Published in final edited form as:

Fertil Steril. 2009 December ; 92(6): 2088–2090. doi:10.1016/j.fertnstert.2009.05.082.

A Strict Infertility Diagnosis Has Poor Agreement with the Clinical Diagnosis Entered Into SART

Thomas A. Molinaro, MD^{a,b}, Alka Shaunik, MD^a, Kathleen Lin, MD^a, Mary D. Sammel, ScD^b, and Kurt T. Barnhart, MD MSCE^{a,b}

^aDepartment of Obstetrics and Gynecology, University of Pennsylvania, Philadelphia PA

^bCenter for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, PA

Abstract

Objective—To investigate if strict infertility diagnoses correlate with clinical judgment and how adjudicated diagnosis may improve accuracy in predicting IVF success.

Design—Cross Sectional

Materials and Methods—Current criteria for infertility diagnoses were determined by literature review. IVF patient's charts between 2004–06 were adjudicated according to these strict criteria. Agreement with patient's clinical diagnosis entered into SART was measured using Kappa statistics to quantify the agreement between criteria. Pregnancy rates were calculated for each diagnosis by clinical and strict criteria. Success rates for diagnoses based upon each criterion were compared using multivariable logistic regression with adjustment for repeated measure.

Results—432 women underwent 590 IVF cycles. Kappa statistics showed only moderate agreement between strict and clinical diagnosis of endometriosis, male, and tubal factor. PCOS diagnosis was less correlated. Uterine, unexplained and diminished ovarian reserve (DOR) diagnoses showed the poorest agreement between diagnostic criteria. There are considerable differences based on the two criteria for diagnosis. There was poor agreement between diagnostic criteria in patients with multiple diagnoses. By strict criteria, these patients were significantly less likely to have a live birth than those with a single diagnosis (OR=0.61, p=0.019). This finding was similar with clinical criteria (OR=0.68, p=0.06).

Conclusions—There is poor correlation between clinical infertility diagnoses and strict criteria. Diagnoses with objective criteria showed higher correlation than those with subjective criteria indicating variability in a clinician's diagnosis. Success rates in some diagnostic categories changed markedly when strict criteria were applied. Patients with multiple diagnoses may have lower success. Accurate infertility diagnoses are important to provide patients with accurate prognosis. Moreover,

© 2009 American Society for Reproductive Medicine. Published by Elsevier Inc. All rights reserved.

Corresponding Author: Thomas A. Molinaro MD, 3701 Market Street, Suite 800, Philadelphia, PA 19104, (p) 215-662-2971, (o) 215-349-5512, tmolinaro@obgyn.upenn.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Disclosures

None

Capsule

There is poor agreement between clinical infertility diagnoses entered in the SART registry and strict criteria which may affect the accuracy of success rates reported.

lack of precision in underlying diagnosis may affect the validity of past and future research using administrative datasets including SART.

Narrative Abstract—A clinical diagnosis of infertility may not agree with strict criteria based on recent review of medical literature. Standardized definitions of diagnostic categories are essential for accurate patient prognosis and future research.

INTRODUCTION

Prior to attempting In Vitro Fertilization (IVF), patients almost always express a desire to know their chance of success. The Society for Assisted Reproductive Technology (SART) makes publicly available the self reported data of participating IVF clinics throughout the United States. Patients are able to access these data via the SART website (www.sart.org) and find specific success rates for their age group and diagnostic category. Clinicians often refer to these rates when counseling patients on their prognosis for pregnancy. Previous authors have demonstrated that age and infertility diagnosis are strong predictors of ultimate success (1,2). In one population based study, older patients were found to be more likely to have “unexplained” and tubal factor infertility, while younger women are more likely to have ovulatory dysfunction or endometriosis(3). Secondary infertility has also been associated with an increased chance of becoming pregnant with IVF(4).

Each participating SART clinic defines the specific criteria for infertility diagnoses given to their patients. In clinical practice, patients may be given one diagnosis when in fact they do not meet the strict criteria for a specific condition. Correct characterization of a patient’s etiology is essential to provide them with their true prognosis for achieving pregnancy. Furthermore, some authors have questioned whether different clinics can be appropriately compared to each other because of differences in populations, number of cycles and methods to determine diagnoses. (5). We hypothesize that the clinical criteria by which many of these diagnoses are made affects the prognostic value of the success rate quoted to patients.

MATERIALS AND METHODS

Current criteria for specific infertility diagnoses that form SART diagnostic categories were reviewed in recent medical literature. Special emphasis was given to position statements from ASRM and ESHRE as well as systematic reviews which analyze the breadth of studies available. Objective criteria for each diagnostic category were determined. IVF patients enrolled for other studies at the University of Pennsylvania between December 2003 and June 2006 had their clinical records reviewed, abstracted and their infertility diagnosis adjudicated according to these strict criteria by trained personnel. Couples were permitted to have multiple diagnoses as long as they met the minimum criteria in each category. Institutional Review Board approval was obtained prior to chart abstraction.

Adjudicated “strict” diagnoses were then compared to clinical as entered into SART. The degree of agreement between clinical and “strict” diagnoses was calculated using Kappa statistics for specific diagnostic categories and evaluated according to the method of Landis, et al (6). Clinical pregnancy rates per transfer were calculated for each stratum and compared using generalized estimating equations models, an extension of logistic regression, which adjusts for repeated measures per subject. All calculations were performed using Stata v.10, College Station, Texas. This study was designed to assess agreement between diagnostic criteria and was not powered to assess for differences between pregnancy rates of the two groups.

RESULTS

Charts for 590 patients were adjudicated according to strict criteria. Live birth rates for each diagnostic criterion are presented in Table 1. The degree of agreement, represented by Kappa coefficients, between clinical and strict diagnoses was poorest among patients with diagnosis of uterine factor and diminished ovarian reserve. Strict criteria for unexplained infertility and PCOS showed slightly improved agreement with clinical criteria. While there was moderate agreement for diagnoses of endometriosis, tubal factor and Male factor, there remained 20% or greater discordance between clinical and strict diagnoses.

There was at least a 3% absolute change in pregnancy rate for every diagnostic criterion when strict and clinical criteria were compared. When pregnancy rates were calculated for each diagnostic category, success rates changed by more than 15 percent for patients with uterine factor, unexplained infertility and diminished ovarian reserve. Pregnancy rates decreased when strict criteria were applied for most diagnostic categories with the exception of diminished ovarian reserve. Patients with multiple factors were less likely to achieve a pregnancy regardless criteria of which were applied; however their likelihood of pregnancy was even lower with adjudicated diagnoses. By strict criteria, these patients were significantly less likely to have a live birth than those with a single diagnosis (OR=0.61, $p=0.019$). This finding was similar with clinical criteria (OR=0.68, $p=0.06$).

CONCLUSIONS

These data provide evidence that there is poor agreement between clinical infertility diagnoses and evidence-based, strict infertility diagnosis. Diagnoses with objective criteria showed higher agreement than those with subjective criteria indicating variability in a clinician's diagnosis. Furthermore, success rates in some diagnostic categories changed markedly when strict criteria were applied. With the exception of diminished ovarian reserve, success rates dropped in all other categories. This discrepancy with DOR patients might reflect that isolated diminished ovarian reserve is actually rare and that it may not carry the same implications as DOR associated with increased age or endometriosis. Furthermore, it is important to note that patients with multiple diagnoses may have lower success than those with a single diagnosis. Given such wide variation in pregnancy rates between clinical and adjudicated diagnoses, we feel it is therefore imperative that clinicians make the most accurate diagnosis when providing their patients with an estimate of their probability of achieving pregnancy.

Previous studies have examined the prognosis for pregnancy associated with specific infertility diagnoses such as tubal factor, endometriosis and PCOS (7-10). Our results may help explain apparent inconsistencies of studies in the literature. According to SART and previous studies, endometriosis patients have no difference in IVF success compared with other groups(12). However, there are conflicting studies which suggest that endometriosis may be associated with a lower chance of success. A meta analysis published from our group confirmed that these patients have a lower chance of pregnancy in IVF and that more severe forms of endometriosis resulted in lower success(9). The discrepancy between previous studies may be due to differences in how endometriosis was diagnosed or coded in the SART database. While our own small sample did not allow for subdivision of endometriosis into minimal, mild, moderate and severe subcategories, instituting these criteria into SART may prove useful in determining patient specific prognosis. Furthermore, lack of precision in underlying diagnosis may affect the validity of past and future research. It is essential that investigators work towards a standardization of diagnostic criteria for all infertility diagnoses in the manner that the Rotterdam Conferences standardized the diagnosis of PCOS(13).

When examining SART clinic specific success rates, it is important to examine diagnosis specific rates(11). The current SART database does not establish specific criteria for each of these diagnoses, but merely offers diagnostic guidelines to participating clinics. In order to accurately compare success rates between clinics, standardization of these criteria are necessary. Careful and critical inspection of published studies in a systematic review of the literature is called for in determining which criteria have the most evidence for affecting outcome. Consensus statements such as the Rotterdam Criteria for PCOS are particularly helpful in bringing together experts in the field to establish definitive criteria for specific diagnoses. We have demonstrated how merely standardizing these criteria in our own practice affected diagnosis specific prognosis. As patients also look at these success rates in order to choose a clinic, standardized and accurate reporting becomes more important. Accurate infertility diagnoses are important to provide patients with accurate prognosis and help them in deciding how and where to best pursue fertility.

REFERENCES

1. Stolwijk AM, Wetzels AM, Braat DD. Cumulative probability of achieving an ongoing pregnancy after in-vitro fertilization and intracytoplasmic sperm injection according to a woman's age, subfertility diagnosis and primary or secondary subfertility. *Hum Reprod* 2000;15:203–9. [PubMed: 10611213]
2. Lintsen AME, Eijkemans MJC, Hunault CC, Bouwmans CAM, Hakkaart L, Habbema JDF, et al. Predicting ongoing pregnancy chances after IVF and ICSI: a national prospective study. *Hum Reprod* 2007;22:2455–62. [PubMed: 17636281]
3. Maheshwari A, Hamilton M, Bhattacharya S. Effect of female age on the diagnostic categories of infertility. *Hum Reprod* 2008;23:538–42. [PubMed: 18308834]
4. Kupka M, Dorn C, Richter O, Felberbaum R, van der Ven H. Impact of reproductive history on in vitro fertilization and intracytoplasmic sperm injection outcome: evidence from the German IVF Registry. *Fertil Steril* 2003;80:508–16. [PubMed: 12969690]
5. Garcia JE. Panel two: reporting and advertising success rates--the Gordian Knot of assisted reproductive technology. The role of professional societies. *Womens Health Issues* 1997;7:188–92. [PubMed: 9203867]discussion 94–6
6. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74. [PubMed: 843571]
7. Lintsen AM, Eijkemans MJ, Hunault CC, Bouwmans CA, Hakkaart L, Habbema JD, et al. Predicting ongoing pregnancy chances after IVF and ICSI: a national prospective study. *Hum Reprod* 2007;22:2455–62. [PubMed: 17636281]
8. Grochowski D, Kulikowski M, Wolczynski S, Kuczynski W, Szamatowicz M. The outcome of an in vitro fertilization program in women with polycystic ovary syndrome. *Gynecol Endocrinol* 1997;11:259–62. [PubMed: 9272422]
9. Barnhart K, Dunsmoor-Su R, Coutifaris C. Effect of endometriosis on in vitro fertilization. *Fertility and Sterility* 2002;77:1148–55. [PubMed: 12057720]
10. Witsenburg C, Dieben S, Van der Westerlaken L, Verburg H, Naaktgeboren N. Cumulative live birth rates in cohorts of patients treated with in vitro fertilization or intracytoplasmic sperm injection. *Fertil Steril* 2005;84:99–107. [PubMed: 16009164]
11. Society for Assisted Reproductive Technology. National Success Rates. 2006. www.sart.org
12. Calhaz-Jorge C, Chaveiro E, Nunes J, Costa AP. Implications of the diagnosis of endometriosis on the success of infertility treatment. *Clin Exp Obstet Gynecol* 2004;31:25–30. [PubMed: 14998182]
13. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. *Fertil Steril* 2004;81:19–25. Rotterdam EA-SPCWG
14. American Society for Reproductive M. Revised American Society for Reproductive Medicine classification of endometriosis: 1996. *Fertility and Sterility* 1997;67:817–21. [PubMed: 9130884]
15. Marcoux S, Maheux R, Berube S, The Canadian Collaborative Group on E. Laparoscopic Surgery in Infertile Women with Minimal or Mild Endometriosis. *N Engl J Med* 1997;337:217–22. [PubMed: 9227926]

16. Pouly JL, Chapron C, Manhes H, Canis M, Wattiez A, Bruhat MA. Multifactorial analysis of fertility after conservative laparoscopic treatment of ectopic pregnancy in a series of 223 patients. *Fertil Steril* 1991;56:453–60. [PubMed: 1832649]
17. Mol BWJ, Collins JA, Burrows EA, van der Veen F, Bossuyt PMM. Comparison of hysterosalpingography and laparoscopy in predicting fertility outcome. *Hum Reprod* 1999;14:1237–42. [PubMed: 10325270]
18. Hulka JF. Adnexal adhesions: a prognostic staging and classification system based on a five-year survey of fertility surgery results at Chapel Hill, North Carolina. *Am J Obstet Gynecol* 1982;144:141–8. [PubMed: 6214187]
19. Strandell A, Lindhard A, Waldenström U, Thorburn J, Janson PO, Hamberger L. Hydrosalpinx and IVF outcome: a prospective, randomized multicentre trial in Scandinavia on salpingectomy prior to IVF. *Hum Reprod* 1999;14:2762–9. [PubMed: 10548619]
20. Jonard S, Robert Y, Cortet-Rudelli C, Pigny P, Decanter C, Dewailly D. Ultrasound examination of polycystic ovaries: is it worth counting the follicles? *Hum Reprod* 2003;18:598–603. [PubMed: 12615832]
21. Vermeulen A, Verdonck L, Kaufman JM. A critical evaluation of simple methods for the estimation of free testosterone in serum. *J Clin Endocrinol Metab* 1999;84:3666–72. [PubMed: 10523012]
22. Practice Committee of the American Society for Reproductive M. Aging and infertility in women. *Fertil Steril* 2006;86:S248–52. [PubMed: 17055834]
23. Smotrich DB, Widra EA, Gindoff PR, Levy MJ, Hall JL, Stillman RJ. Prognostic value of day 3 estradiol on in vitro fertilization outcome. *Fertil Steril* 1995;64:1136–40. [PubMed: 7589666]
24. Barnhart K, Osheroff J. Follicle stimulating hormone as a predictor of fertility. *Curr Opin Obstet Gynecol* 1998;10:227–32. [PubMed: 9619346]
25. 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. [PubMed: 11756373]
26. Scott RT, Opsahl MS, Leonardi MR, Neall GS, Illions EH, Navot D. Life table analysis of pregnancy rates in a general infertility population relative to ovarian reserve and patient age. *Hum Reprod* 1995;10:1706–10. [PubMed: 8582965]
27. Frattarelli JL, Levi AJ, Miller BT, Segars JH. A prospective assessment of the predictive value of basal antral follicles in in vitro fertilization cycles. *Fertil Steril* 2003;80:350–5. [PubMed: 12909498]
28. Bukulmez O, Arici A. Assessment of ovarian reserve. *Curr Opin Obstet Gynecol* 2004;16:231–7. [PubMed: 15129052]
29. Organization WH. WHO Laboratory Manual for the Examination of Human Semen and Sperm-Cervical Mucus Interaction. Vol. 34th ed. Cambridge University Press; Cambridge: 1999.
30. Guzick DS, Overstreet JW, Factor-Litvak P, Brazil CK, Nakajima ST, Coutifaris C, et al. Sperm morphology, motility, and concentration in fertile and infertile men. *N Engl J Med* 2001;345:1388–93. [PubMed: 11794171]
31. Guzick DS, Sullivan MW, Adamson GD, Cedars MI, Falk RJ, Peterson EP, et al. Efficacy of treatment for unexplained infertility. *Fertil Steril* 1998;70:207–13. [PubMed: 9696208]
32. Guzick DS, Carson SA, Coutifaris C, Overstreet JW, Factor-Litvak P, Steinkampf MP, et al. National Cooperative Reproductive Medicine Network. Efficacy of superovulation and intrauterine insemination in the treatment of infertility. *N Engl J Med* 1999;340:177–83. [PubMed: 9895397]
33. Practice Committee of the American Society for Reproductive M. Effectiveness and treatment for unexplained infertility. *Fertil Steril* 2006;86:S111–4. [PubMed: 17055802]
34. Athaullah N, Proctor M, Johnson NP. Oral versus injectable ovulation induction agents for unexplained subfertility. *Cochrane Database Syst Rev*. 2002CD003052
35. Pandian Z, Bhattacharya S, Nikolaou D, Vale L, Templeton A. The effectiveness of IVF in unexplained infertility: a systematic Cochrane review. 2002. *Hum Reprod* 2003;18:2001–7. [PubMed: 14507813]

Table 1**Strict Diagnostic Criteria**

Diagnosis	Criteria
Endometriosis	• Scores based on Revised ASRM Classification of Endometriosis (14,15)
Tubal Factor	• Bilateral tubal occlusion on HSG or chromotubation • H/o BTL or bilateral salpingectomy • Unilateral or bilateral hydrosalpinx • H/o ectopic pregnancy (excluding IVF) (16-19)
PCOS (2 out of 3 criteria)	• Oligo- or anovulation $\Rightarrow$ intermenstrual periods ≥ 45 days OR ≤ 8 menses/yr • Clinical or biochemical hyperandrogenism • Polycystic ovaries (≥ 12 follicles per ovary, 2-9mm) • Exclusion of other etiologies (13,20,21)
Diminished Ovarian Reserve	• Day 3 FSH ≥ 11.4 mIU/ml • Day 3 E2 ≥ 80 pg/ml • Age ≤ 37 • Total antral follicle count ≤ 4 (22-28)
Male Factor	• Count < 13.5 mil/mL • Motility < 32 % • Morphology < 9 % (29,30)
Unexplained	• At least one tube patent with no tubal abnormalities • Age ≤ 37 • Normal ovarian reserve • Normal semen analysis (31-35)
Uterine	• Presence of uterine fibroids/Adenomyosis