoop D, Vloeberghs V, et al. Live birth rates following natural cycle IVF in women with poor ovarian response according to the Bologna criteria. Hum Reprod 2012;27:3481–6.
8. Polyzos NP, Nwoye M, Corona R, Blockeel C, Stoop D, Haentjens P, et al. Live birth rates in Bologna poor responders treated with ovarian stimulation for IVF/ICSI. Reprod Biomed Online 2014;28:469–74.
9. Esposito MA, Coutifaris C, Barnhart KT. A moderately elevated day 3 FSH concentration has limited predictive value, especially in younger women. Hum Reprod 2002;17:118–23.
10. Bancsi LF, Broekmans FJ, Eijkemans MJ, de Jong FH, Habbema JD, teVelde ER. Predictors of poor ovarian response in in vitro fertilization: a prospective study comparing basal markers of ovarian reserve. Fertil Steril 2002; 77:328–36.
11. Demirol A, Gurgan T. Comparison of microdose flare-up and antagonist multiple-dose protocols for poor-responder patients: a randomized study. Fertil Steril 2009;92:481–5.
12. Barrenetxea G, Agirregoikoa JA, Jimenez MR, de Larruzea AL, Ganzabal T, Carbonero K. Ovarian response and pregnancy outcome in poor-responder women: a randomized controlled trial on the effect of luteinizing hormone supplementation on in vitro fertilization cycles. Fertil Steril 2008;89:546–53.
13. Massin N, Cedrin-Durnerin I, Coussieu C, Galey-Fontaine J, Wolf JP, Hugues JN. Effects of transdermal testosterone application on the ovarian response to FSH in poor responders undergoing assisted reproduction technique- a prospective, randomized, double-blind study. Hum Reprod 2006; 21:1204–11.
14. Detti L, Williams DB, Robins JC, Maxwell RA, Thomas MA. A comparison of three downregulation approaches for poor responders undergoing in vitro fertilization. Fertil Steril 2005;84:1401–15.
15. Weissman A, Farhi J, Royburt M, Nahum H, Glezerman M, Levran D. Prospective evaluation of two stimulation protocols for low responders who were undergoing in vitro fertilization-embryo transfer. Fertil Steril 2003; 79:886–92.

16. ASRM. Female age-related fertility decline: Committee Opinion No. 589. Fertil Steril 2014;101:633–4.

17. Baker TG. A quantitative and cytological study of germ cells in human ovaries. Proc R Soc Lond B Biol Sci 1963;158:417–33.

18. Block E. Quantitative morphological investigations of the follicular system in women; variations at different ages. Acta Anat 1952;14:108–23.

19. Faddy MJ, Gosden RG, Gougeon A, Richardson SJ, Nelson J. Accelerated disappearance of ovarian follicles in mid-life: implications for forecasting menopause. Hum Reprod 1992;7:1342–6.

20. Levi A, Raynault M, Bergh P, Drews M, Miller B, Scott R. Reproductive outcomes in patients with diminished ovarian reserve. Fertil Steril 2001;76:666–9.

21. Harris I, Missmer S, Hornstein M. Poor success of gonadotropin-induced controlled ovarian hyperstimulation and intrauterine insemination for older women. Fertil Steril 2010;94:144–8.

22. Hendriks D, Mol B, Bancsi L, teVelde E, Broekmans F. Antral follicle count in the prediction of poor ovarian response and pregnancy after in vitro fertilization: a meta-analysis and comparison with basal follicle-stimulating hormone level. Fertil Steril 2005;83:291–301.

23. Broekmans F, Kwee J, Hendriks D, Mol B, Lambalk C. A systematic review of tests predicting ovarian reserve and IVF outcome. Hum Reprod Update 2006;12:685–718.

24. Broer S, Mol B, Hendricks D, Broekmans F. The role of antimullerian hormone in prediction of outcome after IVF: comparison with the antral follicle count. Fertil Steril 2009;91:705–13.

25. Barad DH, Weghofer A, Gleicher N. Comparing anti-Mullerian hormone (AMH) and follicle-stimulating hormone (FSH) as predictors for ovarian function. Fertil Steril 2009;91:1553–5.

26. Gleicher N, Weghofer A, Barad DH. Anti-Mullerian hormone (AMH) defines, independent of age, low versus good live-birth chances in women with severely diminished ovarian reserve. Fertil Steril 2010;94:2824–7.

27. Quinn P. The development and impact of culture media for assisted reproductive technologies. Fertil Steril 2004;81:27–9.

28. Scott L. Embryological strategies for overcoming recurrent assisted reproductive technology treatment failure. Hum Fertil (Camb) 2002;5:206–14.

29. Karaki R, Samarraie S, Younis N, Lahloub T, Ibrahim M. Blastocyst culture and transfer: a step toward improved in vitro fertilization outcome. Fertil Steril 2002;77:114–8.

30. Gardner DK, Vella P, Lane M, Wagley L, Schlenker T, Schoolcraft W. Culture and transfer of human blastocysts increases implantation rates and reduces the need for multiple embryo transfers. Fertil Steril 1998;69:84–8.

31. Glujovsky D, Blake D, Farquhar C, Bardach A. Cleavage stage versus blastocyst stage embryo transfer in assisted reproductive technology. Cochrane Database Syst Rev 2012:CD002118.

32. Butts SF, Owen C, Maingini M, Senapati S, Seifer DB, Dokras A. Assisted hatching and intracytoplasmic sperm injection are not associated with improved outcomes in assisted reproduction cycles for diminished ovarian reserve: an analysis of cycles in the United States from 2004 to 2011. Fertil Steril 2014;102:1041–7.

33. Yu B, Mumford S, Royster GD, Segars JH, Armstrong AY. Cost-effectiveness analysis comparing continuation of assisted reproductive technology with conversion to intrauterine insemination in patients with low follicle numbers. Fertil Steril 2014;102:435–9.

34. Papathanasiou A. Implementing the ESHRE ''poor responder'' criteria in research studies: methodological implications. Hum Reprod 2014;29:1835–8.

35. Venetis CA. The Bologna criteria for poor ovarian response: the good, the bad and the way forward. Hum Reprod 2014;29:1839–41.

36. Xu B, Li Z, Yue J, Jin L, Li Y, Ai J, et al. Effects of dehydroepiandrosterone administration in patients with poor ovarian response according to the Bologna criteria. PLoS One 2014;9:e99858.

37. Hu L, Bu Z, Guo Y, Su Y, Zhai J, Sun Y. Comparison of different ovarian hyperstimulation protocols efficacy in poor ovarian responders according to the Bologna crteria. Int J Clin Exp Med 2014;7:1128–34.

38. Polyzos NP, Devos M, Humaidan P, Stoop D, Ortega-Hrepich C, Devroey P, et al. Corifollitropin alfa followed by rFSH in a GnRH antagonist protocol for poor ovarian responder patients: an observational pilot study. Fertil Steril 2013;99:422–6.

39. Schattman G. SART newsletter fall-winter 2012 issue. Message from the president. Available at: http://www.sart.org/uploadedFiles/Affiliates/SART/SART_Links/SART%20Newsletter%20Fall%20Winter%202012.pdf. Accessed August 25, 2014.

40. Iliodromiti S, Anderson RA, Nelson SM. Technical and performance characteristics of anti-Mullerian hormone and antral follicle count as biomarkers of ovarian response. Hum Reprod 2014 Dec 8. pii: dmu062. [Epub ahead of print].

41. Bancsi LF, Broekmans FJ, Mol BW, Habbema JD, teVelde ER. Performance of basal follicle-stimulating hormone in the prediction of poor ovarian response and failure to become pregnant after in vitro fertilization: a meta-analysis. Fertil Steril 2003;79:1091–100.

42. Klinkert ER, Broekmans FJ, Looman CW, Habbema JD, te Velde ER. Expected poor responders on the basis of an antral follicle count do not benefit from a higher starting dose of gonadotrophins in IVF treatment: a randomized controlled trial. Hum Reprod 2005;20:611–5.

43. Scott RT, Elkind-Hirsch KE, Styne-Gross A, Miller K, Frattarelli JF. The predictive value for in vitro fertility delivery rates is greatly impacted by the method used to select the threshold between normal and elevated basal follicle-stimulating hormone. Fertil Steril 2008;89:868–78.

SUPPLEMENTAL TABLE 1

Breakdown of criteria used to define POR and outcomes among POR cycles.[a]

Variable	2004 Cycles		2011 Cycles		Combined 2004 and 2011	
No. of POR cycles	13,088	8,299	13,508	7,549	26,596	15,848
Canceled for poor response, n (% of POR)	8,526[b] (65)	5,612[b] (68)	7,452[b] (55)	4,372[b] (58)	15,978 (60)	9,984 (63)
Fewer than four oocytes at retrieval, n (% of POR)	4,562[b] (35)	2,687[b] (32)	6,056[b] (45)	3,177[b] (42)	10,618 (40)	5,864 (37)
Live birth, n (% of all POR) [% of POR by low oocyte yield]	428 (3)[b] [9]	354 (4)[b] [13]	613 (4)[b] [10]	456 (6)[b] [14]	1,041 (4) [10]	810 (5) [14]

Note: POR = poor ovarian response.
[a] Cycles with a daily FSH dose = 0 or missing were excluded.
[b] $P < .0001$, 2004 vs. 2011.

Devine. Overdiagnosis of DOR in SART ART cycles. Fertil Steril 2015.

SUPPLEMENTAL TABLE 2

Factors associated with clinical diminished ovarian response assignment.

DOR	Unadjusted				Adjusted[a]			
	RR	95% CI		P value	RR	95% CI		P value
Age (y), 40+ vs. <40	3.58	3.52	3.64	<.0001	2.76	2.71	2.82	<.0001
Prior fresh cycles								
0	0.61	0.59	0.62	<.0001	0.86	0.84	0.88	<.0001
1	0.76	0.75	0.78	<.0001	0.94	0.92	0.96	<.0001
2+	Reference							
Prior gonadotropin cycles								
0	0.92	0.90	0.95	<.0001	1.00	0.97	1.02	.7198
1	1.09	1.05	1.14	<.0001	1.05	1.01	1.09	.0102
2	1.02	0.98	1.06	.2502	1.00	0.97	1.04	.8594
3+	Reference							
Gravidity								
0	0.79	0.78	0.81	<.0001	1.05	1.03	1.07	<.0001
1	0.91	0.89	0.93	<.0001	1.03	1.01	1.05	.0094
2+	Reference							
Basal FSH, >12 vs. <12 mIU/mL	3.15[a]	3.09	3.20	<.0001	1.97	1.94	2.01	<.0001

Note: CI = confidence interval; DOR = diminished ovarian reserve; FSH = follicle-stimulating hormone; RR = relative risk.
[a] Cycles with basal FSH not entered or entered as 0 or >50 mIU/mL excluded from analysis.

Devine. Overdiagnosis of DOR in SART ART cycles. Fertil Steril 2015.

SUPPLEMENTAL TABLE 3

Factors associated with poor ovarion response in stimulated[a] cycles.

POR	Unadjusted				Adjusted[b]			
	RR	95% CI		P	RR	95% CI		P
Age, 40+ vs. <40	2.37	2.32	2.42	<.0001	1.60	1.55	1.64	<.0001
Prior fresh cycles								
0	0.80	0.78	0.83	<.0001	1.02	0.99	1.05	.1800
1	0.86	0.83	0.88	<.0001	0.99	0.95	1.02	.4472
2+		Reference						
Prior gonadotropin cycles								
0	1.00	0.97	1.04	.9221	1.07	1.03	1.10	.0005
1	1.06	1.01	1.11	.0306	1.05	0.99	1.10	.1046
2	0.98	0.93	1.03	.4141	0.98	0.92	1.03	.3978
3+		Reference						
Gravidity								
0	0.88	0.86	0.90	<.0001	1.03	1.00	1.06	.0976
1	0.93	0.90	0.96	<.0001	1.01	0.98	1.04	.5320
2+		Reference						
Basal FSH, >12 vs. <12	2.91	2.84	2.99	<.0001	1.96	1.91	2.02	<.0001
DOR, yes vs. no	2.88	2.82	2.94	<.0001	1.94	1.88	2.00	<.0001

Note: CI = confidence interval; FSH = follicle-stimulating hormone; POR = poor ovarian response; RR = relative risk.
[a] Cycles with a daily FSH dose = 0 or missing were excluded.
[b] Cycles with basal FSH not entered or entered as 0 or >50 mIU/mL were excluded from analysis.

Devine. Overdiagnosis of DOR in SART ART cycles. Fertil Steril 2015.