Is meaningful reporting of national IVF outcome data possible?

Hossam I. Abdalla1,4, Siladitya Bhattacharya2, and Yacoub Khalaf3

1Lister Fertility Clinic, The Lister Hospital, Chelsea Bridge Road, London SW1W 8RH, UK 2Division of Applied Health Sciences, School of Medicine and Dentistry, University of Aberdeen, Aberdeen Maternity Hospital, Foresterhill, Aberdeen AB25 2ZD, UK 3Assisted Conception Unit, Guy’s and St Thomas’ Hospital Foundation, Trust 11th Floor, Tower Wing Guy’s Hospital, London SE1 9RT, UK

4Correspondence address. E-mail: sam@easynet.co.uk

The traditional use of live birth per IVF cycle started as the sole indicator of success can be potentially misleading. Different policies regarding reporting IVF cycles started, variations in the number of embryos transferred and associated multiple births have a profound effect on success, such that results from clinics or countries with similar expertise may appear significantly different. To account for these differences, we recommend the use of live birth per embryo—calculated as the number live birth events per 100 embryos transferred—as an outcome measure. This method of reporting can correct for under reported cycles started, adjust for differences in embryo transfer policies and provides an objective and reproducible international benchmark. Combining live birth outcomes from fresh and frozen cycles in the same reporting period per oocyte collection is also recommended. These data should be published as a range related to the national average without a mean or central point. Furthermore, for proper interpretation of results, it would be helpful if the policies regarding patient inclusion and cycle cancellation at all clinics are published.

Key words: IVF / outcome measure / national data / live birth / embryo

Introduction

The purpose of publishing national data on the outcome of IVF is to provide stakeholders with information that addresses their needs. For clinicians, this allows opportunities for audit and evaluation whereas for consumers it should provide reliable estimates of treatment related success and complications. National IVF data give purchasers, regulators and politicians an insight into the scale of service provision in IVF, trends in service delivery and outcomes.

In some countries, this form of reporting is obligatory, whereas in others, data are provided by clinics on a voluntary basis. In the UK, the Human Fertilization and Embryology (HFE) Act, 1990, requires the regulatory authority (The Human Fertilization and Embryology Authority, or HFEA) to provide information on IVF to the public. However, the Act does not specify the amount or the format of the data released into the public domain. This is determined by the HFEA in consultation with other stakeholders. This article examines trends in the way UK outcome data are presented to the public, discusses the limitations of existing systems and offers an alternative way of reporting headline success rates.

Models of reporting national IVF outcomes

The HFE Act (1990) ensured that the UK was the first country to present national data from all IVF clinics. The main outcomes chosen for reporting was live birth rate per IVF cycle—a statistic which is easy to generate and is understood by all stakeholders. In the first 4 years of its existence, the HFEA published summary national data rather than individual clinic outcomes. Comparison of live birth rates for individual IVF clinics allows them to be ranked in terms of their success, but this exercise is not free from bias. There are a number of key factors which affect the outcome of IVF treatment rates including female age, previous failed IVF attempts, reduced ovarian reserve, duration of infertility and number of embryos replaced within the uterus. Thus, higher rates of live birth may not always reflect the quality of a clinic but may be due to a population of patients with a more favourable prognosis. Even within the same clinic, the demographics of patients can influence the outcome of their treatment (Johnson et al., 2007). This consideration may have been behind the HFEA’s decision in 1995 to publish a ‘patients’ Guide’ which contained adjusted live birth rates for each clinic along
with appropriate confidence intervals (Deech, 1996). The public were advised that the ‘Guide’ was, in no way, intended to be interpreted as a league table of IVF Centres. Although many stakeholders welcomed this approach, some pointed out that listing individual clinic success rates in this way would result in an inevitable league table that ranks clinics in a way that allowed the subtleties of clinical practice to be subsumed within a single headline outcome figure (Jacobs and Abdalla, 1996).

This format, was, however, abandoned after 3 years. Realizing that the ‘Guide’ was, perhaps predictably, being used as a league table by the public, the Authority decided to publish crude rather than risk adjusted outcomes for each IVF clinic. The feeling was that the perceived inaccuracy of raw data would be enough to discourage the public from viewing the results in ‘Guide’ as valid for a meaningful comparison between IVF clinics. Marshall and Spiegelhalter had shown in 1998 that confidence intervals around crude live birth rates from UK IVF clinics showed a substantial degree of overlap. They concluded that ‘Institutional ranks are extremely unreliable statistical summaries of performance’ and stated that ‘any performance indicator should always have an associated statistical sampling variability’.

In the USA, the Wyden Act (1993) resulted in the publication of individual IVF clinic results in 1997. In the absence of a single regulatory Authority (as in the UK), this was the product of a collaborative exercise involving the American Centres for Disease Control (CDC), the American Society for Reproductive Medicine (ASRM), the Society for Assisted Reproductive Technology (SART) and a patient support group (RESOLVE). The format of reporting has since remained unchanged. Clinical outcomes, such as live birth rates per cycle started, are stratified by women’s age and accompanied by confidence intervals. The report also contained details of the number of IVF cycles cancelled and embryos replaced within the uterus after each treatment. The dangers of using the published data to rank clinics are highlighted in the report. Each page of the published annual report contains a clear warning which states that, ‘a comparison of clinic success rate may not be always easy to generate and to understand. Conventionallly, all outcome measures include a numerator which represents the primary clinical parameter of success (e.g. live birth) and a denomi- nator which defines the clinical context in which the outcome is measured (e.g. each IVF treatment cycle, each woman). For reasons of convenience outcomes in IVF have been traditionally presented as live birth per cycle of treatment initiated. Although live birth per woman is desirable, the need to have a time horizon for annual reports from national datasets means that live birth per treatment cycle is the commonly chosen outcome published by most registries.

This view has been challenged in recent years. Recognizing the poor prognosis of many preterm twins and other multiples conceived as a result of IVF, Min et al. (2004) proposed that the ideal outcome in an IVF cycle was a healthy singleton baby born at full gestation. Others have felt that the best strategy was to report singleton live birth per initiated IVF cycle (Schieve and Reynolds, 2004). Another option is to report singleton live birth per oocyte retrieval procedure (i.e. singleton birth following the replacement of all fresh or frozen embryos generated from all oocytes collected at a single procedure). A limitation of this latter approach is that it is necessary to wait for some time in order to ensure full data capture. Furthermore, with this approach, there could be more than one delivery per patient.

The debate about the ideal way of reporting outcomes in IVF is enlivened by suggestions of how results can be presented in a more favourable light by manipulating either the numerator, or more commonly, the denominator. Some IVF cycles are abandoned after com- mencing drug treatment for ovarian stimulation because of poor ovarian response such that the chance of retrieving oocytes is deemed to be very low. If these are excluded from the denominator outcomes, such as live birth per cycle, will appear to be higher. In an attempt to encourage comprehensive reporting of all started cycles in the denominator for live birth per cycle the HFSA has introduced mandatory Electronic Data Interchange, and Intention to treat forms in the UK. Despite the fact that the literature suggests that cycle can- cellation owing to an inadequate ovarian response is an inevitable part of IVF treatment, occurring in 11% of all reported cycles in the USA in 2006 (CDC, 2006), as Fig. 1 shows, the calculated cancellation rate for patients under the age of 35 years in 2002, 2005 and 2007 can vary markedly among clinics. It is interesting to note that in 2002, only one clinic claimed to have had no cancellations, however, the number of clinics making such a claim has increased exponentially since (5 clinics in 2005 and 14 in 2007).

The use of live birth per cycle started as the sole indicator of success can be potentially misleading. It clearly does not take into account the possibility of underreporting cycle cancellations, differences in case mix (such as whether potential poor responders or patients with repeated IVF failures are offered treatment). Furthermore, it does not take into account the differences in policies between clinics regarding the number of embryos transferred or pol- icies and outcomes following frozen thawed embryo transfers. This underlines some of the limitations of the present system of reporting in the UK.

![Figure 1](image-url) Distribution of the percentage of abandoned cycle rates for all UK Clinics in women under the age of 35 years for the years 2002, 2005 and 2007 (Abandoned cycle = difference between number of cycles started and number of cycles in which oocyte collection was performed).
Alternative approaches to presenting outcomes

Live birth per embryo

An alternative approach would be to calculate the number of embryos required to achieve a live birth (or a singleton live birth) (Nygren and Nyboe Andersen, 2002). This can be turned around to calculate the number live birth events (twins or triplets counted as one) per 100 embryos transferred. If, for example, one embryo is transferred and a singleton live birth resulted, the live birth per embryo (LB_Emb) is counted as one. On the other hand if three embryos were transferred and the patient has triplets, the LB_Emb would be 0.33. This approach, therefore, could be a valuable method of assessment especially when the objective of any fertility treatment is to achieve a live birth event (preferably singleton) and since an embryo is the final product of an IVF laboratory. LB_Emb can be viewed as an assessment of the capacity of what we produce to achieve our objective. This approach is of some interest when applied to UK results for the calendar year 2006 (http://www.hfea.gov.uk/).

Figure 2 (A) Ranking of clinics based on live birth per cycle started for women under 35 years of age as taken from The Human Fertilization and Embryology Authority’s find a clinic (http://www.hfea.gov.uk/) with confidence interval calculated for 2006 data. (B) Results for live birth rate per embryo (with corresponding confidence interval) for the same patient group appearing in the same order of that seen in (A). Solid line represents national average.
Figure 2A illustrates the ranked live birth rate per cycle, with the corresponding confidence intervals in women under the age of 35 years for all clinics. Apart from one, whose result is significantly above the rest, all other clinics that have live birth rates that are significantly above the national average, were not significantly different from each other.

Figure 2B shows LB_Emb for the same patient group. All clinics appear in the same order as shown in Fig. 2A. However, when analysed according to the LB_Emb, the difference between any of these clinics is no longer significant—as shown by the overlapping confidence intervals.

Table I ranks the top 10 IVF clinics in the UK according to LB_Emb, alongside their rankings calculated according to live birth rate per cycle started (as currently presented in HFEA reports). There is a variation between clinics in the mean number of embryos transferred per procedure, ranging from 1.6 to 2.0. Indeed, the higher live birth rate per cycle in the top clinic (compared with its nearest competitor) (Fig. 2A and Table I) is associated with an increase in number of embryos replaced at each attempt and a higher multiple pregnancy rate. The number of embryos needed to achieve a live birth (calculated by dividing the total number of live birth events, by the total number of embryos used) is also shown and suggests a less dramatic difference.

In the USA in women under the age of 35 years the live birth rate per cycle is 37%, which is significantly higher than in the UK where it is 30%. However, this difference almost disappears if we consider the incidence of LB_Emb transferred; in the USA this is 18.1% as opposed to 17.6% in the UK (with a multiple rate in the USA of 36% as opposed to 30% in the UK). Thus, the apparent superiority of the American results is almost totally explained by the fact that a higher number of embryos are transferred in the USA with the inevitable rise in multiple rates (Table II) (Abdalla, 2009).

The advantages of the LB_Emb system include the ability to go beyond the conventional denominator of number of started cycles. It takes into account the number of embryos replaced at each attempt. An additional advantage of using LB rate/embryo as a denominator to report IVF outcome is its potential to be used as an objective and reproducible international benchmark. It bypasses the potential flaw of under reporting started cycles. It does adjust for differences in embryo transfer policies and is a complex measure which is more likely to convey outcomes relevant to the competence of a clinic rather than the prognosis of women treated.

### Table I

<table>
<thead>
<tr>
<th>Rank on LB_EMB*</th>
<th>Rank on LB_CYC**</th>
<th>LB_emb (%)</th>
<th>LB_Cyc (%)</th>
<th>Multiple (%)</th>
<th>Average number of embryos transferred</th>
<th>Emb_LB***</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>30.4</td>
<td>59</td>
<td>34</td>
<td>1.989</td>
<td>3.3</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>29.7</td>
<td>44</td>
<td>18</td>
<td>1.588</td>
<td>3.4</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>25.0</td>
<td>44</td>
<td>29</td>
<td>1.946</td>
<td>4.0</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>24.6</td>
<td>38</td>
<td>23</td>
<td>1.732</td>
<td>4.1</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>23.1</td>
<td>42</td>
<td>21</td>
<td>1.932</td>
<td>4.3</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>22.7</td>
<td>41</td>
<td>28</td>
<td>1.916</td>
<td>4.4</td>
</tr>
<tr>
<td>7</td>
<td>21</td>
<td>22.2</td>
<td>34</td>
<td>26</td>
<td>1.737</td>
<td>4.5</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>22.1</td>
<td>39</td>
<td>29</td>
<td>1.974</td>
<td>4.5</td>
</tr>
<tr>
<td>9</td>
<td>19</td>
<td>22.0</td>
<td>35</td>
<td>31</td>
<td>1.931</td>
<td>4.5</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>21.7</td>
<td>37</td>
<td>16</td>
<td>1.882</td>
<td>4.6</td>
</tr>
</tbody>
</table>

*LB_EMB = Live Birth per embryo.
**LB_CYC = Live Birth per Cycle started.
***Emb_LB = Number of embryos needed to achieve a live birth event.

### Table II

<table>
<thead>
<tr>
<th></th>
<th>USA</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cycles Started</td>
<td>41 302</td>
<td>13 947</td>
</tr>
<tr>
<td>Live birth</td>
<td>15 406</td>
<td>41 311</td>
</tr>
<tr>
<td>Live birth per cycle (%)</td>
<td>37.3</td>
<td>29.6</td>
</tr>
<tr>
<td>Multiple rate (%)</td>
<td>36</td>
<td>30</td>
</tr>
<tr>
<td>Triplets or more (%)</td>
<td>4.4</td>
<td>0.2</td>
</tr>
<tr>
<td>Cycles reaching embryo transfer</td>
<td>35 497</td>
<td>12 135</td>
</tr>
<tr>
<td>Average number of embryos transferred</td>
<td>2.4</td>
<td>1.93</td>
</tr>
<tr>
<td>Total number of embryos transferred</td>
<td>85 193</td>
<td>23 413</td>
</tr>
<tr>
<td>LB_Emb (%)</td>
<td>18.1</td>
<td>17.6</td>
</tr>
<tr>
<td>Emb_LB</td>
<td>5.5</td>
<td>5.7</td>
</tr>
</tbody>
</table>

*CDC, 2005.
†http://www.hfea.gov.uk.

Combined fresh and frozen live birth per oocyte collection

Analysis of UK national data shows that the total number of episodes involving treatment using frozen thawed embryos per annum amounts to a fifth of all fresh IVF treatments. As such, the annual live birth rate for a clinic could be calculated by adding the live births resulting from fresh cycles to the live births resulting from frozen cycles, and dividing.
the sum by the number of oocyte collections performed in the same
year. This is the method used by the International Committee Moni-
toring Assisted Reproductive Technologies (Adamson et al. 2006) [For
example: if a clinic performs 100 oocyte collections, resulting in 30
pregnancies. The same clinic is also expected to perform 20 frozen
embryo transfers, resulting in five pregnancies. In this case, the preg-
nancy rate of that clinic would be 35% (30 fresh + 5 frozen/100
oocyte collections)]. The advantage of this system is that it may encou-
rage more frozen transfers: this can be beneficial in terms of minimiz-
ing the need for another full fresh treatment, which can be invasive and
expensive, as well as the number of embryos transferred at each
attempt—thus reducing the risk of multiple pregnancies.

Regardless of the choice of outcome measure, all clinic specific data
will be prone to bias because of the different inclusion and exclusion
criteria of different clinics. In addition clinics may use different policies
or reporting techniques which can influence their ranking (Sharif and
Afnan, 2003). If the objective is to compare clinics, it is recognized
that a ‘perfect’ reporting method remains elusive unless all clinics
have patients of similar characteristics and similar cycle cancellation
and embryo transfer policies. Furthermore, for accurate reporting of
cancelled cycles, reported data have to be verified through unfettered
access to any licensed clinic at random and without prior warning not
only to verify the accuracy of what has been reported but also to
investigate what is potentially not reported. This is unfortunately not
consistently enforced; thus it may be appropriate to consider a
system of reporting, such as LB_Emb, to minimize bias. The traditional
live birth (preferably including frozen live birth) per cycle could be
retained, alongside LB_Emb, for the benefit of individual patients
who want a patient centred outcome as long as it is made perfectly
clear that these data should not be used to rank clinics. Furthermore,
live birth per oocyte collection may also be used to report the live
birth resulting from both fresh and frozen cycles performed in the
same observation period. Results should be expressed per age
group, per attempt or per any other denominator in patient character-
istics that may affect the outcome. This data should be published as a
range related to the national average without a mean or central point.
The observer will therefore be able to determine whether the success
rate of a specific clinic lies within the national range.

For any comparison to be meaningful, all clinics should publish their
policies regarding inclusion and exclusion of patients and cycle cancel-
lation. Additionally, as there is a healthy drive to reduce the incidence
of multiple births, the incidence of elective single embryo transfer as
well as elective two embryo transfers should also be published
(there is a higher chance of pregnancy if embryos were electively
transferred than if they were transferred because they were the
only embryos available). Further information may include number of
oocytes collected, fertilization rate, average birthweight and incidence
of preterm delivery. This information, although not recorded as
primary outcome measures, is still important for service providers,
researchers and the lay public.

In conclusion, abstracted data publication could be misleading. In
assisted conception this may have a profound effect on the way that
patients are treated, such that some may be denied treatment and
others may be treated inappropriately in pursuance of simplistic
measures of excellence. It is uncertain as to whether a single headline
outcome measure can capture the complete spectrum of clinical and
scientific distinction, regulatory compliance and safety, all of which
constitutes excellence in IVF. LB_Emb, however, provides a more
sophisticated measure that comes a long way in addressing some of
these uncertainties.

References

Abdalla H. Are the Americans more competent than the Europeans? A
new perspective on data from the USA and Europe. 2009; Abstract
accepted for ESHRE 25th Annual Meeting, Amsterdam, the
Netherlands from 28 June to 1 July.

Adamson G, de Mouzon J, Lancaster P, Nygren K, Sullivan E,
Zegers-Hochschild F. World collaborative report on in vitro

ART success rate, National summary of fertility Clinic Reports. U.S.
Department of Health and Human Services, Centre for Disease

ART success rate, National summary of fertility Clinic Reports 2006, U.S.
Department of Health and Human Services, Centre for Disease Control
and Prevention (CDC), January 2009.

Deech R, A patient’s guide to donor insemination and IVF clinics. Hum

Jacobs HS, Abdalla HI. A patient’s guide or a government league table?

Johnson A, El-Toukhy T, Sunkara S, Khairy M, Coomarasamy A, Ross C,
Bora S, Khalaf Y, Braude P. Validity of the In vtro fertilisation league
tables: influence of patients’ characteristics. BJOG 2007;114:
1569–1574.

Marshall EC, Spiegelhalter DJ. Reliability of league tables of in vitro

Min JK, Breheny SA, MacLachlan V, Healy DL. What is the most relevant
standard of success in assisted reproduction? The singleton, term
gestation, live birth rate per cycle initiated: the BESST endpoint for

Nygren KG, Nyboe Andersen A. Assisted reproductive technology in
Europe, 1999. Results generated from European registers by ESHRE.

Schieve LA, Reynolds MA. Challenges in measuring and reporting success
rates for assisted reproductive technology treatments: what is optimal?


Submitted on June 3, 2009; resubmitted on July 31, 2009; accepted on August 5,
2009.