The clinical effectiveness of preimplantation genetic diagnosis for aneuploidy in all 24 chromosomes (PGD-A): systematic review

Evelyn Lee1,*, Peter Illingworth2, Leeanda Wilton3, and Georgina Mary Chambers1

1National Perinatal Epidemiology and Statistics Unit, School of Women’s and Children’s Health, University of New South Wales (UNSW), Level 2, McNevin Dickson Building, Randwick Hospitals Campus, Sydney 2031, Australia 2IVF Australia Pty Ltd, 176 Pacific Highway, Greenwich, Sydney 2065, Australia 3Melbourne IVF, Victoria Parade, East Melbourne, VIC 3002, Australia

*Correspondence address. Tel: +61293821014; E-mail: evelyn.lee@unsw.edu.au

Submitted on June 20, 2014; resubmitted on October 14, 2014; accepted on October 22, 2014

STUDY QUESTION: Is preimplantation genetic diagnosis for aneuploidy (PGD-A) with analysis of all chromosomes during assisted reproductive technology (ART) clinically and cost effective?

SUMMARY ANSWER: The majority of published studies comparing a strategy of PGD-A with morphologically assessed embryos have reported a higher implantation rate per embryo using PGD-A, but insufficient data has been presented to evaluate the clinical and cost-effectiveness of PGD-A in the clinical setting.

WHAT IS KNOWN ALREADY: Aneuploidy is a leading cause of implantation failure, miscarriage and congenital abnormalities in humans, and a significant cause of ART failure. Preclinical evidence of PGD-A indicates that the selection and transfer of euploid embryos during ART should improve clinical outcomes.

STUDY DESIGN, SIZE AND DURATION: A systematic review of the literature was performed for full text English language articles using MEDLINE, EMBASE, SCOPUS, Cochrane Library databases, NHS Economic Evaluation Database and EconLit. The Downs and Black scoring checklist was used to assess the quality of studies. Clinical effectiveness was measured in terms of pregnancy, live birth and miscarriage rates.

PARTICIPANTS/MATERIALS, SETTINGS, METHODS: Nineteen articles meeting the inclusion criteria, comprising three RCTs in young and good prognosis patients and 16 observation studies were identified. Five of the observational studies included a control group of patients where embryos were selected based on morphological criteria (matched cohort studies).

MAIN RESULTS AND ROLE OF CHANCE: Of the five studies that included a control group and reported implantation rates, four studies (including two RCTs) demonstrated improved implantation rates in the PGD-A group. Of the eight studies that included a control group, six studies (including two RCTs) reported significantly higher pregnancy rates in the PGD-A group, and in the remaining two studies, equivalent pregnancy rates were reported despite fewer embryos being transferred in the PGD-A group. The three RCTs demonstrated benefit in young and good prognosis patients in terms of clinical pregnancy rates and the use of single embryo transfer. However, studies relating to patients of advanced maternal age, recurrent miscarriage and implantation failure were restricted to matched cohort studies, limiting the ability to draw meaningful conclusions.

LIMITATIONS, REASONS FOR CAUTION: Relevant studies may have been missed and findings from RCTs currently being undertaken could not be included.

WIDER IMPLICATIONS OF THE FINDINGS: Given the uncertain role of PGD-A techniques, high-quality experimental studies using intention-to-treat analysis and cumulative live birth rates including the comparative outcomes from remaining cryopreserved embryos are needed to evaluate the overall role of PGD-A in the clinical setting. It is only in this way that the true contribution of PGD-A to ART can be understood.

STUDY FUNDING/COMPETING INTERESTS: No specific funding was used to undertake this study. Evelyn Lee does not report any conflict of interest. Associate Professor Illingworth is an employee of and a shareholder in Virtus Health which is a provider of clinical services.
Introduction

Aneuploidy is the most common type of chromosome abnormality and the leading cause of implantation failure, miscarriage and congenital abnormalities in humans (Hassold et al., 1996; Vialard et al., 2011). In the context of assisted reproductive technologies (ART), preimplantation genetic diagnosis (PGD) is the practice of obtaining a cellular biopsy of an embryo to evaluate the genetic composition, allowing the selection of a genetically unaffected embryo for transfer (Brezina et al., 2012). Initially, PGD was used to detect embryos at risk for inheriting sex-linked monogenic diseases (Handyside et al., 1990). More recently, the indications for PGD have expanded to include comprehensive aneuploidy screening of all 24 chromosomes in patients with a presumed normal karyotype (PGD-A). In theory, this approach should enhance ART clinical outcomes by improving implantation rates and reducing miscarriage rates, particularly in those patients at an increased risk of producing aneuploid embryos, such as women of advanced maternal age and those with recurrent implantation failure or recurrent miscarriage (Kahraman et al., 2000; Rubio et al., 2003; Vouillaire et al., 2007). PGD-A also encourages ART best clinical practice by supporting the selection of a single euploid embryo for transfer, thus minimizing iatrogenic multiple births and the concurrent risk to mothers and babies (Helmerhorst et al., 2004; Ombrelet et al., 2006; Lemos et al., 2013).

One of the first PGD-A techniques introduced into clinical practice used comparative genomic hybridization (CGH), with the first baby born from this technique in 2001 (Wilton et al., 2001). Over the last decade more advanced and automated techniques have become widely available, including array CGH (aCGH), single nucleotide polymorphism (SNP) arrays and real-time quantitative polymerase chain reaction (qPCR). However, as previous experience with aneuploidy screening using fluorescence in situ hybridization (FISH) of cleavage stage embryos has shown, caution is needed before the widespread adoption of PGD-A into routine clinical practice. While initially promising, FISH-based PGD-A was limited by its capacity to assess only a subset of chromosomes and suffered from a lack of standardized techniques and interpretation which lead to substantial variation in reproducibility of results between laboratories (Colls et al., 2007; Munné et al., 2007; Hardarson et al., 2008, Northrop et al., 2010). Several prospective studies showed that PGD-A using FISH failed to improve pregnancy rates and in some studies worsened the outcomes in the clinic setting (Staessen, 2004; Blockeel et al., 2008; Staessen et al., 2008; Garrisi et al., 2009; Schoolcraft et al., 2009; Debrock et al., 2010; Ajduk and Zernicka-Goetz, 2012). However, it is worth noting that the most recent prospective RCT demonstrated an increase in the live birth rate in older women after PGD-A using FISH (Rubio et al., 2013).

Comprehensive aneuploidy screening of all 24 chromosomes using more automated approaches should overcome the limitations of FISH. However, limited evidence exists on the clinical and cost-effectiveness of this more comprehensive approach in the fertility clinic setting. Most studies to date have sought to evaluate the scientific and preclinical validity of various PGD-A techniques, using different cell types (polar body, blastomere and trophoectoderm), rather than directly evaluating its clinical effectiveness in terms of patient outcomes (Fragouli et al., 2011; Mamas et al., 2012).

This paper will review the available evidence regarding the effectiveness of comprehensive PGD-A techniques in different patient groups and using different cell types for biopsy. A number of narrative reviews (Brezina et al., 2012; Fiorentino, 2012; Fragouli and Wells, 2012; Gleicher and Barad, 2012; Gleicher et al., 2014) have attempted to assimilate the clinical outcomes of PGD-A into an overall assessment, but to our knowledge this is the first systematic review to focus on the clinical validity of PGD-A and to objectively assess the quality of published studies. This review is timely given that the scientific principles of comprehensive aneuploidy screening are widely accepted. However, considerable controversy still remains about the clinical and, particularly, economic effectiveness of this approach (Hellani et al., 2008; Treff et al., 2010; Gutiérrez-Mateo et al., 2011; Capalbo et al., 2013a,b).

Materials and Methods

Search strategy and study selection

A computer assisted search to identify relevant articles was performed using the OVID interface to MEDLINE, EMBASE, SCOPUS, Cochrane Library, NHS Economic Evaluation Database and EconLit. The search was limited to studies of humans, English language text, with no date restriction and included the MeSH keywords: embryo transfer, preimplantation genetic diagnosis, PGD, prenatal diagnosis, aneuploidy screening, comprehensive chromosomes, pregnancy, IVF, in-vitro fertilisation, polar bod$, cleavage, blastocyst$, comparative genomic hybridisation, arrayCGH, single nucleotide polymorphism arrays, SNPs$, quantitative polymerase chain reaction, qPCR. The comprehensive PGD-A techniques included were CGH (array and metaphase), SNP and qPCR. All article abstracts were reviewed for relevance and the reference lists of all identified studies were cross-checked for additional articles. The search covered all eligible articles published up until August 2014. The inclusion and exclusion criteria for retaining the identified studies in this review are shown in Table I. Studies had to report clinical or ongoing pregnancy, live birth or miscarriage rates to be included in the review. The primary review was performed by the first author (E.L.), with the other three authors reviewing selected publications. The first author also contacted individual study authors to clarify study details where required.
Table I Exclusion and inclusion criteria.

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>A cohort of patients using one of CGH, aCGH, SNP array or qPCR for PGD-A</td>
<td>Patients who had PGD for single gene disease, translocations or monogenic diseases (To avoid additional selection bias, studies were also excluded if they had recruited predominately patients with a known genetic disorder or who had donated oocytes for analysis)</td>
</tr>
<tr>
<td>Study with 20 and more patients</td>
<td>Narrative, commentaries, letters to editor or reviews</td>
</tr>
<tr>
<td>Papers written in English</td>
<td>Studies deficient of an outcome measure</td>
</tr>
<tr>
<td>Experimental and observational studies</td>
<td>Published in peer reviewed journal</td>
</tr>
<tr>
<td>Inclusion of outcome measures that include either rates of clinical pregnancy, ongoing pregnancy, miscarriage or live births</td>
<td>Published in non-peer reviewed journal, abstract, conference proceeding</td>
</tr>
</tbody>
</table>

Assessment of study quality

Critical appraisal of the articles was conducted using the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) Statement (Moher et al., 2009). Additionally, the methodological quality of the studies that met the inclusion criteria were assessed using a modified version of Downs and Black checklist which is considered valid and reliable for assessing randomized and non-randomized studies of healthcare interventions and covers reporting, external validity, internal validity (bias and confounding) and statistical power (Downs and Black, 1998). Because not all of the items in the Downs and Black checklist are relevant to non-randomized studies, a modified checklist was developed using of 22 of the 27 items. For this review, each item in the checklist was scored either 0 or 1, with a score of < 8 considered a poor-quality study, a score of 8 – 15, a moderate-quality study and a score of more than 15, a high-quality study.

Results

Figure 1 outlines the results of the search strategy which resulted in 19 articles meeting the inclusion criteria for this systematic review. Studies were excluded if they were designed primarily to validate technical aspects of comprehensive PGD-A, used predominately donated oocytes, or lacked clinical outcome measures. Details of each study are included in Supplementary data Table SI.

The three RCTs identified (Yang et al., 2012; Forman et al., 2013b; Scott et al., 2013a) sought to compare the clinical outcomes in young, good prognosis patients assigned to embryo selection based on either PGD-A or morphological assessment alone. The remaining 16 studies used observational designs, with only five studies including a control group for comparison (matched cohort studies) (Wilton et al., 2003; Sher et al., 2007, 2009; Fragouli et al., 2010; Schoolcraft, 2010; Fishel et al., 2011; Geraedts et al., 2011; Schoolcraft et al., 2011; Traversa et al., 2011; Forman et al., 2012, 2013a,b; Hodes-Wertz et al., 2012; Adler et al., 2013; Harton et al., 2013; Keltz et al., 2013; Mir et al., 2013).

Study characteristics

Supplementary data, Table SI summarizes the characteristics and outcomes of the 19 eligible studies. The studies were published between 2003 and 2014, with the majority (n = 15) published between 2010 and 2013. There was little geographical spread in the studies, with most (n = 12) being performed in the USA.

Significant heterogeneity existed among the studies in terms of study design, size, comparators and patient characteristics. The number of study participants ranged from 20 to 320 patients with resultsfrom 2983 patients included in the review. Three studies (Fishel et al., 2011; Hodes-Wertz et al., 2012; Harton et al., 2013) presented the results based on ART cycles and provided little detail on the number of patients in their study. Overall, the mean age of study patients in the review ranged from 31.2 to 40 years. Most studies (n = 14) reported outcomes on patients with a poor prognosis, i.e. advanced maternal age, recurrent implantation failure or recurrent pregnancy loss.

Eight studies (Schoolcraft, 2010; Schoolcraft et al., 2011; Traversa et al., 2011; Forman et al., 2012; Yang et al., 2012; Forman et al., 2013a,b; Scott et al., 2013a) assessed trophectoderm biopsy, four blastomere biopsy (Wilton et al., 2003; Sher et al., 2009; Keltz et al., 2013; Mir et al., 2013) and three polar body biopsy (Sher et al., 2007; Fishel et al., 2011; Geraedts et al., 2011). Three studies directly compared trophectoderm with blastomere biopsy (Hodes-Wertz et al., 2012; Adler et al., 2013; Harton et al., 2013) and one study directly compared trophectoderm with polar body biopsy (Fragouli et al., 2010). Seven studies reported outcomes for fresh embryo transfers only (Sher et al., 2007; Fishel et al., 2011; Yang et al., 2012; Forman et al., 2013a; Keltz et al., 2013; Mir et al., 2013; Scott et al., 2013a), six studies reported on fresh and frozen embryo transfers (Geraedts et al., 2011; Forman et al., 2012; Hodes-Wertz et al., 2012; Adler et al., 2013; Forman et al., 2013b; Harton et al., 2013) and the remaining studies reported outcomes based for frozen embryo transfers only (Wilton et al., 2003; Sher et al., 2009; Fragouli et al., 2010; Schoolcraft, 2010; Schoolcraft et al., 2011; Traversa et al., 2011). Twelve studies reported live birth rates, but only six used the number of initiated cycles as the denominator in line with intention-to-treat principles (Sher et al., 2009; Schoolcraft, 2010; Fishel et al., 2011; Schoolcraft et al., 2011; Forman et al., 2013b; Scott et al., 2013a). No cost-effective analysis studies on comprehensive PGD-A techniques were identified.

Quality evaluation

The Downs and Black scores for the non-randomized studies ranged from 6 and 15 (mean: 12.8) indicating studies of poor to moderate quality, while the scores for the three randomized studies ranged from 18 to 20 indicating studies of high quality. The main weaknesses included lack of randomization, lack of patient relevant outcome measures, failure to analyse data according to intention-to-treatment principles, failure to include the outcome of subsequent frozen embryo cycles to calculate cumulative live birth rates, and failure to control for confounding variables either within the study design or analysis (e.g. variable numbers of embryos transferred between case and controls). Only 8 of the 19 studies included a control group. Moreover, while a number of the observational studies were well designed, such studies tend to overestimate treatment effects, and generally can only infer causation (Sroup et al., 2000). Furthermore, a number of studies, including two of the RCTs recruited...
patients from the same centre (Schoolcraft, 2010; Schoolcraft et al., 2011; Forman et al., 2012, 2013a,b; Scott et al., 2013a).

Clinical outcomes of PGD-A in young patients with good prognosis

Three RCTs (Yang et al., 2012; Forman et al., 2013b; Scott et al., 2013a) investigated the clinical outcomes of using PGD-A on young and good prognosis patients demonstrating a benefit in terms of implantation and pregnancy rates in this group (Fig. 2).

Yang et al.’s RCT (2012) evaluated aCGH-based PGD-A to select a single blastocyst for transfer in a fresh cycle (mean age 31.2 years). The ongoing pregnancy rate was significantly higher in the PGD-A group compared to the morphologically assessed embryo group (69.1 and 41.7%, respectively, $P = 0.017$). Although patients in the study were allocated to their intervention using a random number table, a power calculation was not reported and assessors (clinical and laboratory staff) were not blinded. Outcomes from subsequent frozen embryo transfer cycles were not included.

Scott et al.’s RCT (2013a) evaluated qPCR-based PGD-A for intended transfer of two Day 6 embryos versus morphologic-based embryo selection for transfer on Day 5. Patients were aged between 21 and 42 years (mean age 32.2 years) with at least two blastocysts available for biopsy. Overall, the results favoured the use of PGD-A producing higher clinical implantation rates (presence of a gestational sac) per transfer (79.8 and 63.2%, $P = 0.002$), and higher delivery rates per cycle (84.7 and 67.5%, $P = 0.01$). Outcomes from subsequent frozen embryo transfer cycles were not included.
Forman et al.’s RCT (2013a) used similar inclusion criteria as Scott et al. (2013a) and Yang et al. (2012) but the comparison was between elective single embryo transfer of an embryo that had been screened by PGD-A against two morphologically assessed embryos. Similar pregnancy rates were observed between the two groups (fresh transfer 63.9% and 70.5%, \( P = 0.4 \)) but there was a small reduction in the number of embryos being transferred in the PGD-A group (Schoolcraft et al., 2010). However, due to limitations with observational studies, the validity and generalizability of the findings makes it difficult to conclude whether PGD-A is clinically effective in this group of patients.

**Clinical outcomes for PGD-A on women of advanced maternal age**

Thirteen observational studies (Sher et al., 2007, 2009; Fragouli et al., 2010; Schoolcraft et al., 2010; Fishel et al., 2011; Geraedts et al., 2011; Schoolcraft et al., 2011; Traversa et al., 2011; Forman et al., 2012; Hodes-Wertz et al., 2012; Adler et al., 2013; Keltz et al., 2013; Mir et al., 2013) examined the use of PGD-A on women of advanced maternal age (mean age >35 years old). Among the studies which included a control group where embryos were selected based on morphology, all demonstrated consistently improved implantation rates in the PGD-A group (Schoolcraft et al., 2010; Fishel et al., 2011; Keltz et al., 2013; Fig. 3). For example, a small matched-cohort study of women with advanced maternal age by Schoolcraft et al., (2010) reported implantation rates after PGD-A of 68.9% compared with 44.8% in the control group. Four of the five studies reported significantly improved pregnancy rates, which, with the remaining study achieving equivalent pregnancies rates despite fewer embryos being transferred in the PGD-A group (Schoolcraft et al., 2010). However, due to limitations with observational studies, the validity and generalizability of the findings makes it difficult to conclude whether PGD-A is clinically effective in this group of patients.

**Effect of cell biopsy stage**

Seven observational studies (Wilton et al., 2003; Sher et al., 2009; Hodes-Wertz et al., 2012; Adler et al., 2013; Harton et al., 2013; Keltz et al., 2013; Mir et al., 2013) investigated the clinical outcomes of blastomere biopsy of cleavage stage embryos. Of these, only two studies had a
Figure 3 Implantation and pregnancy rates for PGD-A on women of advanced maternal age. "Data were extracted from text and tables of each of the included studies. Implantation rates: Schoolcraft et al. (2010). The implantation rates were 68.9% (62/90 embryos transferred) in the PGD-A group and 44.8% (134/299 embryos transferred) in morphological assessment group. The numerator was number of fetal hearts, the denominator was the number of embryos transferred. Fishel et al. (2011) included two control groups of patients with embryo selection based on morphological assessment. Morphological assessment Group 1 included patients with at least two miscarriages and morphological assessment Group 2 included patients with no previous successful pregnancy and at least two failed implantations after previous IVF attempts. The implantation rates were 27.7% (31 fetal hearts/112 embryos transferred) in the PGD-A group and 8.9% (7/78 embryos transferred) and 7.3% (8/109 embryos transferred) in the two morphological assessment groups, respectively. The numerator was number of fetal hearts, the denominator was the number of embryos transferred. Keltz et al. (2013): the implantation rates were 52.6% (30/57 embryos transferred) in PGD-A and 19.2% (253/1321 embryos transferred) in the morphological assessment group. The numerator was number of gestational sacs, the denominator was the number of embryos transferred. Forman et al. (2012) did not report implantation rate. Sher et al. (2009) did not report implantation rate. Pregnancy rates: Schoolcraft et al. (2010) reported biochemical pregnancy per cycle of 82.2% (37/45 cycles) in the PGD-A group and 84.0% (95/113 cycles) in the morphological assessment group. The numerator was positive pregnancy tests, the denominator was initiated cycles. Mean number of embryos transferred; PGD-A group = 2, Morphological assessment group = 2. Fishel et al. (2011) reported biochemical pregnancy per embryo transfer as 35.2% (38/108 embryo transfers) and 29.2% (14/48 embryo transfers) and 19.6% (11/56 embryo transfers) in the morphological assessment Group 1 and 2, respectively. Mean number of embryos transferred: PGD-A group = 1.04, Morphological assessment groups 1 and 2 = 1.63 and 1.94, respectively. Sher et al. (2009) included two morphological assessment groups of patients with embryo selection based on morphological assessment. Morphological assessment Group 1 included patients who had frozen—warmed transfer cycles and morphological assessment Group 2 included patients who had fresh blastocyst transfers. The clinical pregnancy rate per transfer was 64% (36/57 embryo transfers) in the PGD-A group, 37% (43/117 embryo transfers) and 41% (132/329 embryo transfers) in morphological assessment Groups 1 and 2, respectively. The numerator was the presence of a gestational sac on ultrasound, the denominator was number of embryo transfers. Mean number of embryos transferred: PGD-A group = 1.7, Morphological assessment groups one and two = 2.2 and 1.9, respectively. Keltz et al. (2013) reported clinical pregnancy per fresh cycle as 69.2% (27/39 cycles) in the PGD-A group and 43.9% (173/394 cycles) in the morphological assessment group. The numerator was sonographically confirmed gestational sac, the denominator was number of fresh cycles. Mean number of embryos transferred: PGD-A group = 1.46, Morphological assessment group = 3.35. Forman et al. (2012) reported the ongoing pregnancy rate per embryo transfer (fresh and frozen transfer) as 55.0% (77/140 embryo transfers) in the PGD-A group and 41.8% (76/182 embryo transfers) in the morphological assessment group. The ongoing pregnancy rate was defined by live births and sustained pregnancies (beyond the first trimester) per embryo transfer. Single embryo transfer was used for all embryo transfer cycles, PGD-A (n = 140) and morphological assessment group (n = 182).
per embryo transfer than those who had morphological assessment alone (24.1 versus 16.7 and 12.5%, P < 0.0001).

Three observational studies (Hodes-Wertz et al., 2012; Adler et al., 2013; Harton et al., 2013) directly compared the clinical outcomes of blastomere versus trophectoderm biopsy-based PGD-A. All studies reported an increase in the ongoing/clinical pregnancy rates with trophectoderm biopsy although there was no adjustment for significant differences between patient characteristics and clinical practice, such as the quality of embryos or the number of embryos transferred. Furthermore, the voluntary reporting of outcomes in two of the studies could have contributed to substantial selection bias (Hodes-Wertz et al., 2012; Harton et al., 2013). A study by Fragouli et al. (2010) which compared the clinical outcomes of CGH on polar body to trophectoderm biopsy in 32 poor prognosis women of advanced maternal age (mean age 38.4–39.8 years), also reported improved pregnancy rates among patients who had trophectoderm biopsy-based PGD-A. However, the lack of prospective randomization and differences in patient and practice characteristics between groups (e.g. number of embryos transferred, variation in biopsy methodology) also limit the validity of these findings.

Discussion

The purpose of this review was to evaluate the best available evidence on the effectiveness of PGD-A to improve clinical outcomes. Nineteen studies comprising three RCTs and 16 observational studies were identified.

The main limitation of the synthesis of the evidence was the paucity of published high-quality studies. The 19 studies identified exhibited significant variation in study design and quality. According to the Downs and Black scoring system, only the three RCTs were scored as high quality, with 14 of the remaining studies scored as moderate quality and two as poor quality.

The first criteria for assessing embryo selection techniques is whether the implantation rate of the selected embryo is higher than the implantation rate when conventional morphological assessment of the embryo is used. Despite the limitations of the studies presented in this review, the data presented in Figs 2 and 3 demonstrate, a consistent effect of PGD-A in improving implantation rates.

The evidence from two of the three RCTs also indicates that PGD-A results in an improved clinical pregnancy rates in the index cycle. However, there is an argument that outcomes from subsequent cycles should be included in any analysis of PGD-A because embryo selection may not improve cumulative live birth rates given recent advances in cryopreservation which allow almost all embryos to be cryopreserved and transferred in subsequent cycles (Mastenbroek et al., 2011), and that a proportion of euploid embryos are likely to be excluded as a result of the PGD-A process (Scott et al., 2012). There is also evidence to suggest that all embryos should be cryopreserved and transferred in an unstimulated cycle to improve perinatal outcomes of ART conceived babies (Pelkonen et al., 2010; Wennerholm et al., 2013).

Despite the promise of PGD-A being able to improve live birth rates in women of advanced maternal age and those suffering from repeated implantation failure and recurrent miscarriage, only observational studies were identified in these patient groups (Wilton et al., 2003; Sher et al., 2009; Fragouli et al., 2010; Schoolcraft et al., 2010; Schoolcraft et al., 2011; Adler et al., 2013). While findings from these studies tend to favour PGD-A, observational studies cannot infer causality.

The optimal cells to biopsy for PGD-A is a source of considerable debate. The majority of studies identified in this review used trophoetoderm biopsy, which is in-line with a recent report by the European Society of Human Reproduction and Embryology (ESHRE) which reported a shift away from blastomere biopsy towards polar body and trophoetoderm biopsy (Harton et al., 2011). This has been fuelled by a report that cleavage stage biopsy damages embryos as evidenced by lower implantation rates (Scott et al., 2013b). Blastocysts are known to have higher implantation rates than cleavage stage embryos so it is of some concern that the implantation rates of control cleavage and blastocyst embryos were the same in this study. This is possibly a vagary of the small number of implantations analysed, particularly in the cleavage embryo group, and raises some questions about the conclusions of this study. Furthermore, while a randomized trial by the same group (Scott et al., 2012) found significantly improved predictive values for implantation for trophoetoderm biopsy compared with blastomere biopsy, a RCT is needed to confirm the clinical relevance of these findings.

Furthermore, it is acknowledged that many patients who are referred for PGD-A treatment have a poor prognosis and may not be able to

Implantation rate

The implantation rate is a measure of the capacity of an individual embryo to implant and is generally expressed as the number of fetal hearts/100 embryos transferred. Five studies, two RCTs (Forman et al., 2013b; Scott et al., 2013a) and three matched cohort studies (Schoolcraft et al., 2010; Fishel et al., 2011; Keltz et al., 2013), presented sufficient data to enable the implantation rate to be compared with a control group and these studies are presented in Figs 2 and 3. In four of the five studies, there was a significant increase in the implantation rate per embryo when the embryo had been screened with PGD-A. The remaining study reported a non-statistically significant improvement (P = 0.08) in the implantation rate in the PGD-A group (Forman et al., 2013b). Despite the heterogeneity of the studies, it is reasonable to conclude from this that an embryo that has been screened with PGD-A has a higher implantation rate than an unscreened embryo.

Pregnancy loss

Fourteen studies reported pregnancy loss (variously defined as miscarriage, biochemical loss, missed abortion) as an outcome measure (Sher et al., 2007; Sher et al., 2009; Fishel et al., 2011; Geraedts et al., 2011; Schoolcraft et al., 2011; Hodes-Wertz et al., 2012; Forman et al., 2012; Yang et al., 2012; Adler et al., 2013; Forman et al., 2013a,b; Harton et al., 2013; Keltz et al., 2013; Mir et al., 2013). Pregnancy loss ranged from 2.6% (Yang et al., 2012) to 31.6% of pregnancies (Fishel et al., 2011) in the PGD-A group. Two of the three RCTs (Yang et al., 2012; Forman et al., 2013a,b) reported miscarriage rates and found no significant difference between patients who had PGD-A and those who had embryos selected by morphology alone. Harton et al. (2013) undertook a retrospective review of the miscarriage rates between biopsy at cleavage and blastocyst stage and by age group, and found no significant difference between blastocysts and cleavage stage euploid embryos transferred in women up to 42 years, averaging 9.9% for Day 3 and 7.9% for blastocyst stage. However, two observational studies with control groups of patients who received morphologically selected embryos reported significantly lower miscarriage rates after PGD-A (Sher et al., 2009; Forman et al., 2012).
produce sufficient embryos for blastocyst biopsy (Gleicher et al., 2014) and it is incumbent on ART clinics to provide both types of biopsy and on PGD laboratories to have the expertise to test both trophectoderm and single cell biopsies. There also exists much debate around the effectiveness of polar body biopsy due to the incidence of post-zygotic aneuploidy (Capalbo et al., 2013a; Christopoulou and Handside, 2013; Fragouli and Wells, 2013) which occurs after the diagnostic testing is performed.

High-quality studies of embryo selection techniques

Good study design coupled with appropriate statistical analysis is the cornerstone of evidence-based clinical decisions in medicine. However, particular challenges exist in designing and analysing efficacy and effectiveness trials of embryo selection techniques, such as PGD-A, metabolomics, time-lapse imaging (Daya, 2006; van Gelder and Nijs, 2011). The RCT remains the best study design to establish the efficacy of an intervention and is the best method for assessing emerging technologies, such as embryo selection techniques, before more generalizable effectiveness trials are undertaken. Key principles underlying the ‘gold-standard’ label given to RCTs include defining a clear a priori test hypothesis based on clinically meaningful questions, obtaining an appropriate sample size to detect a difference in outcomes, and using blind randomization to ensure similarity among intervention groups. In addition, in embryo selection trials it is important to minimize bias by defining a cycle as the being from the beginning of stimulation (not at oocyte retrieval or embryo transfer), to report success as birth of a live born singleton child, and to control for confounders such as number of embryos transferred and previous experience of ART in either the study design or analysis. Intermediate end-points such as implantation rates are informative but the cumulative rate of live born singleton children from fresh and subsequent frozen embryo cycles provides the most meaningful outcome measures from a clinical and cost-effectiveness perspective. The time required to achieve a live born child is also of importance to patients. Many of these principles were not observed in the three RCTs reviewed or in previous FISH RCTs. Furthermore, ensuring that the clinical and laboratory protocols are highly standardized among centres and quality assurance programs are incorporated into all trial aspects is needed to minimize bias associated varying expertise among centres and single centre trials. Such measures also provide confidence in the reproducibility of results, a problem with previous RCTs using FISH for PGD-A.

Data analysis using intention-to-treat principles is also critical in preserving the unbiased composition of the randomized groups, such that all patients randomized to a group are analysed together. In intention-to-treat analysis, any non-compliance with the original assignment due, for example, to withdrawing from treatment or crossing over to the alternate intervention, are reflected in outcomes measures assigned to the original group. This analytical approach reflects the clinical effectiveness of the interventions and tends to provide a more conservative and cautious estimation of the treatment effect.

Cost-effectiveness studies of embryo selection techniques

No cost-effectiveness analyses of PGD-A using comprehensive techniques were identified. This is an important consideration as PGD-A technologies are currently expensive, and have potentially significant economic implications for funders and patients. Theoretically PGD-A has the potential to be cost-effective if it reduces the number of ART cycles needed by a couple to achieve a live birth. This is particularly true if it supports single embryo transfer (Polinder et al., 2008; Chambers et al., 2014) in fresh and subsequent less expensive thaw cycles with lower miscarriage rates. PGD-A, therefore has the potential to maximize both the clinical and economic benefits of fewer ART cycles resulting in a singleton baby in a shorter period of time.

PGD-A has been criticized on the grounds that it is an expensive intervention, however, it may also limit costs and emotional burden by preventing the storage and later transfer of aneuploid, non-viable embryos and theoretically reduce the time to pregnancy. Therefore, assessing the cost-effectiveness of an embryo selection technique requires not only consideration of the cost of the PGD-A technique and success rates following an the first embryo transfer cycle, but must also include the likelihood of any transfer occurring in the stimulation cycle, the number of remaining embryos, the costs and effectiveness of the cryostorage programme and impact of time to pregnancy. In other words any cost-effectiveness analysis should consider the incremental differences in societal costs and singleton live births between a strategy of repeated cycles transferring untested embryos and a strategy of transferring euploid only embryos.

The available comparative data on PGD-A are very limited in these respects. The three studies from the New Jersey group (Forman et al., 2012, 2013b; Scott et al., 2013a) report remarkably small changes in the number of available vitrified blastocysts after embryo selection with PGD-A. In contrast, the data of Schoolcraft et al. (2010) suggest that the number of remaining embryos after the initial transfer could change from 4.4 embryos to 1.1 embryos through the use of PGD-A. Of the other five studies with control groups, none report comparative data on the number of vitrified embryos with and without PGD-A. None of the studies, so far published, in the clinical evaluation of PGD-A provide sufficient data to fully understand the clinical and economic effectiveness of this technology.

An important aspect of cost-effectiveness of PGD-A that needs to be considered is the potential for promoting single embryo transfer. The study by Forman et al. (2013b) demonstrated similar pregnancy rates when transfer of a single, PGD-A screened embryo was compared with transfer of two morphologically screened embryos. This raises the possibility that PGD-A may be useful in promoting single embryo transfer in parts of the world where this is not yet established practice, thus reducing the considerable morbidity and economic burden of multiple pregnancy (Chambers and Ledger, 2014; Chambers et al., 2014).

Conclusion

This review did reveal potential benefits of using PGD-A techniques over morphology-based selection of embryos, in particular, demonstrating that a PGD-A screened embryo has a higher implantation rate than a morphologically screened embryo. Three RCTs in young good prognosis patients demonstrated clinical benefit in terms of clinical pregnancy rates and the use of single embryo transfer, however, studies in other patient groups were limited to observational studies.

The previous enthusiasm for adopting FISH based on observational studies which was later dampened by several RCTs (Mastenbroek et al., 2007; Blockeel et al., 2008; Staessen et al., 2008; Schoolcraft et al., 2011).
et al., 2009; Debrock et al., 2010) should serve as a cautionary tale against using observational data to guide clinical practice. Well-designed studies that take into account all the elements necessary to evaluate the clinical and cost-effectiveness of PGD-A techniques in the clinic setting are overdue.

Authors’ roles
E.L. was involved in the design and acquisition of data, analysis and interpretation of data, construction of tables and figures, and drafting and critical revisions of the manuscript. G.M.C. led the conception and design of the study, analysis and interpretation of data, drafting and critical revisions of the manuscript. P.I. and L.W. were involved in the study concept, analysis of data and critical revisions of the manuscript. All authors approved the final manuscript.

Funding
No specific funding was used to undertake this study.

Conflict of interest
P.I. is the Medical Director of the IVF Clinic, IVF Australia and has a financial interest in the parent group, Virtus Health. L.W. is an employee of and a shareholder in Virtus Health which is a provider of PGD services in Australia. G.M.C previously received grant support to her institution from the Australian Government, Australian Research Council (ARC) Linkage Grant No. LP 1002165; ARC Linkage Grant Partner Organisations were IVF Australia, Melbourne IVF and Queensland Fertility Group. E.L. does not report any conflict of interest.

References
Ajduk A, Zernicka-Goetz M. Advances in embryo selection methods. F1000 2010;
Garriss JG, Colls P, Ferry KM, Zheng X, Garriss MG, Munné S. Effect of infertility, maternal age, and number of previous miscarriages on the
blastocest biopsy does not: a randomized and paired clinical trial. *Fertil Steril* 2013b;**100**: 624–630.


