Human Reproduction, Vol.30, No.2 pp. 315–322, 2015
Advanced Access publication on November 28, 2014 doi:10.1093/humrep/deu319

human reproduction

ORIGINAL ARTICLE *Infertility*

A retrospective evaluation of prognosis and cost-effectiveness of IVF in poor responders according to the Bologna criteria

Andrea Busnelli[1,2,*], Enrico Papaleo[3], Diana Del Prato[3], Irene La Vecchia[1,2], Eleonora Iachini[3], Alessio Paffoni[1], Massimo Candiani[3], and Edgardo Somigliana[1]

[1]Infertility Unit, Fondazione Ca' Granda, Ospedale Maggiore Policlinico, Milan, Italy [2]Università degli Studi, Milan, Italy [3]Obstetrics and Gynecology Unit, San Raffaele Scientific Institute, Milan, Italy

*Correspondence address. Infertility Unit, Fondazione Ca' Granda, Ospedale Maggiore Policlinico, Via M. Fanti, 6, 20122 Milan, Italy. Tel: +39-02-55034303; Fax: +39-02-55034302; E-mail: andreabusnelli@live.it

Submitted on July 4, 2014; resubmitted on October 29, 2014; accepted on November 7, 2014

STUDY QUESTION: Do the Bologna criteria for poor responders successfully identify women with poor IVF outcome?

SUMMARY ANSWER: The Bologna criteria effectively identify a population with a uniformly low chance of success.

WHAT IS ALREADY KNOWN: Women undergoing IVF who respond poorly to ovarian hyper-stimulation have a low chance of success. Even if improving IVF outcome in this population represents a main priority, the lack of a unique definition of the condition has hampered research in this area. To overcome this impediment, a recent expert meeting in Bologna proposed a new definition of poor responders ('Bologna criteria'). However, data supporting the relevance of this definition in clinical practice are scanty.

STUDY DESIGN, SIZE, DURATION: Retrospective study of women undergoing IVF-ICSI between January 2010 and December 2012 in two independent infertility units. Women could be included if they fulfilled the definition of poor ovarian response (POR) according to Bologna criteria prior to initiation of the cycle. Women were included only for one cycle. The main outcome was the live birth rate per started cycle. The perspective of the cost analysis was the one of the health provider.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Three-hundred sixty-two women from two independent Infertility Units were selected. A binomial distribution model was used to calculate the 95% CI of the rate of success. Characteristics of women who did and did not obtain a live birth were compared. A logistic regression model was used to adjust for confounders. The economic analysis included costs for pharmacological compounds and for the IVF procedure. The benefits were estimated on quality-adjusted life years (QALY). To develop the model, we used the local life-expectancy tables, we applied a 3% discount of life years gained and we used a 0.07 improvement in quality of life associated with parenthood. Sensitivity analyses were performed varying the improvement of the quality of life and including/excluding the male partner. The reference values for cost-effectiveness were the Italian and the local (Lombardy) gross domestic product (GDP) pro capita per year in the studied period and the upper and lower limits suggested by NICE.

MAIN RESULTS AND THE ROLE OF CHANCE: Overall, 23 women had a live birth (6%, 95% CI: 4–9%), in line with the previous evidence. This proportion did not significantly differ in the different subgroups of poor responders. Positive predictive factors of success were previous deliveries (adjusted OR = 3.0, 95% CI: 1.1–8.7, $P = 0.039$) and previous chemotherapy (adjusted OR = 13.9, 95% CI: 2.5–77.2, $P = 0.003$). Age, serum AMH, serum FSH and antral follicle count were not significantly associated with live birth. The total cost per live birth was 87 748 Euros, corresponding to 49 919 Euros per QALY. This is above both the limits suggested by NICE for cost-effectiveness and the Italian and local GDP pro capita. Sensitivity analyses mainly support the robustness of the conclusion.

LIMITATIONS, REASONS FOR CAUTION: We lack a control group and we cannot thus exclude that an alternative definition of poor responders may be equally if not more valid. Moreover, independent validations are warranted prior to concluding that IVF is not cost-effective.

Women should thus not be denied treatment based on our findings. Noteworthy, there is also not yet a consensus on the most appropriate economic model to be used.

WIDER IMPLICATIONS OF THE FINDINGS: We recommend the use of the Bologna criteria when designing future studies on poor responders. Large multi-centred international studies are now required to draw definite conclusions on the economic profile of IVF in this situation.

STUDY FUNDING/COMPETING INTERESTS: None.

TRIAL REGISTRATION NUMBER: Not applicable.

Key words: poor responder / Bologna criteria / cost-effectiveness

Introduction

Poor ovarian response (POR) to controlled ovarian hyperstimulation (COH) for IVF is a frustrating condition that represents a topic of utmost clinical and scientific relevance (Pandian *et al.*, 2010; Ferraretti *et al.*, 2011; Kamble *et al.*, 2011; Oudendijk *et al.*, 2012). The management of POR is demanding and controversial. So far, there is no shared vision about the most suitable treatment. Difficulties in drawing firm conclusions are due firstly, to the low rate of success of the procedure in affected women (Oudendijk *et al.*, 2012). In order to avoid type II errors, extremely large RCTs are needed to obtain reliable data. Secondly and most importantly, research in this field has been hampered by a lack of consensus on the diagnosis of POR. More than 40 different definitions have been reported (Surrey and Schoolcraft, 2000; Polyzos and Devroey, 2011). Prognosis varies substantially according to the definition used (Oudendijk *et al.*, 2012) and the heterogeneity of the studied populations hampers the validity of meta-analyses (Pandian *et al.*, 2010; Polyzos and Devroey, 2011).

In order to overcome the difficulties consequent to the lack of a unique and shared definition, an expert meeting was organized in Bologna in 2010 aimed at drawing a consensus on the criteria to be used for diagnosis. Results from that effort were published in 2011 (Ferraretti *et al.*, 2011). This consensus potentially represents an outstanding contribution in the area and this paper has received much attention. However, data supporting the relevance of the use of this definition in clinical practice are still scanty (Polyzos *et al.*, 2012, 2014; Ke *et al.*, 2013).

Therefore, in order to evaluate the Bologna criteria in clinical practice, we set up a retrospective study reporting on women with a condition of POR prior to initiating an IVF cycle. The primary aim of the study was to provide a precise estimate of the live birth rate in women fulfilling the criteria. As secondary aims, we investigated predictive factors of success and performed an economic evaluation.

Materials and Methods

Patients who underwent IVF-ICSI between January 2010 and December 2012 at the Infertility Unit of the Fondazione Cà Granda Ospedale Maggiore Policlinico (FCG) and at the Infertility Unit of the Fondazione Ospedale San Raffaele (HSR) were retrospectively reviewed. Our focus was the population of women who were already diagnosed with POR prior to entering an IVF cycle. Women were included in the study if they fulfilled the definition of POR according to the Bologna criteria *prior* to the initiation of the study cycle. They were included only for one cycle (the first cycle after fulfilling the criteria). Specifically, the Bologna criteria used were as follows (Ferraretti *et al.*, 2011):

Presence of at least two of the following three features:

(i) Anamnestic risk factors: advanced maternal age ($\geq$40 years), evidence of ovarian endometrioma at the basal ultrasound, previous ovarian surgery, previous chemotherapy, genetic abnormalities, shortening of the menstrual cycle.
(ii) A previous POR cycle ($\leq$3 oocytes retrieved or a previous cycle cancelled because of $\leq$3 developing follicles with a conventional stimulation protocol using at least 150 IU FSH per day).
(iii) Abnormal ovarian reserve tests: antral follicle count (AFC) $<$5 or anti-mullerian hormone (AMH) $<$0.5 ng/ml.

Moreover, women could be included if they had at least two previous episodes of POR after maximal stimulation (450 IU FSH per day).

The exclusion criteria were as follows: (i) severe male factor infertility, i.e. $<10^6$ spermatozoa/ml or spermatozoa obtained through MESA-TESE procedures; (ii) IVF cycles for genetic indication (preimplantation genetic diagnosis—PGD); (iii) IVF cycles for fertility preservation; (iv) total number of previous IVF cycles $>$3. The study was accepted by the local institutional review boards. A written informed consensus was not requested because this is a retrospective study. However, all women in both centres are routinely requested for their data to be used for research purposes and those denying this consensus were excluded from this study.

Only fresh cycles were included and women were treated and monitored as previously reported in detail (Busnelli *et al.*, 2013; Papaleo *et al.*, 2014). The regimen to be used and the dose of gonadotrophins was determined on an individual basis according to age, Day-3 serum FSH, serum AMH, AFC and information from previous cycles (if available). Treatment cycles were monitored by serial transvaginal ultrasound and blood hormonal assessments. Human chorionic gonadotrophin (hCG) was administered when at least one follicle had a mean diameter of 18 mm. Cycles were cancelled in the absence of follicular growth, if follicular growth arrested or if premature luteinization (serum progesterone $>$2 ng/ml) occurred. Embryo transfer was performed 48–72 h after oocyte collection. Clinical pregnancy was defined as ultrasonographic demonstration of an intrauterine gestational sac with a vital embryo 4–5 weeks after embryo transfer. An active follow-up of the pregnancy course was systematically performed in the study units. The live birth rate refers to the birth of at least one viable child.

Costs analyses were performed as reported in detail elsewhere (Ragni *et al.*, 2012; Somigliana *et al.*, 2013). The perspective of the cost analysis was the one of the health provider. All considered costs were estimations. The study included costs for pharmacological compounds and for the IVF procedure. The formers were obtained through the website of the official Italian institute for drugs, Agenzia Italiana del Farmaco (http://www.agenziafarmaco.gov.it). They relate to the price of entire boxes (the costs of unused ampoules were not deducted). Costs for the IVF cycles were derived from the regional drugs-related group costs (Bollettino Ufficiale Regione Lombardia, 2010). They were as follows: €225 for cycle preparation and monitoring, €2232 for oocyte collection and €2194 for embryo

transfer. Only drugs taken at home (obtained outside the hospital) were included in the costs. Those administered in the hospital were excluded since they were already included in the drugs-related group reimbursement. Costs of the ultrasound scans and the serum tests were excluded for the same reason. Costs related to pregnancy assistance were excluded from the model. The direct and indirect costs supported by the women and their partners for referrals were also excluded. The benefits were estimated on the quality-adjusted life years (QALY) obtained with the achievement of a live birth. To this aim, we used the local life-expectancy tables (www.istat.it), we applied a 3% discount of life years gained (Drummond and Sculpher, 2005) and we applied the reported 0.07 improvement of the quality of life associated with parenthood (Scotland *et al.*, 2011). The sensitivity analysis was performed varying the improvement of the quality of life from 0.07 to 0.05 (Scotland *et al.*, 2011) and including/excluding the male partner. The reference value for cost-effectiveness was the Italian and the local (Lombardy) gross domestic product (GDP) pro capita per year in the studied period (23 470 and 30 342, respectively) (www.istat.it) and the upper and lower limits suggested by the UK National Institute of Clinical Excellence (NICE) (20–30 000 £, corresponding to 25 700 and 38 500 Euros, respectively) (Appleby *et al.*, 2007).

Data analysis was performed using the Statistics Package for Social Sciences (SPSS 18.0, Chicago, IL, USA). A binomial distribution model was used to calculate the 95% confidence interval (CI) of proportions. Data were compared using Student's *t*-test, Wilcoxon test for unpaired data, or Fisher exact test, as appropriate. Statistical significance was set at $P < 0.05$. A multivariate logistic regression model including age and variables found to associate significantly in univariate analyses was used to calculate the adjusted odds ratios (ORs).

Results

Three-hundred sixty-two women were ultimately included (FCG $n =$ 247 and HSR $n = 115$). Baseline characteristics of the studied subjects are shown in Table I. Table II illustrates the distribution of the variables used for POR definition according to the Bologna criteria. The IVF outcome of the fresh cycles is shown in Table III. Overall, 96 women failed to retrieve at least one suitable oocyte (27%, 95% CI: 22–31%) and 161 did not receive an embryo transfer (44%, 95% CI: 39–50%). Eighty-seven women retrieved four or more oocytes (24%, 95% CI: 20–29%). Subsequent cycles using frozen embryos were possible in five women and resulted in one additional live birth. Overall, 23 women had thus a live birth (6%, 95% CI: 4–9%). The relative frequency of the five subgroups of POR and the corresponding live birth rates are shown in Table IV. The live birth rate in the two participating centres was similar (15/247 corresponding to 6% in FGC and 8/115 corresponding to 7% in HSR, $P = 0.82$). During the study period, the cumulative live birth rate per cycle in a group of age-matched normal responders (retrieving >3 oocytes with the use of conventional doses of gonadotrophins) in the two centres was 23% (56/247) and 21% (24/115), respectively ($P = 0.70$).

We compared baseline characteristics in women who did ($n = 23$) and did not ($n = 339$) achieve a live birth. Results are shown in Table V. A statistically significant difference emerged for previous deliveries and previous chemotherapy. The crude ORs for live birth were 3.5 (95% CI: 1.3–9.5, $P = 0.021$) and 16.8 (95% CI: 3.2–88.6, $P = 0.004$), respectively. The adjusted ORs were 3.0 (95% CI: 1.1–8.7, $P = 0.039$) and 13.9 (95% CI: 2.5–77.2, $P = 0.003$), respectively. Finally, we evaluated the live birth rate according to the centre where the treatment was performed, the regimen of stimulation and the type and initial dose of gonadotrophins used (data not shown) but failed to document any statistically significant difference ($P = 0.82$, $P = 0.79$, $P = 1.00$ and $P = 0.99$, respectively).

Table I Baseline characteristics of the study population ($n = 362$).

Characteristics	Mean ± SD, median (IQR) or number (%)
Age (years)	38.9 ± 3.2
BMI (Kg/m^2)	23.0 ± 6.7
Duration of infertility (years)	3.7 ± 2.5
Menstrual cycle length <28 days	162 (45%)
Day 3 serum FSH (IU/ml)	10.0 ± 4.8
Anti-Mullerian hormone (ng/ml)	0.4 (0.2–0.8)
Total antral follicle count	4 (3–5)
Previous miscarriages	62 (17)
Previous voluntary abortions	20 (6)
Previous ectopic pregnancies	14 (4)
Previous deliveries	37 (10)
Previous IVF-ICSI cycles	194 (54)
Indication	
Unexplained/reduced ovarian reserve	138 (38)
Endometriosis	107 (30)
Tubal factor/pelvic inflammatory disease	33 (9)
Male factor	84 (23)

Statistically significant differences between the two participating centres emerged for the duration of infertility, total AFC, previous miscarriages and indications.
IQR, interquartile range.

Table II Variables for definition of poor ovarian response (POR) in the study population ($n = 362$).

Risk factors	Number (%)
Anamnestic factors	
Age ≥40 years	178 (49)
Previous chemotherapy	6 (2)
Shortening of the menstrual cycle	18 (5)
Abnormal karyotype	0 (0)
Previous ovarian surgery	130 (36)
Previous uni or bilateral salpingectomy	23 (6)
Presence of ovarian endometriomas	55 (15)
Previous POR cycles	
0	206 (57)
1	132 (36)
2	24 (7)
Abnormal ovarian reserve tests	
Anti-Mullerian hormone <0.5 ng/ml	216 (60)
Antral follicle count <5	240 (66)

Statistically significant differences between the two participating centres emerged for the presence of ovarian endometriomas and for the proportion of women with AFC <5.

Table III IVF outcome in the study cycles.

Characteristics	Mean ± SD, median (IQR) or number (%)
Stimulation regimen	
Long protocol	79 (22)
Short protocol	147 (40)
GnRH antagonist	136 (38)
Medications	
FSH	279 (77)
FSH + LH (or hMG)	83 (23)
FSH starting dose (IU)	450 (300–450)
Cancelled cycles	23 (6)
FSH total dose (IU) (including cancelled cycles)	3600 (2900–4500)
FSH total dose (IU) (excluding cancelled cycles)	3600 (2900–4500)
Duration of stimulation (days)[a]	9.6 ± 2.6
Follicles ≥11 mm at hCG administration[a]	4.6 ± 2.7
Oocytes retrieved[a]	
0	43 (13)
1	89 (26)
2	57 (17)
3	63 (19)
≥4	87 (25)
No suitable oocytes available[a]	73 (22)
Oocytes used[b]	2.0 ± 1.8
Number of viable embryos[b]	
0	65 (24)
1	103 (39)
2	57 (21)
≥3	41 (16)
Number of embryos transferred[c]	
1	105 (52)
2	68 (33)
3	30 (15)
Clinical pregnancies	34 (9)
Live births[d]	23 (6)

Statistically significant differences between the two participating centres emerged for the stimulation regimen, the medications, the starting dose, the duration of stimulation and the number of embryos transferred.
IQR, interquartile range.
[a]Refers to women who underwent oocyte retrieval ($n = 339$).
[b]Refers to women who retrieved at least one suitable oocyte ($n = 266$).
[c]Refers to women who underwent embryo transfer ($n = 201$).
[d]Includes one twin pregnancy.

Table IV Live birth rates in the different subgroups of poor ovarian response (POR) women.

POR subgroups	Number of women (%)[a]	Number of live births	Live birth rate % (95% CI)
Anamnestic risk factors for POR and one previous POR cycle	40 (11)	4	10 (3–22)
One previous POR cycle and an abnormal ovarian reserve test	52 (14)	2	4 (1–12)
Anamnestic risk factors for POR and an abnormal ovarian reserve test	190 (52)	11	6 (3–10)
Anamnestic risk factors for POR, one previous POR cycle and an abnormal ovarian reserve test	73 (20)	6	8 (3–16)
Two episodes of POR after maximal stimulation	7 (2)	0	0 (0–31)

Live births among the different subgroups did not differ ($P = 0.65$).
POR subgroups were similarly distributed in the two participating centres.
LBR, Live birth rate. 95% CIs were calculated based on a binomial distribution model.
[a]Percentages refer to the whole cohort ($n = 362$).

The total costs for the IVF procedures in the studied population are shown in Table VI. Overall, the total cost per live birth was 87 748 Euros, corresponding to 49 919 Euros per QALY. Results from the sensitivity analysis are shown in Fig. 1. IVF in POR women is not cost-effective for any model when considering the Italian GDP pro capita or the NICE suggested lower limit as thresholds. The referral model (i.e. the one that included exclusively the woman with a 0.07 impact on QALY) is not cost-effective regardless of the thresholds used. The procedure results in borderline cost-effectiveness only when including also the male partner in the model and when using as thresholds for cost-effectiveness the GDP pro capita in Lombardy or the NICE upper limit.

Discussion

The present study confirms that the Bologna criteria define a population with a low rate of success. The live birth rate was 6% and, based on the calculated 95% CI (4–9%), it can be confidently concluded that the chances of success is at least below 10%. The number of women needed to be treated is 16. Of further interest here is that we failed to observe main differences among the five different subgroups of women that can be identified applying the Bologna criteria. This observation suggests that these criteria select a population with a uniformly low chance of success and thus supports the validity of the definition. Interestingly, our results are quite compatible with those emerging from three previous large and similar contributions (Polyzos *et al.*, 2012, 2014; Ke *et al.*, 2013). Ke *et al.* presented data on 479 women undergoing 737 cycles and documented a live birth rate per started cycle of 8% (Ke *et al.*, 2013). In the study from Polyzos *et al.* that included 485 women and 823 cycles, the live birth per cycle was 6% (Polyzos *et al.*, 2014). Finally, the same study group also published an additional contribution in women treated with natural cycle IVF and reported a 3% live birth rate per cycle (136 women, 390 cycles) (Polyzos *et al.*, 2012). Overall, the similarity in the rate of success across centres (at least when using hyper-stimulation) and populations strongly supports the validity of the Bologna criteria. Furthermore, it is noteworthy that the live birth rate was similar between the two centres involved in the present study.

Table V Baseline characteristics in women who did and did not obtain a live birth. Data are mean ± SD, median (IQR) or number (%).

Characteristics	Live birth *n* = 23	No live birth *n* = 339	*P*
Age (years)	38.9 ± 3.0	39.0 ± 3.3	0.89
Age ≥40 years	10 (44)	168 (50)	0.67
BMI (Kg/m^2)	22.1 ± 3.5	23.0 ± 6.9	0.53
Duration of infertility (years)	3.6 ± 2.2	3.8 ± 2.5	0.77
Menstrual cycle length <28 days	14 (61)	148 (44)	0.13
Shortening of the menstrual cycle	0 (0)	18 (5)	0.62
Day 3 serum FSH (IU/ml)	9.8 ± 5.3	10.0 ± 4.8	0.86
AMH (ng/ml)	0.4 (0.2–1.6)	0.4 (0.2–0.8)	0.62
AMH <0.5 ng/ml	14 (61)	202 (60)	1.00
Total AFC	4 (3–6)	4 (3–5)	0.55
AFC <5	15 (65)	225 (67)	1.00
Previous miscarriages	4 (17)	58 (17)	1.00
Previous voluntary abortions	3 (13)	17 (5)	0.12
Previous ectopic pregnancies	1 (4)	13 (4)	0.61
Previous deliveries	6 (26)	31 (9)	0.021
Previous chemotherapy	3 (13)	3 (1)	0.004
Previous ovarian surgery	9 (39)	121 (36)	0.82
Previous uni or bilateral salpingectomy	2 (9)	21 (6)	0.65
Presence of ovarian endometriomas	5 (22)	50 (15)	0.37
Previous IVF-ICSI cycles	13 (56)	181 (53)	0.83
Previous POR cycles			0.92
0	13 (56)	193 (57)	
1	8 (35)	124 (37)	
2	2 (9)	22 (6)	
Indication			0.78
Unexplained/reduced ovarian reserve	8 (35)	130 (38)	
Endometriosis	8 (35)	99 (29)	
Tubal factor/PID	3 (13)	30 (9)	
Male factor	4 (17)	80 (24)	

Table VI Total costs of the IVF cycles.

Items	Costs (€)
Medications	
GnRh agonists	26 888
GnRh antagonists	38 359
Gonadotrophins	737 133
Progesterone	2826
hCG	10 170
Subtotal	815 376
Technical procedures	
Cycle preparation and monitoring[a]	5175
Oocyte retrieval	756 648
Embryo transfer	440 994
Subtotal	1 202 817
Total costs	2 018 193
Mean cost per patient	5575
Mean cost per live birth	87 748

Costs were calculated by multiplying the cost of the treatment regimen by the number of women so treated. Means were calculated by dividing the total by the total number of women (362) or live births (23). Cancelled cycles were included in the analysis. GnRH, gonadotrophin-releasing hormone; hCG, human chorionic gonadotrophin.
[a]Refers to women who did not undergo oocyte retrieval (*n* = 23).

Predictive factors of success prior to cycle initiation were poorly investigated in the three preceding contributions. All studies actually exclusively evaluated the impact of women age. Ke *et al.* reported a statistically significant reduction in the rate of success with age. The live birth rates in women aged ≤35, 36–39 and ≥40 years were 12, 8 and 6%, respectively (Ke *et al.*, 2013). In contrast, Polyzos *et al.* failed to detect any impact of age in their two studies (Polyzos *et al.*, 2012, 2014). Our results corroborate this latter view since age was similar in women who did and did not achieve a live birth. Conversely, our study identified two interesting predictive factors that were not previously investigated. Specifically, we observed a positive association with previous deliveries and with a history of chemotherapy. The demonstration of an enhanced chance of success in women who previously delivered could be expected and is of clinical relevance. Indeed, the chances of delivery are up to three folds higher and, from an epidemiological point of view, the proportion of women in this situation is low but significant (10%). This information needs to be confirmed in independent analyses but may be of potential interest when counselling affected women. The observation of a positive impact of a history of chemotherapy is intriguing. Surprisingly, the magnitude of the association is extremely high. Again, if confirmed, this information may be of utmost interest for women with a history of chemotherapy. Nonetheless, it has to be pointed out that only six women reported this condition (three had a live birth). This low sample size exposes the result to type I error and does not allow the potential role of confounders to be addressed. In particular, previous chemotherapy may be associated with some more favourable characteristics (such as younger age). Overall, albeit statistically significant, this result thus warrants confirmation. Moreover, from an epidemiological point of view, this aspect is of doubtful relevance in identifying significant subgroups of women at better prognosis given the rarity of this condition. Finally, it is interesting to underline that our study also failed to detect any association between live birth and biomarkers of ovarian reserve (serum FSH, serum AMH and AFC). In this regard, it has to be underlined that our results do not deny the important predictive role of biomarkers of ovarian reserve and age in general (Oudendijk *et al.*, 2012; Somigliana *et al.*, 2013). This conclusion is valid only in the very particular population of poor responders according to the Bologna criteria. Indeed, these variables are linked one another in the definition, thus hampering the evaluation of their independent impact on the outcome.

Bologna criteria identify a homogeneous population at low chances of live birth and, from a public health perspective, the opportunity to treat these women may be questioned. Noteworthy, our economic analyses

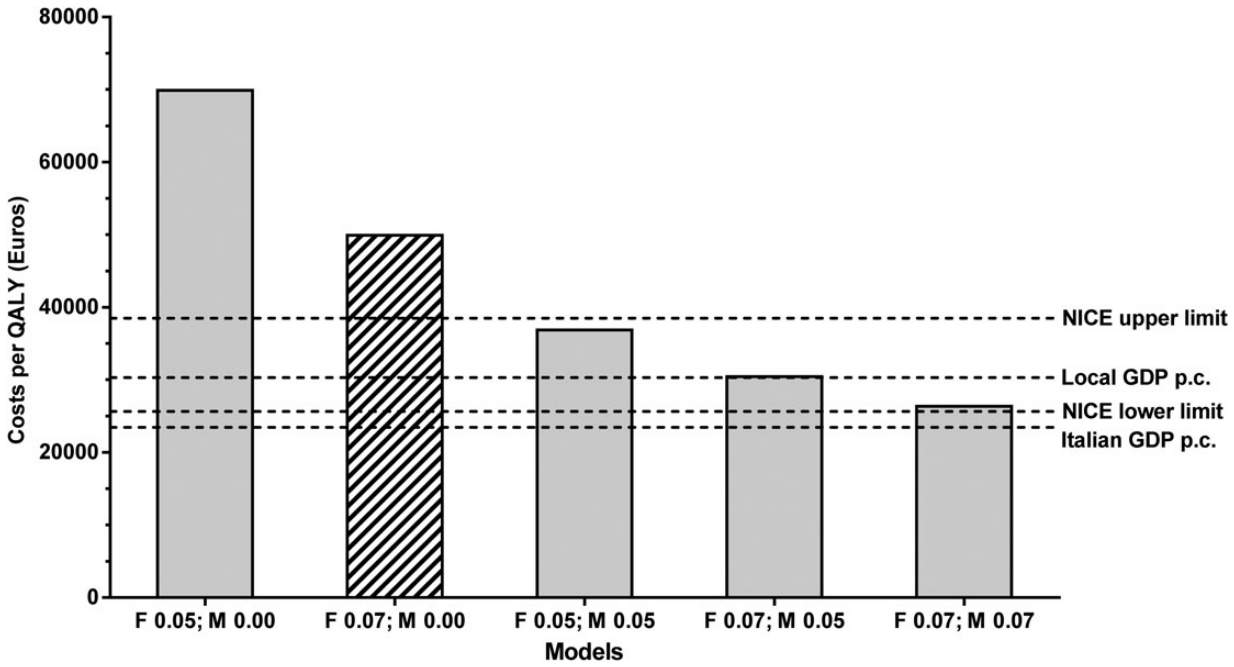

Figure 1 Cost-effectiveness sensitivity analysis. The reference model (represented with a stripped column) included only the woman's quality of life and is based on a 0.07 improvement of quality adjusted life years (QALY) associated with parenthood. Four different models including/excluding the male counterpart and varying the impact on QALY from 0.07 to 0.05 are also represented. A 3% discount of life years was applied for all models. The costs per QALY in the five models, from left to right are 69 887, 49 919 (reference model), 36 881, 30 453 and 26 341 Euros, respectively. The horizontal lines correspond to the thresholds for cost-effectiveness, i.e. the Italian and the local (Lombardy) gross domestic product pro capita (GDP p.c.) that were during the study period 23 470 and 30 342 Euros, respectively and the upper and lower NICE thresholds (20 000 and 30 000 £ equivalent to 25 700 and 38 500 Euros, respectively). F: Female, M: Male.

showed that IVF in these cases is not cost-effective. The cost per live birth was 87 748 Euros, corresponding to 49 919 Euros per QALY with the most accepted model (Scotland *et al.*, 2011; NICE, 2012). This is clearly above the NICE recommended thresholds as well as the Italian and local GDP pro capita. Sensitivity analyses tend to support the robustness of this conclusion (Fig. 1). The intervention becomes cost-effective only when including the male partner in the model and when referring to the upper NICE threshold or to the local GDP pro capita. Non-cost-effective interventions should not be supported by public health policies. However, despite this unfavourable economic profile, we do not believe that our data should be used to deny IVF to women fulfilling the Bologna Criteria. Given the potential substantial impact of cost-effectiveness analyses on public health policies, there is indeed the need for further and independent validations. The characteristics of the populations fulfilling the criteria for poor responders, the therapeutic regimens used for their treatment and the costs of the IVF procedures vary substantially among centres and worldwide (Pandian *et al.*, 2010; Chambers *et al.*, 2013). Moreover, results from economic analyses are markedly influenced by the basal assumptions and results may differ substantially with the use of different models (Polyzos *et al.*, 2011). Definitely concluding on the unfavourable economic profile of IVF in POR women is thus not justifiable based on this single study. More robust evidence from other contexts is warranted. Moreover, some ethical concerns may arise. For instance, the Ethics Committee of the American Society for Reproductive Medicine (ASRM) does not justify declining treatment also in women with a predicted chance of delivery below 5% (Ethics Committee of ASRM, 2012). Finally, the models used in this study exclude the potential economic benefits of the newborns. This is commonly done in cost-effectiveness analyses regarding infertility. However, the complete exclusion of this potential stakeholder is arguable (Connolly *et al.*, 2009; Chambers *et al.*, 2013). For instance, Connolly *et al.* estimated that the discounted net tax revenue of an ART singleton in the UK in 2005 was ~208 400 USD, representing an 8-folds return on investment (Connolly *et al.*, 2009). More in general, our model might be too conservative.

Some further limitations of our study need to be acknowledged. Firstly, the two participating centres differed in some important baseline characteristics (Tables I–III). The populations studied and the management strategies actually differed. Data from a more homogeneous population would have facilitated inferences. On the other hand, despite several differences in baseline clinical characteristics and treatment modalities between the two centres, the proportion of the different subgroup of POR and the live birth rate did not differ. This is in line with the available evidence showing that none of the several treatment options proposed for POR have consistently been demonstrated to be more effective (Pandian *et al.*, 2010). However, there is undoubtedly the need for validation of our findings in other contexts. Even if the live birth rates resulted similar in the other two available studies using conventional/maximal doses of gonadotrophins for hyper-stimulation (Ke *et al.*, 2013; Polyzos *et al.*, 2014), the characteristics of POR

women, specific therapeutic regimens, local costs and thresholds for cost-effectiveness may substantially differ. A large multicentre international study including subgroup analyses may ultimately provide a definite answer.

Secondly, some of our analyses are exposed to a significant type II error. In particular, even if we failed to document any statistically significant difference among the different POR subgroups, we cannot draw definite conclusions. For instance, only seven women were included in the group of women with two prior episodes of POR after maximal stimulation. Given that we deem it important to demonstrate definitely that POR women according to the Bologna criteria have a uniformly low expectation of success, confirmation in larger studies is required.

Thirdly, we took some necessary albeit debatable decisions to cope with some intrinsic uncertainties of the definition of POR (Ferraretti *et al.*, 2011). In particular, in the original definition, the term 'maximal stimulation' is not precisely defined and the thresholds for serum AMH and AFC are given as a range (0.5–1.1 ng/ml and 5–7, respectively). Overall, we opted for conservative and stringent choices. We defined maximal stimulation as 450 IU of FSH daily and the thresholds for AMH and AFC were set at 0.5 and 5 ng/ml, respectively. We were thus more confident on the inclusion of women with a frank condition of POR but these choices have to be taken into consideration for inferences. We cannot indeed exclude that results would have been partly different if we used less strict criteria. On the other hand, it has to be acknowledged that we excluded women who previously underwent >3 cycles and this may have led to overestimating the benefits of IVF.

Finally, we lack a control group. We did not indeed compare the validity of alternative definitions of poor responders. Noteworthy, there is an ongoing burning debate on the Bologna criteria (Ferraretti and Gianaroli, 2014; Papathanasiou, 2014; Venetis, 2014). Of relevance here is that the diagnosis depends on poor response to ovarian stimulation in a prior cycle in half of the cases and is therefore retroactive and subject to biases (dosages given in the previous cycle differ, obese patients may absorb identical dosages poorer than normal weight patients, oocytes retrieval could be hampered for technical difficulties…). Given the important recent progresses in predicting ovarian responsiveness with the use of biomarkers, there is the need for future studies comparing the effectiveness of the Bologna criteria to algorithms that include age-specific FSH, AMH and AFCs but do not require data from a previous cycle.

In conclusion, our study supports the validity of the Bologna criteria. We thus recommend the use of this definition when designing future studies on poor responders. Noteworthy, these studies are urgently warranted since there is the important need to improve IVF outcome in affected women considering in particular that the intervention is currently of doubtful cost-effectiveness.

Authors' roles

E.S., E.P. and M.C. planned and implemented the study. A.B. and E.S. wrote the first draft of the manuscript. A.B. and A.P. did the statistical analyses. A.B., D.D., I.L. and E.I. collected the data. All authors participated to the discussion of the findings and revised the manuscript.

Funding

No external funding was used.

Conflict of interest

None declared.

References

Appleby J, Devlin N, Parkin D. NICE's cost effectiveness threshold. *BMJ* 2007;**25**:358–359.

Bollettino Ufficiale Regione Lombardia. Determinazioni in ordine alla gestione del servizio socio sanitario regionale per l'esercizio 2011. Bollettino Ufficiale Regione Lombardia, N.50 S2. 2010. http://www.regione.lombardia.it/.

Busnelli A, Somigliana E, Benaglia L, Leonardi M, Ragni G, Fedele L. *In vitro* fertilization outcomes in treated hypothyroidism. *Thyroid* 2013;**23**:1319–1325.

Chambers GM, Adamson GD, Eijkemans MJ. Acceptable cost for the patient and society. *Fertil Steril* 2013;**100**:319–327.

Connolly M, Gallo F, Hoorens S, Ledger W. Assessing long-run economic benefits attributed to an IVF-conceived singleton based on projected lifetime net tax contributions in the UK. *Hum Reprod* 2009;**24**:626–632.

Drummond M, Sculpher M. Common methodological flaws in economic evaluations. *Med Care* 2005;**43**:5–14.

Ethics Committee of American Society for Reproductive Medicine. Fertility treatment when the prognosis is very poor or futile: a committee opinion. *Fertil Steril* 2012;**98**:e6–e9.

Ferraretti AP, Gianaroli L. The Bologna criteria for the definition of poor ovarian responders: is there a need for revision? *Hum Reprod* 2014; **29**:1842–1845.

Ferraretti AP, La Marca A, Fauser BC, Tarlatzis B, Nargund G, Gianaroli L; ESHRE working group on Poor Ovarian Response Definition. ESHRE consensus on the definition of 'poor response' to ovarian stimulation for *in vitro* fertilization: the Bologna criteria. *Hum Reprod* 2011; **26**:1616–1624.

Kamble L, Gudi A, Shah A, Homburg R. Poor responders to controlled ovarian hyperstimulation for *in vitro* fertilisation (IVF). *Hum Fertil (Camb)* 2011;**14**:230–245.

Ke H, Chen X, Liu YD, Ye DS, He YX, Chen SL. Cumulative live birth rate after three ovarian stimulation IVF cycles for poor ovarian responders according to the bologna criteria. *J Huazhong Univ Sci Technolog Med Sci* 2013;**33**:418–422.

NICE – National Collaborating Centre for Women's and Children's Health. Fertility: assessment and treatment for people with fertility problems (update): draft for stakeholder consultation—May 2012. www.nice.org.uk/nicemedia/live/12157/59278/59278.pdf.

Oudendijk JF, Yarde F, Eijkemans MJ, Broekmans FJ, Broer SL. The poor responder in IVF: is the prognosis always poor?: a systematic review. *Hum Reprod Update* 2012;**18**:1–11.

Pandian Z, McTavish AR, Aucott L, Hamilton MP, Bhattacharya S. Interventions for 'poor responders' to controlled ovarian hyper stimulation (COH) in *in-vitro* fertilisation (IVF). *Cochrane Database Syst Rev* 2010(1):CD004379.

Papaleo E, Corti L, Vanni VS, Pagliardini L, Ottolina J, De Michele F, La Marca A, Viganò P, Candiani M. Basal progesterone level as the main determinant of progesterone elevation on the day of hCG triggering in controlled ovarian stimulation cycles. *Arch Gynecol Obstet* 2014;**290**:169–176.

Papathanasiou A. Implementing the ESHRE 'poor responder' criteria in research studies: methodological implications. *Hum Reprod* 2014;**29**:1835–1838.

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**:1058–1061.

Polyzos NP, Valachis A, Mauri D, Ioannidis JP. Industry involvement and baseline assumptions of cost-effectiveness analyses: diagnostic accuracy of the Papanicolaou test. *CMAJ* 2011;**183**:E337–E343.

Polyzos NP, Blockeel C, Verpoest W, De Vos M, Stoop D, Vloeberghs V, Camus M, Devroey P, Tournaye H. Live birth rates following natural

cycle IVF in women with poor ovarian response according to the Bologna criteria. *Hum Reprod* 2012;**27**:3481–3486.

Polyzos NP, Nwoye M, Corona R, Blockeel C, Stoop D, Haentjens P, Camus M, Tournaye H. Live birth rates in Bologna poor responders treated with ovarian stimulation for IVF/ICSI. *Reprod Biomed Online* 2014;**28**:469–474.

Ragni G, Levi-Setti PE, Fadini R, Brigante C, Scarduelli C, Alagna F, Arfuso V, Mignini-Renzini M, Candiani M, Paffoni A *et al.* Clomiphene citrate versus high doses of gonadotropins for *in vitro* fertilisation in women with compromised ovarian reserve: a randomised controlled non-inferiority trial. *Reprod Biol Endocrinol* 2012;**10**:114.

Scotland GS, McLernon D, Kurinczuk JJ, McNamee P, Harrild K, Lyall H, Rajkhowa M, Hamilton M, Bhattacharya S. Minimising twins in *in vitro* fertilisation: a modelling study assessing the costs, consequences and cost-utility of elective single versus double embryo transfer over a 20-year time horizon. *BJOG* 2011;**118**:1073–1083.

Somigliana E, Paffoni A, Busnelli A, Cardellicchio L, Leonardi M, Filippi F, Ragni G, Fedele L. IVF outcome in poor responders failing to produce viable embryos in the preceding cycle. *Reprod Biomed Online* 2013; **26**:569–576.

Surrey ES, Schoolcraft WB. Evaluating strategies for improving ovarian response of the poor responder undergoing assisted reproductive techniques. *Fertil Steril* 2000;**73**:667–676.

Venetis CA. The Bologna criteria for poor ovarian response: the good, the bad and the way forward. *Hum Reprod* 2014;**29**:1839–1841.