What Is the Best Strategy for Presenting ART Results? A Controversial Comment

Doctors love conventions. They have them in exotic places and they frequently use them to defend what would otherwise be inexplicable calculations. Thus 38 weeks of true gestation is, of course, 40 weeks of amenorrhoea, which may also be viewed as 38–40 weeks of pregnancy if timing is based on a “conventional” menstrual cycle. Even if this leaves the outside world to wonder, the clinicians understand. Many of the proponents of such ambiguity have moved away from obstetrics into reproductive medicine but have carried with them the propensity for such double-speak. As yet, however, they have been unable to agree on a conventional, or more uniform, way of producing their ART results. These

authors are also unsure whether many understand their own data, but in their defense it appears to be more a matter of confusion than collusion.

There are really two issues. The first relates simply to the ways of presenting data, both for publication and for patient information, and the second, perhaps more important, is defining the variables inherent in arriving at the raw data to allow comparison between clinics. A graphic example of the latter was seen in a review by the scientific subgroup of the Fertility Society of Australia (Scientists in Reproductive Technology; SIRT) of methods used in Australian laboratories. Among 36 units, there were 13 types of oocyte collection needles, with 13 per-

forming flushing of aspirated follicles; 6 different gaseous atmospheres in which oocytes were cultured, with some under oil; 6 different culture media, containing 6 variations of protein additives; 8 types of culture plasticware; 5 times of cryopreserving gametes/embryos; and 14 types of embryo transfer catheters attached to 5 different delivery systems, delivering volumes ranging between 10 and 40 μ l on days 1 through 5 after oocyte collection. One wonders whether this truly represents evidence-based medical science!

The survey also showed that four different assay systems were used to monitor peripheral circulating hormones, and although the number of different ultrasound machines was not known, a pilot study of all New South Wales clinics has demonstrated variation significantly worse (i.e., >10%) than that tolerated for a biochemistry laboratory (1). Furthermore, while all pathology laboratories in Australia are required to be accredited by the National Authority of Testing Agencies (NATA) and show documented evidence of participation in an external quality assurance scheme, these are directed primarily at estrogen concentrations of less than 1000 pM, where the dynamics of the assay system have been set by the manufacturers to meet the requirements of the majority of general pathology users. Concentrations of estradiol seen in assisted reproductive technology (ART) patients far exceed these, and depending upon the clinic, the concentration at which staff become concerned about potential ovarian hyperstimulation can range between 8000 and 20,000 pM. There is as yet no published comparative study between commercially available methods of determining high concentrations of estrogen, and while it appears from a pilot study of Australian laboratories that there is relative consistency, even a small variation at elevated concentrations can represent a large mass change. Couple with this the variation seen in performing follicular ultrasound and it is immediately obvious that the oocytes given to embryologists can come from hugely varying environments and thus have considerable theoretical differences in potential.

Some would argue that only a few of the above factors really contribute to the final pregnancy outcome since it is the embryology laboratory where this is really determined (i.e., when pregnancy rates "fall," where is the "blame" initially laid?). However, we are yet to be convinced that they do not play as significant a part as other aspects of the prepara-

tion of patients before the collection of oocytes. For example, by applying a successful weight loss program, improved outcomes were achieved before ART was undertaken (2), and more stressed patients have a "poorer biologic response" than those who are less stressed (3). Similarly some clinics may raise their success rates by offering ART procedures to patients who might have become pregnant with less technologically advanced treatment, not necessarily through any overt deception by the clinic, but because patients, bombarded by the mass media about the technology available, will not accept simpler, less expensive (and thus in their perception not as successful) alternatives. Patient selection undoubtedly effects results and can be easily controlled if the clinic director so decides (4).

Thus simple strategies can be used to improve ART rates for public scrutiny. An analysis of the 1996 national data from Australasia (5) shows a positive relationship between the pregnancy rate per 100 oocyte collections and the cancellation rate of patients before collection is attempted (Fig. 1). Notwithstanding the huge variability, this trend was also seen in previous annual data compilations. Those patients canceled do not adversely affect the results, as the clinics report pregnancies per retrieval, or beyond. It is also evident from the national data, as one would expect in any statistical analysis, that the lower the number of treatments a clinic performs, the greater the margin of error and the more variability in success rates from year to year. Good medical practice (i.e., cancellation if it is felt that things are not going well) may indeed be better for a unit's results, but what of other concerns for the patient? True cumulative pregnancy rates can be accurate only when this is known. In Australia the government will partially support up to six attempts of stimulated ART that go beyond 9 days of treatment. Beyond that time the patient has "lost" an attempt, so even if the chances of achieving a pregnancy are low, the tendency is to proceed to oocyte retrieval anyway—just in case!

Other basic human factors can also impinge on pregnancy rates. In one study a more than two-fold increase could be achieved simply by choosing the person performing embryo transfer (6). Similarly, those cynical enough to comment occasionally note that the results presented in national data compilations do not appear to match data in published studies from individual units. However, the national compilations represent overall perfor-

Fig. 1. The relationship (linear regression and 95% confidence interval) between pregnancy rate and patient cancellation rate.

mance rather than smaller cohorts used in controlled studies, and it would be somewhat atypical for a unit to present poor figures at a peer gathering. What is usually presented represents the best that can be achieved, not necessarily the norm. Furthermore, a clinic's success rate may have changed, for better or worse, in the 2 years since national data were last collected, the time delay being necessary to include infant delivery details. The North American website (<http://www.cdc.gov/nccdphp/drh/arts/>) discussing the interpretation of ART data also has a large section on how results should be interpreted and calculated, so it is a case of "buyer beware" if a country chooses to identify units and their "success" rates.

The approach taken in Australia is to accredit units regularly and monitor those clinics that fall below a reasonable success rate by offering education and expert advice, etc. This is laudable, and by comparison, the level of monitoring and scrutiny

that occurs in ART is still significantly greater than in most other areas of medical practice, often involving equally difficult lifestyle decisions. A Medscape search on the Internet for "IVF and results" yielded 1962 articles, whereas "vasectomy reversal and results" and "tubal reanastomosis and results" yielded 48 and 0 publications, respectively. This is perhaps not a scientific comparison but the Internet is where many patients begin their search for information. Since the costs of each of the above treatments are similar in Australia, and the desired goal of a child, the same, then ART is already well ahead in its willingness to disseminate information.

Currently ART success rates can be quoted in a number of ways (see Table I), and much like interpreting share market performance, a relatively minor difference of a few percentage points could influence, for better or for worse, patient decisions if such information were readily available. When the data from our clinic were analyzed for 1996 of 381 women who started their first ART treatment cycle in that year, 73.5% reached embryo transfer and 26.1% of these achieved a clinical pregnancy. However, by the end of 1997, of those women starting treatment for the first time the previous year, 57.5% were pregnant as a result of either more attempts or the use of frozen embryos (and some embryos still remain cryopreserved). Thus, depending on the presentation of data, our success rates varied between 20.9 and 57.7%! This is not a true cumulative pregnancy rate but is perhaps what patients

Table I. Pregnancy Rate Expressed per Different Numerators

	No. patients	"Success" rate
Patients starting first treatment between 1/1/96 and 1/1/97	381	
Failed stimulation	29	
Failed oocyte collection	2	
Failed fertilization	27	
All embryos frozen	36	
Patients with embryo transfer	280	73.5% of women starting treatment
Pregnant at first attempt	73	19.2% of women starting treatment
		20.9% of women with oocytes
		22.6% of women having embryos
		26.1% of women with transfer
Pregnant by 1/1/98	219	57.5% of women initially starting

want, or need, to hear in discussions with their doctor.

Embryologists have also attempted to define uniform ways of expressing data to make comparisons between laboratories. Thus rates can be expressed as the embryo utilization rate (i.e., the number of embryos used in that cycle or frozen for later use per the number of oocytes fertilized) or the implantation rate (i.e., the number of fetal sacs seen on ultrasound per embryos transferred). Both these methods give a gross indication of embryo "quality." Other approaches have been the calculation of a female age-standardized live birth rate and its associated 95% confidence interval across all age groupings (<http://www.cdc.gov/nccdphp/drh/arts/>)

or a female age-specific (i.e., women less than 35 years of age) embryo implantation rate (RTAC). None have provided more than a gross comparison between clinics, and this is not unexpected considering that the average number of embryos transferred in the United States in 1995 varied between 1.7 and 7.3 per clinic. And this was in women younger than 35 years old! In an article in *Human Reproduction* (4) an appeal was made for honesty in producing ART results, but this is impossible without standardization. The author also provided some interesting insights into medical publications (publicity?).

To conclude, while we applaud disclosure and the best practice methods regarding all aspects

SUCCESS MEANS DIFFERENT THINGS TO DIFFERENT PEOPLE

Fig. 2. Some perceptions of success in ART.

of patient care, we contend that the multitude of variables and human emotions in ART make the publication of league tables of "success" a daunting prospect. Success means different things to different people (Fig. 2). Is clinic X, with a 2% lower pregnancy rate, 2 years ago, which cares for its couples with empathy and undivided attention, a worse clinic than the 2% "better" clinic Y, which offers little, or no, additional support to its "failures," consistently transfers three, four, or even five embryos, and has a high multiple pregnancy rate, with its inevitable attendant difficulties? We make a plea for conformity but wonder how, and if, the commercialization of ART will ever allow standard "conventions" to be proposed and agreed on.

REFERENCES

1. Driscoll GL, Tyler JPP, Carpenter D: Variation in the determination of follicular diameter: An interunit pilot study using an ultrasonic phantom. *Hum Reprod* 1997;12: 2465-2468
2. Clark AM, Ledger W, Galletly C, Tomlinson L, Blaney F, Wang X, Norman RJ: Weight loss results in significant improvement in pregnancy and ovulation rates in anovulatory obese women. *Hum Reprod* 1995;10:2705-2712
3. Bovin J, Takefman JE: Stress level across stages of in vitro fertilisation in subsequently pregnant and nonpregnant women. *Fertil Steril* 1995;64:802-810
4. Schoysman R: Plea for realism and honesty in results of infertility treatment. *Hum Reprod* 1996;11:695-697
5. Hurst T, Shafir E, Lancaster P: Assisted Conception, Australia and New Zealand 1996 (Assisted Conception Series, No 3). Sydney: AIHW National Perinatal Statistics Unit, 1997
6. Naaktgeboren N, Broers FC, Heijnsbroek I, Oudshoorn E, Verburg H, Van der Westerlaken L: Hard to believe, hardly discussed, nevertheless very important for the IVF/ICSI results: Embryo transfer technique can double or halve the pregnancy rate. *Hum Reprod* 12, Abstract Book 1, 1997, Abstract P-061, p 149

G. L. Driscoll

J. P. P. Tyler

City West IVF

12 Caroline Street

Westmead, NSW 2145, Australia