Published in final edited form as:

Fertil Steril. 2011 June ; 95(7): 2279–2282. doi:10.1016/j.fertnstert.2011.02.055.

Proficiency in oocyte retrieval: how many procedures are necessary for training?

Kara N Goldman, MD^{a,b,*}, Kimberly S Moon, MD^{a,*}, Belinda J Yaeger, MD^a, Mark D Payson, MD^{a,c}, James H Segars, MD^a, and Barbara J Stegmann, MD, MPH^{d,1}

^aProgram of Reproductive and Adult Endocrinology, National Institute of Child Health and Human Development, Building 10, CRC, Rm 1E-3140, 10 Center Drive, National Institutes of Health, Bethesda, Maryland 20892

^bDepartment of Obstetrics and Gynecology, Northwestern University, Prentice Women's Hospital, 250 E. Superior St. Chicago, Illinois 60611

^cART Program, Division of Reproductive Endocrinology and Infertility, Walter Reed Army Medical Center, 6900 Georgia Avenue NW, Washington, DC 20307

^dDepartment of Obstetrics and Gynecology, Division of Reproductive Endocrinology and Infertility, University of Iowa Hospitals and Clinics, 200 Hawkins Drive, 31324 PFP, Iowa City, Iowa 52246.

Abstract

Objective—To determine a minimum number of procedures required for proficiency in oocyte retrieval and to characterize skill acquisition.

Design—Retrospective analysis.

Setting—Reproductive Endocrinology and Infertility fellowship training program.

Study subjects—Fellows in training from 2005-2007 and 2008-2010.

Interventions—None.

Main outcome measures—Proficiency scores (PS) were calculated by dividing the number of oocytes retrieved by oocytes expected (PS=oocytes retrieved/expected). 2005-2007 trainees were grouped and proficiency scores calculated as an average during specific points in their training. 2008-2010 trainees were compared individually to determine differences in individual skill acquisition.

Results—A greater number of oocytes were retrieved than expected, 8.6 vs. 7.6 ($p < 0.05$). A relatively lower proficiency score (PS) was noted during the first 10 trainee aspirations (PS=1.1) compared to subsequent aspirations (PS=1.25 retrievals 11-20, PS=1.21 retrievals 21-30 and >31). When individual fellows' scores were calculated, the majority achieved proficiency by 20 aspirations and all but one trainee achieved the mean staff PS by 50 retrievals.

Conclusions—Regardless of a trainee's initial proficiency in oocyte retrieval, there are no statistically significant differences in the learning curve between trainees. The majority of individual fellows in training demonstrate proficiency in follicular aspirations within 20 procedures; however, a minority may require 50 procedures to achieve the proficiency of an attending physician.

¹Corresponding author for reprints 319-356-3143 (office), 319-384-9367 (fax). Barbara-Stegmann@uiowa.edu. .

*The first two authors contributed equally to this manuscript

Capsule—The majority of fellows in training demonstrate proficiency in follicular aspirations within 20 procedures; however, a minority may require 50 procedures to achieve the proficiency of an attending physician.

Keywords

Follicular aspiration; fellowship; competency; learning curve

INTRODUCTION

The traditional apprenticeship model of surgical training is evolving into an integrative model valuing formal surgical education in the form of didactic sessions, bench laboratories, and animal laboratories conducted outside of the operating room (1,2). Despite implementation of formal surgical curricula in many training programs, evaluation of surgical trainees remains largely informal and subjective. A survey of residency programs in obstetrics and gynecology indicated that only 74% of programs formally evaluate surgical skills, with the majority reporting subjective faculty evaluations as the primary method of assessment (3). There is no established method to assess when a trainee achieves proficiency in performing a specific procedure or operation. Surgical educators may utilize structured assessments including the objective structured assessment of technical skills (OSATS), the observational clinical human reliability assessment (OCHRA), or the operative performance rating system (OPRS) to assess a trainee's competency in a specific procedure or operation, but such tools are not yet universally accepted or utilized (4-7). Chou et al (8) developed and validated the Hopkins Assessment of Surgical Competency (HASC) to evaluate resident surgical competency, but this scale relied on faculty observation and resident self-assessment rather than measurable outcomes. The evaluation of technical skills is challenging without objective measures, and assessment relies on observation by a supervisor or senior faculty member. Most methods of surgical assessment are time-intensive and lack objective criteria necessary to deem a trainee 'proficient' in a particular skill.

Surgical trainees are expected to learn by experience under direct supervision, but few standards exist to ensure that a surgeon is proficient prior to practicing without supervision. For example, breast surgeons recently published the number of breast ductal endoscopy procedures required to designate a trainee 'proficient' at this task. The authors determined that 50% of trainees had achieved proficiency at the 13th procedure or earlier, but they noted that their study lacked a precise definition of success or failure. Proficiency was determined by the ability to perform four consecutively successful ductoscopies, but only subjective accounts of proficiency were documented. No objective criteria of success were used to deem a trainee 'proficient' (9).

In procedures with measurable outcomes, such outcomes can be used to assess clinical proficiency. Fellowship in reproductive endocrinology and infertility trains physicians in assisted reproductive technologies (ART). Multiple studies have suggested a minimum number of procedures required to obtain competency in embryo transfer, the final step involved in ART (10-12). In one such study, success of embryo transfer was measured by clinical pregnancy rates, and this objective measure was used to identify a minimum number of procedures required for proficiency (10). Oocyte retrieval is an important step in ART with similarly measurable outcomes. Practice standards require that physicians perform greater than twenty follicular aspirations under direct supervision prior to independent practice (13); however, to our knowledge no data exist to support this minimum number of retrievals. The number of oocytes retrieved can be used as a marker for proficiency. The purpose of this study was to use objective criteria to determine a minimum number of oocyte

retrievals required for proficiency in order to answer the following two questions: how many oocyte retrievals are sufficient for proficiency, and are there provider-dependent differences in the learning curve?

MATERIALS AND METHODS

Two distinct retrospective analyses examined trainees in the Federal Reproductive Endocrinology and Infertility (REI) Fellowship over 2 two-year periods from July 2005 to June 2007 and from October 2008 to September 2010. All information was extracted from an existing database and the number of mature follicles was counted prior to knowing the number of oocytes retrieved. Institutional review and approval was obtained from the Office of Human Subjects Research at the National Institutes of Health.

Inclusion criteria included fellows in training at the National Institutes of Health (NIH) Federal training program in REI from July 2005 to June 2007 and from October 2008 to September 2010 performing oocyte retrieval procedures at Walter Reed Army Medical Center (WRAMC). Aspiration procedures were performed by one faculty member and one fellow on contralateral ovaries when appropriate.

Exclusion criteria included greater than one provider attempting oocyte retrieval on the ipsilateral ovary, improperly recorded or missing data regarding fellow or attending provider, absence of a follicle scan on the day of trigger with human chorionic gonadotropin (hCG), and inadequate ultrasound visualization.

We approached the study question in two parts. In the first portion of the study, a group of four fellow trainees were compared to a group of seven attending physicians. Attending physicians were defined as active staff that had successfully completed a Reproductive Endocrinology Fellowship certified by the American Board of Obstetrics and Gynecology. A total of 1,349 sides were retrieved during the study period, with 1,209 meeting entry criteria. One hundred forty retrievals were excluded. Sixty retrievals were excluded because greater than one provider attempted retrieval on one ovary, 33 were excluded for improperly recorded information, 24 were excluded due to absence of one ovary, and 23 lacked a follicular scan on the day of hCG. Overall, attending physicians performed more retrievals than fellows. Attending physicians performed a total of 707 aspirations to be included in the analysis, and fellow trainees performed a total of 502 oocyte aspirations from July 2005 through June 2007. An attending physician retrieved follicles from one ovary, and when appropriate, a trainee retrieved follicles from the contralateral ovary. A proficiency score (PS) was determined by dividing the number of oocytes retrieved by the number of oocytes expected based on all ultrasound-visualized follicles ≥ 12 mm ($PS = \text{oocytes aspirated} / \text{expected}$). Trainees were grouped by their first ten aspiration attempts (1-10), their second ten attempts (11-20), their third ten attempts (21-30), and their remaining attempts (>31), and results were pooled for all trainees (Table 1). No individual statistics were calculated. All attending retrieval attempts were grouped and averaged and a proficiency score was calculated.

In the second part of the study, 634 oocyte retrievals were performed by six trainees during their first year on the IVF service between October 1, 2008 and September 30, 2010. Proficiency scores (PS) were calculated for individual trainees. Attending physicians performed a total of 417 aspirations during this same time period and these were used to calculate the target proficiency score. A total of 1,084 retrievals were performed during this time period. For this portion of the study, the first ovary could be retrieved by a third year fellow instead of an attending; however, retrievals by the third year fellows were not included in the analysis. Four hundred and fifty retrievals were excluded, 318 because no

first year fellow participated in the retrieval (the retrieval was performed by a third year fellow/attending pairing), 68 had improperly recorded information, 40 did not have a scan on the day of hCG trigger, and 24 had a single ovary.

Proficiency scores (PS) for fellow trainees were again calculated by dividing the number of oocytes retrieved by the number of oocytes predicted based on the total number of follicles ≥ 12 mm measured by ultrasound on the day of hCG trigger. Proficiency scores were used as the outcome of interest in linear regression modeling to generate “learning curves” for each individual over time, and scores for the six individual trainees were compared to the mean proficiency score of eight experienced staff in the same time period.

Oocyte retrieval

Trainees performed oocyte retrievals under the direct supervision of an attending physician. Patients underwent controlled ovarian stimulation using a combination of long-term GnRH agonist (Lupron, 1.0 mg/d; TAP Pharmaceuticals, Deerfield, IL) or microdose flare GnRH agonist (Lupron 20 mcg/20U or 50mcg/20U) and gonadotropins (Fertinex; Sero Laboratories, Norwell, MA), or a combination of Fertinex and Humegon (Organon, West Orange, NJ) as described (14). Cycles were monitored by performing serial follicular ultrasonography and measuring serum estradiol levels. Administration of human chorionic gonadotropin (hCG) and oocyte retrieval was performed when follicle size exceeded 16 mm-18 mm as described (15,16).

Statistical analysis

Data were analyzed retrospectively, and sample size was defined by the study interval and number of trainees enrolled. Aspiration attempts were divided into groups of ten based on the Society for Assisted Reproductive Technology (SART) practice committee guidelines suggesting a minimum requirement of twenty follicular aspirations for proficiency (13). All calculations were completed using Stata 9.2 (College Station, TX).

For the first portion of the study, the main outcome variable was percentage of oocytes retrieved. One-way ANOVA and Student's t-test with unequal variance were used for statistical analysis. An alpha (α) error of <0.05 was considered statistically significant.

For the second portion of the study, proficiency scores for individual fellow trainees were compared to the mean proficiency score for attending physicians. Linear regression was used to graph the rate of change for individual trainees over the first 80 retrievals, with the PS being the outcome of interest. The PS goal was the average proficiency score of the attending physicians. The fellow was considered to be “proficient” when their PS reached the average PS of the staff.

RESULTS

During the first part of the study, 502 oocyte retrievals were performed by four fellow trainees and 707 retrievals were performed by seven attending physicians from July 2005 through June 2007. The number of retrievals per fellow ranged from 88 to 159, and the number of retrievals per attending ranged from 7 to 249. Trainees' oocyte retrieval attempts were grouped by their first ten aspiration attempts (1-10), their second ten aspirations (11-20), their third ten aspirations (21-30), and all subsequent aspirations (>31). In all groups regardless of training status, the number of oocytes retrieved was greater than the number expected, 8.6 vs. 7.6 (p -value <0.05). A proficiency score (PS) was designated by dividing the number of oocytes retrieved by the number expected based on all visualized follicles ≥ 12 mm ($PS = \text{oocytes retrieved/expected}$). A relatively lower proficiency score was noted during the first ten trainee aspirations ($PS=1.1$) compared to the subsequent trainee

aspirations (PS=1.25 for retrievals 11-20, PS=1.21 for retrievals 21-30, and PS=1.21 for all retrievals over 30) and compared to the proficiency score for attending physicians (PS=1.21). For trainees, the proficiency score did not vary significantly over the training period (Fig. 1). There were no statistically significant differences between any of these groups when compared to the mean percentage of oocytes retrieved by attending physicians.

The intent of the second portion of the study was to assess possible individual differences in skill acquisition between individual trainees. Oocyte retrieval data were collected for six fellow trainees between 2008 and 2010 during their first year of clinical ART training. Fellows performed between 25-90 retrieval procedures. Eight experienced staff performed a total of 417 oocyte retrieval procedures during the same time period, and the staff physicians' scores were averaged to calculate a mean proficiency score which would be used as the 'proficiency target' for trainees. As in the first portion of the study, a greater number of oocytes were retrieved than were predicted based on the number of follicles ≥ 12 mm visualized on ultrasound. The average staff proficiency score was 1.07. When individual fellows' proficiency scores were calculated, 2 fellows scored above the attendings' score at the first retrieval, 2 fellows achieved the attendings' PS by 20 retrievals, 1 fellow achieved the proficiency target within 50 retrievals, and 1 fellow did not achieve the proficiency target (Fig. 2). Fellow 4 reached the proficiency target at the first retrieval attempt (PS=1.16) and the proficiency score subsequently fell with each aspiration attempt, but this fellow performed only 25 retrieval procedures. Fellow 5 did not reach the proficiency target but maintained a stable PS throughout all retrievals (PS=0.98). There was no statistical difference in the learning curve between fellows, and there was no statistical difference between the average proficiency score of the fellows and the attendings ($p=0.84$).

DISCUSSION

The first portion of the study aimed to determine a minimum number of oocyte retrieval procedures necessary for proficiency in a group of trainees. During the first 10 oocyte retrievals performed by trainees, the proficiency score (PS=oocytes retrieved/expected) was slightly less than for all subsequent attempts. While not statistically significant, this difference demonstrated that a learning curve was present for the group of trainees during the first 10 procedures. The remaining retrievals performed by trainees after the 10th attempt were comparable to the proficiency score of experienced attending physicians. This confirms the previous assumption that early proficiency is obtained for the majority of trainees in the skill of follicular aspiration. The data from this portion of the study support the currently enforced SART practice committee guidelines requiring that trainees perform a minimum of 20 follicular aspirations under direct supervision (13).

We next characterized the learning curve of individual trainees, reasoning that skill acquisition could vary between individuals. When individual fellows' proficiency scores were calculated, 2 fellows scored above the attendings' score at the first retrieval, 2 fellows achieved the attendings' PS by the 20th retrieval, 1 fellow achieved the proficiency target within 50 retrievals, and 1 fellow did not achieve the proficiency target. Fellow 4 reached the proficiency target with the first retrieval attempt (PS=1.16) but the proficiency score subsequently fell with each aspiration attempt. This fellow performed only 25 retrieval procedures and may have required additional procedures to again approach the proficiency target. The fellow who never achieved the proficiency target maintained a stable proficiency score throughout all procedures, and there was no statistically significant difference between this proficiency score and that of the other trainees or staff. Four of six trainees achieved the proficiency target by 20 retrievals, and one fellow by 50 retrievals. While this evidence largely supports the currently enforced SART practice guidelines, the data suggest that

individual trainees may benefit from greater than the currently recommended 20 oocyte retrievals to ensure competence.

We were unable to identify published reports detailing training experiences in oocyte retrieval. Early proficiency in oocyte retrieval has been suggested by practice standards requiring trainees to perform twenty follicular aspirations for proficiency (13), but to our knowledge there are no data to establish a minimum number of oocyte retrievals required for trainees. We sought to determine if there were a minimum number of procedures required for proficiency.

Limitations of this study include the relatively small sample size. The training program analyzed in this study trains 3 fellows annually, greater than any other REI fellowship training program in the United States, but a multi-centered trial including a greater numbers of trainees could be considered to better answer the questions posed in this study.

An additional limitation is that experienced staff may have chosen to perform the oocyte retrieval procedure on the more difficult ovary allowing the trainee to perform the less challenging retrieval. The attending proficiency score may be falsely lowered by this confounding factor.

In conclusion, oocyte retrieval is an important skill that is quickly mastered by fellows in training. Regardless of a trainee's initial proficiency when learning to perform oocyte retrieval, there were no statistically significant differences in the learning curve between trainees. The majority of fellows in training demonstrate proficiency in follicular aspirations within 20 procedures; however, a minority may require 50 procedures to achieve the proficiency of an attending.

Acknowledgments

The authors gratefully acknowledge and thank Dr. Alan DeCherney, the laboratory staff at the National Institutes of Health, Dr. Melinda Henne, Dr. Aidita N. James, and the fellows in training in Reproductive Endocrinology and Infertility at the National Institutes of Health.

Financial Support: This research was supported, in part, by The Iowa Women's Reproductive Health Research Career Development Award (K12 HD063117)

References

1. Goff BA. Changing the Paradigm in Surgical Education. *Obstet Gynecol.* 2008; 112(2):328–332. [PubMed: 18669731]
2. Fenner DE. Training of a Gynecologic Surgeon. *Obstet Gynecol.* 2005; 105(1):193–196. [PubMed: 15625162]
3. Mandel LP, Lentz GM, Goff BA. Teaching and evaluating surgical skills. *Obstet Gynecol.* 2000; 95(5):783–785. [PubMed: 10775747]
4. Distlehorst, LH.; Dunnington, GL.; Folse, JR. Teaching and Learning in Medical and Surgical Education: Lessons Learned for the 21st Century. Lawrence Erlbaum Associates, Inc.; New Jersey: 2000. p. 241-242.
5. Talebpour M, Alijani A, Hanna GB, Moosa Z, Tang B, Cuschieri A. Proficiency-gain curve for an advanced laparoscopic procedure defined by observation clinical human reliability assessment (OCHRA). *Surg Endosc.* 2009; 23:869–875. [PubMed: 18810544]
6. Larson JL, Williams RG, Ketchum J, Boehler ML, Dunnington GL. Feasibility, reliability and validity of an operative performance rating system for evaluating surgery residents. *Surgery.* 2005; 138(4):640–649. [PubMed: 16269292]

7. Goff BA, Lentz DL, Houmard B, Mandel LS. Development of an objective structured assessment of technical skills for obstetric and gynecology residents. *Obstet Gynecol.* 2000; 96(1):146–150. [PubMed: 10862857]
8. Chou B, Bowen CW, Handa VL. Evaluating the competency of gynecology residents in the operating room: validation of a new assessment tool. *Am J Obstet Gynecol.* 2008; 199:571.e1–571.e5. [PubMed: 18799153]
9. Zagouri F, Sergeantanis T, Giannakopoulou G, Panopoulou E, Chrysikos D, Bletsas G, et al. Breast ductal endoscopy: how many procedures qualify? *BMC Research Notes.* 2009; 2:115. [PubMed: 19566939]
10. Papageorgiou TC, Hearn-Stokes RM, Leondires MP, Miller BT, Chakraborty P, Cruess D, et al. Training of providers in embryo transfer: what is the minimum number of transfers required for proficiency? *Hum Reprod.* 2001; 16(7):1415–1419. [PubMed: 11425822]
11. Wittenberger MD, Catherino WH, Armstrong AY. Role of embryo transfer in fellowship training. *Fertil Steril.* 2007; 88:1014–1015. [PubMed: 17412339]
12. Dessolle L, Fréour T, Barrière P, Jean M, Ravel C, Daraï E, et al. How soon can I be proficient in embryo transfer? Lessons from the cumulative summation test for learning curve (LC-CUSUM). *Hum Reprod.* 2010; 25:380–386. [PubMed: 19923166]
13. The Practice Committee of the Society for Assisted Reproductive Technology and the Practice Committee of the American Society for Reproductive Medicine. Revised minimum standards for practices offering assisted reproductive technologies. *Fertil Steril.* 2008; 90:S165–S168. [PubMed: 19007618]
14. Leondires M, Escalpes M, Segars J, Scott RT Jr, Miller BT. Microdose follicular phase gonadotropin-releasing hormone agonists (GnRH-a) compared with luteal phase GnRH-a for ovarian stimulation at in vitro fertilization. *Fertil Steril.* 1999; 72(6):1018–1023. [PubMed: 10593374]
15. Wittmaack FM, Kreger DO, Blasco L, Tureck RW, Mastroianni L Jr, Lessey BA. Effect of follicular size on oocyte retrieval, fertilization, cleavage, and embryo quality in in vitro fertilization cycles: a 6-year data collection. *Fertil Steril.* 1994; 62(6):1205–10. [PubMed: 7957985]
16. Andersen CY. Characteristics of Human Follicular Fluid Associated with Successful Conception after in vitro Fertilization. *J Clin Endocrinol Metab.* 1993; 77(5):1227–1234. [PubMed: 7521343]

Figure 1.

Y-axis: Proficiency score (PS): oocytes retrieved divided by oocytes expected for follicles $\geq 12\text{mm}$.

X-axis: Aspiration attempts. All attending retrievals were grouped and one proficiency score (PS) designated. Fellows' retrieval attempts were pooled and divided in increments of 10: Retrieval attempt 1-10, attempt 11-20, attempt 21-30, and all attempts >31 .

Figure 2.

Y-axis: Proficiency score (PS): oocytes retrieved divided by oocytes expected for follicles ≥ 12 mm.

X-Axis: Aspiration attempts. All attending retrievals were grouped and one proficiency score (PS) calculated (PS=1.07) to be used as the proficiency target. Proficiency scores (PS) were calculated for individual fellows (fellows 1-6) for the first 80 oocyte retrieval attempts. Linear regression was used to graph the rate of change for individual fellows over all attempts.

Table 1

Provider Characteristics and Proficiency Scores

PS= Proficiency score. Number of oocytes retrieved divided by number of oocytes expected based on ultrasound-visualized follicles ≥ 12 mm on day of hCG trigger.

Group	Provider level of training	No. of providers	Aspiration attempts	No. of retrievals ^a	PS ^b (%)	Std. dev.
1	Attending	7	707	707	1.206	0.666
2	Fellow ^c	4	1-10	40	1.101	0.565
3	Fellow	4	11-20	40	1.253	0.622
4	Fellow	4	21-30	40	1.214	0.499
5	Fellow	4	>31	382	1.205	0.539
Total		11	1209	1209		

^aNumber of retrievals per attending ranged from 7 to 249. Number of retrievals per fellow ranged from 88 to 159.

^bOne-way ANOVA and Student's t-test with unequal variance were used for statistical analysis. No statistically significant differences were noted between groups.

^cData for four fellows were pooled and no individual statistics were calculated.