Chapter 10
General discussion and future perspectives
Chapter 10

General discussion and future perspectives

This general discussion offers a brief reflection on the results of the studies included in this thesis and their implications for clinical practice. Next, changes in fertility practice and their consequences for follow-up studies are discussed, followed by methodological considerations and suggestions for future research. The general discussion ends with concluding remarks.

Reflection on the results and implications for clinical practice

This thesis has provided evidence for parental subfertility leading to a higher risk of some specific congenital anomalies (chapter 2) and worse perinatal outcomes (chapter 3). The detected specific congenital anomalies include abdominal wall defects, penoscrotal hypospadia, right ventricular outflow tract obstruction, and methylation defects causing imprinting disorders. The worse perinatal outcomes include a shorter duration of pregnancy, lower birthweight, fetal growth restriction, and perinatal mortality. These findings are in line with the increasing amount of literature suggesting it is the underlying subfertility rather than assisted reproductive techniques (ART) that lead to problems in the offspring (Basso and Baird 2003, Davies et al. 2012b, Draper et al. 1999, Jaques et al. 2010, Raatikainen et al. 2010, Thomson et al. 2005, Zhu et al. 2006). At present, it is unknown how subfertility may lead to suboptimal child health and development. The search for this answer is complicated by the large variety of types of subfertility, for example, causes of subfertility are found in semen abnormalities (genetic quality, volume or motility), blocked or absent Fallopian tubes, hormonal dysregulation, endometriosis, uterine problems and subfertility with unknown cause. It is possible that acquired subfertility due to blocked Fallopian tubes after pelvic inflammatory disease will have different consequences for the health of offspring than subfertility of unknown cause. In the latter case epigenetics could play a more pronounced role. Perhaps the gametes of subfertile couples are more prone to epigenetic instability, resulting in both subfertility of the parents and adverse health outcomes in their children (Horsthemke and Ludwig 2005, Ludwig et al. 2005). The severity of the subfertility, as reflected by a longer time to pregnancy (TTP), is also related with worse outcomes in the offspring. This thesis describes a positive relation between a longer TTP and congenital abnormalities (chapter 4) and less optimal neurodevelopment at the age of 2 years (chapter 5).

Nevertheless, this thesis also provides some evidence that ART itself plays a role in increasing the risk of fetal growth restriction (chapter 3) and suboptimal cardiometabolic outcome (chapters 6 and 7) in the offspring. These findings are also in line with several other studies, especially a link between the use of ART and cardiometabolic alterations has been suggested before (Belva et al. 2007, Belva et al.
This thesis, for the first time, provides evidence that the effect of ART is caused by a direct (non-mediated) effect of ovarian hyperstimulation on higher systolic blood pressure (BP) levels and thicker subscapular skinfolds in 4-year-old children (chapters 6 and 7). Small differences in BP are not important at an individual level, but as up to 5% of newborns in Europe are now conceived through ART (Ferraretti et al. 2013), these differences may well be important at a population level. Furthermore, as it is known that BP tracks into adulthood, the differences may become more pronounced in later life (Chen and Wang 2008). It is not known how ovarian hyperstimulation might induce cardiometabolic alterations in the resulting offspring. The ‘developmental origins of health and disease hypothesis’ states that an adverse environment in early life may induce adaptive changes that may result in diseases in later life (Roseboom et al. 2000, Watkins and Fleming 2009). Perhaps ovarian hyperstimulation changes the environment of the oocyte and/or embryo in a way that epigenetic modifications can occur in systems that regulate BP (Fleming et al. 2004). It is noteworthy that widespread epigenetic modifications have been demonstrated in phenotypically normal children born following IVF (Batcheller et al. 2011).

The implications for clinical practice of these findings mainly involve the counselling of subfertile couples requesting ART. At present, the evidence for ovarian hyperstimulation having a role in poorer cardiometabolic outcome is not firm enough to adjust ART treatments. It is also too early to acknowledge this specific association in the counselling. The following three topics are recommended for counselling:

1. Children born following ART are generally healthy. It should be noted that solid information on the health of ART offspring beyond adolescence is lacking, but that ART itself does not seem to affect neurological, mental and behavioural development.
2. The risk of perinatal adversities, birth defects and cardiometabolic adversities is slightly increased after ART, although this message should also reassure the prospective parents that the absolute risk increases are small.
3. The increased risks are, to a large extent, mediated by factors associated with the underlying subfertility. This information is not only of concern for couples consulting a fertility specialist. A more general message for couples in their reproductive years is that they should not postpone starting a family for too long, since advanced maternal age is strongly related to diminishing fecundity.

This thesis did not demonstrate an effect of embryo biopsy as performed in preimplantation genetic screening (PGS) on BP or anthropometrics (chapter 9). Although PGS is no longer recommended, this finding is meaningful in the counselling of couples requesting preimplantation genetic diagnosis (PGD). PGD is offered to couples at risk of having a child with a hereditary disease and the procedure also involves embryo biopsy. It is reassuring for these couples that embryo biopsy is not related to alterations in BP or anthropometrics of 4-year-old children.
Finally, this thesis describes a project in which we studied the relationship between dysmorphic features and the development of 2-year-old children. We found associations between the presence of an abnormality or the presence of ≥ 3 minor anomalies and worse neurological performance. As dysmorphic features emerge in early pregnancy, these associations suggest an early ontogenetic origin of minor neurodevelopmental problems. Hence, clinicians involved in the health care of a child with an abnormality or ≥ 3 minor anomalies should be aware of an increased vulnerability for subtle neurodevelopmental problems, since early detection could offer opportunities for early intervention.

The above recommendations for counselling are based on the findings in this thesis as well as on recent reviews on child health after ART (Hart and Norman 2013a, Hart and Norman 2013b). However, when giving implications for clinical practice it should be noted that ART treatment protocols have changed over the years. It is therefore difficult to estimate the effects of current ART. In studies where the dates of birth of participants had a large range, we corrected for year of birth in the multivariable analyses (chapters 2 and 3). However, it remains uncertain whether our findings hold true for ART babies who are born nowadays. As long as treatment protocols are changing, it will be a challenge to extrapolate long-term follow-up data to modern day practice. Some of the recent and current changes in fertility practice will now be discussed and their influence on follow-up research will be considered.

**Changes in fertility practice and consequences for follow-up studies**

Since the introduction of IVF in 1978, treatment protocols have changed considerably and continue to change. Fertility treatment policies are influenced by many factors, including new medical-technical possibilities, insights from scientific research and ethical, social and financial considerations. The goal of this section is to give an impression of the direction and progress in reproductive medicine and its consequences for the need of new follow-up studies.

In 2012, the yearly costs of fertility treatments were estimated at €50 million in the Netherlands and Edith Schippers, Minister of Health, Welfare and Sport in the Netherlands, advocated a cost reduction of €30 million per year. On average, an ART treatment costs €3,000 per attempt. In the Netherlands, three attempts are covered by the compulsory, nationwide health insurance. The transfer of cryopreserved embryos does not count as an attempt and only in case of successful oocyte pickup, an attempt is counted. In order to realize cost reductions, the Dutch government introduced five new measures for fertility treatment in January 2013:
1. In women aged 38 years or younger, elective single embryo transfer (eSET) is obligatory for the first two attempts.
2. In couples with a reasonable chance of conceiving naturally, the policy of ‘watchful waiting’ is in order.
3. Fertility treatment (including ovulation induction and intra-uterine insemination) is covered by the health insurance for women up to age 43 years, with the exception of transferring cryopreserved embryos, which is covered up to age 45 years.
4. Medication is to be provided by the hospital pharmacist to reduce the waste of expensive medication.
5. The less expensive urinary FSH should be used instead of recombinant FSH.

The reason that eSET is embraced is that multiple pregnancies increase the burden for health care systems because of the increased risk for pregnancy and perinatal complications (Avraham and Seidman 2012). In eSET, a single fresh embryo is transferred even though more than one high-quality embryo is available, and if no live birth results, the transfer of a thawed embryo subsequently takes place (Practice Committee of Society for Assisted Reproductive Technology and Practice Committee of American Society for Reproductive Medicine 2012). The lower risk of a multiple pregnancy after eSET needs to be balanced against the risk of jeopardizing the overall live birth rate. A double-blind RCT demonstrated that in women under 36 years of age, transferring one fresh embryo and, if needed, another frozen and thawed embryo clearly reduces the rate of multiple births while live births rates are not substantially lower than those seen after a double embryo transfer (DET) (Thurin et al. 2004). The use of eSET also simplifies some research questions as the transfer of multiple embryos is related to a higher incidence of vanishing twins. It is known that singleton survivors of a vanished co-twin have worse health outcomes than singletons from pregnancies with only one initial gestation (Pinborg et al. 2005, Pinborg et al. 2007, Sullivan et al. 2012). Singletons born after eSET may therefore be healthier than singletons born after DET. Hence, an ideal follow-up study would be a prospective study that only includes patients who underwent eSET.

With the increasing use of eSET, the number of cryo cycles is also increasing, making it important to study the health of children born following frozen and thawed embryo transfer (FET). On the one hand, studies reported less prematurity and fewer babies being small-for-gestational age (SGA) or born with a low birthweight after FET (Pinborg et al. 2012, Wennerholm et al. 2013). On the other hand, studies have also reported more large-for-gestational age (LGA) babies after FET (Wennerholm et al. 2013). As sibling-ship analyses that took age, parity, child sex, year of birth, and birth order into account demonstrated that children born after FET were more often LGA than their siblings born after fresh embryo transfer (Pinborg et al. 2014), it was suggested that it is the freezing and thawing procedure that
is involved in leading to LGA babies after FET. As LGA babies have worse perinatal outcomes, including a higher risk for stillbirth, asphyxia, shoulder dystocia and perinatal mortality (Henriksen 2008), it is important that future research tries to unravel the causal pathways of the association between FET and LGA, and assesses the effect of FET on the offspring’s long-term health and development.

The second cost reduction measure – watchful waiting if there is a reasonable chance of a natural pregnancy – was introduced to prevent the overzealous use of ART. ART was developed for women with tubal disease, in which the indication for ART is clear. Soon however, indications for ART began to grow and they now include other types of subfertility such as mild male subfertility, endometriosis and unexplained subfertility, for which the effectiveness of ART is less clear (Kamphuis et al. 2014). Economic modelling studies indicated that ART is not cost effective in younger women with unexplained subfertility and a time to pregnancy of less than three years (Mol et al. 2000). Apart from these financial motives, the revised policy is also supported by the information provided in this thesis regarding the lack of sufficient knowledge on the long-term health of ART offspring. Carefulness in indicating ART is therefore warranted in couples who have a reasonable chance of natural conception. Research on the effectiveness of ART for new indications is necessary to reveal who should undergo ART and when.

The third measure – covering the cost of fertility treatment up to age 43 years and transfer of cryopreserved embryos up to age 45 – takes the decline of success rates with advanced maternal age into account. Between the ages of 43 and 45 years, women may undergo ART at their own expense, thereafter, according to the Dutch law, ART should not be performed (Embryowet, 2002). As subfertility has a major impact on the quality of life and health experienced by affected couples, the decision-making processes involved in who can receive fertility treatment are extremely difficult. What makes it even more complicated is that the regulations differ across countries and continents and that ‘fertility tourism’ is becoming a major problem. In several countries, ART is available after the age of 45 and multiple embryos are often transferred. As a result, the home country has to deal with the cost of potential problems related to a multiple pregnancy in an older ‘fertility tourist’ upon her return home after. Future research should address the question whether chronological age indeed is the decisive factor to determine whether a woman may opt for ART, or whether biological age is the critical factor. This question may be answered by assessing whether parameters such as anti-müllerian hormone, follicle-stimulating hormone, inhibin B and antral follicle count are better predictors of live birth after ART than the woman’s chronological age.

Another distinct change in recent fertility practice is the development and implementation of new PGS techniques. PGS was introduced in the 1990s on the basis of promising case reports, cohort studies, and non-randomized comparative
studies with small numbers. The results of RCTs appeared almost a decade after introducing PGS into the clinical setting. PGS unexpectedly resulted in lower pregnancy rates and is therefore no longer routinely recommended (Mastenbroek et al. 2007). The main reasons for the inefficacy of this older type of PGS are believed to be the mosaic nature of the embryo at the cleavage stage (day three after fertilization) (Van Echten-Arends et al. 2011) and the FISH analysis of a limited number of chromosomes tested (eight chromosomes were tested for aneuploidies instead of 22 pairs of autosomes and one pair of sex chromosomes). New PGS methods have been developed in which the embryo is biopsied either at the polar body (a by-product of the meiotic cell cycle) or at the trophectoderm (a precursor of the placenta that can be biopsied at the blastocyst stage, i.e. day five after fertilization). These techniques may reduce problems associated with mosaicism and are considered less detrimental as the embryo’s integrity remains unaffected (Brezina et al. 2012, Capalbo et al. 2013, Ly et al. 2011). Just as the early PGS test, using FISH at the cleavage stage, these new PGS methods are being introduced into the clinical setting without reliable information on their safety and efficiency (Mastenbroek and Repping 2014). IVF clinics compete with each other to be innovative, making PGS commercially attractive as patients are, of course, eager to increase their chances of having a baby. Well-designed RCTs on the safety, efficiency and cost-effectiveness of PGS must be performed in order to prevent us making the same mistake as in the 1990s: performing expensive and less effective treatments. In addition, the long-term health of the offspring born following these new PGS techniques should be closely monitored.

**Methodological considerations**

As stated in the introduction to this thesis, research of good methodological quality is a challenge in this field. This thesis describes studies with different methodological approaches: chapters 2 and 3 cover registry-based studies, chapters 4 to 7 cover a prospective follow-up cohort study, chapter 8 reports on the follow-up of an randomized controlled trial (RCT), and chapter 9 covers a cross-sectional study. The different designs and study questions mean that our studies faced their own strengths and limitations, which will be briefly discussed below.

**Part 1. The Eurocat Subfertility Project**

One major strength of the Eurocat Subfertility Project is that it included all types of births regardless of gestational age, i.e. terminations due to a fetal anomaly were also included. This is important because women who have undergone ART may be less willing to abort their much desired pregnancy so that studies excluding termina-
tions are prone to bias. Another strength is that it included children with congenital anomalies up to the age of 10 years. Since children born following ART may be monitored more closely than children conceived naturally, congenital anomalies may be detected earlier and ascertainment bias might influence the results of studies with a short registration or follow-up period. The total subfertile group consisted of 340 children in total; this may seem a small group, but as all the children had a congenital anomaly, this selected sample allowed for an adequate evaluation of most of the specific anomalies considered. The number of participants also made it possible to look at specific and pathogenetically similar types of congenital anomalies, instead of larger heterogeneous subgroups of congenital anomalies. Furthermore, the detailed questionnaires filled in by the parents enabled us to correct for a large variety of confounders and we verified their subfertility status by searching in their medical fertility records.

Exclusion of participants with unconfirmed subfertility may, on the other hand, be seen as a limitation, as it reduced the power of the study. Another limitation was that we were unable to distinguish different causes of subfertility and by combining all the different causes of subfertility into one group, we may have missed relations between specific types of subfertility and specific congenital anomalies. Furthermore, we were not certain that fertile couples in the control group were indeed fertile according to the definition of a time to pregnancy of less than 12 months. This could have induced an underestimation of the effect of subfertility on the presence of specific congenital anomalies in the offspring. It should also be noted that our power to detect associations with very rare congenital anomalies and syndromes was likely to be insufficient.

**Part 2. The Netherlands Perinatal Registry Sibling Project**

The major strength of this registry-based study was the use of two different designs: the sibling-ship design and the inter-sibling design. The sibling-ship design minimized confounding factors from maternal characteristics, so that differences could primarily be attributed to the ART treatment. The inter-sibling design prevented confounding due to parity, while keeping maternal characteristics constant between the comparisons. Furthermore, the inter-sibling design provided information about the effects of maternal characteristics including the underlying subfertility on perinatal outcomes. The large number of participants in this registry was a huge advantage for this study. The longitudinal linkage of first- and secondborn children over a period of nine years resulted in the registration of 272,551 women who had conceived two subsequent singletons. A final strength was that we were able to correct for many potential confounders as the PRN collects information on maternal characteristics, pregnancies and deliveries.
The main limitation of this study was that we were uncertain whether children who shared the same mother also shared the same father. Since the PRN does not collect information on paternal characteristics, we could not to investigate this. Another limitation is that ART was not always registered, because mode of conception is not an obligatory field to be filled out in the PRN. This could have resulted in an overestimation of the adverse effects of ART, if ART was only registered when maternal or neonatal complications occurred, or in an underestimation of the adverse effects of ART if problematic ART pregnancies were not recorded as such. Furthermore, we may not be able to generalize our findings to all women undergoing IVF, since we studied women who had had two children in the study period. These women are thought to be less reproductively compromised than subfertile women who could not conceive twice, thereby possibly introducing a positive selection bias. Other limitations were that we had no information on the use of intracytoplasmic sperm injection (ICSI), the cause of subfertility, maternal BMI, the stimulation protocol, the use of fresh or cryopreserved embryos, or the culture conditions.

**PART 3. THE GRONINGEN ART COHORT STUDY**

The major strength of the Groningen ART cohort is the composition of its three study groups. Most studies compare controlled ovarian hyperstimulation-IVF (COH-IVF) children with naturally conceived children from the general population. This comparison reflects the combined effect of ovarian hyperstimulation and the *in vitro* procedure and is confounded by parental and subfertility-related factors. The inclusion of the modified natural cycle-IVF (MNC-IVF) group in The Groningen ART cohort allowed us to study the effect of ovarian hyperstimulation separately by comparing this group to the COH-IVF group. In addition, a group of naturally conceived children born to subfertile couples served as our reference group, minimizing the role of some potential confounders. By comparing the MNC-IVF group to a group of children born to subfertile couples who eventually conceived naturally (Sub-NC group), we studied mainly the effect of the *in vitro* procedure. This unique study design was better than one in which children born following ovulation induction were included in an attempt to study the effect of ovarian hyperstimulation, since lower doses of hormones are applied in ovulation induction in order to stimulate the growth of only a few follicles. Other strengths were the prospective and assessor-blinded design. In addition, there was a very low attrition rate of only 10% from inclusion at the start of the study to the follow-up assessment at age 4 years. A final strength was the application of sensitive and age-specific methods to assess dysmorphic features and neurological development.

The main limitation of the Groningen ART cohort was the size of the groups. Power calculations were based on neurological outcome at age 18 months (Mid-
delburg et al. 2009), and not on dysmorphic features or BP and anthropometrics. The low numbers of children meant we could not stratify the data according to ICSI or gender. Subanalyses did not, however, reveal any effect of gender or ICSI on the outcome measures described in this thesis. Furthermore, the Groningen ART cohort does not allow for proper investigation of the role of subfertility on child development. To study the role of subfertility, a second control group of children born to fertile couples is needed. This group was not prospectively recruited and retrospective inclusion is prone to selection bias. We therefore used time to pregnancy (TTP) as a proxy for the severity of subfertility. Twins were not included, limiting the generalizability of the results, and survivors of a vanished co-twin were included as singletons. The latter made it necessary to correct for vanishing twins in the multivariable analyses since Pinborg et al. demonstrated that IVF singletons with a vanished co-twin more often had adverse obstetric outcomes and neurological sequelae than ‘true’ singletons (Pinborg et al. 2005, Pinborg et al. 2007).

In chapters 6 and 7, our focus was on the effects of ART on BP and anthropometrics. We measured BP twice on one day only, whereas multiple BP measurements or a 24-hour telemetric approach would have been better. Finally, the limitations of the statistical analyses should be kept in mind. In general, multivariable regression analyses were used in this thesis, which is currently the standard statistical approach for research questions like ours. Using this approach, one corrects for covariates without knowing which covariates are true confounders. This makes it hard to draw conclusions with respect to the underlying causal mechanisms. Moreover, no distinction can be made between direct and indirect effects, and the influence of latent variables cannot be taken into account. In chapter 7, our data on BP and anthropometrics were further explored using causal inference search algorithms in order to tackle these hurdles. As this statistical approach is explorative, the results should be interpreted with some caution. A detailed description of the advantages and disadvantages of causal inference search algorithms is given in chapter 7.

**PART 4. THE PGS FOLLOW-UP STUDY**

The major strength of this study is that it involves the follow-up of an RCT with a prospective, assessor-blinded design. We studied children born to women who were randomly assigned to ART with PGS (PGS+) or ART without PGS (PGS-), resulting in comparable background variables between the two groups. We were the first to study blood pressure (BP) levels in PGS offspring. We also performed the first study on anthropometrics and the medical care given to children born following only PGS, instead of in a group of children born following either PGS or PGD. Another strength is the inclusion of both singletons and twins, which contributes to the generalizability of the results. However, as twins are more often born preterm or with
a low birthweight, it could be argued that singletons and twins should not be ana-
yzed together (Rogers 2003). We therefore first checked whether PGS had a different
effect on twins than on singletons: it did not, so we analyzed singletons and twins
together, taking into account the possible correlation structure for twins.

The main limitation of the PGS Follow-Up Study is the number of BP mea-
surements performed. As in the Groningen ART Cohort Study, we measured BP
twice on one day rather than on several separate occasions, which is necessary to re-
veal a more accurate BP. It may also be regarded as a limitation that we analyzed IVF
and ICSI offspring together, although, sensitivity analyses did not reveal any signifi-
cant differences in outcome measures for IVF or ICSI children. Another limitation is
that the study was not designed to detect differences in BP or anthropometrics, but
rather to detect differences in ongoing pregnancy rates (Mastenbroek et al. 2007).
However, a post hoc power analysis indicated that we had sufficient power to detect
clinically relevant differences in systolic BP (SBP) and diastolic BP (DBP).

PART 5. AETIOLOGICAL PATHWAYS OF NEURODEVELOPMENT

The main strength of this project was the extensive assessment of 2-year-old chil-
dren. Dysmorphic features and child development were determined using sensitive
and age-specific tests. We were therefore able to evaluate associations between dys-
morphic features and child development across a broad spectrum of neurodevelop-
mental functions. In addition, the long-standing follow-up of the children included
in this project allowed us to correct for a large variety of confounders. The inclusion
of both singletons and twins contributed to the generalizability and, in order to mod-
el the possible correlation between twins, we used the mothers as a cluster variable.

The main limitation of this project was that the participants were part of two
other studies: the Groningen ART Cohort Study and the PGS Follow-Up Study. This
means that all the children were born to subfertile parents and our results cannot
be readily generalized to the entire population of 2-year-old children in the Neth-
erlands. As mean parental age at child birth in the Groningen ART cohort is higher
than in the general population (mean ages for our cohort vs. the general population:
mothers 33.4 vs. 31.1 years, fathers 35.6 vs. 34.1 years), we corrected for parental age
and time to pregnancy in the multivariable analyses.

SUGGESTIONS FOR FUTURE RESEARCH

Considering the methodological limitations discussed in above, it is clear that de-
signing a sound methodological follow-up study is difficult but of utmost impor-
tance. We offer several suggestions for future research below, based on the studies
presented in this thesis. Secondly, we provide a brief overview of the effect of ART on health outcome parameters not discussed in this thesis (i.e. asthma, childhood cancer and pubertal developmental), followed by suggestions for future research focusing on these outcome measures.

It is known that subfertile couples are at increased risk of giving birth to a child with a congenital anomaly, but it is unknown why. Studying the association between ART and de novo mutations or microdeletions in the offspring would contribute to our knowledge on how parental subfertility is related to an increased risk of congenital anomalies. Data from the first project in this thesis, the Eurocat Subfertility Project, provide excellent means to study this issue. The relation between parental characteristics and congenital anomalies should also be further investigated using inter-sibling analyses and sibling-ship analyses.

The second project, the Perinatal Registry Netherlands Sibling Project, demonstrated how inter-sibling analyses and sibling-ship analyses help in disentangling ART-related effects and effects from parental characteristics, including the underlying subfertility on perinatal outcomes. Not only perinatal outcomes, but all kinds of outcome measures — such as neurodevelopmental and cardiometabolic outcomes — can be studied with these approaches. The simultaneous use of both approaches is encouraged for future studies that try to study the separate effects of ART and parental characteristics on child health.

The third study in this thesis, the Groningen ART Cohort Study, provides the opportunity to study separately the effects of ovarian hyperstimulation and the in vitro procedure on several health outcome parameters at later ages. The children in the Groningen ART cohort are currently being assessed at the age of 9 years, an age at which subtle dysfunctions in more complex neural functions emerge, reflected by a relatively high prevalence of minor neurological dysfunction in the general population. It is possible that an effect of ovarian hyperstimulation and the in vitro procedure on neurodevelopment may first become clear at this age. Furthermore, as most 9-year-olds have not started puberty, tracking of BP can be studied in our cohort. It would be interesting to see whether the differences found between COH-IVF and MNC-IVF in SBP are still present, have increased, or have attenuated. A 24-hour telemetric approach would be the most refined measurement option at this age. In the future, lipid profiles, glucose, insulin, and cortisol levels could be determined next to the standard anthropometric assessment in order to gain more knowledge on the metabolic profiles of ART children.

As the Groningen ART cohort is a non-randomized cohort, other less clear differences between the groups may have caused the difference in BP between COH-IVF and MNC-IVF. The differences might also be confounded by the vanishing twin phenomena. Follow-up of a randomized study in which only eSET is performed, like
the INeS, is therefore needed. The INeS is a Dutch multicentre RCT that aims to determine the safest and most cost-effective treatment for couples with unexplained subfertility or mild male subfertility. The INeS study compares three treatments:
1. Six cycles of intra-uterine insemination with controlled ovarian hyperstimulation (COH-IUI)
2. Six cycles of MNC-IVF
3. Three cycles with eSET after COH-IVF plus cryo-cycles

Comparing BP levels and anthropometrics of the children born in the INeS study (Bensdorp et al. 2009) may confirm or refute that ovarian hyperstimulation has a role in causing worse cardiometabolic outcomes. The Groningen ART cohort also cannot show us whether the potential adverse effect of ovarian hyperstimulation is through an effect on the oocyte or the endometrium. Studies comparing cryo cycles with fresh cycles, for example, by performing subanalyses in the INeS study, may provide this insight as cryo cycles are characterized by the absence of a recent effect of ovarian hyperstimulation on the endometrium. A final suggestion for future research in the INeS is to study the relationship between embryo quality and child health. Insight into this may have valuable consequences for clinical practice. The association between embryo quality and health outcome at later ages could not be properly studied in our data as DET was often performed in the COH-IVF group, and it was therefore not known which embryo resulted in the ongoing pregnancy.

The children in the final study of this thesis, the PGS Follow-Up Study, should also be reassessed at later ages. Our finding of an increased use of para-medical care in the ART with PGS group compared to the ART without PGS group may indicate an adverse effect of PGS on subtle health outcome parameters. Furthermore, we found that neurodevelopment was slightly less good in 2-year-old children born after PGS (Middelburg et al. 2011). At the age of 4 years, PGS was associated with altered neurodevelopment in twins (Schendelaar et al. 2013). These findings and the lack of knowledge on developmental disorders that emerge at later ages and the health of adults born after embryo biopsy justify further follow-up.

**ART and asthma**

This thesis did not investigate whether ART has an effect on the prevalence of asthma, but as asthma is one of the most common chronic childhood diseases, it is important to explore the potential association. Several studies have reported an association between ART and asthma (Carson et al. 2013, Ericson et al. 2002, Finnstrom et al. 2011, Guibas et al. 2013, Kallen et al. 2013, Koivurova et al. 2007), whereas others found no association (Bonduelle et al. 2005, Cetinkaya et al. 2009, Klemetti
et al. 2006, Pinborg et al. 2003). The inconsistencies could be related to attrition, the way asthma was diagnosed (based on medication prescription, self-report, or doctor’s diagnosis) and confounding factors.

There is limited understanding of the aetiology of asthma but risk factors include perinatal, environmental and genetic factors. Perinatal factors such as preterm birth, low birthweight and Caesarean section are associated with ART and are risk factors for asthma, possibly explaining the putative association between ART and asthma (Carson et al. 2013). Interestingly, mothers with asthma more often undergo ART treatment (Sheiner et al. 2005). This association makes it even more important to correct for the mother’s history of asthma, which also reflects a genetic component in the development of asthma in the child. It has also been suggested that subfertility increases the risk for asthma in the offspring (Kallen et al. 2013). The possible mechanisms underlying this association are not known, but one could speculate that parental asthma is a risk factor for subfertility and childhood asthma. Furthermore, it has been suggested that anti-asthmatics may reduce ovulation and fertility (Grodstein et al. 1993), although there does not seem to be an effect of asthma on final fertility rates (Tata et al. 2007).

It is clear that the above confounders complicate the interpretation of the association between ART and asthma in the offspring. Most registers do not allow for proper correction of all known confounders but offer the advantage of large numbers of patients. As asthma is a relatively common childhood disease, its occurrence could be studied in a prospective cohort study. This future study should include both a subfertile and a fertile control group, and should study potential causality between ART, subfertility and asthma. The current reassessment of the children in the Groningen ART cohort includes a validated parental and child questionnaire on respiratory disorders and symptoms.

ART and childhood cancer

Another health outcome parameter that deserves attention is the prevalence of childhood cancer after ART. Some studies reported no increased overall childhood cancer risk (Bruinsma et al. 2000, Klip et al. 2001, Sundh et al. 2014), whereas others reported an increased risk for early onset acute lymphoblastic leukaemia (OR: 4.29, 95%CI: 1.49-12.37) (Petridou et al. 2012) or Langerhans histiocytosis (5 out of the 29 cases with cancer) in children born following ART (Kallen et al. 2005b). In Kallen et al.’s study, the overall risk for cancer was not increased in children born between 1982 and 2001 after ART (29 cases of childhood cancer, 21 cases expected), but new
analyses of the data from the Swedish Cancer Register five years later resulted in a total cancer risk estimate of 1.42 (95%CI: 1.09-1.87): 53 cases of cancer were identified against 38 expected cases in 26,692 children born following ART between 1982 and 2005. In comparison to the 2005 study, one new case of Langerhans histiocytosis occurred, resulting in six cases against one expected case (Kallen et al. 2010a).

A meta-analysis on all types of fertility treatments and childhood cancer, for which a literature search was performed in 2010, indicated an increased Relative Risk (RR) for all cancers: 1.33 (95%CI: 1.08-1.63), and for haematological cancers: 1.59 (95%CI: 1.32-1.91), central nervous system/neural cancers: 1.88 (95%CI: 1.02-3.46) and other solid cancers: 2.19 (95%CI: 1.26-3.80) (Hargreave et al. 2013). However, after the appearance of this worrisome meta-analysis, two large population-based studies were published that did not confirm the overall increase for childhood cancer after ART. One of these studies consisted of 106,013 children born in Britain between 1992 and 2008 and had a mean follow-up of 6.6 years. The 108 cases of cancer were compared with the 110 expected cases (standardized incidence ratio 0.98, 95%CI: 0.81-1.19). ART was however associated with two subtypes of cancer: rhabdomyosarcoma and hepatoblastoma, but the absolute risk increases were small (Williams et al. 2013). The other large population-based study followed 91,796 ART children born in the Nordic countries for a mean follow-up time of 9.5 years and compared them to 358,419 naturally conceived controls matched for parity, year of birth, and country. Again, no overall cancer risk increases were found (adjusted Hazard Ratio 1.08 (95%CI: 0.91-1.27) after ART, but two types of cancer occurred more frequently than expected: central nervous system tumours and malignant epithelial neoplasms (Sundh et al. 2014).

From the above, it is clear that further research in this area is warranted. As childhood cancers are rare, register-based studies provide a good option for evaluating an association between ART and childhood cancer, on the condition that they can adjust for high birthweight, premature delivery, the presence of a respiratory diagnosis, and a low Apgar score, as these variables are identified risk factors for childhood cancer (Kallen et al. 2010a). The next step, investigating possible explanations such as treatment-related procedures, the parents’ underlying subfertility, mediation by perinatal adversities or altered imprinting, is more difficult. Large registers should seek the opportunity to adapt inter-sibling and sibling-ship analyses in order to disentangle ART-related effects from effects related to parental characteristics including subfertility. If specific types of childhood cancer are repeatedly reported after ART, studying the effect of ART on the imprinting status of the corresponding cancer genes could help us understand the underlying mechanisms.
ART AND PUBERTAL DEVELOPMENT / REPRODUCTIVE CAPACITY

A different point of concern is the reproductive health of children born following ART. The first ART baby, Louise Brown, is now 36-years-old and she conceived spontaneously in 2006, at age 27. Very little is known about the reproductive success of her peers born following ART, although the developing endocrine systems in the fetus and maturation of endocrine-control systems can be influenced by the hormonal environment of the fetus. In theory, ART could induce alterations in intrauterine hormonal levels. In addition, hereditary aspects of infertility are of concern. In response to the reproductive concern, several studies focused on the effect that ART may have on specific features that may predict reproductive success, like pubertal development. Precocious puberty after ART was reported by Rojas-Marcos et al., who found sex-steroid and hormone levels in the prepubertal range in 7 infants aged between 5 and 21 months who presented with breast development and/or pubic hair (Rojas-Marcos et al. 2005). Precocious puberty was not confirmed by others (Belva et al. 2012c, Ceelen et al. 2008a, Sakka et al. 2010). Yet, Ceelen et al. (2008a) did find that pubertal IVF girls (n = 19) had higher serum luteinizing hormone (LH) and dihydroepiandrosteredione (DHEAS) concentrations than age- and gender-matched naturally conceived girls born to subfertile parents (n = 20). Furthermore, the bone age of a larger subset of the pubertal IVF girls (n = 72) appeared to be advanced compared with the control group (n = 75) (Ceelen et al. 2008a). In addition, Sakka et al. did find a significantly increased incidence of precocious puberty and higher serum DHEAS levels in small-for-gestational age (SGA) children conceived via IVF when compared to SGA children conceived spontaneously. These differences were not found between IVF and naturally conceived children who were born with an adequate size for gestational age (Sakka et al. 2010). Belva et al. reported normal pubertal development in ICSI boys and girls (mean age 14 years), but did note that ICSI girls (n = 101) had less advanced breast development than naturally conceived girls (n = 108), also after adjusting for current, early life and social factors (aOR: 0.5, 95%CI: 0.3-0.8). The authors hypothesized that the less advanced breast development might be caused by a disturbed hypothalamic reactivation of gonadotrophin-releasing hormone (GnRH) release (Belva et al. 2012c).

Another point of concern are the semen parameters of boys conceived through ART, especially ICSI, as the main indication for ICSI is severe male infertility. A study on 87 ICSI boys showed their testosterone serum levels (a measure of Leydig cell function) to be 27% lower than normal at the age of three months. No differences in penile length were found (Mau Kai et al. 2007). De Schepper et al. (2009) provided reassuring information on early testicular development by studying anti-müllerian hormone (AMH, n = 59) and serum inhibin B (n = 59) (both are measures of Sertoli cell function/spermatogenesis) and testicular size (n = 88) in 8-year-old ICSI boys and
comparing them to reference values. The testicular size of ICSI boys turned out to be independent of the degree of oligozoospermia in their fathers (De Schepper et al. 2009). The re-assessment of 50 boys from this cohort at the age of 14 years revealed that their serum inhibin B levels were significantly increased compared to the levels at 8 years of age and that the inhibin B levels corresponded with the stage of pubertal development (Belva et al. 2010). Further follow-up of these children is needed to examine whether normal Sertoli cell markers will be followed by normal spermatogenesis.

Overall, pubertal timing after ART seems to be normal, but the need for future research assessing the reproductive abilities of ART offspring is clear. Existing cohort studies, including the Groningen ART cohort, should therefore organize a follow-up assessment during puberty and monitor the reproductive success and problems of individuals born after ART.

**Concluding remarks**

In the literature, an association between assisted reproductive techniques (ART) and subtle adverse health outcomes in the offspring has been repeatedly described. The main goal of this thesis was to investigate whether it is the ART treatment or the parents’ underlying subfertility that is the culprit. We further assessed the role of several components of the ART treatment, i.e. ovarian hyperstimulation and the *in vitro* procedure, and of preimplantation genetic screening on the offsprings’ health.

The seven main findings of this thesis are:

1. Subfertility is associated with some specific congenital anomalies (abdominal wall defects, penoscrotal hypospadia, right ventricular outflow tract obstruction and methylation defects causing imprinting disorders).
2. Parental characteristics, including their underlying subfertility, are associated with worse perinatal outcomes (shorter duration of pregnancy, lower birthweight and a higher risk of fetal growth restriction and perinatal mortality), and a reduced risk of pregnancy-induced hypertensive diseases.
3. A longer time to pregnancy, reflecting more severe subfertility, is associated with congenital abnormalities and suboptimal neurodevelopment of 2-year-olds.
4. ART itself is associated with fetal growth restriction and polydactyly.
5. Ovarian hyperstimulation, as applied in COH-IVF, has been suggested as a causative factor for higher blood pressure levels and thicker skinfold thickness in 4-year-olds.
6. Preimplantation genetic screening does not have an adverse effect on the blood pressure or anthropometrics of 4-year-olds.
7. Minor alterations in neurodevelopment may have an early ontogenetic origin because an association with dysmorphic features was established.
Children born following ART are generally healthy. ART itself does not seem to affect neurological, mental and behavioural development. The risk of perinatal adversities, birth defects and cardiometabolic adversities is slightly increased after ART, but the absolute risk increases are only small. The increased risks are, to a large extent, mediated by factors associated with subfertility. As there is still very little information on the acute and chronic illnesses of adults born after ART, carefulness in the indication for ART is warranted. We hope that the scientific attention for the follow-up of ART offspring will increase and allow for the elucidation of the important questions on the long-term health outcomes after ART.