Preimplantation genetic screening: back to the future

Sebastiaan Mastenbroek* and Sjoerd Repping

Center for Reproductive Medicine, Academic Medical Center, University of Amsterdam, Q3-119, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands

*Correspondence address. E-mail: s.mastenbroek@amc.uva.nl

Submitted on April 1, 2014; resubmitted on May 9, 2014; accepted on June 3, 2014

ABSTRACT: All agree that in hindsight the rapid adoption of preimplantation genetic screening (PGS) using cleavage stage biopsy and fluorescence in situ hybridization (FISH) in routine clinical practice without proper evaluation of (cost-)effectiveness basically resulted in couples paying more money for a less effective treatment. Now, almost 20 years later, we are on the verge of a new era of PGS. But have things really changed or are we simply going back to the future?

Key words: IVF/ICSI / PGS / aneuploidy / efficacy / randomized controlled trials

The discovery of PGS

The main goal of PGS, which is the biopsy of one or more cells from a pre-implantation embryo followed by the ploidy analysis of these cells and finally transfer of those embryos deemed to be euploid, has always been the improvement of IVF success rates. The rationale for PGS results from a combination of (i) an increasing number of aneuploidies in clinical miscarriages of women of advanced maternal age, (ii) the fact that aneuploidies are mostly incompatible with life (Hassold and Hunt, 2001; Lintsen et al., 2007). The first pregnancies after PGS were reported in 1995, and in the following decade numerous positive results for PGS were published (Verlinsky et al., 1995). Its logical rationale and these initial positive publications resulted in widespread introduction of PGS in routine clinical practice (Moutou et al., 2014). PGS was increasingly used not only for advanced maternal age, for which the above explained rationale applies, but also for women with repeated implantation failure and even for women with a good prognosis, as aneuploidies were detected in the embryos of these women as well (Mastenbroek et al., 2011a). The common methodology employed for PGS in the first 15 years of its existence was blastomere aspiration of embryos on the third day after fertilization, followed by fluorescence in situ hybridization (FISH) analysis of the aspirated blastomere(s).

Initial analysis of effectiveness of PGS

Although the rationale for PGS seemed logical at the time, in hindsight the initial publications certainly did not justify its widespread adoption. The available literature consisted of case reports, cohort studies and non-randomized comparative studies; the included numbers of patients were small; and often suboptimal outcome measures such as implantation rate were used as the primary outcome for these studies (Gianaroli et al., 1997; Munne et al., 1999, 2003; Kahraman et al., 2000). It needs no discussion that, when investigating the efficacy of subfertility treatment, ongoing pregnancy rate or preferably live birth should be the main outcome measure as this is the ultimate goal of the woman undergoing treatment (Barlow, 2003; Daya, 2003; Vail and Gardener, 2003). Analyses should of course be based on the number intended to treat, and should use the right unit of analysis, e.g. per woman or per started cycle, as the design of the study should fit the clinical decision moment, which is before initiation of treatment. Implantation rate is calculated by dividing the number of implanted embryos by the number of transferred embryos, which makes it an inappropriate outcome measure in studies of PGS, because with PGS many women do not have a transfer or if they do, there are fewer embryos to transfer (Mastenbroek et al., 2005). Whereas PGS, or any other embryo selection technique could indeed select better embryos, resulting in improved implantation rates or pregnancy rates per transfer, it could at the same time result in less transfers or less embryos to transfer by selecting out embryos, resulting in lower pregnancy rates per woman or per started cycle.

It took almost a decade after its clinical introduction before the first properly designed randomized controlled trials (RCTs) on PGS were published. To the shock of many, these did not indicate a benefit of PGS, but a significantly decreased chance on ongoing pregnancy in comparison with IVF without PGS (Staessen et al., 2004; Mastenbroek et al., 2007). These findings started a fierce debate on the use of PGS with proponents of PGS claiming malpractice of PGS in the conducted trials and opponents claiming that these were the first properly designed studies.
and therefore the best evidence available. Soon thereafter, more RCTs were published, and none showed a positive outcome after PGS (Twisk et al., 2006; Mastenbroek et al., 2011a). Policy statements were issued by the American Society for Reproductive Medicine (ASRM), the British Fertility Society (BFS), and the American College of Obstetricians and Gynecologists (ACOG), all stating that there was no evidence that supported the routine clinical use of PGS (Practice Committee of the Society for Assisted Reproductive Technology and Practice Committee of the American Society for Reproductive Medicine., 2007; American College of Obstetricians and Gynecologists, 2009). This all led to a decline in the use of PGS, although the technique was never completely abandoned (Moutou et al., 2014).

Now, several years later, it is widely recognized, even by the most firm critics of the first trials, that PGS as it had been applied for over a decade, with FISH analysis of blastomeres aspirated on Day 3 of embryo development, was of no use (Brown, 2014). The mosaic nature of human preimplantation embryos at this developmental stage, i.e. the observation that not all cells have the same chromosomal constitution, and the technique used, i.e. FISH analysis of a limited number of chromosomes with limited accuracy, are considered the main reasons for the inefficacy of PGS (Scriven and Bossuyt, 2010; van Echten-Arends et al., 2011).

The rediscovery of PGS

Was this then the end for PGS? No, it seems not. The disappointing results of the first generation of PGS methodology led to the development of new PGS methods. Biopsy is now either at the polar body or at the blastocyst stage of development, following the assumption that this is less detrimental to the embryo and that this avoids the negative effect of mosaicism on efficacy. Ploidy status is now determined using CGH arrays and SNP arrays, allowing the analysis of all chromosomes with proclaimed greater accuracy than FISH (Wells et al., 2008). And again, just as in the 1990s, these new forms of PGS are rapidly being introduced into routine clinical practice. It is difficult to find exact numbers, as the esteemed data collections always lag behind a few years due to logistic reasons (Centers for Disease Control and Prevention et al., 2012; Moutou et al., 2014). However, it has been suggested that over the past 3 years, the number of PGS cases in the USA has increased by >30% per year, resulting in the application of PGS in about 8% of all IVF treatments (Brown, 2014). One company that provides array analysis of biopsied embryos claims on its website to have performed over one quarter of a million clinical PGS biopsies with their array platform alone (Website Blue Gnome, 2014).

Effectiveness of new PGS methods

The apparent rapid clinical introduction of the new PGS methods again appears to be based on a few case reports, cohort studies and non-randomized comparisons. And yet again, these studies are unfortunately very small and often use inappropriate outcome measures such as implantation rate (Schoolcraft et al., 2010, 2011; Rubio et al., 2013). But in addition, and in contrast to the early days of PGS, three RCTs on these new PGS methods have been published, allegedly showing a positive outcome after PGS (Yang et al., 2012; Forman et al., 2013; Scott et al., 2013a). These trials, as well as some of the data from the other comparative non-randomized studies, are actively being used by proponents of PGS and commercial companies involved in PGS to widely stimulate the use of the new PGS methods. But how solid is the claim that the new PGS methods are really of clinical benefit? The truth is that the community seems not to have learned from the PGS experience in the 1990s and we are simply going back to the future.

Of course by itself an RCT is considered to be high quality evidence (Guyatt et al., 2008, 2011; Higgins and Green, 2009). However, the level of evidence of an RCT decreases in case of serious limitations of study quality (such as lack of allocation concealment, lack of blinding, large loss to follow-up, no intention to treat analysis, failure to report proper outcomes), in case of important inconsistencies in comparison to other studies (heterogeneity across studies), in case of uncertainty about directness (evidence comes from a different research question), or in case there is a high probability of reporting bias (Guyatt et al., 2008, 2011; Higgins and Green, 2009). Unfortunately, there are some significant issues with the current RCTs on the new PGS methods that seriously downgrading the level of evidence they provide. We limit ourselves here to briefly summing up the most relevant issues to support this statement. An extensive review and possible meta-analysis will soon be published in an updated version of the PGS Cochrane review (Twisk et al., 2006).

The first RCT (Yang et al., 2012) was not registered in a trial register, performed no power calculation and included a seemingly small number of patients, which are factors that seriously downgrade the level of evidence this trial provides. All patients received a transfer and only good prognosis patients with a mean age of 31 years were included, thereby limiting the generalizability of the trial to other patient groups. The number of embryos cryopreserved and the result of subsequent cryo-cycles were not reported. This is a potential cause of bias as it can be expected that more embryos are cryopreserved in the control group, and these embryos are likely to contribute to the cumulative pregnancy rate, i.e. the pregnancy rate per started cycle including fresh and cryo-transfers.

The second RCT (Forman et al., 2013) included patients under 43 years of age with at least two blastocysts on Day 6 and was designed as a non-inferiority trial. One of the reasons for this design was that the trial did not only evaluate the effect of selection by PGS, but also the effect of different transfer strategies: single embryo transfer (SET) in the PGS group and double embryo transfer (DET) in the control group. The latter limits the generalizability of the outcomes of the trial in regard to PGS efficacy. Also, the design of a non-inferiority trial is inappropriate as PGS is a costly and invasive (and potentially harmful) addition to an IVF treatment, so it should at least add benefit to the IVF treatment. What is more is that the study was designed to exclude a difference in favour of the control group by >20%, so PGS could practically result in almost 20% less pregnancies, and it would still be concluded that outcomes in both groups are equal. Also in this trial, the results of the cryo-cycles were not taken into account.

The third RCT (Scott et al., 2013a) included good prognosis patients with a mean age of 32 years with two or more high-quality blastocysts available, thereby limiting the generalizability of the trial to other patient groups. The day of transfer differed between groups, there appeared to be a slightly skewed randomization, all patients received a transfer, and also in this trial the outcome of subsequent cryo-cycles was not included. The primary end-point was sustained implantation rate, which, as stated before, is an inappropriate outcome in PGS efficacy...
trials. The study was also powered on a difference in implantation rates, causing the trial to be severely underpowered in regard to an outcome such as pregnancy rate.

Based on just these three trials and their significant limitations, it is simply too early to claim that there is enough evidence to justify the use of any new PGS method in clinical practice for any group of patients. As medical professionals, we need to learn from our joint past experiences with PGS and obtain high-quality data before clinical implementation.

Moral obligation

The massive reintroduction of PGS shows that the medical professionals offering PGS either are unaware of the true value of the available data or are driven by other motives. We feel it is important to acknowledge the financial aspects associated with PGS. In many centres worldwide, IVF is a commercial business and PGS is commercially very attractive as it can significantly increase the turnover of a clinic. Centres have a natural tendency to simply copy what they think are successful innovations and patients are eager to attempt anything that might increase their likelihood on a healthy child. Competition is fierce, which often leaves no time or money for extensive research prior to clinical introduction, and forces the centre to do what the patient asks, as otherwise the couple leaves for another centre and business is lost. And if a centre wants to do a proper evaluation, funding is difficult to obtain and regulatory rules for conducting clinical trials are increasingly stringent. The liability in this is often left to the patient by means of ‘informed consent’. But it is simply too easy to just hide behind the demand of the patient. The problem here is that a patient could agree with being treated with a technique of unknown effectiveness, but the clinician still remains responsible for what he or she does (Dondorp and de Wert, 2011). Besides, it can be questioned whether all patients will ever be able to understand all of the complexities concerning PGS. The moral obligation of making innovative treatment the object of research as soon as possible is also reflected by authoritative documents, including the Helsinki declaration and the American Belmont report (Dondorp and de Wert, 2011). Apart from this responsibility towards the consenting patient, in the case of reproductive medicine the responsibility also extends to the welfare of any children who will be born as a result of the treatment (Pennings et al., 2007).

Is there any future for PGS?

Besides the absence of proper evidence that current day PGS is of benefit, an even greater concern with PGS is that it is unlikely to ever work at all. As stated before, the biological rationale of PGS does sound very logical. But is it really? Recent insights and developments, especially in the effectiveness of cryopreservation, seriously challenge the concept of PGS, or even more broadly, the value of any selection method, when it comes to the improvement of pregnancy rates.

Cryopreserved embryos for a long time had a reduced chance of implanting compared with fresh embryos, making proper selection of embryos for fresh transfer a necessity to optimize success rates in IVF. But with increasing success rates using cryopreserved embryos in recent years, evidence is now accumulating that all embryos in an IVF cycle can be cryopreserved and transferred in subsequent cycles without impairing, and maybe even improving, the cumulative pregnancy rate of that IVF cycle (Mastenbroek et al., 2011b, Wong et al., 2014). In such a freeze-all scenario, no selection method will ever lead to improved live birth rates, because, by definition, the live birth rate per stimulated IVF cycle can never be better than after serial transfer of all available embryos in that cycle. In fact, selecting out and discarding embryos with less than 100% accuracy, as is done in PGS, could only lower the live birth rate after IVF (Mastenbroek et al., 2011b). Embryo selection should therefore not be used to select out embryos, but only to determine the order in which the embryos will be transferred, as the time to pregnancy can be improved by embryo selection, if embryos with the highest implantation potential are transferred first.

Then, there are other uncertainties. A recent Cochrane review on Day 2/3 versus Day 5/6 transfer in IVF showed an increased live birth rate in favour of Day 5/6 transfers (Glujovsky et al., 2012). But when the meta-analysis included frozen cycles, a favourable live birth rate for Day 2/3 transfers was observed (Glujovsky et al., 2012). This finding needs further confirmation as data were limited, but if true then this would not plea in favour of PGS as the new PGS methods predominantly make use of Day 5/6 transfers. Of additional concern is the potential harm to the embryo caused by the biopsy procedure, as this harm cannot yet be fully excluded (De Vos and Steirteghem, 2001; Scott et al., 2013b). Also, the exact prevalence of mosaicism at the blastocyst stage using the new methods for analysis is yet unknown, and with that a potential detrimental effect of mosaicism on PGS effectiveness cannot be fully excluded (van Echten-Arends et al., 2011). Large studies on re-analysis of embryos after PGS using the new methods have still to be published. And for most new methods, diagnostic accuracy has yet to be determined. PGS is a black and white test that divides embryos into those that can be transferred and those that should be discarded. When this is done with <100% accuracy, either due to technical failure or mosaicism, potentially viable embryos are discarded and this will potentially even lower the pregnancy rate after IVF. For PGS using FISH, this appeared to be one of the problems (Scriven and Bossuyt, 2010). For the new methods of analysis, which are generally assumed to be more accurate, surprisingly limited data on diagnostic accuracy is available considering the fact that they are already being used clinically. For most methods, it still has to be determined what the sensitivity (correctly identified embryos), specificity (correctly identified diploid embryos), positive predictive value (proportion of aneuploid embryos) and negative predictive value (proportion of diploid embryos) actually are.

These doubts make high-quality evidence even more needed before routine clinical application of PGS. PGS should still only be offered by means of rigorously designed, conducted and reported randomized trials, if at all.

Are such trials currently being conducted? No, not really. The trial registers show a cumulative total of ten registered trials on PGS efficacy ever conducted or on their way (see the Supplementary data for details). Six of those are on the original PGS method, using cleavage stage biopsy and FISH, and have been completed and were discussed above. The other four trials are on the new PGS methods. Two of these have been completed and their relevancy is discussed above (Forman et al., 2013; Scott et al., 2013a). One trial was terminated due to slow enrolment and transitioning to a new method for the analysis of the aspirated cells. This leaves only one trial that is still on its way. Notwithstanding the relevancy of this trial which is conducted under the auspices of ESHRE, the trial evaluates PGS on polar bodies, where currently almost all PGS treatments make use of blastocyst biopsy, a fundamentally different method.
Healthcare evaluation in a broader scope

Reproductive medicine is not the only field of medicine in which evidence on the effectiveness of commonly used treatments is lacking. Recent analysis on the effectiveness of 3000 treatments by the BMJ Evidence Center indicates that about half of these treatments are of ‘unknown effectiveness’ (Clinical Evidence from the BMJ Evidence Centre, 2014). This urges medical professionals to conduct proper studies to determine whether these treatments are of value and should thus be continued to be offered to patients, or if they are not and should no longer be offered. Such effectiveness studies should be conducted with extreme scientific rigour and precision as weaknesses in design, conduct, analysis, and reporting can lead to incorrect conclusions and wastes of essential initiatives and funding (Ioannidis et al., 2014; Glasziou et al., 2014). In this light, additional attention has to be paid to stakeholders with potential commercial interests as economic forces are well known to have an effect on the design and reporting of clinical studies with the aim to (mis)use science for marketing purposes (Gotzsche et al., 2007; Psaty and Kronmal, 2008; Ross et al., 2008; Krumholz et al., 2011; Macleod et al., 2014). In our view evaluating the (cost-)effectiveness of medical treatments is by far the greatest challenge in current day medicine, especially in an era where health care costs continue to increase to the extent where they are the number one item of expense for many governments across the globe.

Conclusion

Let us not be mistaken. An estimated 5 million people walk the earth as a result of assisted reproduction. Great progress has been made in the 35 years that have gone by since the first birth after IVF, and this would not have been without the vision and hard work of the many professionals in our field of work. But our field of work also comes with great responsibility. Professionals in reproductive medicine are responsible for what they do on a day-to-day basis, and are responsible for proper introduction and use of new treatments. In hindsight it is clear that an error was made by quickly and massively adopting PGS using cleavage-stage biopsy and FISH in routine clinical practice without proper evaluation of effectiveness and cost-effectiveness, basically resulting in women paying more money for a less effective treatment. The word on the usefulness of the new PGS methods is still out, as the rationale of PGS is all but self-evident, and as proper trials on the efficacy of these new methods are lacking. But the new PGS methods are increasingly being used in routine clinical practice, and proper evaluation of their effectiveness and cost-effectiveness is being forgotten or ignored once again.

Supplementary data

Supplementary data are available at http://humrep.oxfordjournals.org/.

Authors’ roles

S.M. and S.R.: conceptualization of the manuscript. S.M.: drafting of the article. S.M. and S.R.: critical revision of the manuscript and approval of the final manuscript.

Funding

No funding was used.

Conflict of interest

None declared.

References


Barlow DH. The design, publication and interpretation of research in Subfertility Medicine: uncomfortable issues and challenges to be faced. Hum Reprod 2003; 18:899–901.


Centers for Disease Control and Prevention, American Society for Reproductive Medicine, and Society for Assisted Reproductive Technology. 2010 Assisted Reproductive Technology Success Rates: National Summary and Fertility Clinic Reports, 2012.


Psaty BM, Kronmal RA. Reporting mortality findings in trials of rofecoxib for Alzheimer disease or cognitive impairment: a case study based on documents from rofecoxib litigation. JAMA 2008;299:1813–1817.


