RISK STRATIFICATION FOR WOMEN UNDERGOING IN-VITRO FERTILISATION TREATMENT

by

Dr Rima Kaur Dhillon

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

School of Clinical and Experimental Medicine

College of Medical and Dental Sciences

University of Birmingham

November 2015

ABSTRACT

The aim of this thesis was to explore three factors that are easily available and contribute important information for women *before* commencing in-vitro fertilisation (IVF) treatment: ethnicity, body-mass index (BMI) and thyroid disease. Results of the systematic review, cohort study and meta-analysis investigating ethnicity and IVF outcome showed South Asian and Black women have lower adjusted live-birth (LB) rates, after fresh cycle treatment, compared with White women. The relationship between BMI and IVF outcome was explored in a prediction model estimating chances of LB following first cycle. The model found BMI has reduced effect on IVF outcome when adjusting for other confounders such as age. The prevalence of thyroid dysfunction and thyroid peroxidase antibodies (TPOAb) was examined across the UK in >7000 women of reproductive age, and a cohort study investigating the effect of subclinical hypothyroidism (SCH) on IVF outcome was also performed. The prevalence of overt thyroid disease was 0.38% and subclinical disease 3.45%. Using an upper limit cut off for thyroid-stimulating hormone of 2.5mU/L the prevalence of SCH was 19.64%. The overall prevalence of TPOAb was 9.11%; this was 7.98% in euthyroid women. Finally, there were no significant differences in LB between euthyroid women and women with SCH.

DEDICATION

I wish to dedicate this thesis to my family and friends for their love and ever continuing support of my career. In particular my husband, Paul Smith, for his love, support, advice and without whom I could not have made it through!

ACKNOWLEDGEMENTS

The work for this thesis was conducted between August 2012 and August 2013 whilst I was based at the Birmingham Women's Hospital. I worked as the lead clinical research fellow for the Thyroid AntiBodies and LEvoThyroxine trial (TABLET); which was funded by the National Institute for Health Research, Efficacy and Mechanism Evaluation (EME) Programme. There are many people who have supported me through my PhD years and I would like to acknowledge them for this.

Thank you firstly to my lead supervisor, Professor Arri Coomarasamy, who's continued support, enthusiasm and inspiration has guided me through the last three years and without whom my work would not have been possible.

Thank you to Professor Jon Deeks, my secondary supervisor, for his insightful input with the statistical elements of my work as well as providing invaluable mentorship when needed.

Thanks must also go to Dr David McLernon and Professor Bhattacharya from the University of Aberdeen. David, for all his time in providing informal teaching on statistics and helping to perform the statistical analyses for the prediction model work, and Professor Bhattacharya for supervising the prediction model work and supporting my presentation at the European Society for Human Reproduction and Embryology (ESHRE) conference in Lisbon 2015.

I would also like to thank the medical directors of CARE (Centres for Assisted Reproduction); Simon Fishel and Ken Dowell for providing the data for chapters 3 and 4 and their continued support of my work. Furthermore, I would also like to thank Ashley Beccles (head of IT department at CARE) for all his assistance with the database.

I would like to thank Rosie Malhas for being the 2nd reviewer for the systematic review on ethnicity and IVF outcome; also Hoda Harb and Ioannis Gallos in their contributions to the work on ethnicity and IVF outcome.

I would like to thank the team at the Birmingham Clinical Trials unit (Krys Baker, Samantha Farrell-Carver and Kirandeep Sunner) who have worked tirelessly with me on the TABLET trial over the years and helped to set up the prevalence studies for this thesis.

Finally I would like to thank my parents, Swinder and Sukhjit Dhillon, my sister Kirandip Cooner and brother in-law Jagdeep Cooner, and also my brother Jaipal Dhillon for their ever-continuing encouragement and support of my career. Last, but not least, I would like to thank my best friend Rakhi Saha and my husband Paul Smith for their endless love, guidance and for never losing faith in me.

ABBREVIATIONS

AACE	American Association of Clinical Endocrinologists
ACU	Assisted conception unit
AFC	Antral follicle count
AMH	Anti-mullerian hormone
aOR	Adjusted odds ratio
ART	Assisted reproductive technologies
ATA	American Thyroid Association
AUROC	Area under receiver operating characteristic
BMI	Body mass index
BTS	British Thyroid Society
EPAU	Early pregnancy assessment unit
ESCPG	Endocrine Society Clinical Practice Guidelines
FET	Frozen embryo transfer
FSH	Follicle stimulating hormone
fT4	Free Thyroxine (T4)
HFEA	Human Fertilisation and Embryology Authority
ICSI	Intra-cytoplasmic sperm injection
IVF	In-vitro fertilisation
LT4	Levothyroxine
NACB	National Academy of Clinical Biochemistry
NEQAS	National External Quality Assessment Service
NHS	National Health Service

NICE	National institute for health and clinical excellence
OR	Odds ratio
REC	Research ethics committee
RR	Relative risk
SART	Society for Assisted Reproductive Technologies
SCH	Subclinical hypothyroidism
SD	Standard deviation
TABLET	Thyroid AntiBodies and LEvoThyroxine trial
TAI	Thyroid autoimmunity
TFT	Thyroid function test
TPOAb	Thyroid Peroxidase Antibodies
TSH	Thyroid stimulating hormone
UK	United Kingdom
US	United States

TABLE OF CONTENTS

CHAPTER 1 1

Thesis overview and objectives 1

Introduction to in-vitro fertilisation (IVF) 2

Ethnicity and IVF outcome 3

BMI and IVF outcome 4

Thyroid dysfunction and thyroid autoimmunity 5

Thesis objectives 8

CHAPTER 2 12

Investigating the effect of ethnicity on IVF outcome: a systematic review and meta-analysis 12

Introduction 13

Aims and objectives 14

Methods 15

Data sources 15

Eligibility criteria for selecting studies 15

Data extraction and synthesis 18

Quality assessment 19

Results 24

Black vs. White 35

Asian vs. White 40

Hispanic vs. White 45

Discussion 47

Main findings....47

Strengths and limitations....48

Conclusions and future work....50

CHAPTER 3 51

Investigating the effect of ethnicity on IVF outcome: an analysis of 13,473 cycles and an updated meta-analysis 51

Introduction 52

Aims and objectives....53

Methods 53

Study Design 53

Statistical analysis 54

Missing data....58

Results 59

Overall description of data 59

Univariate and multivariate analyses for live birth for all cycles 66

Univariate and multivariate analyses for clinical pregnancy for all cycles 69

Sensitivity analysis accounting for fibroids....72

Analysis of frozen cycles....72

Discussion 73

Main findings....73

Strengths and limitations.....75
Comparison of results with existing literature and an updated meta-analysis.....77
Exploring reasons for differences 83
Conclusions and future work.....84

CHAPTER 4 86

Body mass index and IVF outcome: a prediction model 86

Introduction 87
Aims and objectives.....92
Methods 92
Derivation cohort.....92
Primary outcome.....94
Model development 94
Missing data.....97
Predictive ability.....97
Model validation.....98
Ethical approval 100
Results 100
Missing data.....104
Univariate and multivariate analyses 106
Predictive ability.....110
Model validation.....113

Discussion **119**

Main findings 119

Strengths and limitations 119

Comparison to existing models 122

Clinical implications 124

Future research 129

CHAPTER 5 131

National study of prevalence of thyroid dysfunction in women of reproductive age 131

Introduction **132**

Aims and objectives 139

Methods **139**

Eligibility criteria 140

Thyroid function test reference ranges 141

Demographic data 142

Sample size 143

Statistical analyses 143

Ethical approval 144

Results **144**

Geographical distribution of population 144

Overall description of data for each demographic subgroup 148

Overall prevalence of each thyroid function group for whole population 149

Prevalence of each thyroid dysfunction group subgrouped into age, BMI, ethnicity and originating population....150
Relationship between TSH concentration and each demographic subgroup....151
Discussion....155
Main findings....155
Strengths and limitations....155
Comparisons to existing literature....157
Conclusions and future work....162

CHAPTER 6 165

National study of prevalence of thyroid autoimmunity in women of reproductive age.... 165

Introduction....166
Aims and objectives....171
Methods....172
Eligibility criteria and recruitment setting....172
Thresholds for thyroid antibody tests....172
Thyroid function grouping and demographic data....173
Statistical analyses....173
Results....174
Distribution of patients in study and regional prevalence's....175
Overall description of data and prevalences of TPOAb grouped by thyroid function....178

Prevalence of TPOAb within demographic subgroups.....180
Relationship between TSH concentration and prevalence of TPOAb positivity 183
Discussion.....184
Main findings.....184
Strengths and limitations.....185
Comparisons with existing literature 186
Conclusions and future work.....190

CHAPTER 7 192

The effect of subclinical hypothyroidism on IVF outcome..... 192

Introduction 193
Aims and objectives.....197
Methods 198
Study design 198
Eligibility criteria 198
Classification of SCH 199
Selective allocation of LT4 treatment.....199
Statistical analysis 202
Results 203
Overall study population 203
SCH vs. euthyroid.....203
Untreated SCH vs. euthyroid (TSH 0.44-3.64mU/L) 209
Untreated SCH vs. euthyroid TSH (0.44-2.5mU/L) 210

Discussion **211**

Main findings 211

Strengths and limitations 214

Comparison to existing literature 216

Conclusions and future work 217

CHAPTER 8 ..... **220**

Discussion **220**

Ethnicity and IVF outcome; summary of findings and future work 221

BMI and IVF outcome; summary of findings and future work 222

Thyroid function and autoimmunity; summary of findings and future work 224

Conclusion 227

APPENDICES **229**

Appendix 1. Complete case multivariate logistic regression model for live birth (n=2911) 230

Appendix 2. Calibration and internal validation for complete dataset prediction model (n=2911) 232

Appendix 3. Prevalence study consent form 233

Appendix 4. Prevalence study protocol 234

Appendix 5. Ethical approval for prevalence study from Berkshire B REC 248

Appendix 6. Prevalence of euthyroid 1 (TSH 0.44-4.50 and T4 10-21) 250

Appendix 7. Prevalence of euthyroid 1a (TSH 0.44-2.49 and T4 10-21) 251

Appendix 8. Prevalence of euthyroid 1b (TSH 2.5-4.5 and T4 10-21) 252

Appendix 9. Prevalence of overt hypothyroidism (TSH >4.50 and T4 <10) ... 253

Appendix 10. Prevalence of overt hyperthyroidism (TSH <0.44 and T4 >21) 254

Appendix 11. Prevalence of moderate SCH (TSH 4.51-10 and T4 10-21) 255

Appendix 12. Prevalence of severe SCH (TSH >10.0 and T4 10-21) 256

Appendix 13. Prevalence of subclinical hyperthyroidism (TSH <0.44 & T4 10-21) 257

References 258

LIST OF FIGURES

Figure 1. Summary of PRISMA flow diagram – summary of selection process of included papers 17
Figure 2. STROBE assessment for the included 16 papers 23
Figure 3. Black vs. White – Live birth (fresh cycles) 37
Figure 4. Black vs. White – Clinical pregnancy (fresh cycles) 38
Figure 5. Black vs. White – Live birth (frozen cycles) 39
Figure 6. Black vs. White – Clinical pregnancy (frozen cycles) 39
Figure 7. Asian vs. White – Live birth (fresh cycles) 42
Figure 8. Asian vs. White – Clinical pregnancy (fresh cycles) 43
Figure 9. South Asian vs. White – Live birth (fresh cycles) 44
Figure 10. South Asian vs. White – Clinical pregnancy (fresh cycles) 44
Figure 11. Hispanic vs. White – Live birth (fresh cycles) 46
Figure 12. Hispanic vs. White – Clinical pregnancy (fresh cycles) 46
Figure 13. Age plotted against the log odds of the probability of live birth. 56
Figure 14. BMI plotted against the log odds of the probability of live birth 57
Figure 15. Duration of infertility plotted against the log odds of the probability of live birth 58
Figure 16. White vs. Black – adjusted odds ratio for live birth (fresh cycles) 79
Figure 17. White vs. Asian – adjusted odds ratio for live birth 82
Figure 18. White vs. South Asian – adjusted odds ratio for live birth 82
Figure 19. Live birth rate per cycle started (1991-2012) 89
Figure 20. AFC plotted against the log odds of the probability of live birth 96
Figure 21. Definition of eligible cohort and analysis samples 101
Figure 22. Receiver Operating Curve for final model 111
Figure 23. Calibration plot for final model 112
Figure 24. Calibration slope plot following external validation 117
Figure 25. Calibration plot following recalibration 118
Figure 26. Geographical distribution of recruiting centres 145
Figure 27. Flowchart of patients in prevalence study 146
Figure 28. Numbers screened by each participating centre; grouped by geographical region 147

Figure 29. The relationship between age and TSH concentration.....151
Figure 30. The relationship between BMI and TSH concentration 152
Figure 31. The relationship between ethnicity and TSH concentration.....153
Figure 32. The relationship between originating population and TSH concentration 154
Figure 33. Flowchart of patients in study.....176
Figure 34. Graphical representation of TSH and probability of TPOAb positivity ...184
Figure 35. Management Pathway for abnormal thyroid function in BWH Assisted Conception Unit.....201
Figure 36. Summary of patients in the study 204

LIST OF TABLES

Table 1. Summary of research studies within the PhD thesis.....9
Table 2. Assessment for quality of a cohort study – Newcastle Ottawa Scale 20
Table 3. Summary of study characteristics.....26
Table 4. Demographic breakdown for included studies.....31
Table 5. Baseline characteristics for each ethnic group 61
Table 6. Cycle characteristics for each ethnic group.....63
Table 7. Outcome data for each ethnic group 65
Table 8. Univariate analysis for live birth (all cycles).....66
Table 9. Multivariate logistic regression model for live birth (n=13473).....67
Table 10. Univariate analysis for clinical pregnancy (all cycles).....69
Table 11. Multivariate logistic regression model for clinical pregnancy (n=13473)...70
Table 12. Baseline characteristics of the cohort for whole dataset and the complete cases dataset.....102
Table 13. Descriptive data for whole dataset (n=9915) 103
Table 14. Display of missing and valid data 105
Table 15. Univariate associations of potential predictors for live birth following IVF for cases that have a non-missing value for each predictor (whole dataset) and cases with complete data for all predictors (complete cases) 107
Table 16. Final multivariate logistic regression model for Live birth (n=9915).....108
Table 17. Baseline characteristics, cycle characteristics and outcome data for external cohort (n=2723).....114
Table 18. Examples of predicted probabilities using the final model and calibrated model 125
Table 19. The effect of BMI on predicted probabilities within the model.....128
Table 20. Overall description of dataset (n=7022) - delete.....148
Table 21. Overall prevalence's for whole dataset.....149
Table 22. Pooled prevalence of thyroid antibodies across various populations.....168
Table 23. TPOAb positivity by geographical region.....177
Table 24. Overall description of dataset.....179
Table 25. Prevalence of TPOAb positivity within each thyroid function group.....180

Table 26. TPOAb positive prevalence within demographic subgroups.....181
Table 27. TSH concentration and probability of being TPOAb positive.....183
Table 28. Baseline characteristics of the cohort.....205
Table 29. Cycle characteristics of the study cohort.....207
Table 30. Outcome data for study cohort.....207
Table 31. Outcome data comparing euthyroid (TSH<3.64) and untreated SCH (3.64-5.99).....209
Table 32. Outcome data comparing TSH <2.5 and untreated SCH (2.50-5.99).....210

CHAPTER 1

Thesis overview and objectives

Introduction to in-vitro fertilisation (IVF)

Since the introduction of IVF treatment in 1978 the technology has improved greatly. Figures from the annual Human Fertilisation and Embryology Authority (HFEA) reports show a trend in success rates over the years due to technological advancements and changes in clinical practice[1]. The popularity of assisted reproduction also continues to grow, the most recent HFEA report found that 49,636 women underwent a total of 64,600 cycles of IVF or intra-cytoplasmic sperm injection (ICSI) in 2013, compared with 62,158 cycles in 2012, this is an increase in the number of treatment cycles of 3.9%[1]. Despite the significantly increased success rates since the introduction of IVF, the average chance of a couple having a baby following IVF treatment (if the female is aged 35 and under) is still only around 33%[1]. Consequently there is constant pressure from patients and policy makers to improve fertility services and outcomes, therefore much research is conducted within the field of assisted reproductive technology (ART).

There are many factors that can affect IVF outcome, which can be split into three main categories: pre-treatment factors (i.e. the age of the female, the ovarian reserve, the cause of infertility etc); embryo factors (e.g. poor embryo quality, having no embryos) and uterine factors (e.g. endometrial thickness, uterine receptivity). The work of this thesis focuses solely on pre-treatment factors and the impact of certain important under-investigated factors on IVF outcome.

While there is an exhaustive list of pre-treatment factors that have been identified as having an effect on IVF outcome, the work in this thesis focuses on three in particular

which are common, easily available and contribute important information for women before embarking on their IVF treatment; ethnicity, body-mass index and thyroid function.

Ethnicity and IVF outcome

Ethnicity is often explored as a prognostic marker in studies in medicine, but the link between ethnicity and assisted reproduction outcome remains unclear. Large studies, using the Society for Assisted Reproductive Technologies (SART) database, in the United States (US) have attempted to explore the relationship between ethnicity and IVF outcome; however, many of these studies have been unable to account for the common confounders[2–5]. The two largest studies, by Seifer et al in 2008 and 2010 (n=44,585 cycles and n=158,693 cycles respectively) only looked at differences between Black women and White women[4,5]. Furthermore, as US populations dominate the existing studies, this means inclusion of ethnic groups such as Hispanic, Pacific Islander and American Indian[2], which makes the findings non-transferrable to a UK population. In addition to this, several large studies have inappropriately combined ethnic groups that do not necessarily behave the same i.e Indian and Chinese grouped into "Asian"[2,3,6]. There is yet to be a large study in the UK to investigate whether ethnicity impacts on IVF outcome. Furthermore there needs to be clear and appropriate definitions of different ethnic groups and finally the common confounders need to be accounted for; such as age, cause of infertility, duration of infertility and so forth. While ethnicity is not a factor that women can change, it is important to determine whether it effects success rates following IVF treatment as it should be taken into consideration when counselling women pre-

treatment. Moreover, if there is an association between ethnicity and IVF outcome we need to work to explain why and find ways to improve IVF outcomes for all women of all ethnicities.

BMI and IVF outcome

The relationship between raised body mass index and poor obstetric outcomes is well known[7,8]. However, the literature regarding the association between raised body-mass index and IVF outcome is conflicting. Some recent studies have shown that raised BMI does not appear to reduce clinical pregnancy or live birth rates and that women with raised BMI do not require higher doses of gonadotrophins in their treatment compared with normal BMI women[9,10]. Whereas other studies, including a large systematic review and meta-analysis by Rittenberg et al, have shown increased cancellation rates for women with raised BMI and lower clinical pregnancy and live birth rates[11–13]. Despite this conflicting evidence linking raised BMI to poorer IVF outcome, because the relationship between raised BMI and adverse pregnancy outcomes is well established this has dictated the UK government funding criteria; with NHS funded treatment only provided for those with a body-mass index under 30. What remains unclear is exactly to what degree body mass index affects IVF outcome, and the interplay of this with other factors. Often when counselling patients clinicians advise women to lose weight to improve their chances of success, however there is no tool or model available that allows for calculation of success rates based on body-mass index and it's association with other important predictive factors, such as age. In fact there is currently no model available for use that predicts the chances of IVF success for women *before* they undergo their first treatment. Frequently in

clinical consultations with couples seeking IVF treatment, clinicians are asked what the couples chances of live birth are; crude age related success rates are provided based on national data. To date there is no model to calculate a more personalised probability of live birth at the point before commencing treatment. Creation of such a model would be important as it would facilitate decision making for couples at the most critical part of their journey and fill a gap in counselling in current clinical practice. There is a need for a relevant prediction model, incorporating body mass index, to be used as an adjunct to counselling and decision-making for clinicians and patients before they embark on IVF treatment.

Thyroid dysfunction and thyroid autoimmunity

Finally, thyroid problems are one of the most prevalent of all medical conditions, especially in women of reproductive age. The most prevalent form of thyroid disease is subclinical hypothyroidism (SCH). SCH is more common in females and in particular those with fertility problems[14]. In the infertility population the incidence of SCH has ranged between 1-43% with a mean of around 13%[15]. Subclinical thyroid problems are often asymptomatic and therefore go undetected, however evidence has shown that subclinical disease can have negative effects on a pregnancy, including increased risk of miscarriage, perinatal loss, preterm birth, pre-eclampsia and low IQ in the offspring[16,17]. Consequently, the Endocrine Society Clinical Practice Guideline (ESCPG) regarding “Management of Thyroid Dysfunction during Pregnancy and post-partum” recommends the use of hormone replacement therapy, in the form of Levothyroxine (LT4) treatment, for pregnant women with subclinical hypothyroidism as well as those with overt disease[15].

One of the central challenges in defining SCH is agreeing on the upper limit of TSH. This has lead to significant debate over whether routine screening should be performed on all pre-conception women; with a view to treating prior to- and during pregnancy in order to optimise obstetric outcomes. The most widely accepted reference range for a "normal" TSH is 0.4-4.5mU/l. However, there is currently a shift in clinical practice, in certain parts of the world, towards routinely treating women with SCH who are trying for a pregnancy, particularly the infertility population, and aiming for a pre-conception target TSH of <2.5mU/l. Despite this suggestion of aiming for a tighter pre-conception TSH threshold below 2.5mU/l, it is not currently recommended routine practice for universal thyroid function screening in women trying for a pregnancy, or even in women who are pregnant, by any of the major endocrine societies. This is based on the limited conclusive evidence to suggest that treatment with Levothyroxine for women with TSH 2.5-4.5mU/l has any benefit on pregnancy outcomes, including fertility outcomes after assisted reproduction, compared with untreated women[18].

Following on from thyroid dysfunction there is also the issue of thyroid autoimmunity. The presence of thyroid auto-antibodies in women, specifically Thyroid Peroxidase antibodies (TPOAb), has been linked to increasing the chances of adverse effects on the pregnancy, including miscarriage and pre-term birth[19]. Furthermore, it is thought that around 1 in 5 women are positive for TPOAb[19] and that this figure is even higher in women with backgrounds of infertility or recurrent miscarriage[19–21]. However as the antibodies alone (without thyroid dysfunction) are asymptomatic, the vast majority of women who have them will never know unless specifically investigated.

Despite the fairly common prevalence and known detrimental effects in pregnancy it is currently not routine practice to check thyroid function for women who are actively trying for a pregnancy, either via natural conception or fertility treatment. Furthermore, existing literature shows that women who carry these antibodies are also at higher risk of developing thyroid abnormalities during pregnancy and should therefore be offered routine thyroid monitoring throughout pregnancy[22]; however without routine screening of pre-conception women, the majority of TPOAb positive women will be missed.

Existing research has yet to identify if there are specific higher risk women, based on demographic such as age, BMI or ethnicity, within the cohort of those of reproductive age who are trying for a pregnancy (including fertility patients), who may benefit from routine TFT and TPOAb screening and/or treatment prior to embarking on a pregnancy. A large-scale prevalence study of thyroid dysfunction and thyroid autoimmunity is required to help identify the women who may benefit the most from thyroid screening. A further important unanswered question is at what severity of subclinical thyroid disease (i.e. what threshold of TSH concentration) should treatment be commenced and will this improve pregnancy success.

Undergoing assisted reproductive technologies can be a very costly and emotional burden for many couples. It is therefore crucial that women are well informed about their chances of success and that they are appropriately stratified, investigated and managed before commencing their fertility treatment. The work presented in this thesis has adopted a mixed methodological approach to explore certain key factors

that affect IVF outcome in order to aid effective counselling and clinical decision-making for both patients and clinicians prior to commencing assisted reproduction treatment.

Thesis objectives

The aim of the work in this thesis is to address in detail the three key factors identified as having an effect on IVF and pregnancy outcome; ethnicity, BMI and thyroid function/autoimmunity. Chapters 2 and 3 will explore ethnicity and IVF outcome; chapter 2 will be a systematic review and meta-analysis of all the existing literature and chapter 3 will be a cohort study and an updated meta-analysis combining the cohort study with the existing studies. Chapter 4 will be the derivation and validation of a prediction model incorporating BMI. Finally, chapters 5-7 will be the thyroid work; prevalence of thyroid dysfunction (chapter 5), prevalence of thyroid autoimmunity (chapter 6) and a cohort study investigating the effect of SCH on IVF outcome (chapter 7).

The summary of research studies and the methodology applied in this thesis, as per PICO format, is displayed in Table 1.

Table 1. Summary of research studies within the PhD thesis

Chapter Number	Population	Intervention or factor of interest	Comparison or reference standard	Outcome	Study Design
	Objective 1: To investigate the effect of ethnicity on IVF outcome				
2	First non-donor cycle of IVF or ICSI treatment for all women. Both fresh and frozen cycles. No year or country restrictions.	Ethnic group; Black, Asian, Hispanic	White population	Live birth rates and clinical pregnancy rates	Systematic review and meta-analysis
3	Women undergoing first non-donor cycle of IVF or ICSI treatment at any CARE clinic across the UK between 2008-2012. Both fresh and frozen cycles.	Ethnic group; Black, South Asian, Chinese, mixed, or other	White population	Live birth, clinical pregnancy and miscarriage.	Observational cohort study
	All studies included in systematic review plus cohort study.	Ethnic group; Black, South Asian, Chinese, mixed, or other	White population	Live birth, clinical pregnancy and miscarriage.	Updated meta-analysis including data from cohort study

Chapter Number	Population	Intervention or factor of interest	Comparison or reference standard	Outcome	Study Design
	Objective 2: To investigate the effect of body mass index (BMI) and it's interplay with other predictors on IVF outcome				
4	Women undergoing their first fresh non-donor cycle of IVF or ICSI treatment at any CARE clinic across the UK between 2008-2012.	Body mass index is the primary factor of interest for the prediction model as this is an important factor that has not previously been used in any prediction tool for IVF outcome. Other predictors built into the model: age, ethnicity, ovarian reserve, cause of infertility, duration of infertility history of previous live birth and history of previous miscarriage.		Live birth rate	Construction of prediction model.
4	Women undergoing their first fresh non-donor cycle of IVF or ICSI treatment at any CARE clinic across the UK between 2013-2014.	As above		Live birth rate	External validation of prediction model.

Chapter Number	Population	Intervention or factor of interest	Comparison or reference standard	Outcome	Study Design
	Objective 3: To investigate the prevalence of thyroid dysfunction and autoimmunity in women of reproductive age				
5	All women who underwent screening blood test for thyroid function as part of the *"Prevalence study"* (sub-study to TABLET trial) across the UK.	Women with abnormal thyroid function (overt or subclinical)	Euthyroid women	Demographic features: age, ethnicity, BMI, originating population	Prospective national multi-centre prevalence study
6	All women who underwent screening blood test for thyroid peroxidase antibodies (TPO) as part of the *"Prevalence study"* (sub-study to TABLET trial) across the UK.	Women who are TPO antibody positive	Women who are TPO antibody negative	Demographic features: age, ethnicity, BMI, originating population	Prospective national multi-centre prevalence study
	Objective 4: To investigate the effect of subclinical hypothyroidism (SCH) on IVF outcome				
7	All women screened as part of the TABLET trial at Birmingham Womens Hospital Assisted Conception unit between June 2012-December 2013.	Women with SCH, TSH >3.63 and normal T4. Women with TSH 3.63-5.9 received no treatment, women with TSH 6.0≤ started on Levothyroxine.	Euthyroid women	Live birth, clinical pregnancy, miscarriage	Prospective observational cohort study

CHAPTER 2

Investigating the effect of ethnicity on IVF outcome: a systematic review and meta-analysis

The work in this chapter was published in Reproductive Biomedicine Online; ***published online 03.06.15***

Introduction

Ethnicity is one of the most investigated prognostic factors in medicine. However, studies investigating the relationship of ethnicity to IVF outcome are often limited in terms of sample sizes and have produced inconclusive findings. Furthermore, the large majority of existing studies are American and so focus on ethnic groups that are specific to their population, such as Hispanic, African American and American Indian, making the findings non-transferable to a UK population.

A study published in 2009 by Dayal et al[23] compared IVF outcomes for 251 females; 180 White and 71 African American. They concluded that there were no differences in pregnancy outcomes following IVF, although they did find that African Americans produced fewer embryos than White women. A larger study published in 2005 by Bendikson et al[24] looked at 1135 women (1039 White, 43 African American, 35 Asian (combining South Asians and South-East Asians) and 18 Hispanic), also concluded that their data showed no significant difference in pregnancy outcomes with IVF among the ethnic groups.

More recent larger studies[4,5,25–29] have collectively provided stronger evidence on the existence of racial disparity in assisted reproductive technology (ART) outcomes. In 2008 and 2010 Seifer et al[4,5], showed that Black race was an independent risk factor for not achieving live birth. Fujimoto et al[6] found a statistically significant decreased odds ratio of achieving live birth amongst Asians, Hispanics and Blacks compared with Whites.

A review by Wellons et al[30] in 2012 found that there were significant racial disparities in IVF outcomes, however they restricted their findings to studies using Society for Assisted Reproductive Technologies (SART) data only and did not perform a meta-analysis of the data. A UK based study[31] recently published data showing that ethnicity was an independent risk factor for IVF and ICSI outcome, although this was not statistically significant. Furthermore, the numbers representing each ethnic group were small, meaning that they had to group all ethnicities into one group.

There appears to be variation in success rates with assisted conception for women from different ethnic groups with several studies producing inconsistent findings. There is a need for accurate information regarding ethnicity and its effect on IVF outcome as this can lead us to investigate the potential biological plausibility.

Aims and objectives

1. To provide a robust synthesis of all available literature on the relationship between ethnicity and IVF outcome.
2. To perform the first meta-analysis of all the published data regarding ethnicity and IVF outcome.

Methods

Our systematic review followed a protocol developed using widely recommended and comprehensive methodology[32–34].

Data sources

This review focused on studies where IVF was performed and the primary outcomes measured were clinical pregnancy or live birth.

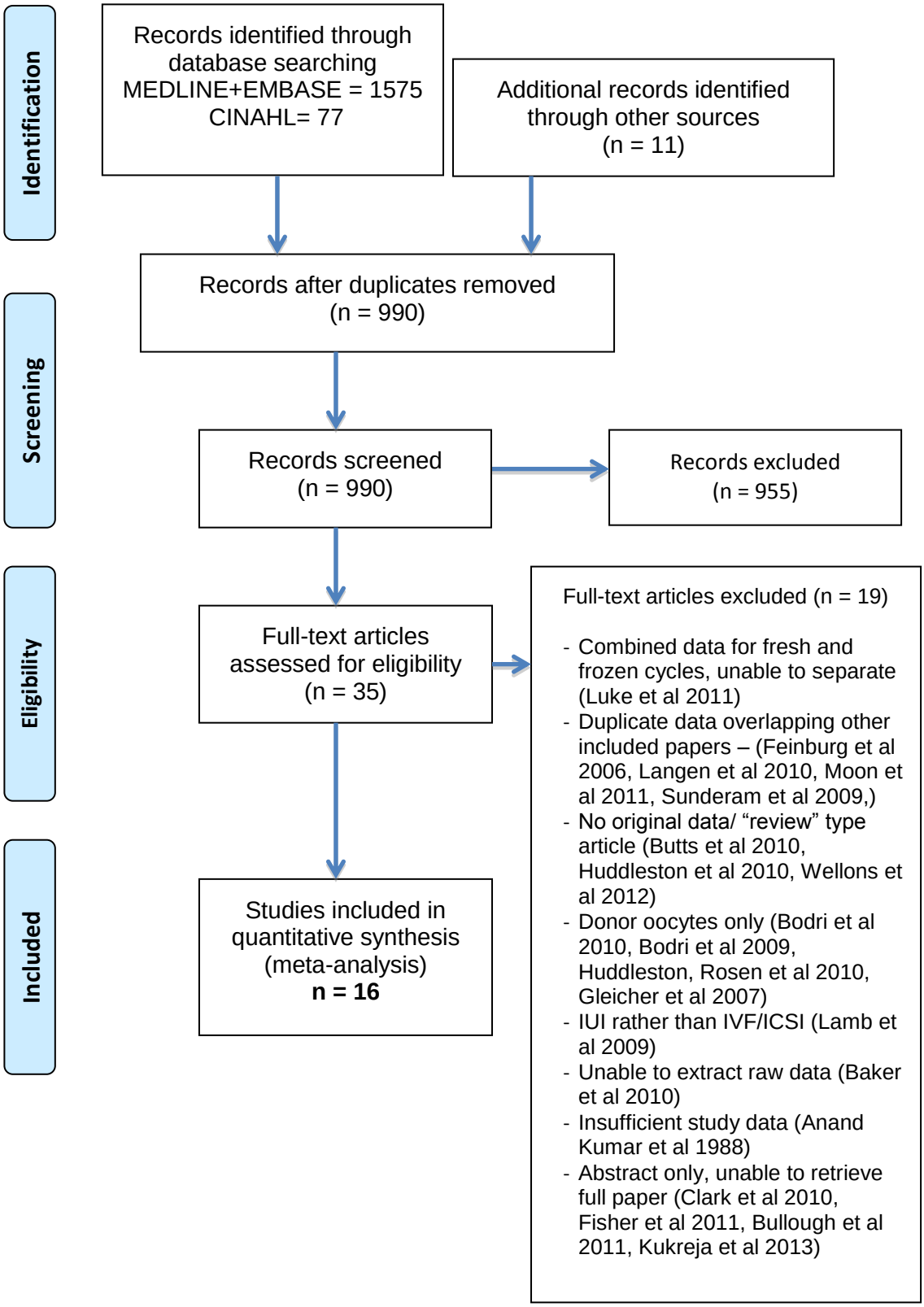
MEDLINE (from inception to Jan 2015) EMBASE (from inception to Jan 2015) and CINAHL (from inception to Jan 2015) databases were searched electronically and Web of Science was used to search for grey literature. The search of MEDLINE and EMBASE and CINAHL captured citations containing the relevant MeSH keywords and word variants for “ethnicity” and “in-vitro fertilisation”; race, assisted reproductive technology. Bibliographies of relevant articles were manually searched to identify papers not captured by the electronic searches. Authors were contacted for completeness of the search. There were no language restrictions in the search or selection of papers.

Eligibility criteria for selecting studies

Studies were selected in a two-stage process. Initially, all abstracts or titles in the electronic searches were scrutinised by two reviewers (R.D. and R.M.) and full manuscripts of potentially eligible citations were obtained. Differences were resolved by discussion with a third reviewer (H.H). Studies were selected if the primary

outcomes measured were clinical pregnancy and or live birth and if IVF (including ICSI) was the method of assisted conception. Studies had to report their success rates by ethnic group to allow data extraction for meta-analysis. Only studies where first cycle was used were selected, this was to eliminate bias from previous cycles. Both fresh and frozen cycles were included with the outcome data analysed separately. Data were analysed separately to reduce treatment type (i.e. fresh vs. frozen) acting as a confounder. A summary of the selection process is displayed in Figure 1.

Figure 1. Summary of PRISMA flow diagram – summary of selection process of included papers



Data extraction and synthesis

Data were extracted by two reviewers (R.D. and R.M.) and verified by a 3rd reviewer (H.H.). For each of the outcomes, data were extracted into tables. Primary outcomes were live birth and clinical pregnancy (also recorded as clinical intrauterine gestation). Data was collected regarding the ethnicities reported, study design, population size, outcomes measured and whether fresh or frozen cycles were included.

Data were extracted from each paper for unadjusted live birth and clinical pregnancy rates per cycle. Baker et al[25] were contacted for the raw numbers, as this was not extractable from the paper, however no correspondence was received and so this paper was not included. Several of the papers published by US authors used data from the same time period and from the SART database. To avoid analysing duplicate data, those papers with the greater sample numbers were included for analysis and any overlapping datasets from the remaining papers were removed[6,27,35]. Any papers where the outcome data for fresh and frozen data were combined, and we were not able to extract separately, were also removed[36].

The meta-analysis was performed using Review Manager (version 5.0 for Windows) to combine and analyse the data; using the generic inverse variance method. For the purpose of the meta-analysis the ethnicities were grouped into four broad categories; White, Black, Asian and Hispanic. For the Asian group this included both South Asians and South-East Asians. Each ethnic group was compared with a White reference population for all outcomes. The outcomes from fresh and frozen cycles

were analysed separately. The adjusted odds ratios were pooled where possible to account for potential confounders. Heterogeneity was assessed by examining the χ^2 statistics and a random-effects model was used where there was statistically significant heterogeneity. Heterogeneity was presented statistically and graphically using forest plot estimates of rates and 95% confidence intervals.

Quality assessment

All articles meeting the selection criteria were assessed for quality using the Newcastle Ottawa Scale[37,38] (Table 2), the exact criteria used to award points for quality are described following Table 2. This tool was selected as it is designed to assess the quality of non-randomised studies; which is in-line with the study design of the included studies in this review. The quality of reporting was assessed using the STROBE checklist[39] (Figure 2).

Table 2. Assessment for quality of a cohort study – Newcastle Ottawa Scale

Study	1. Selection				2. Comparability (max 2 stars available)	3. Outcome			Total no. of stars
	(i) Represen-tativeness of the interven-tion cohort	(ii) Selection of the non-interven-tion cohort	(iii) Ascertain-ment of interven-tion	(iv) Demonstration that outcome of interest was not present at start of study	(i) Comparability of cohorts on the basis of the design/ analysis	(i) Assess-ment of outcome	(ii) Was follow up long enough for outcomes to occur	(iii) Adequacy of follow up of cohorts	
Bendikson et al 2005	*	*	*	*	---	*	*	*	7
Csokmay et al 2011	*	*	*	*	---	*	*	*	7
Dayal et al 2009	*	*	*	*	---	*	*	*	7
Fujimoto et al 2010	*	*	*	*	---	*	*	*	7
Jayapraka-san et al 2014	*	*	*	*	---	*	*	*	7
Lashen et al 1999	*	*	*	*	*	*	*	*	8

Mahmud et al 1995	*	*	*	*	*	*	*	*	8
McCarthy-Keith et al 2010	*	*	*	*	---	*	*	*	7
Nichols et al 2001	*	*	*	*	---	*	*	*	7
Purcell et al 2007	*	*	*	*	---	*	*	*	7
Seifer et al 2008	*	*	*	*	---	*	*	*	7
Seifer et al 2010	*	*	*	*	---	*	*	*	7
Shahine et al 2009	*	*	*	*	---	*	*	*	7
Sharara and McClamrock 2000	*	*	*	*	---	*	*	*	7
Sharara et al 2012	*	*	*	*	---	*	*	*	7
Shuler et al 2011	*	*	*	*	---	*	*	*	7

Key: Each * represents if individual criterion within the subsection was fulfilled (maximum 9 stars).

1. (i) Studies received a * if the cohort of interest included all non-White women undergoing their first non-donor cycle of IVF

   (ii) Studies received a * if the reference cohort included all White women undergoing their first non-donor cycle of IVF

   (iii) Studies received a * if the ascertainment of intervention was in a secure record or via a structured interview

   (iv) Studies received a * if the women were non-pregnant at the start of the study

2. (i) Studies received a * if the women were matched by at least age and body mass index.

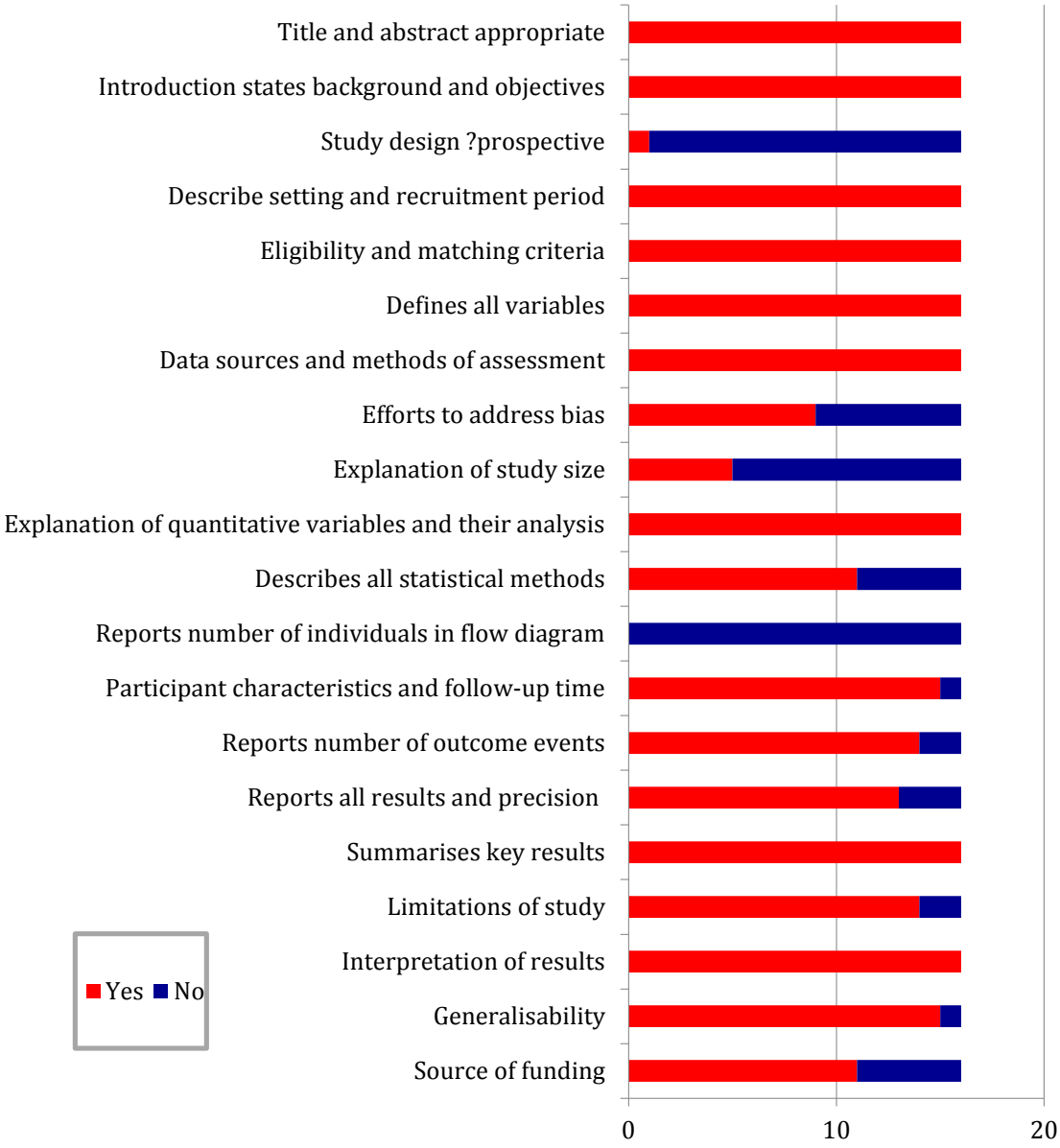
   Studies received ** if they were matched for additional factors (e.g. socioeconomic status, education)

3. (i) Studies received a * if there were independent blinded assessments

   (ii) Studies received a * if the women were followed up to the end of their first IVF cycle of treatment

   (iii) Studies received a * if all subjects were accounted for or if those lost to follow up were unlikely to introduce bias (i.e. number lost <=20%, or description of those lost suggesting no different from those followed)

Figure 2. STROBE assessment for the included 16 papers



Results

Sixteen studies were included for the meta-analysis[2–6,23,24,26,28,29,31,40–43]. Thirteen of the studies had data for fresh cycles only[2,3,6,23,24,28,29,31,40–44] , 1 for frozen only[26] and 2 studies included both[4,5]. All papers used data for non-donor cycles and first treatment cycles only were included. A summary of each of the included study characteristics is shown in Table 3.

The demographic breakdown for each paper is shown in Table 4. Where possible, data has been extracted for age, BMI and fertility diagnosis. Of the 8 papers which reported BMI only two found statistical significant differences; Dayal et al[23] and Nichols et al[41] both reported the White population as having a lower mean BMI compared with Black (African American).

When comparing age; Seifer et al[4] found a statistically significantly higher proportion of White women under the age of 35 years compared with Black women, and Fujimoto et al[6] also showed the highest mean age amongst Black women compared with White, Asian and Hispanic. Both Sharara et al[42] and Shahine et al[28] found Asian women to have a statistically significant mean younger age compared with White women, whereas Purcell et al[3] found Asian women to be older than Caucasian from the SART dataset.

Nine of the 16 papers[2,4–6,23,24,26,41,43] found that Black women have a statistically significantly higher likelihood of tubal and/or uterine factor compared with White

women, whereas White women were found to have a higher likely diagnosis of endometriosis. Polycystic ovarian syndrome (PCOS) was found to be more common amongst Asians than White women[40,42]. There was also a statistically significantly increased duration of fertility seen in Asian women compared with White women[40,44].

Table 3. Summary of study characteristics

Reference, year and study design	Sample population	Outcomes measured	Fresh/frozen cycles	Patient numbers
Bendikson et al 2005 Retrospective Cohort study	Women undergoing first IVF cycle between August 1994 and March 1998 at Boston IVF, Brigham Womens Hospital and Boston Reproductive Science Centre (USA)	Live birth, chemical and ectopic pregnancies, miscarriage	First cycle, fresh non-donor transfer	Total 1135 cycles White = 1039 Black (African American) = 43 Asian = 35. Hispanic = 18
Csokmay et al 2011 Retrospective Cohort Study	All patients who underwent frozen blastocyst transfer between 2003 and 2008 in a University-based ART program. University of California.	Clinical pregnancy and live birth	Frozen embryo cycles with autologous oocytes	Total 169 women White (Caucasian) = 119 Black (African American) = 50
Dayal et al 2009 Retrospective Cohort Study	All African-American women and Caucasian women who underwent IVF cycles between 01-Jan-2004 and 31-Dec-2005 at George Washington Fertility & IVF Centre (USA)	Biochemical and clinical pregnancy, live birth and implantation	Initial fresh non-donor IVF cycles/embryo transfer	Total 251 women White (Caucasian) = 180 Black (African American) = 71

Fujimoto et al 2010 Retrospective Cohort Study	Cycles between 2004 and 2006 in White, Asian, Black and Hispanic women, identified using the SART database (USA).	Clinical pregnancy and live birth given as an adjusted odds ratio compared to the white women group. Also stillbirth rate and plurality of live-born pregnancies	Fresh non-donor cycles	Total 139,027 cycles. White = 107484 Asian = 13671 Black = 8903 Hispanic = 8969 Black data not used for analysis as duplicate data from Seifer et al 2010. White data used for reference purposes.
Jayaprakasan et al 2014 Retrospective cohort study	All women undergoing first cycle of ART between 2006 and 2011 at Nottingham University Research and Treatment Unit in Reproduction (NURTURE), UK.	Biochemical pregnancy rate, clinical pregnancy rate and live birth rate	First cycle, fresh, non-donor	Total 1571 women White (Caucasian) = 1291 South Asian = 182 African-Caribbean = 30 Middle Eastern = 14
Lashen et al 1999 Nested case-control study	Patients undergoing IVF between 1994-1997 at Birmingham Womens Hospital, Assisted conception unit (UK).	Implantation rate and clinical pregnancy	First cycle, fresh, non-donor	Total 324 women White (Caucasian) = 216 Asian = 108 (58 Pakistani, 34 Indian, 16 Bangladeshi)

Mahmud et al 1995 Controlled comparative clinical study	Patients selected prospectively from a Oxford (UK) IVF database from April 1987- Dec 1993. "Indian" women (Pakistani, Bangladeshi and Indian). Matched by age, BMI and year of treatment.	Cumulative pregnancy rates over 3 cycles. Rates of abandoned cycles, egg retrievals, endometrial thickness, clinical pregnancies, miscarriages and live births from the first IVF cycle.	First fresh IVF cycles	Total 132 women. White = 88 Asian (Indian) = 44
McCarthy-Keith et al 2010 Retrospective Cohort Study	Women undergoing cycles from 2000 - 2005 in federal assisted reproduction programmes (USA). Walter Reed Army Medical Center, Wilford Hall Medical Center and Tripler Army Medical Center.	Assisted reproduction technique utilisation rate, clinical pregnancy and live birth	First cycle fresh non-donor transfer	Total 2050 women. White =1280 Black (African American) = 353 Asian = 110 Hispanic = 81 American Indian = 8 Pacific Islander = 12 Multi-racial = 85 Not stated = 121 Data for LB only available for Whites, African-American and Hispanics.
Nichols et al 2001 Retrospective cohort study	Women undergoing IVF between Nov 1996 and June 2000 in a Hospital based IVF practice (Greenville Hospital, South Carolina, US).	Implantation rate and clinical pregnancy	Non-donor, multiple cycles. Only data for 1 cycle per patient analysed	Total cycles 297 White = 273 Black (African American) = 24

Purcell et al 2007 Retrospective Cohort Study	Caucasian and self-identified Asian women. Clinics reporting to SART for years 1999-2000, and University based clinic – University of California (USA) for years Jan 2001 – December 2003.	Clinical pregnancy and live birth	First Fresh non-donor cycles	Two data sets: 1) SART dataset; White (Caucasian) = 25,843 cycles Asian = 1,429 cycles 2) University database; Asian = 197 cycles. White (Caucasian) = 370 cycles
Seifer et al 2008 Retrospective Cohort Study	Cycles performed between 1999 and 2000 by Society of Assisted Reproductive Technology member clinics who perform >50 cycles/ year and have race/ethnicity reported in >95% of cycles.	Live birth per cycle started.	Fresh and frozen embryos	Total 44,585 IVF 1st cycles. (34,042 cycles not used for analysis as not 1st cycle) Fresh cycles: White = 32049. Black = 1839 Frozen cycles: White = 10147. Black = 550
Seifer et al 2010 Retrospective Cohort Study	Non-donor IVF cycles between 2004 and 2006 in White and Black women, identified using the SART database (USA).	Live birth per cycle started	Fresh and frozen non-donor IVF cycles	Total 158,693 cycles. (50,143 cycles not used for analysis as not 1st cycle) Fresh cycles: White = 120,994. Black = 10,354 Frozen cycles: White = 25,412. Black = 1,933

Shahine et al 2009 Retrospective Cohort Study	Indian and Caucasian Women undergoing blastocyst transfer between Jan '05 – July '07 in Stanford University fertility centre, USA	Live birth per cycle started	Initial Fresh cycle, blastocyst transfer	Total 225 women White (Caucasian) = 145 Asian (Indian) = 80
Sharara and McClamrock 2000 Retrospective Cohort Study	Women undergoing IVF at an inner city, university-based IVF programme (University of Maryland, USA) between April 1997 and July 1999 <40years.	Implantation rate, clinical pregnancy and ongoing/delivered pregnancy	Fresh non-donor IVF cycles	Total 168 cycles White = 121 cycles Black = 47 cycles
Sharara et al 2012 Retrospective Cohort Study	All white and South Asian Women <40years undergoing blastocyst transfers at Virginia Centre for Reproductive Medicine, USA.	Clinical pregnancy and live birth	Non-donor, initial fresh cycle, blastocyst transfer	Total = 292 White = 238 cycles Asian (South Asian) = 54
Shuler et al 2011 Retrospective Cohort Study	Patients who self identified as Hispanic or as non-Hispanic white undergoing IVF cycles at South Texas Fertility Centre (USA) between 1998 and 2008.	Clinical intrauterine gestation, spontaneous abortion and live birth	First fresh embryo cycles with autologous oocytes	Total 435 cycles Non-Hispanic White = 301 Hispanic = 134

Table 4. Demographic breakdown for included studies

Study	Mean age	Mean BMI	Infertility diagnosis/duration – Statistically significant differences only
Bendikson et al 2005	White = 35 ±4 Black (African American) = 34 ±5 Hispanic = 35 ±6 Asian = 35 ±5	White = 24 ± 5 Black = 26 ±4 Hispanic = 25 ±4 Asian = 23 ±3 NS	Black women more likely than white women to have tubal factor (51.2% vs 22.0%, p<0.0001).
Csokmay et al 2011	Black (African American) = 34.1 ±3.6. White = 34.7 ±4.2 P=0.31 (NS)	Not reported.	Black women had a significantly higher likelihood of tubal factor (64% vs. 31% p<0.0001) and uterine factor (40% vs. 10% p<0.001) and a lower likelihood of anovulation (4% vs. 22% p=0.005) when compared with White women.
Dayal et al 2009	Black (African American) = 37.1 ±3.8 White = 36.5 ±4.1 p=0.28 (NS)	Black = 26.5 ±5.2 White = 23.7 ±4.8 P<0.001	Black women were more likely to have tubal infertility than White women (23% vs. 9%, p=0.007). White women were more likely to have unexplained infertility (53% vs. 32%, p=0.004)
Fujimoto et al 2010	White = 35.3 ±4.6 Asian = 35.8 ±4.6 Black = 35.9 ±4.7 Hispanic = 35.0 ±4.8 p<0.001	Not reported.	Black women less likely to be diagnosed with endometriosis, PCOS and unexplained infertility and more likely to have tubal or uterine factors (p<0.0001 for all) Hispanic women more likely to have tubal factors and less likely to have unexplained factors p<0.0001

Lashen et al 1999	Asian = 32.3 ±0.5 White = 32.3 ±0.3	Not reported.	Significantly longer duration of infertility (P<0.001) and higher incidence of polycystic ovaries in Asian group compared to White (OR = 2.7, 95% CI 1.25-5.8 P=0.01).
Jayaprakasan et al 2014	"Ethnic group" (South-East Asian, Afro-Caribbean, Middle Eastern) = 33.3± 4.5 White = 34.4± 4.3 p<0.001	"Ethnic Group" = 25.8± 4.2 White = 24.3± 3.5	Male factor as a cause of infertility was found to be higher amongst the ethnic group compared to the White (92 (40.7%) compared with 441 (34.2%) p<0.02)
Mahmud et al 1995	White = 32.5 ±4.8. Asian (Indian) = 32.5 ±4.8 (Matched by Age)	White 23.1 ±3.1 Asian (Indian) 22.8 ±3.5 (Matched by BMI)	More Asian women had experienced infertility for ≥8years than White women; 40.9% vs. 21.6% (OR 2.5 [95% CI 1.1, 5.5]) p<0.05.
McCarthy-Keith et al 2010	Reported as no significant difference	Not reported	Tubal factor was more common in Black (African American) women compared against White women (65.4% vs. 32.9%, RR 1.99, 95% CI 1.78-2.22) Black women less likely than White women to have; - Male factor (25.8% vs. 37.6%, RR 0.69, 95% CI 0.57-0.83) - Endometriosis (8.8% vs. 14.2%, RR 0.62, 95% CI 0.43-0.89) - Anovulation (5.7% vs. 9.8%, RR 0.51, 95%CI 0.34-0.78)

Nichols et al 2001	Black (African American) = 32.8 ±3.7 White = 32.7 ±4.1 p=0.84	AA = 27.1 ± 4.6 White = 24.8 ±5.3 p=0.004	African American women more likely than White women to have had tubal factor 67% vs. 27% (p=0.001).
Purcell et al 2007	*Women SART database* Asian = 34.7±4.54 White = 33.7±4,52 p<0.001 *Women UCSF Clinic* Asian = 36.1 ±4.09 White = 36.6 ±4.08 p=0.24 (NS)	Not reported.	SART Database; Asian women showed statistically significant higher rate of diminished ovarian reserve compared with White (11.4% vs 7.9% P<0.01) UCSF Clinic; White women significantly more likely to have diminished ovarian reserve (38.5% vs. 29.1%, p=0.03) whereas Asian women were significantly more likely to have ovulatory dysfunction (15.8% vs. 8.8%, p=0.01) and unexplained infertility (4.6% vs. 1.4%, p=0.02)
Seifer et al 2008	No significant differences between Black or White women in age for either fresh or frozen cycles	Not reported.	Black women twice as likely to have tubal disorder than White women Fresh cycles; 62.8% vs. 28.2%, p<0.001, Frozen; 60.2% vs. 31.6% p<0.001 Black women more likely to have uterine factor (Fresh; 11.4% vs. 4.8%, p<0.001, Frozen; 11.6% vs. 4.3% p<0.001). For both fresh and frozen cycles; White women have significantly higher likelihood of male factor, endometriosis and ovulatory disorders (p<0.001)

Seifer et al 2010	Significantly higher proportion of White women <35years compared with Black; p<0.001	Not reported.	Black women significantly more likely to have tubal factor and uterine factor than White women (p<0.001). White women significantly more likely to have male factor, endometriosis and ovulatory disorders (p<0.001)
Shahine et al 2009	White = 36.7 ±3.9 Asian (Indian) = 34.03 ±4.09 p=0.03	White = 24.4 ± 4.6 Asian (Indian) = 25.2 ±3.7 p=0.2	There were no significant differences in infertility diagnosis.
Sharara and McClamrock 2000	White = 33.0 ±3.7 Black = 32.1 ±3.8 No significant difference	White = 26.7 ±7.0 Black = 28.6 ± 7.5 p=0.38 (NS)	Black women were more likely to have tubal factor than White women (73% vs. 40%, p<0.001). Duration of infertility, in years, was significantly longer in Black women than White women (5.9 ±3.7 vs. 4.6 ±3.2, p=0.03)
Sharara et al 2012	White = 33.5 ±3.6 Asian (South Asian) = 30.5 ±3.5 p<0.001	Not reported.	50% of South Asians had a diagnosis of PCOS compared with 29% of White women p=0.004
Shuler et al 2011	Reported as no significant difference in age	Stated as no difference between groups	Hispanic women more likely to have tubal factor infertility (p <.001) White women higher likelihood of endometriosis (p= 0.02).

Black vs. White

Data from 8 studies[2,4,5,23,24,31,41,43] were combined to compare the Black population with a White population for live birth and/or clinical pregnancy rates (Figures 3 and 4). Results showed Black women were found to have a statistically significant reduction in live births (OR 0.62 [95% CI 0.55-0.71] p<0.001) and clinical pregnancy (OR 0.74 [95% CI 0.64-0.87] p<0.001) compared to White women. The results showed moderate heterogeneity for live birth data and statistically significant high heterogeneity for clinical pregnancy rate.

Of the 8 studies, the two studies by Seifer et al in 2008[5] and 2010[4] contributed to around 70% of the whole data analysed. With the results dominated by data from these studies, to look to see if the findings were driven by these studies we removed them from the meta-analyses. Results showed that there was still a statistically significant reduced chance of live birth with Black women compared to White women (OR 0.63 [95% CI 0.41-0.96] p=0.03) however there was no difference seen with clinical pregnancy (OR 0.86 [95% CI 0.50-1.48] p=0.6).

There were three studies which recorded data for frozen cycles[4,5,26]. For the frozen cycles there was no statistically significant difference in live birth or clinical pregnancy between the Black and White population (Figures 5 and 6); (OR 0.90 [95%CI 0.80, 1.01]) and (OR 0.92 [95% CI 0.54, 1.55]) respectively.

Three papers calculated adjusted relative risks (RR) or odds ratios of the chances of live birth following a first fresh cycle[4–6]. Each paper adjusted for multiple factors,

these included; maternal age, body-mass index, number of embryos transferred, diagnosis of male factor, endometriosis, PCOS, diminished ovarian reserve, tubal factors, uterine factors and other factors. All three papers found that Black women were statistically significantly less likely to have a live birth after first cycle IVF than White women. Both papers by Seifer et al[4,5] (in 2008 they analysed 33,888 first cycles and in 2010 they analysed 131,348 first cycles) looked at the relative risk of *not* achieving a live birth for Black women vs. White; Seifer et al in 2008[5] (aRR 1.31 [95% CI 1.26-1.37] p<0.001) and Seifer et al in 2010[4] (aRR 1.24 [95% CI 1.12-1.36] p<0.001). Fujimoto et al[6] (n=116,387 first cycles) had overlapping data with that of Seifer et al 2010, and therefore was not used in the unadjusted meta-analysis, however their results were consistent with both Seifer papers in showing a statistically significant reduction in live birth for Black women compared with White women after adjusting for confounders; (aOR 0.62 [95% CI 0.56-0.68] p<0.001).

Seifer et al in 2008 and 2010 also calculated an adjusted relative risk analysis for frozen cycles separately; they calculated the aRR of *not* achieving a live birth. Results showed that even after adjusting for confounders there were no significant differences in live birth for Black women compared to White women; Seifer et al 2008 (aRR 0.94 [95% CI 0.70-1.25] p=0.7) and Seifer et al 2010 (aRR 1.10 [95% CI 1.00-1.21] p=0.05).

Figure 3. Black vs. White – Live birth (fresh cycles)

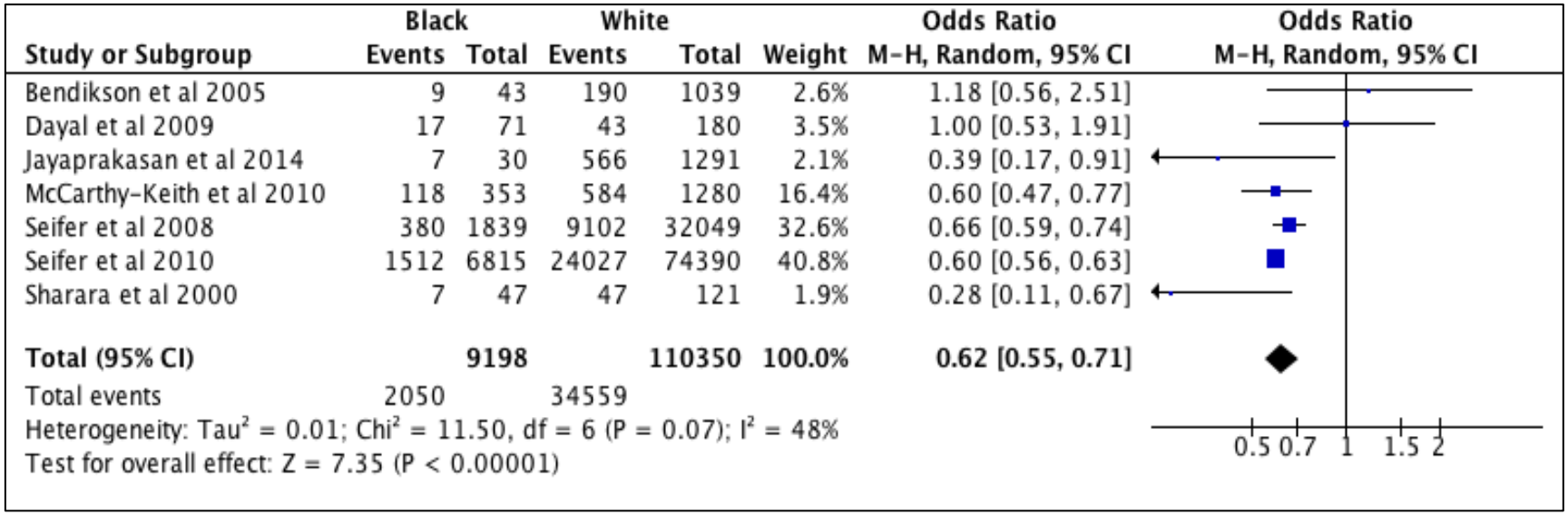


Figure 4. Black vs. White – Clinical pregnancy (fresh cycles)

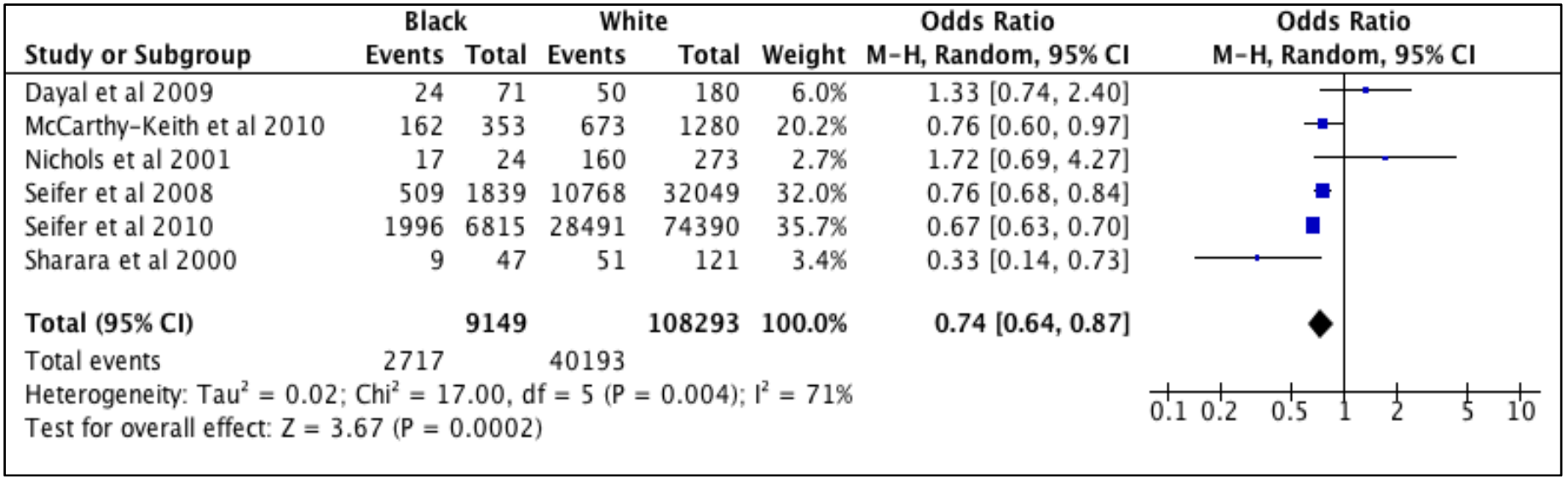


Figure 5. Black vs. White – Live birth (frozen cycles)

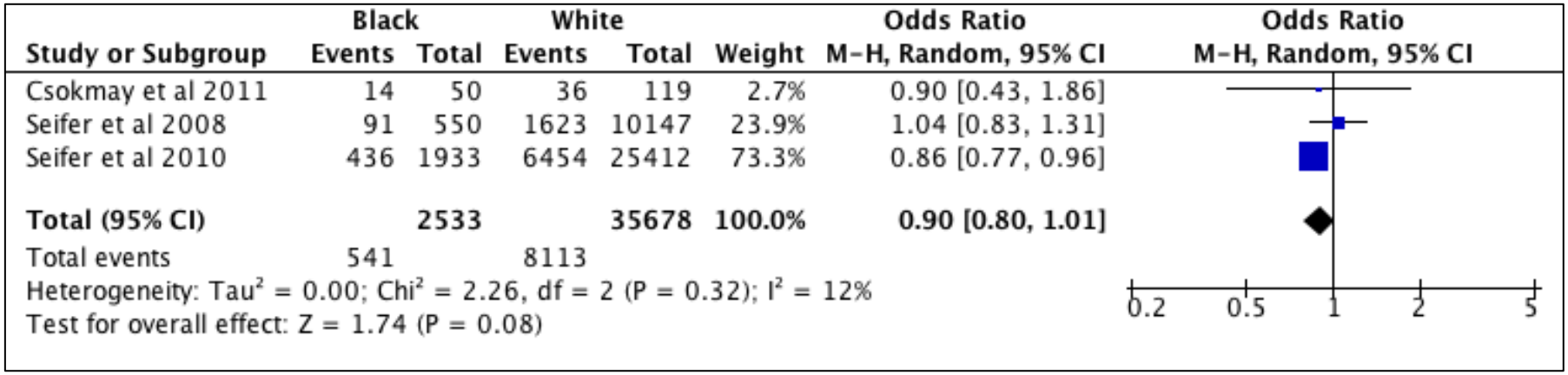
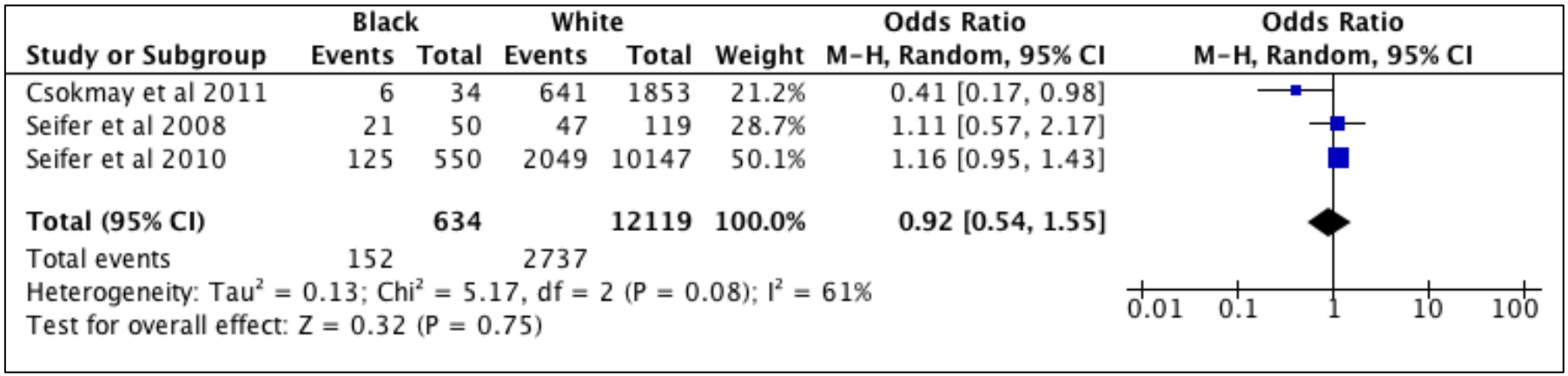


Figure 6. Black vs. White – Clinical pregnancy (frozen cycles)



Asian vs. White

There were 8 studies[3,6,24,28,31,40,42,44] comparing Asian ethnicity against a White reference group (Figures 7 and 8). These studies comprised of women both from South Asian and Chinese ethnic groups, the meta-analysis showed these women had a statistically significant reduction in both live birth (OR 0.67 [95% CI 0.64-0.69] p<0.001) and clinical pregnancy rate (OR 0.67 [95% CI 0.65-0.70] p<0.001) compared with White women. The data for live birth and clinical pregnancy showed very low heterogeneity, however this was not statistically significant.

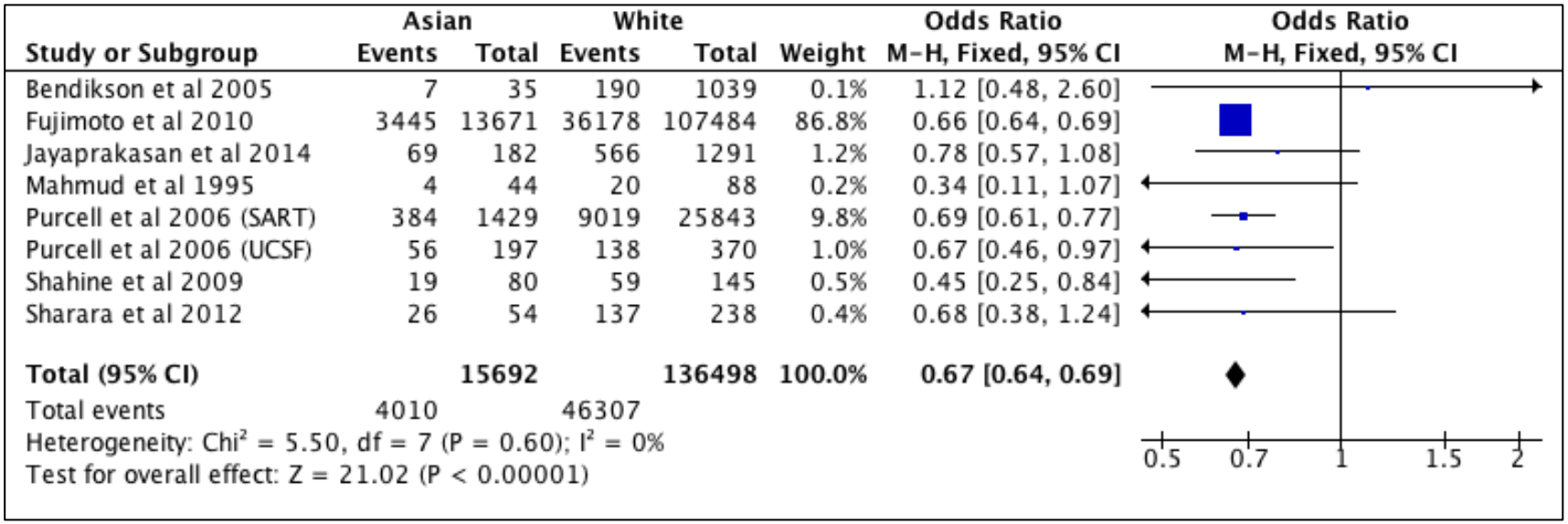
The results of the meta-analysis were dominated by one study by Fujimoto et al; which contributed to over 85% of the data analysed. When we removed this study from the analysis there was still a statistically significant lower odds of live birth and clinical pregnancy in Asian women than in White women (OR 0.69 [95% CI 0.62-0.76] p<0.001) and (OR 0.70 [95% CI 0.63-0.78] p<0.001) respectively.

Five papers[28,31,40,42,44] specified a cohort of South Asian or Indian women. These data were meta-analysed separately and showed a statistically significant reduction in live birth and clinical pregnancy: (OR 0.66 [95% CI 0.52-0.85] p=0.001) and (OR 0.65 [95% CI 0.47-0.90] p=0.008) respectively (Figures 9 and 10).

Three studies[3,6,28] calculated adjusted odds ratio for Asian women accounting for multiple confounders including age, BMI and diagnosis of PCOS; the results of these were pooled. Even after adjusting for potential confounders there remained a statistically significant reduction in live birth rate for Asian women compared with

White (aOR 0.70 [95% CI 0.55, 0.86]). Only one of these studies specified a cohort of South Asian women[28]; the adjusted odds ratio for this study (using age, PCOS and BMI as confounders) was (aOR 0.56 [95% CI 0.40-0.79]).

Figure 7. Asian vs. White – Live birth (fresh cycles)

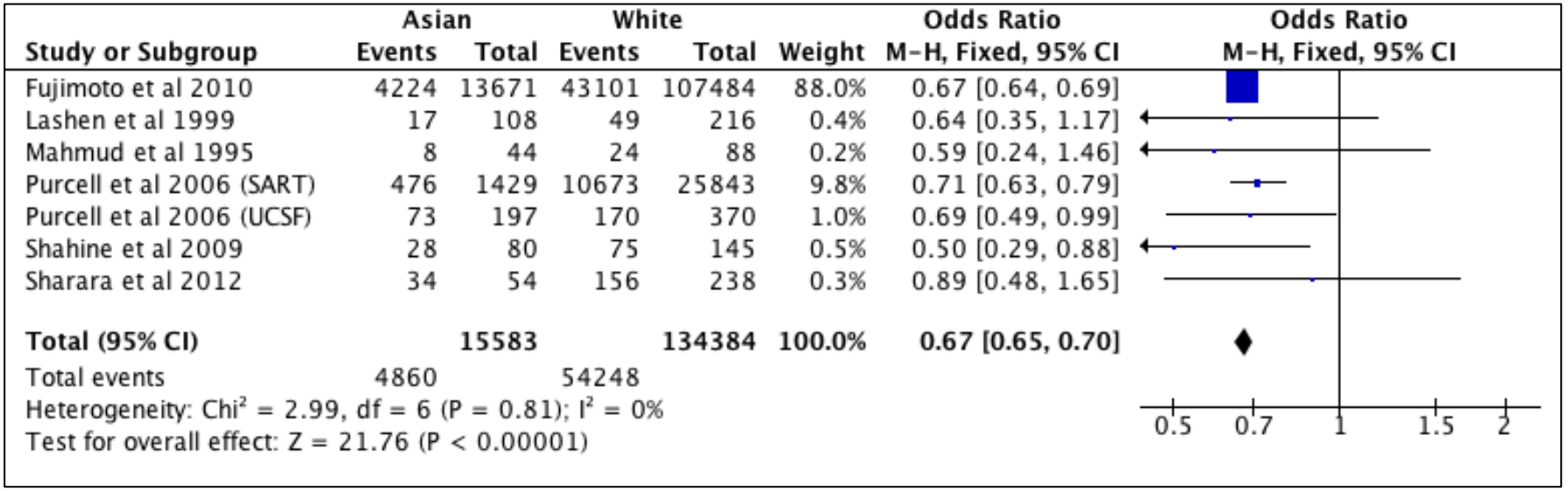


Study or Subgroup	Asian Events	Asian Total	White Events	White Total	Weight	Odds Ratio M-H, Fixed, 95% CI
Bendikson et al 2005	7	35	190	1039	0.1%	1.12 [0.48, 2.60]
Fujimoto et al 2010	3445	13671	36178	107484	86.8%	0.66 [0.64, 0.69]
Jayaprakasan et al 2014	69	182	566	1291	1.2%	0.78 [0.57, 1.08]
Mahmud et al 1995	4	44	20	88	0.2%	0.34 [0.11, 1.07]
Purcell et al 2006 (SART)	384	1429	9019	25843	9.8%	0.69 [0.61, 0.77]
Purcell et al 2006 (UCSF)	56	197	138	370	1.0%	0.67 [0.46, 0.97]
Shahine et al 2009	19	80	59	145	0.5%	0.45 [0.25, 0.84]
Sharara et al 2012	26	54	137	238	0.4%	0.68 [0.38, 1.24]
Total (95% CI)		**15692**		**136498**	**100.0%**	**0.67 [0.64, 0.69]**
Total events	4010		46307			

Heterogeneity: $Chi^2 = 5.50$, df = 7 (P = 0.60); $I^2 = 0\%$

Test for overall effect: Z = 21.02 (P < 0.00001)

Figure 8. Asian vs. White – Clinical pregnancy (fresh cycles)



Study or Subgroup	Asian Events	Asian Total	White Events	White Total	Weight	Odds Ratio M-H, Fixed, 95% CI
Fujimoto et al 2010	4224	13671	43101	107484	88.0%	0.67 [0.64, 0.69]
Lashen et al 1999	17	108	49	216	0.4%	0.64 [0.35, 1.17]
Mahmud et al 1995	8	44	24	88	0.2%	0.59 [0.24, 1.46]
Purcell et al 2006 (SART)	476	1429	10673	25843	9.8%	0.71 [0.63, 0.79]
Purcell et al 2006 (UCSF)	73	197	170	370	1.0%	0.69 [0.49, 0.99]
Shahine et al 2009	28	80	75	145	0.5%	0.50 [0.29, 0.88]
Sharara et al 2012	34	54	156	238	0.3%	0.89 [0.48, 1.65]
Total (95% CI)		**15583**		**134384**	**100.0%**	**0.67 [0.65, 0.70]**
Total events	4860		54248			

Heterogeneity: $Chi^2 = 2.99$, $df = 6$ ($P = 0.81$); $I^2 = 0\%$

Test for overall effect: $Z = 21.76$ ($P < 0.00001$)

Figure 9. South Asian vs. White – Live birth (fresh cycles)

Study or Subgroup	South Asian Events	South Asian Total	White Events	White Total	Weight	Odds Ratio M-H, Fixed, 95% CI
Jayaprakasan et al 2014	69	182	566	1291	55.2%	0.78 [0.57, 1.08]
Mahmud et al 1995	4	44	20	88	7.7%	0.34 [0.11, 1.07]
Shahine et al 2009	19	80	59	145	20.3%	0.45 [0.25, 0.84]
Sharara et al 2012	26	54	137	238	16.7%	0.68 [0.38, 1.24]
Total (95% CI)		**360**		**1762**	**100.0%**	**0.66 [0.52, 0.85]**
Total events	118		782			

Heterogeneity: $Chi^2 = 3.82$, df = 3 (P = 0.28); $I^2 = 21\%$

Test for overall effect: Z = 3.23 (P = 0.001)

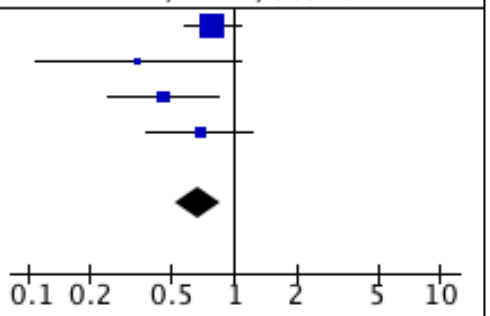
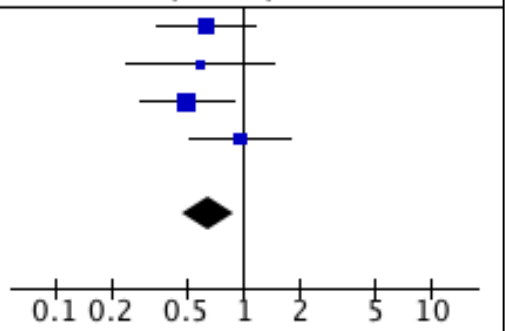


Figure 10. South Asian vs. White – Clinical pregnancy (fresh cycles)

Study or Subgroup	South Asian Events	South Asian Total	White Events	White Total	Weight	Odds Ratio M-H, Fixed, 95% CI
Lashen et al 1999	17	108	49	216	28.8%	0.64 [0.35, 1.17]
Mahmud et al 1995	8	44	24	88	13.7%	0.59 [0.24, 1.46]
Shahine et al 2009	28	80	75	145	36.3%	0.50 [0.29, 0.88]
Sharara et al 2012	35	54	156	238	21.2%	0.97 [0.52, 1.80]
Total (95% CI)		**286**		**687**	**100.0%**	**0.65 [0.47, 0.90]**
Total events	88		304			

Heterogeneity: $Chi^2 = 2.44$, df = 3 (P = 0.49); $I^2 = 0\%$

Test for overall effect: Z = 2.64 (P = 0.008)

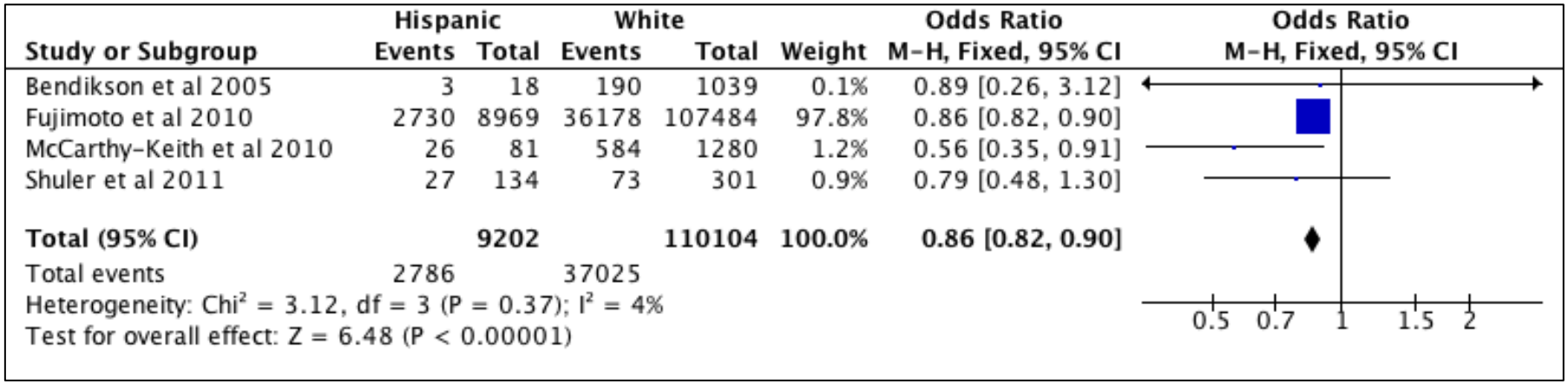


Hispanic vs. White

The findings for the Hispanic population were consistent with those for Black and Asian women, four studies[2,6,24,29] had data for live birth rate, while only three studies[2,6,29] also reported clinical pregnancy rate. The findings for the Hispanic population were consistent with those for Black and Asian women showing a statistically significant reduction in live birth and clinical pregnancy rate when compared to a White population: (OR 0.86 [95% CI 0.82-0.90] p<0.001) and (OR 0.89 [95% CI 0.85-0.93] p<0.001) respectively (Figures 11 and 12). Both the live birth and clinical pregnancy data for the Hispanic population showed very low heterogeneity although this was not statistically significant. Only one of the four papers[6] calculated an adjusted odds ratio for the live birth outcome. They adjusted for maternal age, number of embryos transferred and diagnosis of male factor, endometriosis, PCOS, diminished ovarian reserve, tubal factors, uterine factors and other factors. This result was consistent in showing that the Hispanic population have a lower live birth rate compared with White women (aOR 0.87 [95% CI 0.79, 0.96] P=0.005).

As seen in the Asian vs. White comparisons, one study (Fujimoto et al) dominated the analysis in the Hispanics vs. White comparisons by contributing around 98% of the data. When we removed this study from the analysis there was still a statistically significant reduced odds of live birth for Hispanic women compared with White women (OR 0.67 [95% CI 0.48-0.94] p=0.02), although this was not the case for clinical pregnancy (OR 0.79 [95% CI 0.57-1.10] p=0.2).

Figure 11. Hispanic vs. White – Live birth (fresh cycles)



Study or Subgroup	Hispanic Events	Hispanic Total	White Events	White Total	Weight	Odds Ratio M-H, Fixed, 95% CI
Bendikson et al 2005	3	18	190	1039	0.1%	0.89 [0.26, 3.12]
Fujimoto et al 2010	2730	8969	36178	107484	97.8%	0.86 [0.82, 0.90]
McCarthy-Keith et al 2010	26	81	584	1280	1.2%	0.56 [0.35, 0.91]
Shuler et al 2011	27	134	73	301	0.9%	0.79 [0.48, 1.30]
Total (95% CI)		**9202**		**110104**	**100.0%**	**0.86 [0.82, 0.90]**
Total events	2786		37025			

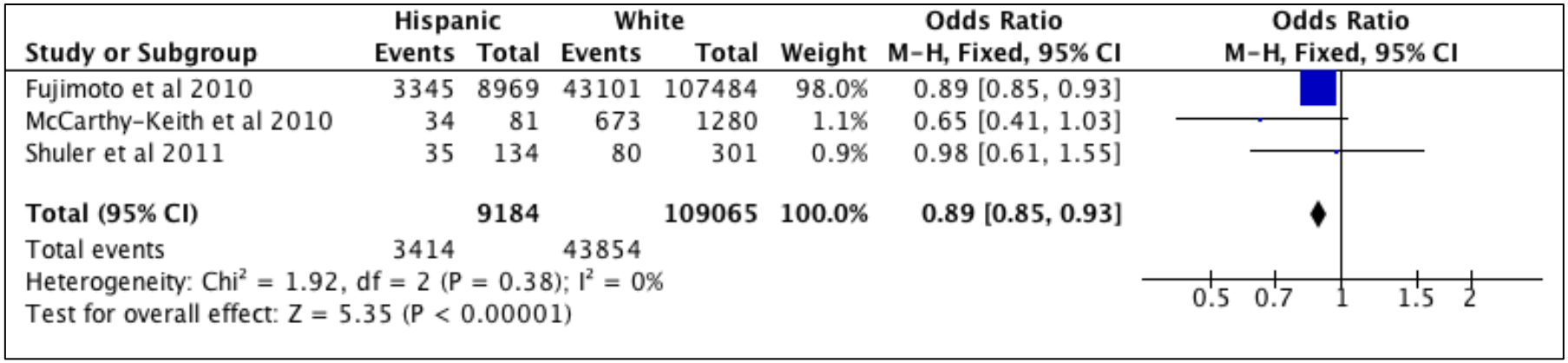
Heterogeneity: $Chi^2 = 3.12$, df = 3 (P = 0.37); $I^2 = 4\%$

Test for overall effect: Z = 6.48 (P < 0.00001)

Odds Ratio M-H, Fixed, 95% CI

0.5 0.7 1 1.5 2

Figure 12. Hispanic vs. White – Clinical pregnancy (fresh cycles)



Study or Subgroup	Hispanic Events	Hispanic Total	White Events	White Total	Weight	Odds Ratio M-H, Fixed, 95% CI
Fujimoto et al 2010	3345	8969	43101	107484	98.0%	0.89 [0.85, 0.93]
McCarthy-Keith et al 2010	34	81	673	1280	1.1%	0.65 [0.41, 1.03]
Shuler et al 2011	35	134	80	301	0.9%	0.98 [0.61, 1.55]
Total (95% CI)		**9184**		**109065**	**100.0%**	**0.89 [0.85, 0.93]**
Total events	3414		43854			

Heterogeneity: $Chi^2 = 1.92$, df = 2 (P = 0.38); $I^2 = 0\%$

Test for overall effect: Z = 5.35 (P < 0.00001)

Odds Ratio M-H, Fixed, 95% CI

0.5 0.7 1 1.5 2

Discussion

Main findings

Our systematic review suggests that significant ethnic disparities in IVF outcomes exist. The results consistently showed a statistically significant reduction in live birth and clinical pregnancy for fresh cycles across all races compared with White women. For live birth the Black population were found to have the poorest outcome followed by Asian and Hispanic, while for clinical pregnancy Asian women had the lowest success. The Hispanic population only showed a small reduction in live birth success compared with White women. Black women were found to have the poorest live birth rate for the fresh cycle data, however interestingly there was no statistical significant difference for the frozen data. This difference between fresh and frozen cycle outcome may suggest that treatment protocol has a role to play in the reduced success rate for fresh cycles. Further studies of tailored treatment protocols across different racial groups could potentially improve success rates.

There is an argument that the difference in IVF success rates could be influenced primarily by socioeconomic factors, such as lack of access to medical treatment leading to higher age at first encounter. As stated in the results only three papers found a statistical difference in mean age; Seifer et al[4] and Fujimoto et al[6] both found a statistically significantly higher proportion of White women under the age of 35 years compared with Black women, while Sharara et al[42] found South Asian women to have a statistically significant mean younger age compared with White women. Further to this, there could be a case for non-White women having a poorer diet and

lifestyle leading to reduced chance of pregnancy success, however only two of the seven studies which reported BMI found statistically significant differences between the racial groups, with the White population having a lower mean BMI compared with Black[23,41].

Strengths and limitations

One of the strengths of this study was the thorough methodological approach used. It met the quality criteria laid down in the MOOSE statement[45]. The analysis contained a good sample size with a total of 195,978 fresh cycles and 38,211 frozen cycles for overall analysis. However the papers were heterogeneous with both prospective and retrospective methodology. There was variation between the papers with regards to country of origin, differing categories of racial groups, year of publication and source of data. The country from which the data was produced was particularly important in this study as it determined which ethnic groups would be included and thus introducing a potential population bias.

One of the main limitations in interpreting the findings of this paper was the differing definitions used for each racial group. For the purpose of combining the data for analysis only four racial groups were defined; White, Black, Asian and Hispanic. The definition of the Asian group proved problematic. The papers using SART data reported patients as "Asian" with no specific definition, therefore incorporating both South Asians (e.g. Indian, Pakistani) and South-East Asians (e.g. Chinese). To truly examine differences in outcomes between South Asians and South-East Asians and

other specific racial groups a more comprehensive definition of race reporting within the SART database and other databases used worldwide needs to be implemented.

A further significant limitation was in the quality of the original data. Only five of the 15 studies provided data for adjusted odds ratios[3–6,28]. Being unable to account for significant confounders i.e. age and body mass index, invariably affects the reliability of the results. As mentioned in the results, two of the largest studies; Seifer et al[4] and Fujimoto et al[6] found that Black women had a statistically significant higher mean age amongst Black women compared to White women. Therefore, one could argue that higher maternal age is the reason for Black women having lower success rates following IVF. However both Seifer et al[4] and Fujimoto et al[6] accounted for age in their adjusted analyses and the results were consistent with that of the unadjusted analysis. Regarding diagnosis of infertility; across nine studies Black women were found to have higher rates of tubal factor and uterine factor infertility and in two studies Asian women were found to have higher rates of PCOS compared to White women. However, cause of infertility was adjusted for (in selected studies[3–6,28]) and the results remained unchanged from the unadjusted analysis.

Finally, the bulk of the results of the meta-analyses were dominated by three large US studies[4–6]. Although the large sample size of these studies adds credibility there are still serious limitations. The Seifer studies only compared outcomes for African American women and White women, and although Fujimoto et al included Asian women they inappropriately combined South Asian and South-East Asian women. A further limitation of these studies is that as they are based on the ethnic distribution of the US and so the findings are not transferrable to a UK population. For example,

the UK does not have a large proportion of Hispanic people and instead South Asians contribute a greater percentage of ethnic minorities.

Conclusions and future work

This meta-analysis provides sound supporting evidence for the hypothesis that there is an association between ethnic background and IVF success, although with some serious limitations. It prompts investigation into the mechanisms behind this to subsequently allow modification of clinical practice to account for ethnicity. Given the limitation in accounting for confounders within this review and the lack of specificity with the ethnic groups, a large cohort study where there are clear definitions of ethnicity and a comprehensive range of variables available for analysis, would help to explore in closer detail the racial differences in IVF success and thus help to begin to investigate ways to minimise this. Furthermore, given that the existing studies, and in particular the largest studies contributing the majority of data to the meta-analyses, are all based on women in the US; there is a need for a large UK based study to determine the relationship between ethnicity and IVF outcome specific to the UK population.

CHAPTER 3

Investigating the effect of ethnicity on IVF outcome: an analysis of 13,473 cycles and an updated meta-analysis

The work in this chapter was published in Reproductive Biomedicine Online; published online 03.06.15

Introduction

Although ethnicity is frequently investigated as a prognostic factor in medicine, as shown in chapter 2, few studies have been able to clearly explore the association between ethnicity and in-vitro fertilisation (IVF) outcomes while accounting for the key confounders. Ethnic minorities contribute to around 13% of the UK population[46]. However, to our knowledge there is no published data on what percentage of couples seeking IVF treatment in the UK comes from ethnic minorities. Nonetheless, it is important for couples undergoing assisted conception to be counselled appropriately and according to their individual backgrounds.

The existing literature on ethnicity and IVF outcomes, as discussed in the systematic review, consists largely of American studies that have used the Society of American Reproductive Technologies (SART) database[4,5]. Such studies have not been able to adjust their findings to key confounders; furthermore, the ethnic mix of the American population is quite different from that of the UK. Therefore, the findings of these studies may not be transferrable, thus prompting the need for a large UK study. In the UK, there have been three studies[31,40,44] exploring the association between ethnicity and in-vitro fertilisation outcome. Two of these were conducted over 10 years ago[40,44], so there is a question about their applicability to today's population given the rapid advances in IVF over the years. The most recent publication[31] in 2014 was limited by its sample size (n=1517) and consequently had to group all ethnic groups into one group in order to perform any meaningful analysis. This, however, significantly limits the findings; as not all ethnic groups will behave the same.

Aims and objectives

1. To investigate the effect of ethnicity on live birth, in a large population (representative of UK ethnicities), following IVF treatment, while adjusting for known important confounders.
2. To investigate the effect of ethnicity on clinical pregnancy, in a large population (representative of UK ethnicities) following IVF treatment, while adjusting for known important confounders.
3. To provide an updated meta-analysis of studies investigating the effect of ethnicity on IVF outcome incorporating the findings of our cohort study.

Methods

Study Design

This observational cohort study included all women undergoing their first non-donor cycle of IVF or Intra-cytoplasmic Sperm Injection (ICSI) at any Centres for Assisted Reproduction (CARE) clinic in the UK and Ireland, from 2008 to 2012. CARE is one of the UK's largest independent providers of fertility services and provides treatment for both NHS and non-NHS patients. Permission to utilise an anonymised dataset of routinely collected electronic data was granted by the Institution Review Board of CARE.

Data were analysed from 12 fertility clinics within the CARE consortium; Nottingham, Manchester, Northampton, Sheffield, Dublin, Bolton, Boston, Derby, Leicester, Mansfield, Milton Keynes and Peterborough. Both fresh and frozen assisted

conception cycle data were included. The primary outcome for the study was live birth (defined as the birth of one of more living infants), and the secondary outcome was clinical pregnancy (defined as the presence of a gestational sac on ultrasound).

All women undergoing treatment at CARE were required to complete their demographic profile. This includes ethnicity definitions that were in line with that of the Human Fertilisation and Embryology (HFEA) coding. There were 17 individual ethnic groups, which we grouped into seven main categories: White (White British, White Irish, any other White), South Asian (Indian, Pakistani, Bangladeshi, any other Asian background), Black (Black Caribbean, Black African, other Black), Chinese, Mixed (White and Black Caribbean, White and Black African, White and Asian, any other mixed), any other and not stated.

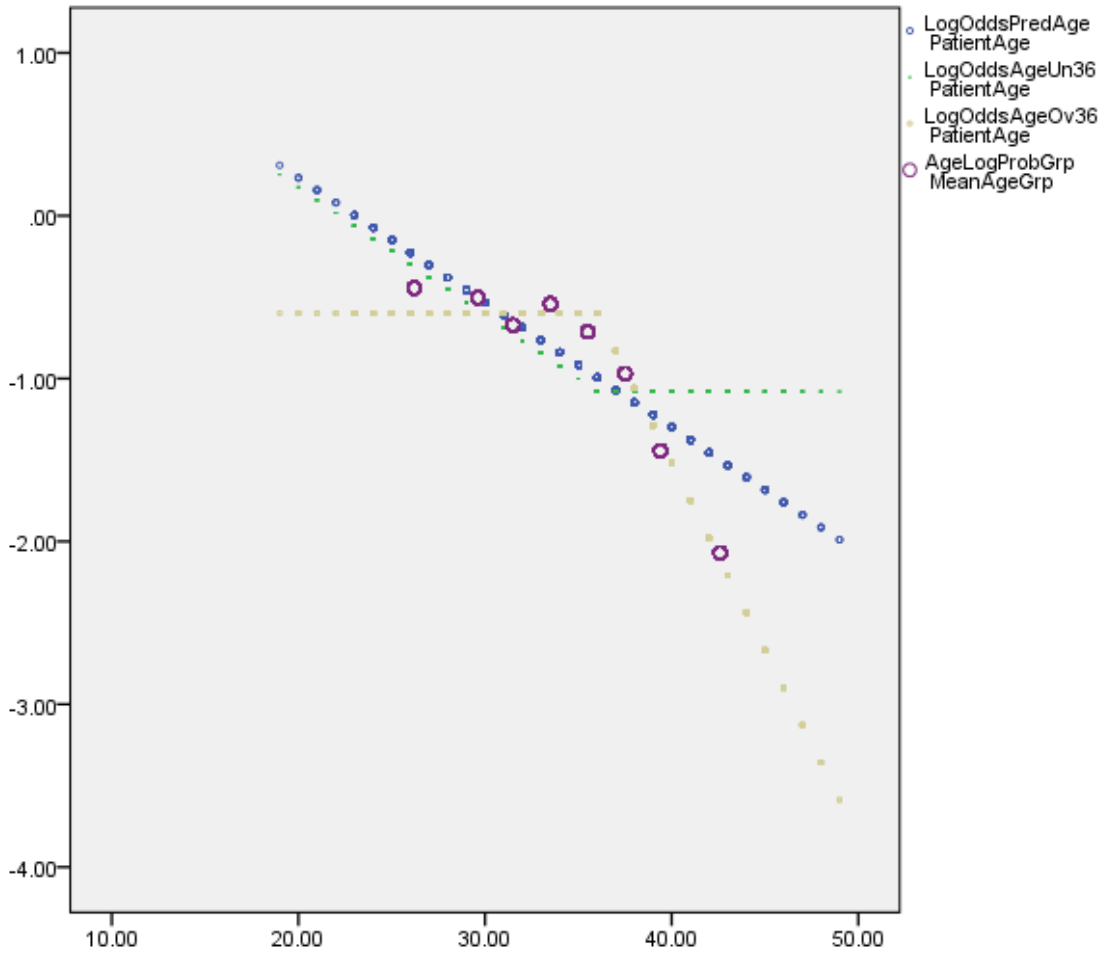
Statistical analysis

The baseline patient characteristics, cycle characteristics and outcome data were described giving frequencies with percentages, or means with standard deviations, as appropriate. To estimate the contribution of ethnicity to live birth rate and clinical pregnancy, univariate and multiple logistic regression analyses were performed to calculate odds ratios and corresponding 95% confidence intervals along with p values. Covariates were preselected when they had a known effect on IVF outcome, based on clinical knowledge and experience. The covariates selected for the multivariate model were: age, body mass index, duration of infertility, cause of infertility, history of previous live birth, history of previous miscarriage and number of embryos transferred. Ideally a measure of ovarian reserve (i.e. Day 2 Follicle

Stimulating Hormone (FSH), Anti-Mullerian Hormone (AMH) or Antral Follicle Count (AFC)) would have been included, however these variables were not well recorded in the database and so were removed from analysis.

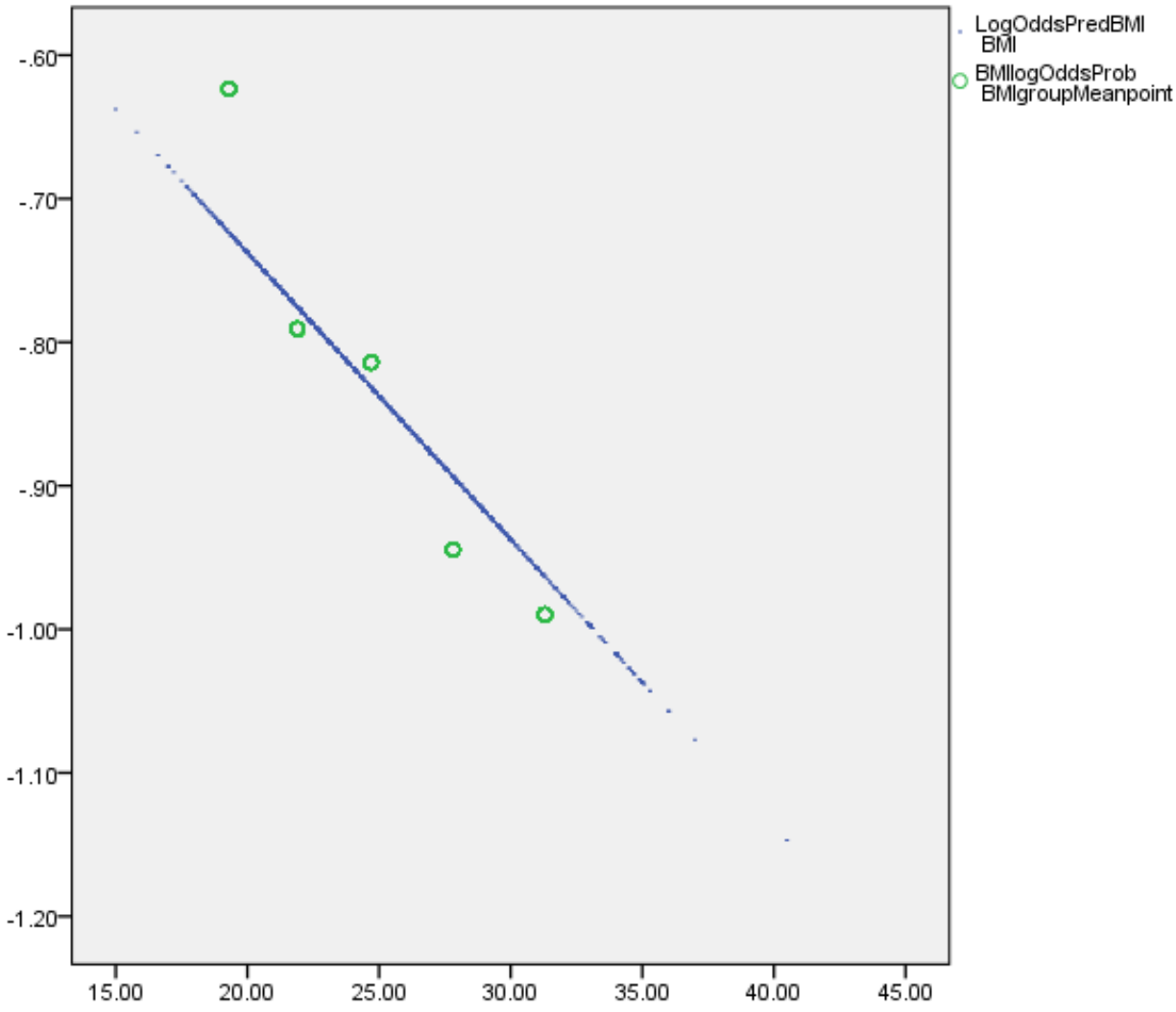
For the continuous variables: age, BMI and duration of infertility, they were assessed for their functional form using plots of the observed log odds for live birth. BMI appeared to have a linear relationship with live birth, however in the case of age, and duration of infertility there was a non-linear relationship. The results for age showed that below 36 years of age the chances of live birth appeared fairly constant, but above 36 there was a sharp linear decline, resulting in two linear variables being created for age. Whereas for duration of infertility from 0-4 years the log odds of live birth appeared static. From 5 years onwards there was a sharp decline, and so it was decided to dichotomise duration of infertility into 0-4 years and >5 years. The graphical representation for each continuous variable and the log odds of live birth is shown in Figures 13 to 15.

Figure 13. Age plotted against the log odds of the probability of live birth.



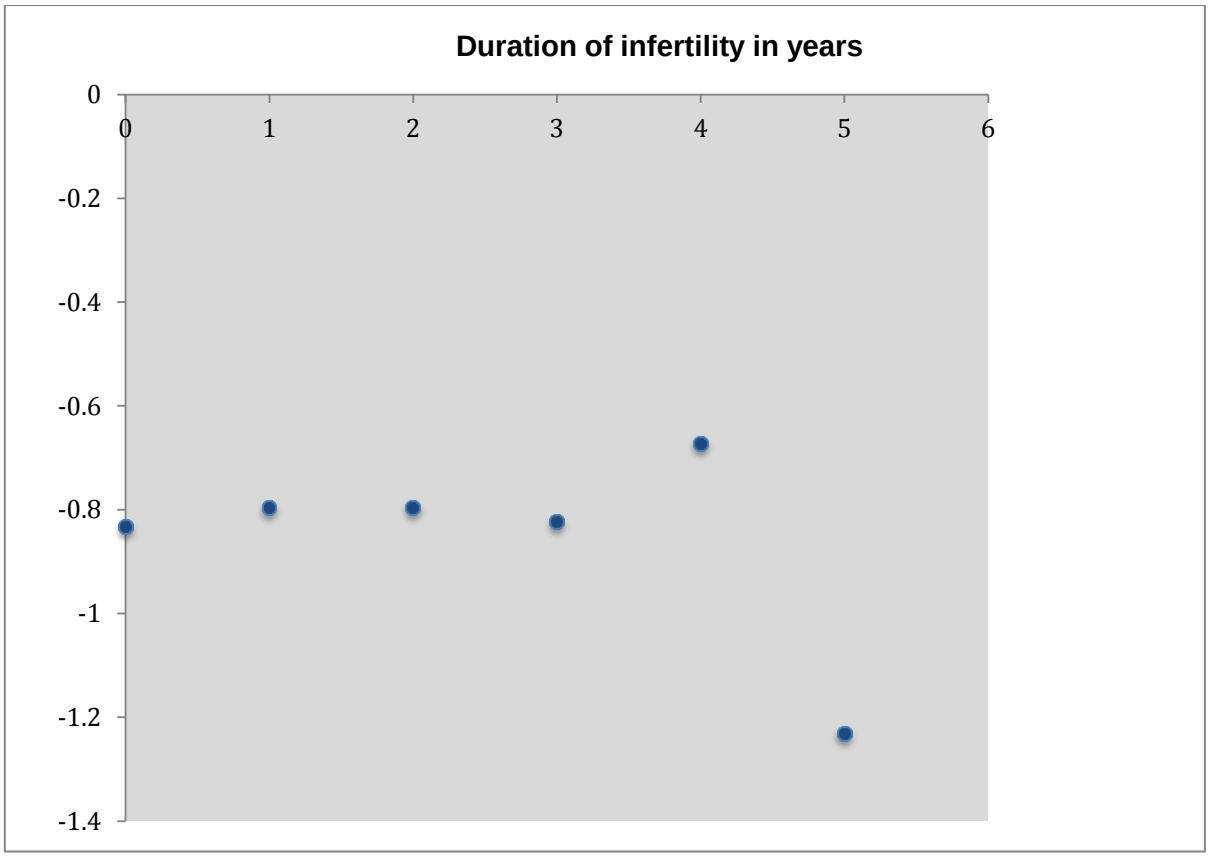
The results for age show that below 36 years of age the chances of live birth appear fairly constant, but above 36 there is a sharp linear decline, this resulted in two linear variables being created for age. The first was a continuous variable for age up until 36 years. Anyone older than 36 years had the value 36 for this variable. A second age variable equalled zero for all ages ≤36, and equalled age minus 36 for all ages >36 years.

Figure 14. BMI plotted against the log odds of the probability of live birth



BMI shows a linear relationship and so was included in the model as a continuous variable.

Figure 15. Duration of infertility plotted against the log odds of the probability of live birth



For duration of infertility from 0-4 years the log odds of live birth appear static. From 5 years onwards there was a sharp decline, and so it was decided to dichotomise duration into 0-4 years and >5 years.

Missing data

All variables included in the regression analysis had 100% of data entry with the exception of BMI. This was only reported in 48% of cases. A multiple imputation

procedure was conducted using an iterative Markov chain Monte Carlo (MCMC) method to impute for the missing BMI data. All predictors and the outcome of live birth were included in the imputation process to maximise the precision of the imputations. All univariable models and the multivariable model were fitted to the 20 imputed datasets arising from the multiple imputation procedure. The parameter estimates and covariance's arising from the models from each imputed dataset were combined to produce inferential results.

A sensitivity analysis was performed analysing the data for frozen cycles separately and breaking down causes of infertility to specifically include fibroids.

Finally, the results of unadjusted and adjusted odds ratios from our cohort study were combined, where appropriate, with the data from the existing studies to provide an updated meta-analysis.

Data were analysed using SPSS (ver. 21.0; SPSS Inc., Chicago) and Review Manager (version 5.0 for Windows).

Results

Overall description of data

A total of 13473 cycles were reported between 2008 and 2012 at the 12 CARE clinics in the UK. The ethnic groupings were as follows; White (10062), Black (212), South Asian (1025), Chinese (83), Mixed (476), Other (148) and Not stated (1467). The

total percentage of ethnic minorities who underwent IVF treatment at the CARE clinics (including “Other”) was 14.4%.

Tables 5-7 display an overall description of the results including baseline patient characteristics, cycle characteristics and cycle outcomes. The number of cycles that had data for each variable is specified within the tables. It is noted that Black women had worse risk factors: they were on average older, with higher BMI, fewer previous live births, more previous miscarriages and a longer duration of infertility than White women; whereas Asian women were on average younger, with lower BMI, greater rates of anovulation, lower rates of previous miscarriage but longer duration of infertility than White women. The group with unstated ethnic group had the highest rates of previous live births, lowest rates of previous miscarriage but the longest duration of infertility.

Table 5. Baseline characteristics for each ethnic group

	White (n=10062)	Black (n=212)	South Asian (n=1025)	Chinese (n=83)	Mixed (n=476)	Other (n=148)	Not stated (n=1467)
Age (in years)	**(n=10062)**	**(n=212)**	**(n=1025)**	**(n=83)**	**(n=476)**	**(n=148)**	**(n=1467)**
<35	5577 (55.4%)	103 (48.6%)	731 (71.3%)	49 (59%)	281 (59.0%)	72 (48.6%)	757 (51.6%)
35.1-40	3166 (31.5%)	59 (27.8%)	223 (21.8%)	25 (30.1%)	133 (27.9%)	61 (41.2%)	459 (31.3%)
40.1-45	1112 (11.1%)	39 (18.4%)	65 (6.3%)	9 (10.8%)	53 (11.1%)	15 (10.1)	188 (12.8%)
>45.1	207 (2.1%)	11 (5.2%)	6 (0.6%)	0	9 (1.9%)	0	63 (4.3%)
Body mass index	**(n=5278)**	**(n=116)**	**(n=527)**	**(n=45)**	**(n= 290)**	**(n=86)**	**(n=132)**
<18.5	89 (1.7%)	3 (2.6%)	15 (2.8%)	2 (4.4%)	16 (5.5%)	0	0
18.6-25	3100 (58.7%)	35 (30.2%)	293 (55.6%)	40 (88.9%)	160 (55.2%)	58 (67.4%)	85 (64.4%)
25.1-30	1625 (30.8%)	48 (41.1%)	178 (33.8%)	2 (4.4%)	81 (27.9%)	25 (29.1%)	32 (24.2%)
30.1-35	421 (8.0%)	28 (24.1%)	33 (6.3%)	0	30 (10.3%)	3 (3.5%)	12 (9.1%)
>35.1	43 (0.8%)	2 (1.7%)	8 (1.5%)	1 (2.2%)	3 (1.0%)	0	3 (2.3%)
Cause of infertility*	**(n=10062)**	**(n=212)**	**(n=1025)**	**(n=83)**	**(n=476)**	**(n=148)**	**(n=1467)**
Male factor	5896 (58.6%)	109 (51.4%)	589 (57.5%)	54 (65.1%)	296 (62.2%)	95 (64.2%)	548 (37.4%)
Tubal factor	1554 (15.4%)	36 (17.0%)	123 (12.0%)	22 (26.5%)	68 (14.3%)	29 (19.6%)	226 (15.4%)
Anovulation	1156 (11.5%)	17 (8.0%)	197 (19.2%)	7 (8.4%)	58 (12.2%)	17 (11.5%)	200 (13.6%)
Female other (e.g. Endometriosis)	3014 (30.0%)	91 (42.9%)	230 (22.4%)	14 (16.9%)	146 (30.7%)	45 (30.4%)	319 (21.7%)
Unexplained	2948 (29.3%)	60 (28.3%)	343 (33.5%)	23 (27.7%)	130 (27.3%)	34 (23.0%)	437 (29.8%)

*Not mutually excl

Prev. live birth	**(n=10062)** 1907 (19.0%)	**(n=212)** 29 (13.7%)	**(n=1025)** 190 (18.5%)	**(n=83)** 11 (13.3%)	**(n=476)** 94 (19.7%)	**(n=148)** 21 (14.2%)	**(n=1467)** 349 (23.8%)
Prev. miscarriage	**(n=10062)** 2047 (20.3%)	**(n=212)** 61 (28.8%)	**(n=1025)** 163 (15.9%)	**(n=83)** 9 (10.8%)	**(n=476)** 98 (20.6%)	**(n=148)** 28 (18.9%)	**(n=1467)** 98 (6.7%)
Duration of infertility in years (Mean ±SD)	**(n=10062)** 2.71 ±2.1	**(n=212)** 3.5 ±2.8	**(n=1025)** 3.4 ±2.7	**(n=83)** 3.3 ±2.8	**(n=476)** 2.6 ±2.3	**(n=148)** 3.1 ±2.5	**(n=1467)** 4.4 ±3.2
Day 2 FSH (Mean ±SD)	**(n=3214)** 8.13 ±21.9	**(n=66)** 7.9 ±3.8	**(n=343)** 7.3 ±6.4	**(n=27)** 5.7 ±2.1	**(n=215)** 6.8 ±2.5	**(n=60)** 6.6 ±2.2	**(n=64)** 6.6 ±1.9
AMH level (Mean ±SD)	**(n=1289)** 16.98 ±18.2	**(n=13)** 20.5 ±27.7	**(n=107)** 24.5 ±33.5	**(n=8)** 25.0 ±34.9	**(n=44)** 9.3 ±11.3	**(n=15)** 13.6 ±9.9	**(n=17)** 26.7 ±24.9
Antral follicle count (Mean ±SD)	**(n=3987)** 20.7 ±12.5	**(n=91)** 18.4 ±13.5	**(n=359)** 20.3 ±14.7	(**n=24)** 15.5 ±7.4	**(n=199)** 19.3 ±12.8	**(n=69)** 18.1 ±13.5	**(n=42)** 27.6 ±16.3

Table 6. Cycle characteristics for each ethnic group

	White (n=10062)	Black (n=212)	South Asian (n=1025)	Chinese (n=83)	Mixed (n=476)	Other (n=148)	Not stated (n=1467)
Treatment:	**(n=10062)**	**(n=212)**	**(n=1025)**	**(n=83)**	**(n=476)**	**(n=148)**	**(n=1467)**
IVF	2704 (26.9%)	60 (28.3%)	252 (24.6%)	26 (31.3%)	96 (20.2%)	38 (25.7%)	359 (24.5%)
ICSI	5010 (49.8%)	106 (50.0%)	556 (54.2%)	30 (36.1%)	270 (56.7%)	81 (54.7%)	598 (40.8%)
FET	1853 (18.4%)	34 (16.0%)	183 (17.9%)	20 (24.1%)	99 (20.8%)	25 (16.9%)	428 (29.2%)
Not recorded	495 (4.9%)	12 (5.7%)	34 (3.3%)	7 (8.5%)	11 (2.3%)	4 (2.7%)	82 (5.5%)
No. of oocytes retrieved (mean ±SD)	**(n=10062)** 7.4 ±6.3	**(n=212)** 8.1 ±9.4	**(n=1025)** 8.1 ±6.8	**(n=83)** 6.9 ±6.8	**(n=476)** 7.8 ±6.5	**(n=148)** 7.9 ±5.9	**(n=1467)** 6.0 ±6.2
No. of mature oocytes (mean ±SD)	**(n=10062)** 5.7 ±5.1	**(n=212)** 5.9 ±7.8	**(n=1025)** 6.2 ±5.5	**(n=83)** 5.4 ±5.6	**(n=476)** 5.9 ±5.2	**(n=148)** 6.1 ±4.9	**(n=1467)** 4.7 ±5.0
No. inseminated (mean ±SD)	**(n=10062)** 6.2 ± 5.5	**(n=212)** 6.4 ±8.3	**(n=1025)** 6.7 ±5.8	**(n=83)** 5.9 ±5.9	**(n=476)** 6.2 ±5.5	**(n=148)** 6.6 ±5.1	**(n=1467)** 5.1 ±5.4
2 Pronuclei	**(n=10062)** 4.01 ± 3.8	**(n=212)** 4.2 ±6.3	**(n=1025)** 4.2 ±3.9	**(n=83)** 3.6 ±3.8	**(n=476)** 4.1 ±4.0	**(n=148)** 4.2 ±3.8	**(n=1467)** 3.4 ±3.7
3 Pronuclei	**(n=10062)** 0.2 ±0.5	**(n=212)** 0.3 ±0.8	**(n=1025)** 0.2 ±0.5	**(n=83)** 0.3 ±0.7	**(n=476)** 0.2 ±0.6	**(n=148)** 0.2 ±0.5	**(n=1467)** 0.2 ±0.6
Total no. of embryos	**(n=10062)** 4.9 ±3.9	**(n=212)** 5.4 ±6.6	**(n=1025)** 5.3 ±4.1	**(n=83)** 4.9 ±3.9	**(n=476)** 5.1 ±4.0	**(n=148)** 5.1 ±3.7	**(n=1467)** 4.5 ±3.7

Fertilisation rate* (mean ±SD)	**(n=7522)** 0.73 ±0.24	**(n=157)** 0.73 ±0.23	**(n=784)** 0.71 ±0.24	**(n=56)** 0.69 ±0.24	**(n=357)** 0.72 ±0.26	**(n=114)** 0.71 ±0.25	**(n=933)** 0.74 ±0.24
No. of embryos transferred	**(n=10062)**	**(n=212)**	**(n=1025)**	**(n=83)**	**(n=476)**	**(n=148)**	**(n=1467)**
0	1395 (13.9%)	48 (22.6%)	128 (12.5%)	12 (14.5%)	60 (12.6%)	20 (13.5%)	183 (12.5%)
1	3157 (31.4%)	55 (25.9%)	302 (29.5%)	25 (30.1%)	160 (33.6%)	46 (31.1%)	222 (15.1%)
2	5250 (52.2%)	102 (48.1%)	580 (56.6%)	46 (55.4%)	242 (50.8%)	81 (54.7%)	1021 (69.6%)
3	260 (2.6%))	7 (3.3%)	15 (1.5%)	0	14 (2.9%)	1 (0.7%)	41 (2.8%)
No. of embryos frozen (Mean±SD)	1.1 ± 2.5	1.9 ±6.1	1.2 ±2.5	0.9 ±2.6	1.1 ±2.4	1.2 ±2.2	0.8 ±2.2

*Fertilisation rate is the number of embryos over the total number of oocytes retrieved

Table 7. Outcome data for each ethnic group

	White **(n=10062)**	Black **(n=212)**	South Asian **(n=1025)**	Chinese **(n=83)**	Mixed **(n=476)**	Other **(n=148)**	Not stated **(n=1467)**
Implantation rate* (Mean ±SD)	**(n=8667)** 0.38 ±0.46	**(n=164)** 0.24 ±0.39	**(n=897)** 0.38 ±0.46	**(n=71)** 0.35 ±0.53	**(n=416)** 0.33 ±0.42	**(n=128)** 0.30 ±0.41	**(n=1284)** 0.36 ±0.44
Biochemical pregnancy rate	4634 (46.1%)	57 (26.9%)	477 (46.5%)	33 (39.8%)	215 (45.2%)	54 (36.5%)	676 (46.1%)
Clinical pregnancy rate**	3970 (39.5%)	48 (22.6%)	409 (39.9%)	27 (32.5%)	175 (36.8%)	48 (32.4%)	591 (40.3%)
Live birth rate[a]	3492 (34.7%)	42 (19.8%)	341 (33.3%)	26 (31.3%)	149 (31.3%)	42 (28.4%)	530 (36.1%)
Other pregnancy outcomes							
Miscarriage[b]	379 (9.5%)	6 (12.5%)	45 (11.0%)	1 (3.7%)	18 (10.2%)	3 (6.3%)	49 (8.3%)
Termination[b]	20 (0.5%)	0	3 (0.7%)	0	1 (0.6%)	0	3 (0.5%)
Still birth[b]	15 (0.4%)	0	4 (1.0%)	0	1 (0.6%)	0	4 (0.7%)
NND[b]	24 (0.6%)	0	2 (0.5%)	0	1 (0.6%)	0	4 (0.7%)

*Defined as the number of foetal hearts divided by the number of embryos transferred, per cycle
** Defined as the presence of a gestational sac by ultrasound during first trimester
[a] Expressed as a percentage of all cycles [b] Expressed as a percentage of clinical pregnancies

Univariate and multivariate analyses for live birth for all cycles

The live birth rate was statistically significantly lower in Black women than White women (19.8% vs. 34.7% p<0.001). The rates in South Asian women and White women were similar (33.3% vs. 34.7% p=0.4). The difference between Black and White women remained statistically significant when differences in age, BMI, cause and duration of infertility, previous live birth, previous miscarriage and number of embryos transferred were adjusted for; (OR 0.53 [95% CI 0.37-0.76] p<0.001). Adjustment for differences, using the same variables, showed that the adjusted odds of having a live birth rate for South Asian women was significantly lower than for White women (OR 0.79 (0.68-0.91) p<0.001). The univariate analysis and multivariate logistic regression model for live birth is shown in Tables 8 and 9 respectively.

Table 8. Univariate analysis for live birth (all cycles)

			95% CI		
Ethnic Group	**No. of cycles**	**Odds ratio**	**Lower**	**Upper**	**P value**
White	10062	Reference	---	---	---
South Asian	1025	0.94	0.82	1.08	0.4
Black	212	0.47	0.33	0.65	<0.001
Chinese	83	0.86	0.54	1.40	0.5
Mixed	476	0.86	0.70	1.05	0.1
Other	148	0.75	0.52	1.07	0.1
Not stated	1467	1.07	0.95	1.19	0.3

Table 9. Multivariate logistic regression model for live birth (n=13473)

	Parameter Estimate	S.E.	P value	Odds Ratio	95% CI	
					Lower	Upper
Age:						
≤36 years (per increasing year)	-0.056993	0.006	<0.001	0.95	0.93	0.96
>36 years (per increasing year)	-0.114606	0.010	<0.001	0.89	0.87	0.91
Body mass index	-0.012935	0.007	0.08	0.99	0.97	1.00
Cause of infertility:						
Male factor	-0.077805	0.048	0.1	0.93	0.84	1.02
Tubal factor	-0.298338	0.059	<0.001	0.74	0.66	0.83
Anovulation	-0.023018	0.063	0.7	0.98	0.86	1.11
Unexplained	-0.153625	0.058	0.008	0.86	0.77	0.96
Other (e.g. Endo)	-0.112451	0.051	0.03	0.89	0.81	0.99
Ethnicity:						
White	0			Reference		
South Asian	-0.240432	0.074	<0.001	0.79	0.68	0.91
Black	-0.640090	0.184	<0.001	0.53	0.37	0.76
Chinese	-0.205110	0.253	0.4	0.82	0.50	1.34
Other	-0.289247	0.193	0.1	0.75	0.51	1.09
Not stated	-0.062196	0.065	0.4	0.94	0.83	1.07
Mixed	-0.220246	0.107	0.04	0.80	0.65	0.99

Table 9. continued

	Parameter Estimate	S.E.	P value	Odds Ratio	95% CI	
					Lower	Upper
Previous Live Birth						
No	0			Reference		
Yes	0.068248	0.051	0.2	1.07	0.97	1.18
Previous Miscarriage						
No	0			Reference		
Yes	0.016775	0.052	0.8	1.02	0.92	1.13
Duration of infertility:						
0-4 years	0			Reference		
≥5 years	-0.167231	0.056	0.003	0.85	0.76	0.94
No. of embryos transferred	1.000064	0.031	<0.001	2.72	2.56	2.89
Constant	0.372027	0.267	0.2	1.45	0.86	2.45

Univariate and multivariate analyses for clinical pregnancy for all cycles

The unadjusted results for clinical pregnancy for Black women compared with White women were similar to that of live birth: 22.6% and 39.5% respectively (p<0.001); the odds ratios for univariate analysis of ethnic group and live birth is shown in Table 10. This difference was maintained after accounting for known confounders (aOR 0.50 [95% CI 0.35-0.71] p<0.001), as shown in the logistic regression model in Table 11. The crude rates for implantation rate were also much lower for Black women compared with White women: 0.24 vs. 0.38.

Unlike for live birth, South Asian women had similar clinical pregnancy rates to White women (39.9% vs. 39.5% clinical pregnancy and 0.38 vs 0.38 for implantation rates). After adjustment in multivariate analyses for differences in confounding variables, there remained no significant difference in clinical pregnancy rates between South Asian women vs. White women (aOR = 0.86 [95% CI 0.75-1.00) p=0.05).

Table 10. Univariate analysis for clinical pregnancy (all cycles)

			95% CI		
Ethnic Group	**No. of cycles**	**Odds ratio**	**Lower**	**Upper**	**P value**
White	10062	Reference	---	---	---
South Asian	1025	1.02	0.89	1.16	0.8
Black	212	0.45	0.33	0.62	<0.001
Chinese	83	0.74	0.47	1.17	0.2
Mixed	476	0.89	0.74	1.08	0.2
Other	148	0.74	0.52	1.04	0.08
Not stated	1467	1.04	0.93	1.16	0.5

Table 11. Multivariate logistic regression model for clinical pregnancy (n=13473)

	Parameter Estimate	S.E.	P value	Odds Ratio	95% CI	
					Lower	Upper
Age:						
≤36 years (per increasing year)	-0.050964	0.006	<0.001	0.95	0.94	0.96
>36 years (per increasing year)	-0.114273	0.010	<0.001	0.89	0.88	0.91
Body mass index	-0.019096	0.007	0.1	0.99	0.98	1.00
Cause of infertility:						
Male factor	-0.074553	0.047	0.1	0.93	0.85	1.01
Tubal factor	-0.246090	0.058	<0.001	0.78	0.70	0.88
Anovulation	0.075815	0.062	0.2	1.08	0.96	1.22
Unexplained	-0.119439	0.057	0.04	0.89	0.79	0.99
Other (e.g. Endo)	-0.086382	0.050	0.09	0.92	0.83	1.01
Ethnicity:						
White	0			Reference		
South Asian	-0.150521	0.072	0.05	0.86	0.75	1.00
Black	-0.690810	0.177	<0.001	0.50	0.35	0.71
Chinese	-0.363738	0.252	0.1	0.70	0.42	1.14
Other	-0.311138	0.188	0.1	0.84	0.69	1.03
Not stated	-0.113980	0.064	0.08	0.89	0.79	1.01
Mixed	-0.173671	0.104	0.1	0.84	0.69	1.03

Table 11. continued	Parameter Estimate	S.E.	P value	Odds Ratio	95% CI	
					Lower	Upper
Previous Live Birth						
No	0			Reference		
Yes	0.075909	0.050	0.1	1.08	0.98	1.19
Previous Miscarriage						
No	0			Reference		
Yes	0.016938	0.051	0.7	1.02	0.92	1.12
Duration of infertility:						
0-4 years	0			Reference		
≥5 years	-0.161055	0.055	0.003	0.85	0.77	0.95
No. of embryos transferred	1.085808	0.030	<0.001	2.96	2.79	3.14
Constant	0.176627	0.261	0.5	1.19	0.71	2.00

Sensitivity analysis accounting for fibroids

The causes of infertility were grouped into tubal, ovulatory, male, unexplained and other. A sensitivity analysis was performed to specifically look at whether fibroids could explain the effects on live birth outcome in the Black population. Fibroids were included in the heterogenous group termed 'other' that included endometriosis and structural abnormalities. We created a separate variable for fibroids alone, adding this to the model including all the other covariates; this actually increased the magnitude of the effect of Black ethnicity on lower live birth rates (Black OR 0.33 [95% CI 0.14-0.77] p<0.001).

Analysis of frozen cycles

When exploring the live birth and clinical pregnancy rates specifically for cryopreserved (frozen) cycles, we performed the same multivariate analysis, using the same covariates on the frozen cycles alone. We found similar significant differences, as seen in the overall analysis, for live birth and clinical pregnancy rates between South Asian women and White women; South Asian women showed a reduced odds of achieving live birth compared to White women (OR 0.69 [95% CI 0.48-0.99]) and a comparable odds of achieving clinical pregnancy (OR 0.88 [95% CI 0.63-1.23] p=0.5).

However, in contrast to the results seen for the overall analysis when the frozen cycles were analysed separately there appeared to be no difference in live birth or clinical pregnancy rates between Black women and White women; live birth (OR 0.42

[95% CI 0.16-1.10] p=0.08) and clinical pregnancy (OR 0.41 [95% CI 0.17-1.02] p=0.06).

Discussion

Main findings

The results of our study show that there are significant disparities between ethnic groups for IVF outcomes.

Within our study both the Black and South Asian population showed a statistically significant reduced chance of live birth for women undergoing their first fresh cycle, after adjustment for confounding factors. When exploring clinical pregnancy rates, the Black population once again showed a statistically significant reduced chance of clinical pregnancy; furthermore implantation rates were much lower for Black women than White women. Interestingly, when looking at implantation rates and clinical pregnancy rates for the South Asian population there was no statistically significant difference compared with White women. This could suggest that although the South Asian population have a similar chance of achieving a pregnancy as the White population, they are more likely to lose the pregnancy (i.e. have a higher miscarriage rate) resulting in a lower chance of live birth. This is consistent with data from a systematic literature review presented recently at an international conference for reproductive medicine (ASRM), which looked at the relationship between ethnicity and miscarriage[47]. Women of mixed race were also found to have statistically significantly reduced odds of live birth compared with White women, however as

seen with South Asian women this difference was not seen for clinical pregnancy. It may be that mixed race women are also at higher risk of pregnancy loss; although this has not been demonstrated in existing literature.

One of the most interesting findings of this study was seen when the cycles for frozen data were removed and analysed separately from the overall dataset. For Black women there were no significant differences in live birth or clinical pregnancy compared with White women, even after adjusting for confounders; although the numbers of cycles included in the frozen analysis was small (n=34). However, for South Asian women the reduced odds of live birth were maintained in the frozen cycle analysis.

Other predictors in the model that showed statistically significant reduced odds of live birth were age, tubal factor and duration of infertility greater than 5years. Increasing age has been long established as strongly correlated to poorer IVF (and natural conception) success. Similarly a review of predictors for IVF success by Loendersloot et al found duration of infertility greater than 5years also to be linked to poorer IVF success[48]. Cause of infertility, specifically tubal factor, has also been implicated in predicting poorer chance of IVF success[48]. The only factor which was significantly associated with increased IVF success was number of embryos transferred; this variable consisted of a scale from 0-3, however current practice guidelines dictate that no more than 2 embryos should be transferred at one time. The model shows that the more embryos transferred the greater the chance of live

birth, this increased odds is most likely highlighting the distinction between transferring one or two embryos compared with none.

Strengths and limitations

One of the main strengths of our cohort study is the sample size (n=13473). With the benefit of this large sample size, the size of the ethnic groups were large enough to analyse individually, thus allowing for detailed exploration into the effects on specific racial groups. Another strength is the specificity of the ethnic groups. No study to date has been able to analyse data for specific ethnic groups in detail. The largest American studies[4,5] compared only Black women with White women. Other studies[2,6,24] only used four main ethnic groups (Black, Asian, Hispanic and White) which meant combining certain racial groups like South Asian with Chinese, who are genetically different and so would not necessarily behave in the same way. Furthermore there has previously been no study that has accounted for the mixed race population. Due to the large number of variables recorded within the database we were able to account for a large majority of the known confounders in the multivariate analysis, which other studies previously have failed to do.

We acknowledge there is significant unequal distribution of cycles amongst each ethnic group, furthermore a substantial number of patients (n=1467) have not stated ethnicity. This group constitutes more than 10% of the study population, plus all the ethnic minority groups are smaller than this 'not stated' group and so this may have influenced the data and added bias to the results.

A further limitation of the study is that we were unable to account for smoking status or alcohol consumption, it could be that these factors play a role in the lower pregnancy success rates seen in certain ethnic groups. In addition we were unable to adjust for ovarian reserve or embryo quality as known confounders when performing multivariate analysis, this was due to the insufficient numbers recorded. There is an argument that the difference in IVF success rates could be influenced primarily by socioeconomic factors, such as lack of access to medical treatment leading to higher age at first encounter. Unfortunately, our cohort study was unable to explore socio-economic factors in detail, furthermore the large majority of the patient population from our cohort study were non-NHS patients (75%) paying for their own treatment, which adds a population bias.

We observed differences in findings between unadjusted and adjusted estimates in our analyses. For example, South Asian women were found to have no significant difference in live birth rate compared with White women in the univariate analysis; however following adjustment for confounders a statistically significant reduced odds of live birth was seen. These differences have arisen because of clear differences in the characteristics of women from different ethnic groups who underwent infertility treatment (Tables 5 and 6). As South Asian women and those with “not stated” ethnicity had less risk factors than White women, adjusting for the risk factors increased the difference between these groups (Tables 9 and 11).

Comparison of results with existing literature and an updated meta-analysis

We compared the results of this cohort study with the meta-analysed existing data, as identified in the systematic review outlined in Chapter 2. We then added the results from our cohort study to the existing study data, to provide an updated meta-analysis.

Black vs. White

Results of the meta-analysis of 9 studies[2,4,5,23,24,31,40,41,43] (in chapter 2) showed Black women were found to have a statistically significant reduction in live birth and clinical pregnancy, these findings were in keeping with those of our cohort study.

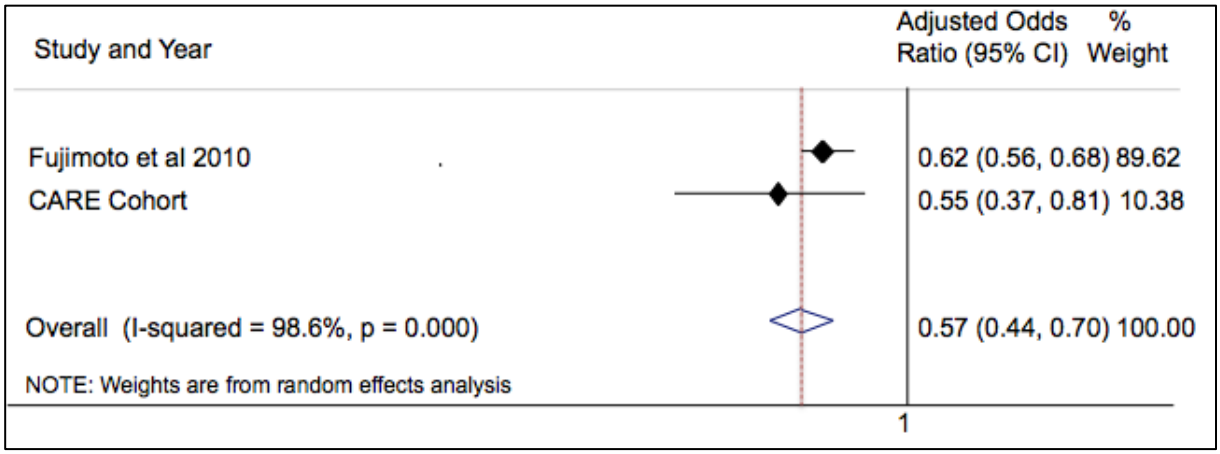
Data from the existing 9 studies[2,4,5,23,24,31,40,41,43], plus the data from our cohort study were combined to provide an updated meta-analysis comparing Black women with White women for live birth and clinical pregnancy rates following fresh cycle of treatment; the results were (OR 0.60 [95% CI 0.52-0.69] p<0.001) and (OR 0.69 [95% CI 0.58-0.82] p<0.001) respectively. The results showed moderate heterogeneity for live birth data and statistically significant high heterogeneity for clinical pregnancy rate.

There were three studies which recorded data separately for frozen cycles[4,5,26] these studies only investigated Black women and White women. The meta-analysed results showed no difference in live birth or clinical pregnancy rates for Black women compared with White women: (OR 0.90 [0.75-1.07] p=0.23) and (OR 0.94 [1.03-1.12] p=0.54) respectively. This finding was consisted with our cohort study. We

combined the three studies which looked at frozen data[4,5,26] with the data from our cohort study to provide an updated unadjusted meta-analysis; live birth (OR 0.89 [95% CI 0.74-1.08] p=0.24) and clinical pregnancy (OR 1.01 [95% CI 0.82-1.24] p=0.92) respectively.

More importantly than the unadjusted live birth rates were the adjusted rates. Three papers[4–6] calculated adjusted live birth rates (for fresh cycles) accounting for multiple factors, these included; maternal age, BMI, number of embryos transferred, diagnosis of male factor, endometriosis, PCOS, diminished ovarian reserve, tubal factors, uterine factors and other factors. Two papers from the systematic review, both by Seifer et al, calculated adjusted relative risks[4,5]; they found a statistically significant increased adjusted relative risk of Black women *not* having a live birth compared with White women. One paper, by Fujimoto et al, calculated an adjusted odds ratio for live birth[6]. As was found with our cohort study, there were statistically significant reduced odds of live birth for Black women compared with White women after adjustment for confounders. We pooled the adjusted odds ratios from Fujimoto et al with our cohort study; this is displayed in Figure 16.

Figure 16. White vs. Black – adjusted odds ratio for live birth (fresh cycles)



The two studies that provided an adjusted analysis for frozen cycles used the adjusted relative risk (aRR) of *not* achieving a live birth and so we were unable to calculate a pooled odds ratio combining our adjusted data for frozen cycles. However; as outlined in chapter 2 both of these studies found no statistical difference in achieving a live birth for Black women compared to White women after adjusting for confounders; this was in keeping with the finding from our adjusted analysis.

In practical terms, the main difference between fresh cycle treatment and frozen cycle is that no stimulation is required for a frozen cycle i.e. no use of follicle stimulating hormone (FSH). It could be that there is something within the stimulation process that does not suit Black women. This finding should prompt further research to investigate the biological plausibility.

Asian vs. White

Regarding the comparison between Asian and White women, eight studies in the systematic review comprised of women both from South Asian and Chinese ethnic

groups[3,6,24,28,31,40,42,44]. Of these eight studies five papers specified a cohort of Indian or South Asian women[28,31,40,42,44]. To directly compare the results of these five studies with our own cohort study, data were meta-analysed separately in a specific "South Asian" group. The meta-analysis showed a statistically significant reduction in live birth for South Asian women compared with White women, this was also seen in our cohort study. For clinical pregnancy, the unadjusted meta-analysed data showed a statistically significant reduction for South Asian women compared with White women, however our cohort study did not; (OR 0.65 [95% CI 0.47-0.90] p=0.008) vs. (OR 1.02 [95% CI 0.89-1.16] p=0.8). We performed the following statistical calculations to see if the difference between these odds ratios was significant:

Se_1 = (Log(upper limit OR) – Log(lower limit OR) / (2 x 1.96)

= (Log0.90 – Log0.47) / (2 x 1.96)

= 0.166

Se_2 = (Log(upper limit OR) – Log(lower limit OR) / (2 x 1.96)

= (Log1.16 – Log0.89) / (2 x 1.96)

= 0.068

diff = $LogOR_1 – LogOR_2$

= Log0.65 – Log1.02

= -0.450

Se(diff) = $\sqrt{(Se_1)^2 + (Se_2)^2}$

= $\sqrt{0.027 + 0.004}$

= 0.179

$$\frac{\text{diff}}{\text{Se(diff)}} = -2.517$$

A value below -1.96 indicates that there is a significant difference between the two odds ratios ($p<0.01$).

We combined the results from our cohort study with the existing studies to provide an updated unadjusted meta-analysis for South Asian women vs. White women; results showed significant reduced odds of live birth for South Asian women (OR 0.73 [95% CI 0.60-0.89] p=0.002) and no difference in clinical pregnancy (OR 0.78 [95% CI 0.56-1.08] p=0.1). The results of our cohort study contributed to around 38% of the total data in the updated meta-analysis, therefore we can be reassured it does not dominate the results. Given that the clinical pregnancy rates are comparable between South Asian women and White women but the live birth rates are much lower for South Asian women; this suggests that South Asian women have higher miscarriage rates. This finding is consistent with results from a systematic review presented at a recent international conference looking at the effect of ethnicity on miscarriage[47].

The more important analysis was the adjusted odds ratios for live birth. Three existing studies[3,6,28] calculated adjusted odds ratio for Asian women vs. White women, accounting for multiple confounders including; age, BMI and diagnosis of PCOS. However, only one of these studies specified a "South Asian" cohort[28]. We pooled the adjusted odds ratios for the South Asian population from our cohort study

with both the broader termed “Asian” group and the specific “South Asian” group. The results of the updated meta-analysis in Figures 17 and 18 show that both “Asian” women and specifically “South Asian” women have reduced adjusted odds of live birth compared with White women. None of the existing published studies calculated an adjusted odds ratio for clinical pregnancy for South Asian vs. White women.

Figure 17. White vs. Asian – adjusted odds ratio for live birth

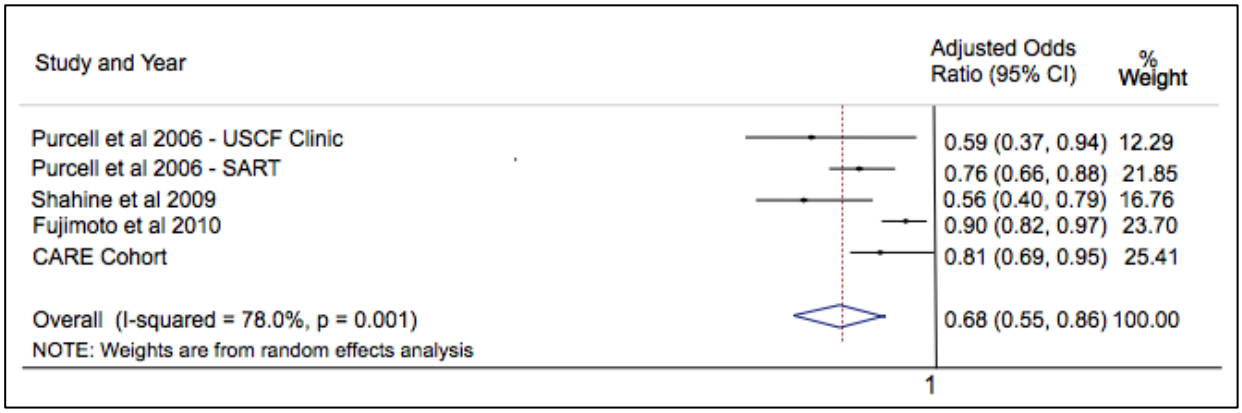
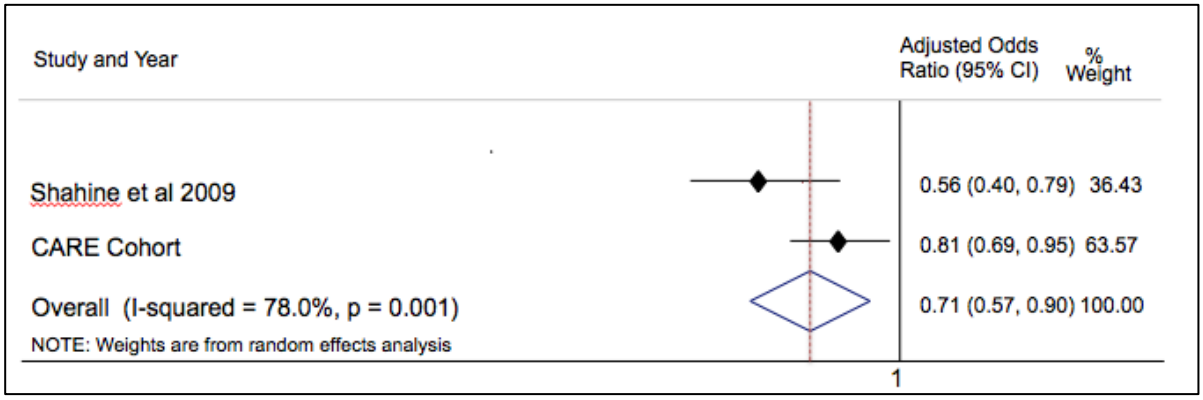


Figure 18. White vs. South Asian – adjusted odds ratio for live birth



Hispanic vs. White

Given the UK population of our cohort study we did not specifically account for the Hispanic population.

Exploring reasons for differences

The data from both our cohort study and meta-analysis of existing studies shows that Black women and South Asian women have the poorest outcomes following fresh IVF treatment. One might argue that these differences could potentially be explained by the different diagnoses of infertility seen in different ethnic populations, which is shown in (Table 4 in chapter 2). However one of the strengths of our cohort study is that we were able to adjust for cause of infertility. It is also well known that fibroids are more common amongst the Black population and so would be the obvious explanation for the lower live birth rates seen in black women. In our analysis fibroids were adjusted for within a heterogenous group of infertility termed 'other', which included endometriosis, structural abnormalities and multiple fibroids. Furthermore, a sensitivity analysis adjusting for fibroids specifically, maintained a lower live birth rate for Black women. Therefore, it is unlikely that causes of infertility alone can explain the differences in live birth seen across ethnic groups. In addition, there were inconsistent findings across the existing papers for any differences in age and body-mass index for each ethnicity (as seen in Table 4 in chapter 2) and so this is also not likely to explain the differences seen in live birth or clinical pregnancy rates.

Regarding the finding that Black women appear to do worse with fresh cycles compared with frozen cycles, this adds support for the growing interest in a "freeze-all" embryo policy in IVF. Such an approach, which would aim to freeze all embryos generated in a fresh IVF cycle, with a view to transferring later in a non-stimulated (natural) cycle. The theory behind this approach is that it would avoid the adverse effects which ovarian stimulation might have on endometrial receptivity during the treatment cycle. The effect of ovarian stimulation on endometrial receptivity could be a contributing factor to why Black women appear to perform worse with fresh (stimulated) cycles. A systematic review of 64 relevant studies (including three randomised trials) showed that the probability of a clinical pregnancy is significantly higher from “freeze-all” cycles than in fresh embryo transfers (a relative risk of 1.31, which was statistically significant)[49]. It may be worth considering a “freeze-all” policy for Black women in the first instance, if not women of all ethnicities.

Conclusions and future work

Research on assisted conception has predominantly been performed on cohorts of White women. Existing published studies have found inconclusive results for assisted conception success rates amongst women from different ethnic backgrounds. The results of the systematic review, cohort study and updated meta-analysis provide robust evidence for the hypothesis that there is an association between ethnic background and IVF success. More importantly, the commonly known confounders cannot explain this.

The findings of the work presented in this thesis should prompt investigation into the mechanisms underpinning such disparities to allow modification of laboratory and or clinical practice to improve IVF outcome for all ethnic groups. In particular, one of the interesting findings was the suggestion that Black women appear to perform worse with fresh cycles than frozen cycles. It may be worth exploration into trialling a change of clinical practice, i.e. routine elective freeze for Black women to then have a frozen embryo transfer (FET) at a later date, followed by evaluation of results to see if this can improve overall success rates for Black women. Furthermore, future research should also look to try and explain the apparent higher miscarriage rate seen in South Asian women compared to White women and look for ways to reduce this.

The work of this thesis has focused solely on the ethnicity of the female, to further investigate the effects of ethnicity on IVF outcome it would be useful to also take into consideration the male ethnicity to see what affect this has on the results.

Finally, there needs to be careful consideration of whether information regarding ethnicity and its potential affect on IVF outcome should be routinely provided to patients as part of pre-treatment counselling. Although this is not a factor that women are able to change, it may still have implications on their decision-making.

CHAPTER 4

Body mass index and IVF outcome: a prediction model

The work in this chapter was presented as an oral presentation at ESHRE Lisbon June 2015 and has been published in Human Reproduction; published on 25.10.15

Introduction

The number of couples seeking in-vitro fertilisation (IVF) in the U.K. continues to rise year upon year ("Fertility treatment in 2013. Trends and figures. HFEA."). Contrary to common perception, IVF does not guarantee success; between 38–49% of couples who start IVF will remain childless, even after undergoing up to six IVF cycles[50]. It is therefore important that subfertile couples are well informed about their chances of success with IVF. Based on their specific probability of success, the couple can decide whether the risks of the treatment and the emotional and, in many cases, financial burden can be justified. To optimise counseling for couples on their chances of a live birth after IVF, clinical prediction models, which estimate the chance of an outcome adjusted for a patient's characteristics, may play a role since clinicians' judgments can often be inaccurate[51,52]. Reliance on annually published validated age-stratified national success rates[53,54] has meant that clinicians often tend to base predictions solely on age.

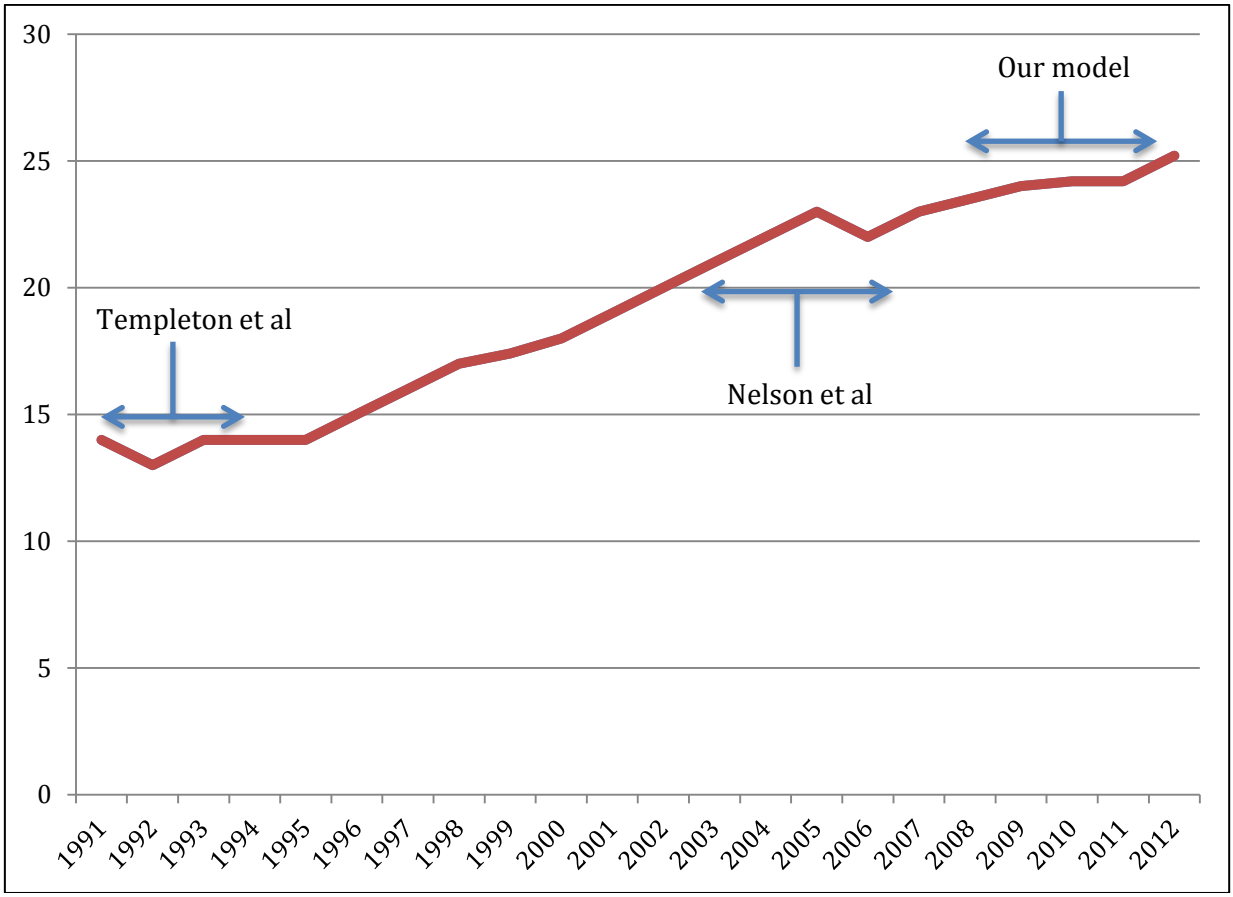
There have been many attempts to build prediction models to aid clinicians in predicting IVF success [56–62]. The two most widely recognised models, which used live birth as the primary outcome, are those by Templeton et al [58] and Nelson et al [61]. A study by a Dutch team, te Velde et al [63], used their cohort to validate both these models to assess the effects of time trends on model performance. They found that the Templeton model underestimated success rates, as one may expect given that it is a much older study and the Nelson model over-estimated success rates. The

study showed that the calibration of both models considerably improved when the models were adjusted for the changing success rates over time.

A recent study by Smith et al also performed external validation of the Templeton and Nelson models using a large dataset of over 130,000 cycles[64]. They found that the discriminative power (assessed using area under receiver operator curve) was comparable between the models; but that the Nelson model had markedly better calibration. They also found both models underestimated the live birth rate, although as seen with te Velde et al [63], this improved when the models were updated to reflect improvements in live birth rates over time.

A recent report by HFEA (Human Fertilisation and Embryology Authority) recognised that IVF practice and outcomes have seen significant changes between 2008-2012, primarily because of the introduction of day 5 (blastocyst) embryo transfer[1]. Given these advancements in technology, and the fact that the existing most noted models were built before 2008, there is a need for a new model to be built from more recent data (Figure 19).

Figure 19. Live birth rate per cycle started (1991-2012)



Data taken from HFEA report **'Fertility treatment 2013: trends and figures'** released 17th December 2014.

Another major further pitfall of the existing models is that they have not been able to account for certain key predictors of IVF treatment outcome. In particular, the most recent of these models[61] built using a large dataset provided by the Human Fertilisation Embryology Authority (HFEA) was not able to account for body-mass index (BMI), any measure of ovarian reserve or ethnicity.

A systematic review in 2011 which looked at 33 studies exploring BMI and IVF outcome concluded that a raised BMI has an adverse effect on pregnancy outcomes for women underdoing IVF treatment[13]. They also found that this negative association was apparent for both obese and overweight women[13]. However, recent studies have challenged this finding and have shown that raised BMI does not appear to reduce clinical pregnancy or live birth rates; nor is it that women with raised BMI require higher doses of gonadotrophins in their treatment compared with normal BMI women[9,10]. BMI is the only pre-treatment factor that can be controlled and changed by the patient to help improve their chances of success. Often in the clinical setting clinicians will advise patients to lose weight, and if the patient is obese the advice is to reach a target BMI of 30 or below; in order to meet government NHS funding criteria. This arbitrary value set for funding criteria is based on the knowledge and experience of knowing that high BMI's result in poorer pregnancy outcomes[7,8]. What is currently lacking in the field of reproductive medicine is a tool that can be used by patients and clinicians to demonstrate how body mass index affects pregnancy success and also the interplay of BMI with other important factors that affect success rates, i.e. female age.

Other important variables that have yet to be included in previous clinical prediction models in assisted reproduction are ovarian reserve and ethnicity. There is strong evidence to suggest that women with a diminished ovarian reserve generally result in a poor response to gonadotropin therapy and therefore the chance of a successful pregnancy[65,66]. A recent study found specifically that antral follicle count (AFC) correlated strongly with the number of mature oocytes retrieved in IVF/ICSI cycles[67],

which can dictate the chances of pregnancy, moreover another study found that AFC significantly added prognostic value to female age in predicting response to ovarian hyperstimulation[68]. Regarding ethnicity and IVF outcome, there have been several large cohort studies that have shown ethnicity has an association with IVF outcome, in particular showing that Black and South Asian races appear to have the poorest outcomes following fresh cycle IVF treatment[4–6,25,36]. The work presented in this thesis thus far also provides robust evidence to show that disparities exist in IVF outcomes amongst women of different ethnicities, both in the UK and worldwide. Despite this evidence, ethnicity is a factor that is yet to be included as a predictor in any model predicting live birth following IVF.

Finally, to-date there is no model available for use *before* a couple embarks on their *first* treatment utilising pre-treatment factors alone. Both the Templeton and Nelson models use data from previous cycles and the Nelson model includes treatment factors such as hormonal preparation[58,61]. Frequently in clinical consultations couples will ask the clinician what their chance of IVF success is; based on personal experience this is more common amongst couples seeking treatment for the first time. The vast majority of clinicians will base their estimates on the age-related success rates produced by national HFEA data; however, these success rates may not be directly applicable to the patient. What is currently lacking in this field is a prediction model to provide a more personalised approach to counseling and allow for a more accurate estimate of success. Consequently, the aim of this study is to derive, assess and validate a novel predictive model that will estimate the chance of live birth for women undergoing their first IVF non-donor cycle. This model will use

only pre-treatment factors and include previously unrecorded predictors such as BMI, ovarian reserve and ethnicity.

Aims and objectives

1. To investigate the relationship between BMI and IVF outcome.
2. To build a novel prediction model, incorporating BMI, to estimate the chance of live birth for women *before* undergoing their first fresh IVF cycle.
3. To incorporate novel factors into prediction the model, such as ovarian reserve and ethnicity.

Methods

Derivation cohort

The study population was derived from a database of all patients who had undergone their first fresh non-donor cycle of IVF (including ICSI) at any of the *Centres for Assisted Reproduction* (CARE) clinics across the UK and Ireland, between 2008 and 2012. CARE is one of the UK's largest independent providers of fertility services, where both NHS and non-NHS patients are treated, approximately 25% of patients are NHS funded and 75% fund themselves. The CARE database consists of routinely collected baseline demographics, cycle data and outcome data for all patients.

Within the variable for previous IVF, any woman with a history of IVF treatment, whether it was at a CARE clinic or elsewhere, were assigned a "1", women without

any history of IVF treatment were assigned “0”. All women with a “1” were excluded from analysis. The reason for this was to exclude previous treatment as a confounder and also because the primary use of the model is for couples seen at their very first clinic appointment, *prior* to embarking on IVF treatment. The decision to include IVF and ICSI as one variable was because the authors agreed that success rates are comparable for the two treatment modalities and so it was reasonable to include them together. Also, the model is designed for use before patients undergo treatment, because occasionally in the cases of “mild male factor” clinicians will often decide to cross-over from IVF to ICSI once the patient has come through for treatment we felt it was better to keep IVF and ICSI as one variable.

Baseline demographics, cycle data and outcome data were retrieved from 12 CARE clinics across the UK. The CARE consortium is composed of five main fertility clinics; Nottingham, Manchester, Northampton, Sheffield and Dublin, and a further seven satellite centres; Bolton, Boston, Derby, Leicester, Mansfield, Milton Keynes and Peterborough. For patients seen initially at the satellite clinics, they are seen up to the point of egg collection; egg collection, all embryology and embryo transfer is then performed at the nearest main clinic. Following the embryo transfer the satellite clinic resumes full care of the patient.

The original database contained information on over 50,000 cycles dating back to 1998. A decision was made to limit the dataset from 2008 onwards due to advances in technology over time and improvements in clinical practice such as greater numbers of blastocyst transfer and single embryo transfer, as detailed in the recent

HFEA report, which in turn have effected success rates ("Fertility treatment in 2013. Trends and figures. HFEA."). Data from the first cycle only were used to eliminate the bias from previous cycle failures. Furthermore, by limiting to only first cycle we were able to express the probability of live birth outcome per individual woman.

Primary outcome

The primary outcome for the model was live birth. The definition was consistent with that of Nelson et al[61]; *"at least one baby was born alive and survived for more than one month"*.

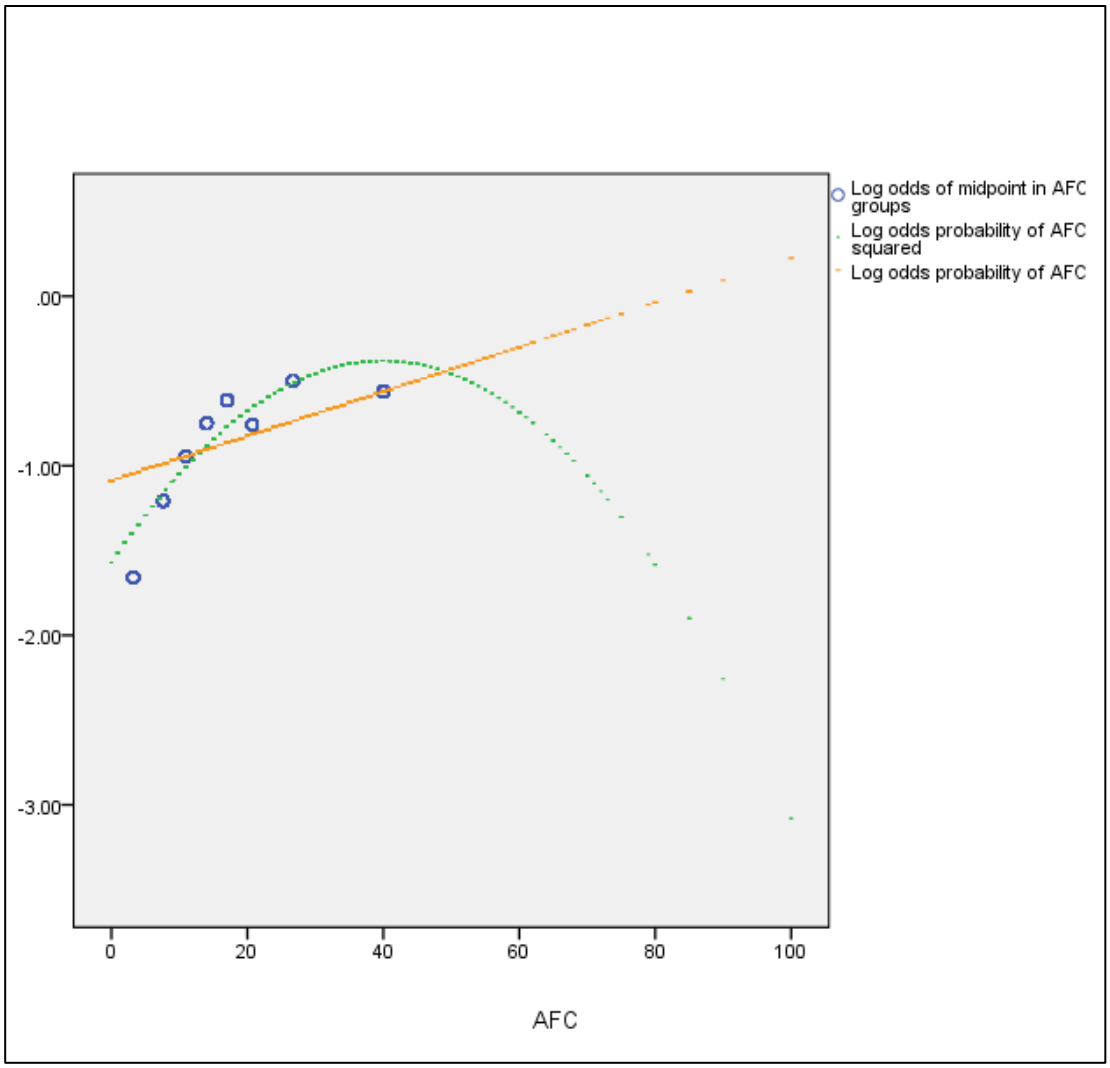
Statistical analyses:

Model development

Univariable logistic regression analyses were performed to assess the association of each of the predictive factors with live birth. A multivariable logistic regression model was used to derive the final prediction model for live birth. The predictors included in the multivariable model were pre-selected based on knowledge from the existing literature [48] and clinical knowledge and were as follows: age, body mass index (BMI), ethnicity, cause of infertility, duration of infertility, antral follicle count (AFC), previous live birth and previous miscarriage. AFC was selected in preference to early follicular follicle stimulating hormone (FSH) as it is a more accurate measure of ovarian reserve[65,68]. Anti-mullerian hormone (AMH) has similar accuracy to AFC[65,68] and is a more objective measure of ovarian reserve. However as AMH is a fairly recent test it was not available for most patients in the derivation cohort, therefore AFC was selected in preference.

The continuous variables: age, BMI, duration of infertility and AFC were assessed for their functional form using plots of the observed log odds. As was shown in chapter 3; BMI had a linear relationship with the log odds of the probability of live birth (Figure 14). In the case of age, duration of infertility and AFC there was a non-linear relationship with live birth. Appropriate transformations were carried out, to produce linear relationships, and subsequently included in the model. For age and duration of infertility these are shown on the same dataset in chapter 3 (Figures 13 and 15), AFC is shown in Figure 20.

Figure 20. AFC plotted against the log odds of the probability of live birth



For AFC a quadratic relationship with the log odds of live birth was observed.

Regarding categorical variables; for ethnicity a reference population of White was selected and each ethnic group was compared to this. For cause of infertility, each cause was categorised as yes or a no, with no reference group; the causes of infertility were not mutually exclusive.

Missing data

The whole dataset contained 9915 women, data entry was complete in all variables except for BMI and AFC. The total number of cases with compete data was 2911. Two separate models were created; one using the entire dataset (n=9915) and a second model using patients with complete data only (n=2911). For the larger dataset we were required to impute the missing data. A multiple imputation procedure was conducted using an iterative Markov chain Monte Carlo (MCMC) method. All predictors and the outcome of live birth were included in the imputation process to maximise the precision of the imputations. All univariable models and the multivariable model were fitted to the 20 imputed datasets arising from the multiple imputation procedure. The parameter estimates and covariance's arising from the models from each imputed dataset were combined to produce inferential results. For the purpose of this study the primary model was that built using the whole dataset (n=9915 women) and will be referred to as the "final model".

Predictive ability

Initially the model was assessed for predictive ability using apparent validation. Apparent validation is when the model performance is assessed directly in the same cohort from which it was derived[69]. The two performance measures used were discrimination and calibration. Discrimination is the ability of the model to correctly discriminate between those who had the outcome and those that did not i.e. correctly distinguish between the women who had a live birth (for whom the model assigns a higher probability) and women who do not have a live birth (for whom the model assigns a lower probability). The area under receiver operating characteristics

(AUROC) curve (also known as a c-statistic) was used as a measure of discrimination; a model with an AUROC curve of 0.5 would have no discriminative power at all, while 1.0 would reflect perfect discrimination. Calibration refers to the agreement between the predicted probabilities of live birth and the observed (actual) probabilities. The predicted probabilities from the final model were assessed for accuracy across increasing tenths of predicted probabilities using calibration plots. The mean observed probability is plotted against the mean predicted probability in each tenth and perfect calibration is displayed as a straight line passing through zero with a gradient of one.

Model validation

Before any prediction model can be utilised in clinical practice to aid decision-making, it is essential to confirm that the developed model also predicts well in a "similar but different" population outside of the development cohort, i.e. external validation (generalisability)[69]. There are three different types of external validation: temporal validation, geographical validation, and domain validation. In temporal validation, the model is validated on new patients that are from the same centre as the development cohort, but in a different time period[69]. This is the form of external validation performed on our final model.

External validation of the model was performed on a cohort of women undergoing their first fresh IVF cycle at any CARE clinic during the year of 2013 (temporal validation)[69,70]. The missing data in the validation cohort were also imputed using the same method as the derivation cohort. For ease of computation and interpretation,

the average measures of the imputed values were taken across all 20 imputed datasets for women who had values imputed, so that validation was performed on only one dataset. The model was fitted to the validation cohort (2013 population) using the same parameter estimates derived from the study cohort (2008-2012 population). The predictive ability of the model was assessed on the external validation cohort. The AUROC under the curve was determined to assess discriminatory ability and calibration plots were presented.

As a formal test of calibration we assessed "calibration-in-the-large" to compare the mean predicted probability of live birth with the mean observed probability of live birth. This is essentially the intercept from the model, which is only adjusted for the linear predictors from the final model, applied to the patients in the external cohort. A significant deviation from zero indicates that predictions are systematically too low or too high[71]. The calibration slope was also calculated, where a perfect slope (i.e. perfect agreement between predicted and observed probabilities) would have a gradient of one. Significant deviations from one would suggest that low predicted probabilities were too low or too high, and high predicted probabilities were too high or too low.

All statistical analyses were performed using SPSS (ver. 21.0; SPSS Inc., Chicago) and SAS (ver.9.2; SAS Institute, Cary, North Carolina).

Ethical approval

Permission for use of the database was granted by the CARE IRB following review of the study protocol. The dataset was anonymised according to the ICO's (Information Commissioner's Office) guide on non-identifiable data. Furthermore the CARE data protection certificate allows for their data to be used for survey and research purposes.

Results

A total of 9915 women were used to build the final model. Figure 21 shows how we established the eligible cohort of IVF (including ICSI) treatment cycles. We used both the whole dataset, which included missing data (n=9915 women), and the complete data only dataset (n=2911 women) to build two independent models; however we did not perform external validation on the model created using the complete data only dataset. Table 12 shows a comparison of the baseline characteristics of the two cohorts. Since the characteristics were similar between the two cohorts we could be satisfied that there were no significant differences between the two cohorts and so it was reasonable to use the larger dataset to build the "final model". A description of the whole dataset including cycle data and overall outcome data can be found in Table 13. The overall rate of at least one live birth, from the whole dataset, was 31.5%.

Figure 21. Definition of eligible cohort and analysis samples

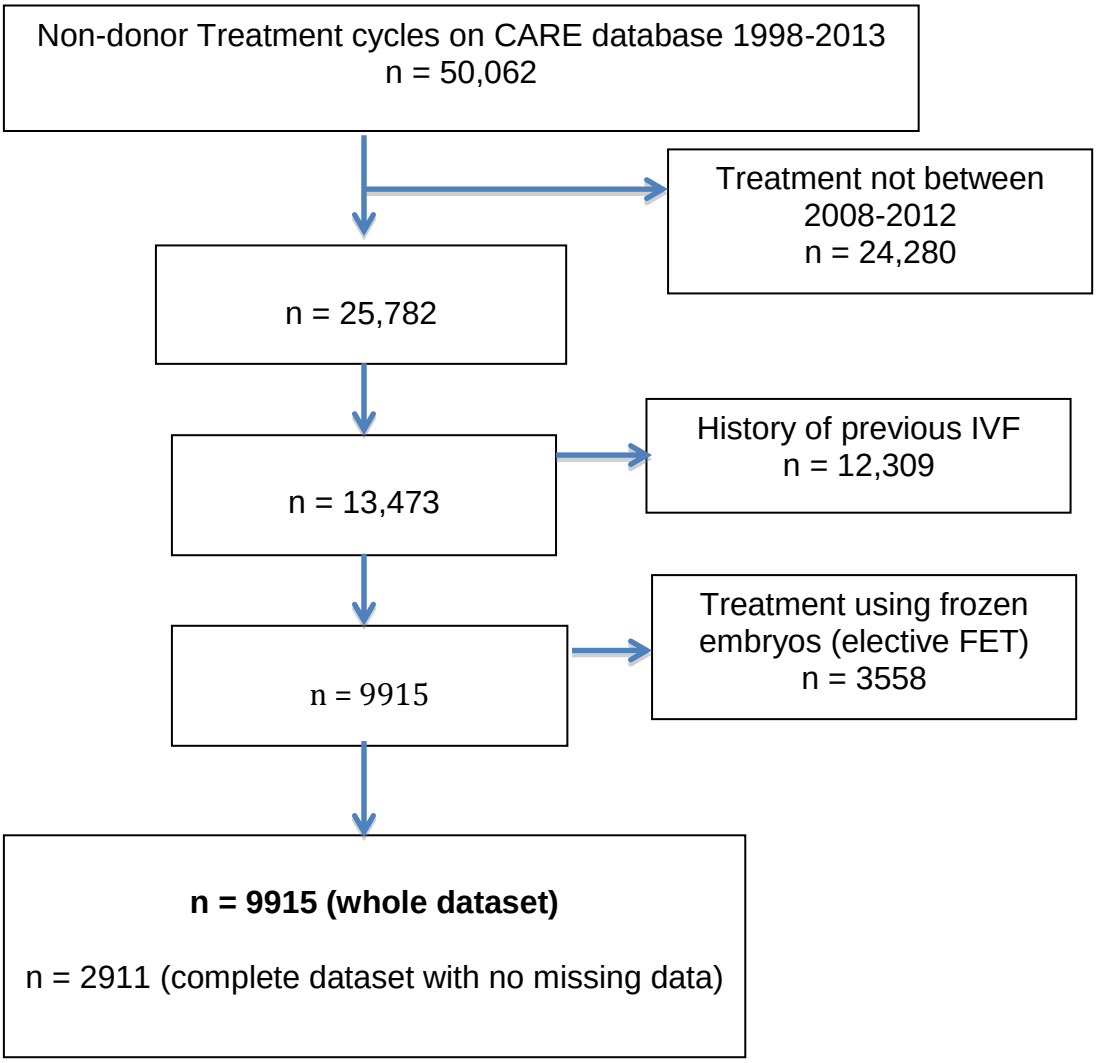


Table 12. Baseline characteristics of the cohort for whole dataset and the complete cases dataset

	Cohort n (%) *or* Mean (SD)	
	Whole dataset (n=9915)	**Complete cases dataset (n=2911)**
Age	34.6 (5.4)	34.2 (5.0)
Duration of infertility (in completed years)	2.0 (2.0)	2.5 (1.8)
BMI*	24.8 (4.0)	24.7 (3.8)
AFC*	18.7 (13.6)	19.2 (14.0)
Previous miscarriage	1818 (18.3%)	621 (21.3%)
Previous live birth	1578 (15.9%)	415 (14.3%)
Cause of infertility		
Tubal factor	1442 (14.5%)	481 (16.5%)
Anovulation	1088 (11.0%)	342 (11.7%)
Unexplained	2950 (29.8%)	681 (23.4%)
Other (E.g Endometriosis, fibroids)	3005 (30.3%)	1060 (36.4%)
Male factor	5611 (56.6%)	1953 (67.1%)
Ethnicity		
White	7530 (75.9%)	2437 (83.%)
Asian	768 (7.7%)	214 (7.4%)
Black	162 (1.6%)	51 (1.8%)
Chinese	60 (0.6%)	13 (0.4%)
Other	115 (1.2%)	47 (1.6%)
Not stated	924 (9.3%)	27 (0.9%)
Mixed	356 (3.6%)	122 (4.2%)

* Variable contains missing data.

Table 13. Descriptive data for whole dataset (n=9915)

	N (%) or Mean (SD)
Other ovarian reserve markers	
AMH (n=1212)	16.2 (18.8)
Day 2 FSH (n=2911)	7.49 (5.2)
Cycle data	
IVF	3829 (38.6%)
ICSI	6086 (61.4%)
No. of oocytes retrieved	9.0 (5.9)
No. of mature oocytes inseminated	6.9 (4.8)
2pn	4.8 (3.8)
3pn	0.2 (0.6)
Total no. of embryos	5.3 (4.1)
No. of embryos transferred	
0	1591 (16.0%)
1	2911 (29.4%)
2	5127 (51.7%)
3	286 (2.9%)
No. of embryos frozen	1.3 (2.7)
Outcome data	
Biochemical pregnancy	4144 (41.8%)
Clinical pregnancy	3514 (35.4%)
Pregnancy outcome	
Live birth	3121 (31.5%) [a]
Miscarriage	328 (9.3%) [b]
Termination	18 (0.5%) [b]
Still birth	7 (0.2%) [b]
Neonatal death	9 (0.1%) [b]

[a] Overall live birth rate for whole cohort
[b] Calculated as a % of those with a clinical pregnancy

Missing data

As mentioned previously, only two of the variables selected for use in the multivariate model had missing data, these were BMI and AFC. Descriptive characteristics of women with missing and non-missing data for BMI and AFC can be found in Table 14. The data across each baseline characteristic were comparable between the two groups. However, more women with a BMI measurement were of white ethnicity (81.7% versus 70.5%) and had partners with male factor infertility (65.1% versus 48.5%) than women without a BMI measurement.

Table 14. Display of missing and valid data

	BMI missing (n=5101, 51.4%)	BMI Valid (n=4814, 48.6%)	AFC Missing (n=6365, 64.2%)	AFC Valid (n=3550, 35.8%)
Age (Mean +SD)	35.0 (5.7)	34.3 (5.1)	34.9 (5.6)	34.2 (5.0)
Duration of inf in years (Mean, SD)	3.0 (2.0)	2.0 (2.0)	3.0 (2.0)	2.0 (2.0)
AFC (Mean, SD)	19.1 (13.3)	18.4 (13.7)	---	18.6 (13.5)
BMI (Mean, SD)	---	24.8 (4.0)	25.0 (4.3)	24.7 (3.8)
Previous miscarriage	840 (16.5%)	978 (20.3%)	1064 (16.7%)	754 (21.2%
Previous live birth	880 (17.3%)	698 (14.5%)	1038 (16.3%)	540 (15.2%)
IVF	1726 (33.8%)	1444 (30.0%)	2099 (33.0%)	1071 (30.2%)
ICSI	2995 (58.7%)	3091 (64.2%)	3772 (59.3%)	2314 (65.2%)
Tubal factor	706 (13.8%)	736 (15.3%)	856 (13.4%)	586 (16.5%)
Anovulation	473 (9.3%)	615 (12.8%)	687 (10.8%)	401 (11.3%)
Unexplained	1670 (32.7%)	1280 (26.6%)	2122 (33.3%)	828 (23.3%)
Other	1458 (28.6%)	1547 (32.1%)	1708 (26.8%)	1297 (36.5%)
Male factor	2476 (48.5%)	3135 (65.1%)	3265 (51.3%)	2346 (66.1%)
Ethnicity - White	3596 (70.5%)	3935 (81.7%)	4562 (71.7%)	2698 (83.6%)
Ethnicity - Asian	372 (7.3%)	396 (8.2%)	502 (7.9%)	266 (7.5%)
Ethnicity - Black	81 (1.6%)	81 (1.7%)	98 (1.5%)	64 (1.8%)
Ethnicity - Chinese	26 (0.5%)	23 (0.7%)	41 (0.6%)	19 (0.5%)
Ethnicity - Other	48 (0.9%)	67 (1.4%)	60 (0.9%)	55 (1.5%)
Ethnicity - Not stated	836 (16.4%)	88 (1.8%)	893 (14.0%)	31 (0.9%)
Ethnicity - Mixed	143 (2.8%)	213 (4.4%)	209 (3.3%)	147 (4.1%)

Univariate and multivariate analyses

The univariate associations of live birth for the original whole dataset and complete cases dataset are shown in Table 15. Results from both datasets showed similar findings. The multivariable logistic regression model predicting live birth for the imputed dataset (final model) is displayed in Table 16. The final model (n=9915) shows that the odds of a successful live birth decrease with age. This reduction in the odds of live birth is greater with each increasing year of age past the age of 36 compared with up to the age of 36. Other variables which showed a statistically significant reduction in odds of live birth in the multivariate final model were; tubal factor, unexplained infertility, and being Asian or Black. The univariate analysis suggested that increasing BMI, duration of infertility greater than 5 years and previous miscarriage was associated with decreased odds of live birth, however these associations became non-significant in the multivariate analysis.

The model built from the complete data only dataset (n=2291) can be found in Appendix 1. Findings were consistent with that of the "final model", although there was also a statistically significant reduction in live birth associated with duration of infertility greater than 5 years (p=0.04).

Table 15. Univariate associations of potential predictors for live birth following IVF for cases that have a non-missing value for each predictor (whole dataset) and cases with complete data for all predictors (complete cases)

	Whole dataset		Complete cases	
	OR (95% CI) n=9915	**p value**	**OR (95% CI) n=2911**	**p value**
Age	0.94 (0.93-0.94)	<0.001	0.93 (0.91-0.94)	<0.001
Duration of infertility (in completed years)	0.97 (0.95-0.99)	0.003	0.94 (0.90-0.99)	0.02
BMI*	0.98 (0.97-0.99)	0.01	0.98 (0.96-1.00)	0.06
AFC*	1.01 (1.01-1.02)	<0.001	1.01 (1.01-1.02)	<0.001
Prev Misc	0.84 (0.74-0.94)	0.002	0.83 (0.68-1.01)	0.07
Prev LB	0.93 (0.83-1.05)	0.2	1.10 (0.88-1.37)	0.4
Cause of infertility				
Tubal factor	0.90 (0.80-1.01)	0.08	0.94 (0.76-1.17)	0.6
Anovulation	1.21 (1.07-1.40)	0.003	1.13 (0.89-1.44)	0.3
Unexplained	1.01 (0.92-1.11)	0.8	1.09 (0.9-1.30)	0.4
Other (E.g Endometriosis, fibroids)	0.70 (0.63-0.77)	<0.001	0.76 (0.64-0.90)	0.001
Male factor	1.02 (0.94-1.11)	0.7	1.08 (0.92-1.28)	0.3
Ethnicity				
White	Reference		Reference	
Asian	0.95 (0.81-1.12)	0.6	1.00 (0.74-1.36)	0.9
Black	0.44 (0.29-0.67)	<0.001	0.36 (0.16-0.80)	0.01
Chinese	0.84 (0.48-1.48)	0.5	1.92 (0.65-5.75)	0.2
Other	0.68 (0.44-1.05)	0.08	0.46 (0.21-0.99)	0.05
Not stated	1.00 (0.86-1.16)	1	1.12 (0.50-2.51)	0.8
Mixed	0.86 (0.68-1.09)	0.2	0.98 (0.66-1.45)	0.9

* Variable contains missing data

Table 16. Final multivariate logistic regression model for Live birth (n=9915)

	Parameter Estimate	S.E.	P value	Odds Ratio	95% CI	
					Lower	Upper
Age:						
≤36 years (per increasing year)	-0.035589	0.008	<0.001	0.97	0.95	0.98
>36 years (per increasing year)	-0.106139	0.012	<0.001	0.90	0.88	0.92
Body mass index	-0.010881	0.009	0.2	0.99	0.97	1.01
Cause of infertility:						
Male factor	-0.085967	0.056	0.1	0.91	0.82	1.02
Tubal factor	-0.254369	0.069	<0.001	0.78	0.68	0.89
Anovulation	-0.138708	0.082	0.09	0.87	0.74	1.02
Unexplained	-0.133782	0.067	0.04	0.88	0.77	0.99
Other (e.g. Endo)	-0.118451	0.062	0.05	0.89	0.79	1.00
Ethnicity:						
White	0			Reference		
Asian	-0.171572	0.084	0.04	0.84	0.71	0.99
Black	-0.683648	0.214	<0.001	0.51	0.33	0.77
Chinese	-0.181580	0.293	0.5	0.83	0.47	1.48
Other	-0.355212	0.222	0.1	0.70	0.45	1.08
Not stated	-0.005533	0.083	0.9	0.99	0.84	1.17
Mixed	-0.192857	0.122	0.1	0.83	0.65	1.05

Table 16. continued

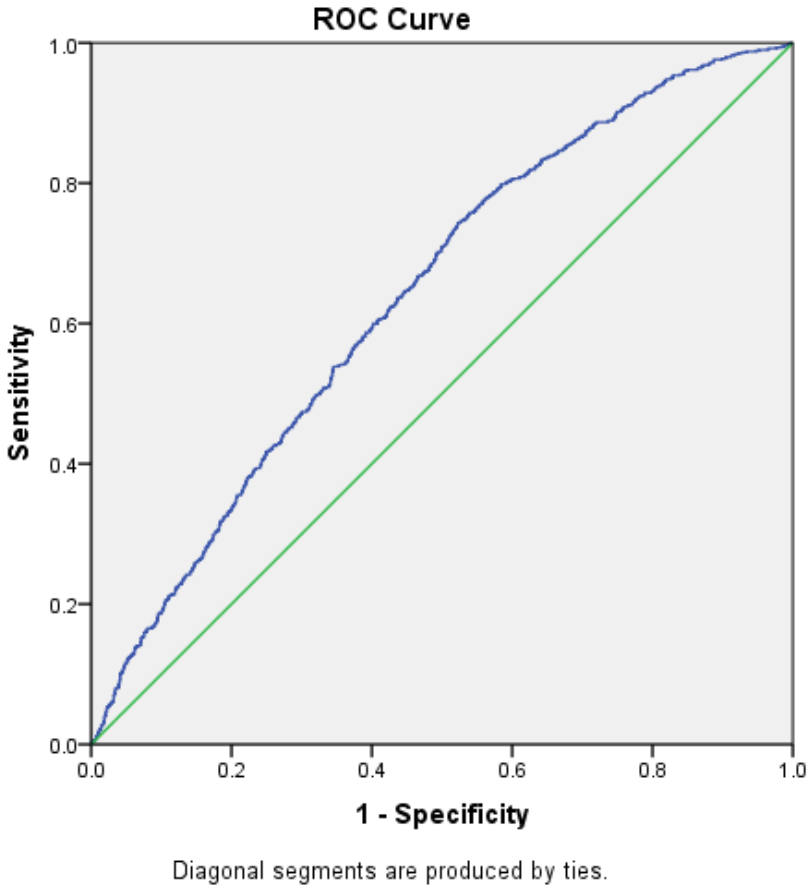
	Parameter Estimate	S.E.	P value	Odds Ratio	95% CI	
					Lower	Upper
Previous Live Birth						
No	0			Reference		
Yes	0.093953	0.063	0.1	1.10	0.97	1.24
Previous Miscarriage						
No	0			Reference		
Yes	-0.023788	0.060	0.7	0.98	0.87	1.10
AFC	0.015095	0.008	0.06	1.02	1.00	1.03
AFC (squared)	-0.000142	0.000	0.2	1.00	1.00	1.00
Duration of infertility:						
0-4 years	0			Reference		
≥5 years	-0.093313	0.066	0.2	0.91	0.80	1.04
Constant	0.811547	0.355	0.02	2.25	1.12	4.54

Predictive ability

The AUROC curve test for discriminatory ability of the final prediction model for odds of live birth was 0.62 (95% CI 0.61-0.63). In general, the *AUROC curve value* for prediction models in reproductive medicine is quite low, ranging between 0.59 and 0.64[72]. In view of this the more reliable and widely accepted measure of performance of a prediction model is the calibration following validation[72]. The ROC curve and calibration plots are displayed in Figures 22 and 23 respectively.

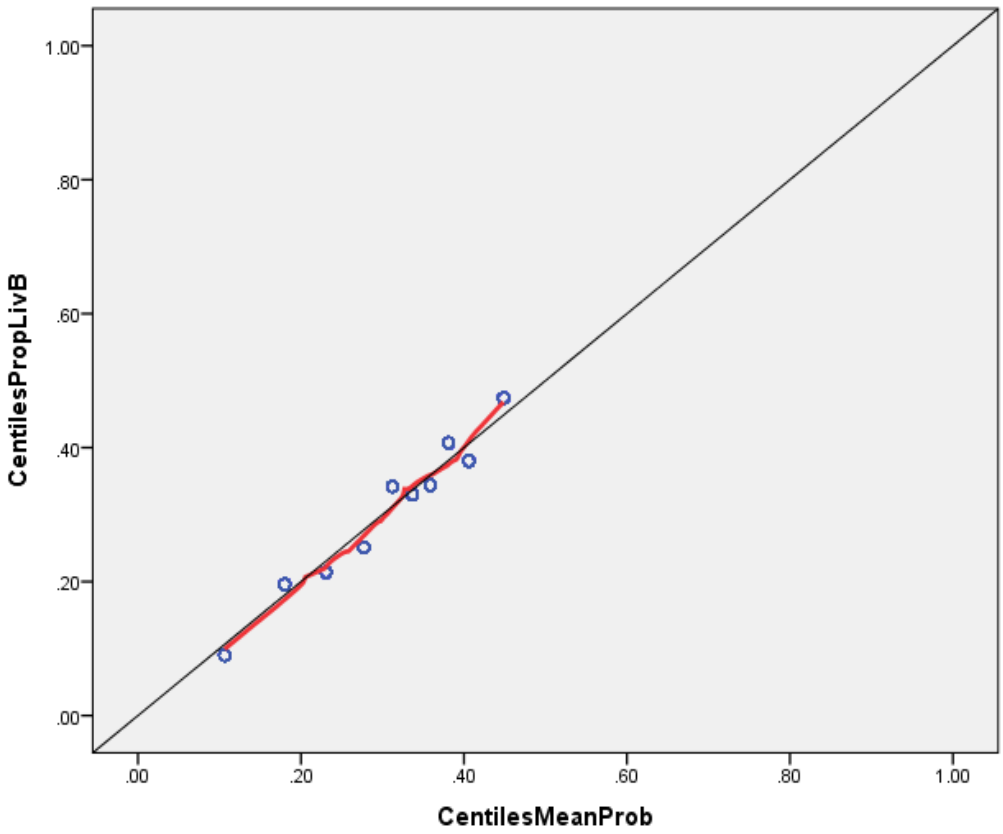
Internal validation was performed on the model built from the complete data only (n=2911 women) using the bootstrapping technique and the AUROC curve was calculated. For this model, built using complete cases only, the performance measures after internal validation were similar to the final model (see Appendix 2).

Figure 22. Receiver Operating Curve for final model



Following apparent validation, the c-statistic, or AUROC curve, for the final model was 0.62 (95% CI 0.61 - 0.63).

Figure 23. Calibration plot for final model



This figure shows the observed (actual) proportion of live births (y axis) plotted against the predicted probability of having a live birth (x axis), as predicted by our model, and split into deciles. This calibration was performed on the derivation cohort (apparent validation).

Model validation

Our external cohort consisted of 2723 patients who had undergone their first fresh assisted treatment cycle at any CARE clinic in the year of 2013.

The baseline characteristics, cycle characteristics and outcome data for the validation cohort are displayed in Table 17. The overall live birth rate for this cohort was 31.7%. The baseline characteristics of the both the derivation and validation cohorts were comparable, as were the overall live birth rates.

Table 17. Baseline characteristics, cycle characteristics and outcome data for external cohort (n=2723)

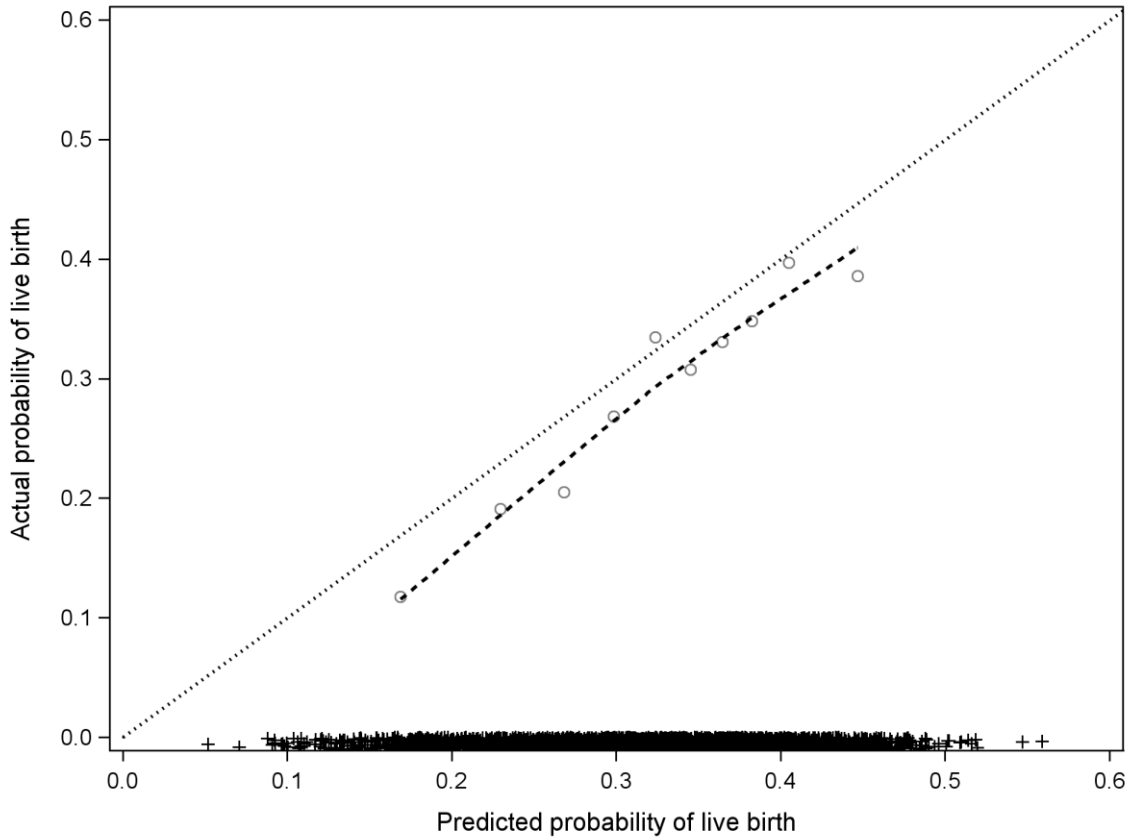
Variable	Validation cohort, n (%) *or* Mean (SD)
Age	34.3 (4.9)
Duration of infertility (in completed years)	2.6 (2.3)
BMI	24.8 (3.9)
AFC	19.7 (13.8)
AMH	17.7 (20.7)
FSH	7.0 (3.3)
Previous miscarriage	658 (24.2%)
Previous live birth	558 (20.5%)
Cause of infertility	
Tubal factor	385 (14.1%)
Anovulation	389 (14.3%)
Unexplained	773 (28.4%)
Other (E.g. Endometriosis, multiple fibroids)	806 (29.6%)
Male factor	1760 (64.6%)
Ethnicity	
White	2190 (80.4%)
Asian	262 (9.6%)
Black	31 (1.1%)
Chinese	21 (0.8%)
Other	33 (1.2%)
Not stated	61 (2.2%)
Mixed	125 (4.6%)
Cycle data	
IVF	1043 (38.3%)
ICSI	1680 (61.7%)
Number of oocytes retrieved	7.0 (6.6)
Number of mature oocytes inseminated	5.3 (5.4)
2pn	3.6 (4.0)
3pn	0.2 (0.6)
Total number of embryos	4.6 (4.1)

Variable	Validation cohort, n (%) or Mean (SD)
Number of embryos transferred	
0	375 (13.8%)
1	1501 (55.3%)
2	799 (29.3%)
3	42 (1.5%)
Number of embryos frozen	1.0 (2.1)
Outcome data	
Biochemical pregnancy	1149 (42.2%)
Clinical pregnancy	914 (33.6%)
Pregnancy outcome	
Live birth	863 (31.7%)[a]
Miscarriage	76 (8.3%)[b]
Termination	7 (0.8%)[b]
Still birth	2 (0.2%)[b]
Neonatal death	5 (0.5%)[b]

[a] Overall live birth rate for whole cohort
[b] Calculated as a % of those with a clinical pregnancy

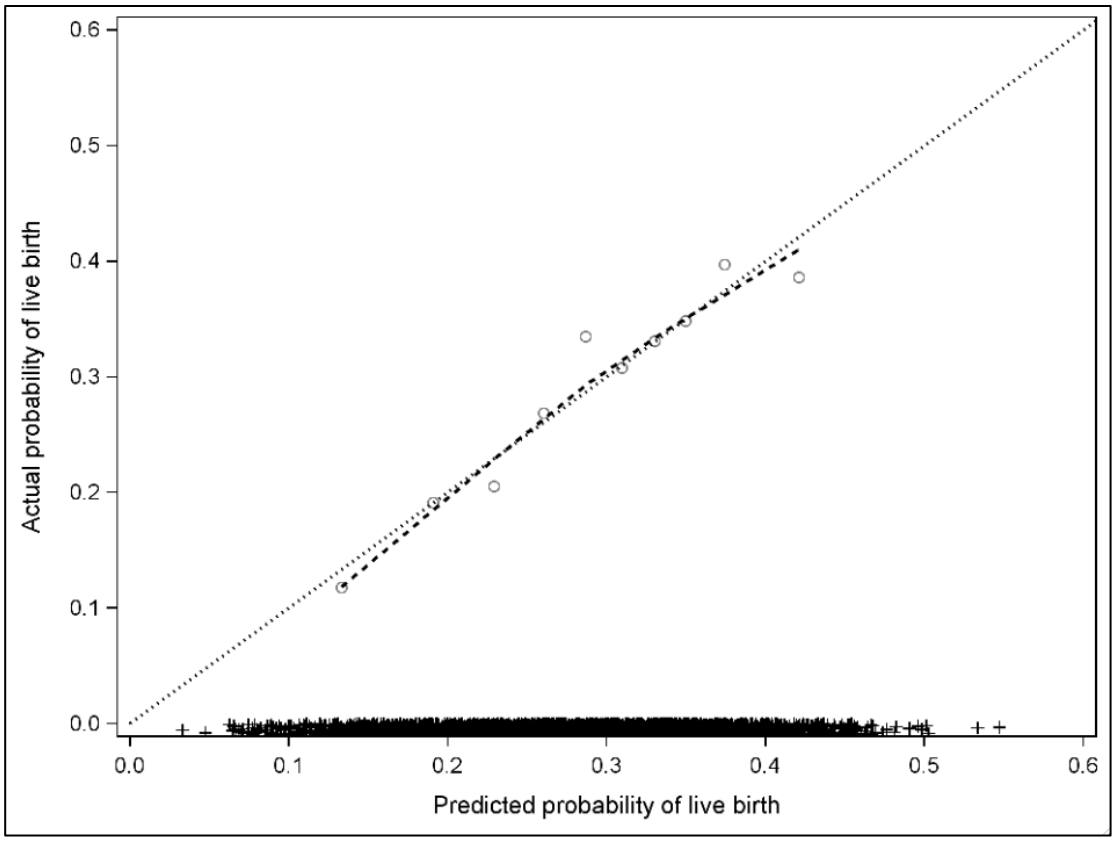
The AUROC for the final model applied to the external cohort was 0.62 (95% CI 0.60 to 0.64). Calibration-in-the-large showed a systematic over-estimation of the predicted probability of live birth (Intercept (95% CI) = -0.168 (-0.252 to -0.084), p<0.001). However, the calibration slope test was not significant (slope (95% CI) = 1.129 (0.893 to 1.365) p=0.28) meaning that the over-estimation was uniform across the range of predicted probabilities (Figure 24). Due to the calibration-in-the-large test being significant we recalibrated the final model. This was done by scaling the linear predictor from the final model, using the slope and intercept (y=-0.078 + 1.129); we then adjusted for the final model linear predictor and applied this to the external cohort. The recalibrated model is shown in Figure 25 and shows a much-improved calibration.

Figure 24. Calibration slope plot following external validation



Note: Patients were ranked into order of predicted probability of live birth and divided into tenths. The circles represent the mean risks for each tenth; the dotted line represents the perfect relationship; the dashed line represents the smooth non-parametric Loess calibration curve fitted through the circles; the plus symbols across the bottom of the graph represent the spread of patients across predicted risks.

Figure 25. Calibration plot following recalibration



Following recalibration; using the intercept and slope from the linear predictor (-0.078 and 1.129 respectively) we can see the calibration is much improved.

Discussion

Main findings

To date, successful prediction of live birth after assisted reproductive technology (ART) has been limited. We have developed a novel model, which encompasses prognostic factors that have not previously been used, such as body mass index, ovarian reserve and ethnicity. The key predictors in our model that have shown to have a significant effect on the chances of live birth are: age, tubal factor, unexplained causes of infertility and being South Asian or Black. Although BMI was shown to be significantly associated with reduced chances of live birth in the univariate analysis, this association was weakened when other confounders were accounted for.

Strengths and limitations

This is the first successfully derived and externally validated prediction model for live birth following assisted conception for women *before* undergoing their first fresh non-donor cycle of treatment. This prediction model is purposefully simple, in that its use is only for women undergoing their first fresh non-donor cycle. We believe this prediction tool holds an important role as an adjunct in the counseling process for women at the critical decision-making point in their journey, i.e. before they embark on their first treatment cycle. The advantage of using data from a first IVF cycle means that the calculated probabilities are expressed per woman/couple and not per cycle.

Our model has highlighted key predictors for IVF success, including ethnicity, which has not been used in any previous prediction models. Ethnicity has been recognised in many American papers[4–6,25,30,36] as a confounding factor in affecting IVF success and we have seen this also in the work presented in this thesis so far. There appears to be a strong association between being South Asian or Black and having a lower chance of live birth even when accounting for the other predictors in the multivariate analysis. The work in chapters 2 and 3 has attempted to explore potential reasons for this ethnic disparity in IVF success. However, no firm conclusions can be drawn, as the differences remain even when known confounders (such as age, fibroids in Black women etc.) have been adjusted for. Despite the addition of ethnicity as a novel key predictor to our model, given the large variation in ethnic groups across the globe, our model is somewhat restricted to representing the ethnic distribution and outcomes for the UK population only. It would be useful to externally validate this model on a dataset from a different country to see if ethnic variability affects the performance measures of the model. A further limitation of the inclusion of ethnicity within the model is that the group with "not stated" ethnicity constitutes more than 10% of the study population, in addition all the ethnic minority groups are smaller than this "not stated" group and so this may have influenced the data and added bias to the results.

In addition to ethnicity, no previous models have accounted for BMI or AFC. As mentioned in the results the univariate analysis for BMI and live birth outcome was statistically significant, showing that increasing BMI reduces the odds of live birth, however this association became non-significant in the final model. This could be

explained by the fact that other predictors in the model carry more weight in influencing live birth when looked at in combination, one of the strongest predictors as we would expect was female age. It appears from our data that BMI increases with increasing age and so this would explain why in multivariate analysis, where age is accounted for, the effect of BMI on live birth is not significant. In addition to this, it appears that in general Black women have higher BMI than White women, knowing that Black ethnicity is a strong predictor for lower chances of IVF success, when performing multivariate analysis this could be another reason why the association between BMI and reduced IVF success is lost.

Similarly as for BMI, the univariate analysis for AFC and live birth was significant, showing that increasing AFC is associated with a higher odds of live birth, however this became non-significant in the final model. Furthermore, we acknowledge that AFC is a subjective measure and therefore open to intra-observer variability, however it has been shown that even with this variability, its ability to predict IVF success is comparable with anti-mullerian hormone (a non-subjective measure of ovarian reserve)[73,74]. Furthermore, recording of AMH was very poor within the database and so in view of using a variable with fewer missing entries AFC was selected over AMH.

Inevitably, in any prediction model, one is unable to account for the residual confounding effect of the unavailable variables. In our prediction model we have been unable to account for confounders such as smoking status, alcohol intake or socioeconomic status. A recent systematic review and meta-analysis on predictive

factors in IVF evaluated nine predictive factors: female age, duration of subfertility, type of subfertility, indication for IVF, basal follicle stimulating hormone, fertilisation method, number of oocytes, number of embryos transferred, and embryo quality [48]. As our model is for pre-treatment counseling only, we did not include any oocyte or embryo factors. We have, however, accounted for the other mentioned factors with the exception of basal FSH, where instead we have used a more accurate ovarian reserve measure in AFC. Given the complexities of assisted conception there are many other confounders that can have an effect at different time points. For example there are prognostic factors which are only determined once a cycle has begun, such as oocyte number and embryo quality. For this reason, this model is restricted to use *prior* to starting treatment only. We appreciate that IVF success rates depend on more than the factors in this model alone. Therefore it is important for clinicians who may use the model, to ensure their patients understand that their probability of having a successful outcome will invariably change as they progress through their treatment and thus should be interpreted as a baseline prediction only.

Comparison to existing models

Using our novel model, one is able to predict the chances of live birth following IVF, and this predictive ability has been assessed by the AUROC curve. Our model is not directly comparable with that of the Templeton and Nelson models, given its inclusion of different predictors and for use at a different clinical time point. Furthermore, the Nelson model predicts live birth for different cycles, whereas our model predicts live birth in the first cycle only, before embarking on the treatment. Therefore, we felt that directly comparing the performance measures (i.e. performing statistical analyses)

was not appropriate; however we have provided a crude comparison. Following apparent validation Templeton et al found the AUROC curve to be 0.62 (95% CI 0.61-0.62) and Nelson et al 0.63 (95% CI 0.62-0.64), whilst our model showed an AUROC curve of 0.62 (95% CI 0.61-0.63). Following external validation the AUROC curve was 0.62 (95% CI 0.60 to 0.64), the recently externally validated Nelson model (IVFpredict) and Templeton model had an AUROC of 0.63 (0.62-0.63) and 0.62 (0.61-0.62) respectively[64], showing that our model has comparable discriminatory ability with these previous models. The Dutch study [63] and the more recent study by Smith et al [64] showed improvements in the performance of the Nelson and Templeton models when taking into account the effect of time trends. However, for our model there was no significant difference in live birth rates between 2008-2013 (p=0.2). Adding treatment year to our model made no difference in the performance (AUROC 0.62 95% CI 0.61-0.62) and so it was not included. A likely explanation is that both the Templeton and Nelson models were built on considerably older datasets compared with our model, pre-dating significant changes in clinical practice that occurred from 2008 onwards, therefore requiring an adjustment for time.

For IVF prediction models, calibration is deemed to be a more important measure of predictive ability than discrimination. A systematic review by Coppus et al concluded that prediction models in reproductive medicine will be limited to an AUROC of no greater than 0.65 due to the relatively homogeneous group of subfertile patients[72]. The calibration assessments for our model showed that there was a small systematic over-estimation in the predicted probabilities. After recalibration to correct for this, the calibration plot was much improved.

Clinical implications

Examples of how our novel prediction model could be utilised in clinical practice to provide couples with a personalised estimated probability of having a live birth is shown in Table 18. We have presented the predicted probabilities for both the original externally validated model and the recalibrated model; reassuringly the results show that the predicted probabilities from the recalibrated model were only slightly lower than those from the original model. An example of how probabilities (including re-calibrated probabilities) are calculated using the model follows Table 18.

In addition, as BMI is the only variable within the model that the patient is able to change, we have explored this further. Table 19 shows an example of a woman/couple from whom we would expect better than average success rates, and how the probability of success changes with altered BMI alone.

Table 18. Examples of predicted probabilities using the final model and calibrated model

Example Couples	Predicted probability of live birth	Recalibrated predicted probability of live birth
A. A 38 year old White woman and her partner have been trying to conceive for over 5 years. She has a body mass index of 35 and an antral follicle count of 14. The couple had a miscarriage in the past following a natural conception. The couple's cause of infertility is male factor infertility.	0.25	0.21
B. If we take the same couple as in A but change the body mass index to 25	0.27	0.24
C. If we take the same couple as in A but change the age to 30 and ethnicity to Black.	0.21	0.18
D. A 28 year old White woman with unexplained fertility. She has a BMI of 22 and AFC of 15. The couple have had a child after a previous natural conception and have been trying for 2 years.	0.43	0.41

Note: These examples are plausible in terms of the types of patients that are seen in IVF clinics, and they show the influence of couple characteristics. Example D shows the characteristics that result in a greater chance of success.

The calculation of the predicted probability of live birth and an example of how to use the model based on example D:

XB = 0.8115 + Previous live birth (Yes=1, No=0) x 0.0940 - Previous miscarriage (Yes=1, No = 0) x 0.0238 - BMI x 0.0109 – tubal infertility (Yes=1, No=0) x 0.2544 – anovulatory infertility (Yes=1, No=0) x 0.1387 - unexplained infertility (Yes=1, No=0) x 0.1338 – male factor infertility (Yes=1, No=0) x 0.0860 - other infertility (Yes=1, No=0) x 0.1185 – Asian ethnicity (Yes=1, No=0) x 0.1716 – Black ethnicity (Yes=1, No=0) x 0.6836 – Chinese ethnicity (Yes=1, No=0) x 0.1816 – Other Non-White ethnicity (Yes=1, No=0) x 0.3552 – No stated ethnicity (Yes=1, No=0) x 0.0055 – Mixed ethnicity (Yes=1, No=0) x 0.1929 - Age1 x -0.0356 - Age2 x 0.1061 + AFC x 0.0151 - AFC squared x -0.00014 - Duration of infertility (0-4 years=0, >5years = 1) x 0.0933

Where Age1 is the woman's age if the age is ≤36 years or is 36 years if the age is >36 years; Age2 is the difference (woman's age - 36 years) if the woman's age is >36 years and zero if ≤36 years.

Probability of live birth = EXP(XB) / (1+EXP(XB))

Example D

The highlighted parts of the model are the only non-zero terms relating to this example:

XB = **0.8115 + Previous live birth (Yes=1, No=0) x 0.0940** - Previous miscarriage (Yes=1, No = 0) x 0.0238 **- BMI x 0.0109** – tubal infertility (Yes=1, No=0) x 0.2544 – anovulatory infertility (Yes=1, No=0) x 0.1387 **- unexplained infertility (Yes=1, No=0) x 0.1338** – male factor infertility (Yes=1, No=0) x 0.0860 - other infertility (Yes=1, No=0) x 0.1185 – Asian ethnicity (Yes=1, No=0) x 0.1716 – Black ethnicity (Yes=1, No=0) x 0.6836 – Chinese ethnicity (Yes=1, No=0) x 0.1816 – Other Non-White ethnicity (Yes=1, No=0) x 0.3552 – No stated ethnicity (Yes=1, No=0) x 0.0055 – Mixed ethnicity (Yes=1, No=0) x 0.1929 **- Age1 x -0.0356 - Age2 x 0.1061 + AFC x 0.0151 - AFC squared x -0.00014** - Duration of infertility (0-4 years=0, >5years = 1) x 0.0933

XB = 0.8115 + 1 x 0.0940 - 22 x 0.0109 - 1 x 0.1338 - 28 x 0.0356 + 15 x 0.0151 - 225 x -0.00014

XB = -0.2696

Predicted probability = EXP(XB)/(1+EXP(XB))

= 0.7637 / 1.7637 = **0.433 = 43.3%**

Recalibrated probability:

Recalibrated XB = -0.2696*1.129 – 0.078 = -0.3824

Recalibrated Pred Prob = EXP(-0.3824)/(1+EXP(-0.3824))

= 0.68222/1.68222 = **0.405 = 40.5**

Table 19. The effect of BMI on predicted probabilities within the model

Example Couples	Predicted probability of live birth	Recalibrated predicted probability of live birth
A. A 28 year old white woman and her partner have male factor infertility. She has a BMI of 20 and AFC of 15. The couple have had a child after a previous natural conception and have been trying for 2 years.	0.45	0.42
B. If we take the same couple as in A but change the body mass index to 25	0.44	0.41
C. If we take the same couple as in A but change the body mass index to 30.	0.42	0.39
D. If we take the same couple as in A but change the body mass index to 35.	0.41	0.38
E. If we take the same couple as in A but change the body mass index to 40.	0.39	0.36

As discussed previously, BMI has less of an effect on IVF outcome when other predictors are accounted for. However, we have shown in Table 19 that for a woman who has a good chance of success with a normal BMI, there is still between a 4-6%

reduced chance of success if this woman had a BMI above 35; keeping all other factors the same.

We have illustrated not only the clinical use of this model but also how a couples characteristics influence their prognosis. This model provides a personalised prediction of a couple's chances of IVF success, in favour of using crude age related success rates based on national HFEA data. The idea would be for clinicians to use the model routinely when counselling couples seen in outpatient clinics for the first time, as the vast majority of UK hospital clinics will have computers with internet access. This should ensure that all patients have the opportunity to use the model at some point, which is particularly important for those patients who may have limited access to the internet or a mobile phone.

Future research

The next step for our model will be to further externally validate by performing geographical validation. We plan to do this using the data collected from the Birmingham Women's Hospital Fertility Centre, as well as other assisted conception units. The subsequent use of our model will be implementation into clinical practice as an up-to-date counseling tool (in the form of a user-friendly freely available web-page and/or mobile application) for use by the relevant clinicians and patients in aiding decision making *before* commencing their first IVF cycle.

The third and final stage in the pathway of producing a clinical prediction model is impact analysis. This establishes whether the prediction model improves decisions,

in terms of quality or cost-effectiveness of patient care[69,70]. None of the existing IVF prediction models have yet reached the impact analysis stage. We intend to evaluate the impact of our model by conducting a feasibility study to explore patient experience of the tool and its impact on their counselling and decision-making.

Collaboration statement:

This work was done as part of a collaboration between the University of Birmingham and the University of Aberdeen. Statistical assistance was provided by Dr David McLernon (post-doctoral researcher in medical statistics). Dr McLernon provided hands-on teaching and performed the external calibration using SAS (ver.9.2; SAS Institute, Cary, North Carolina).

CHAPTER 5

National study of prevalence of thyroid dysfunction in women of reproductive age

Introduction

Thyroid disorders are one of the most prevalent of all medical conditions, especially in women of reproductive age. Thyroid disease comprises a spectrum of disorders, broadly speaking this can be categorised into three main groups; abnormal thyroid function, thyroid autoimmunity and thyroid tumours (benign or malignant). The focus of the work in this chapter will be on thyroid dysfunction. Thyroid dysfunction involves an abnormality of hormone levels (i.e. free thyroxine (fT4), thyroid stimulating hormone (TSH) or free tri-iodothyronine (fT3)). The two most common problems with the thyroid are; overactive thyroid (hyperthyroidism) and underactive thyroid (hypothyroidism).

All types of hyperthyroidism are due to an overproduction of thyroid hormones (in particular an increase of fT4); active hyperthyroidism is present in around 0.1-0.4% of pregnant women. The most common cause of hyperthyroidism is Graves disease[75] (an autoimmune condition) but the condition can develop in several ways. It has been shown that Graves disease is associated with adverse pregnancy outcomes such as miscarriage, pre-eclampsia, pre-term birth, placental abruption and foetal hyperthyroidism[76,77]. Other causes of hyperthyroidism include; toxic adenomas, subacute thyroiditis, pituitary gland malfunctions or growths in the thyroid gland.

Hypothyroidism, in contrast, stems from an underproduction of thyroid hormones. The prevalence of overt hypothyroidism is around 0.5-0.7% in women of reproductive age[78]. In women of reproductive age the most common cause of hypothyroidism is

an autoimmune thyroiditis and Hashimoto's disease[15] (excluding iodine deficient populations). As with hyperthyroidism, hypothyroidism has also been associated with adverse pregnancy outcomes such as miscarriage, placental abruption, higher rates of neonates being admitted to intensive care units and lower intelligence scores (IQ) in the offspring[31].

Subclinical hyperthyroidism and subclinical hypothyroidism (SCH) are diagnoses that are essentially based on laboratory reference ranges; the majority of patients do not have clinical signs or symptoms. Subclinical hyperthyroidism is defined as a decrease in serum TSH below the reference range with normal serum fT4 and fT3 concentrations; subclinical hypothyroidism (SCH) is defined as an elevation in serum thyroid-stimulating hormone (TSH) above the upper limit of the reference range with normal serum fT4 concentration. The "normal reference range" is dependent on the reference range for the assay used. It is known that these conditions represent the earliest stages of thyroid dysfunction, more so for subclinical hypothyroidism and it's progression to overt disease[79]; however the benefits of detecting and treating subclinical thyroid disease are not well established.

Subclinical hyperthyroidism has been found to be linked to several conditions such as atrial fibrillation, reduced bone mineral density, cardiac dysfunction, and progression to overt hyperthyroidism[80–83]; however the association with progression to overt disease is much less than that seen with hypothyroidism[84]. Although it has been shown that that treatment of subclinical hyperthyroidism can slow the loss of bone mineral density there is no strong evidence to show any benefits from treating it

in pregnancy. Due to evidence showing that there are no adverse effects associated with subclinical hyperthyroidism in pregnancy it is not screened for or treated pre-conception or in pregnancy[85].

The prevalence of subclinical hypothyroidism (SCH) is thought to be around 9% in all adults (including non-pregnant women); the prevalence is around 3-5% in pregnant women[15]. This figure is higher in women with infertility[15]. In contrast to subclinical hyperthyroidism, evidence has shown that subclinical hypothyroidism is linked to negative pregnancy outcomes such as miscarriage, pre-term birth, pre-eclampsia, gestational hypertension and peri-natal mortality[17,86]. Consequently, the Endocrine Society Clinical Practice Guideline (ESCPG) regarding "Management of Thyroid Dysfunction during Pregnancy and post-partum" recommends the use of hormone replacement therapy, in the form of Levothyroxine (LT4) treatment, for pregnant women with subclinical hypothyroidism as well as those with overt disease. The guideline reports that there is "fair" evidence on improvements with LT4 replacement in SCH for most pregnancy outcomes, although no differences have been seen for neurological outcomes in the offspring[15].

One of the main challenges in defining SCH is agreeing on the upper limit of TSH, consequently this has also lead to debate over when to treat SCH. The most widely accepted reference range for a "normal" TSH in non-pregnant women is 0.4-4.5mU/l. For women who have known overt hypothyroidism (TSH >4.5mU/l and fT4 <21mU/l), and are already on LT4 treatment pre-conception, it is generically accepted that a target TSH of 0.1-2.5mU/l, 0.2-3.0mU/l and 0.3-3.0mU/l should be maintained in the

first, second and third trimester respectively (in the absence of laboratory specific reference ranges). These recommended reference ranges are taken from the American Endocrine Society (AES)[87] and American Thyroid Association (ATA)[16] guidelines and are endorsed by most other international endocrine societies, including the British Thyroid Association[88].

There is much less clarity and consensus on when to treat SCH; both pre-conception and in pregnancy. The British Thyroid Association, recommends the following: *"If the serum TSH concentration is above the reference range but <10mU/L, then serum thyroid peroxidase (TPO) antibodies should be measured. If the serum antibody concentration is high, then serum TSH should be measured annually or earlier if symptoms develop; thyroxine (LT4) therapy should be started if the serum TSH concentration rises above 10mU/L. If the serum antibody concentration is not raised, then repeat measurement of serum TSH approximately every three years is all that is required. There is no evidence to support the benefit of routine early treatment with thyroxine in non-pregnant patients with a serum TSH above the reference range but <10mU/L. Physicians may wish to consider the suitability of a therapeutic trial of thyroxine on an individual patient basis"* [88].

In contrast the recent ATA and AACE (American Association of Clinical Endocrinologists) guideline states: *"treatment with L-thyroxine should be considered in women of childbearing age with serum TSH levels between 2.5mU/l and the upper limit of normal or a given laboratory's reference range if they are in the first trimester of pregnancy or planning a pregnancy, including assisted reproduction in the*

immediate future" [89]. However, the American Congress of Obstetricians and Gynaecologists have not supported these recommendations.

This debate regarding pre-conception thresholds of TSH is of particular growing interest in the assisted reproduction setting. In the United States recent guidelines proposed by the National Association of Clinical Biochemistry (NACB) have stated that it is likely that in the future the upper limit serum TSH euthyroid reference range will be reduced to 2.5mU/L; but that this will be for all adults including pre-conception women in the fertility setting. However, in the UK the British Thyroid Association maintains that a TSH level up to 4.5mU/L is normal.

As mentioned, there is currently a shift in clinical practice, in certain parts of the world, towards routinely treating women with SCH who are trying for a pregnancy, particularly the infertility population, and aiming for a pre-conception target TSH of <2.5mU/l. Despite this suggestion of aiming for a rigid pre-conception TSH threshold below 2.5mU/l, it is not currently recommended routine practice for universal thyroid function screening in women trying for a pregnancy, or even in women who are pregnant, by any of the major endocrine societies (ATA, AACE, ACOG or the Endocrine society). This is because there is limited conclusive evidence to suggest that treatment with Levothyroxine for women with TSH 2.5-4.5mU/l has any benefit on pregnancy outcomes compared with untreated women[18]. However, it should be noted that other endocrine societies, including the Spanish society of endocrinology and nutrition (SEEN), argue that universal screening is warranted based on the impact of undiagnosed thyroid dysfunction on pregnancy. The 2012 Endocrine

society guidelines reflected the lack of agreement over universal versus "case finding" screening, by producing two different recommendations as the committee could not reach a unanimous decision[87]. Furthermore, in 2011 the AES/ATA recommendations for thyroid function screening was that it should be universal, but this was retracted in the 2012 guideline; this was almost certainly in part to do with the results of the Antenatal Thyroid Screening and Childhood cognitive function (CATS) study[90] published in the New England Journal of Medicine which showed no difference in cognitive function of offspring born to women with treated hypothyroidism compared with untreated.

There are certain "high risk" groups whom most major international bodies agree should be considered for routine thyroid screening, these are; 1) women with a history of risk factors for thyroid dysfunction (i.e. cervical irradiation, 1st degree family history), 2) women with poor obstetric history which could be attributable to thyroid dysfunction (i.e. recurrent miscarriage, previous still birth, baby with congenital defects) and 3) women with a history of metabolic disorders (i.e. diabetes). The AES/ATA guideline also includes women with infertility as a "high risk" group of people who should be screened routinely prior to pregnancy[16]; however this is not endorsed by British guidelines. Outside of these agreed "high risk" groups there is little known about the benefit to screening asymptomatic women with potentially undiagnosed subclinical disease.

One of the largest epidemiological studies of thyroid disease, which assessed thyroid function and thyroid autoimmunity in over 17,000 people was conducted over

20years ago and was restricted to the geographic and ethnic distribution of the U.S. population, furthermore it did not look specifically at women of reproductive age[91]. Another large study was the Whickham study in the UK, this survey looked at the prevalence of thyroid disorders in a randomly selected sample of 2779 adults, which represented the population of Great Britain in age, sex and social class[92]. To date there have been no large-scale epidemiological studies in the UK identifying the prevalence of undiagnosed thyroid disease in women of reproductive age who are actively trying for a pregnancy. Given the effects of thyroid disease in pregnancy this is an important area of research to establish.

Women with symptoms of thyroid disease will generally present in primary care and thus

be investigated by their general practitioner (GP), however many women with overt hypothyroidism and almost all with SCH are asymptomatic; and so there is a cohort of undiagnosed women with thyroid disease that will not routinely have their thyroid function tested prior to conception. It is therefore important to investigate if there may be a specific subset of women with undiagnosed thyroid disease who are likely to be at higher risk and in whom we could provide screening and/or treatment to optimise their pregnancy outcomes.

As increasing numbers of UK clinicians (endocrinologists, fertility specialists and obstetricians) are moving towards empirically treating subclinical thyroid disease, and adhering to stricter thresholds for reference ranges, there is a need to establish the

prevalence of undiagnosed thyroid disease, using these different thresholds, and identify which women will benefit most from pre-conception screening and subsequent treatment.

Aims and objectives

1. To study the distribution of undiagnosed thyroid dysfunction in women of reproductive age in the UK.
2. To study the distribution of undiagnosed thyroid dysfunction in demographic subgroups; age, body-mass index, ethnicity and originating population (i.e. history of 1 or 2 miscarriages (early pregnancy unit setting); recurrent miscarriage; infertility).
3. To examine the relationship between demographic characteristics (age, BMI, ethnicity and population) and TSH concentration.

Methods

This was a large prospective epidemiological prevalence study conducted at 42 hospitals across the UK between December 2013 and February 2015.

This study directly linked to an on-going large multi-centre randomised controlled trial called TABLET. The aim of the TABLET trial is to determine if 50micrograms of LT4, started pre-conceptually in women who are euthyroid and positive for TPO antibodies (TPOab), can reduce the risk of miscarriage and pre-term birth. All women gave

written consent to have their blood taken (for thyroid function and TPOab) and used for research purposes, prior to their blood samples being taken. The consent form can be found in Appendix 3. Pre-screening logs were completed for all women who were approached and asked to take part.

Eligibility criteria

The eligibility criteria were as follows:

- Aged between 16-41
- Trying for a pregnancy in the next 12 months
- Not known to have any history of or current thyroid problems
- Not known to have any cardiac problems
- Not taking Amiodarone or Lithium

All patients were screened from any one of the following clinical settings:

- Early pregnancy assessment unit (EPAU) – women with a recent diagnosis of miscarriage, including women who may have had one miscarriage previously.
- Recurrent miscarriage clinics - recurrent miscarriage defined as three consecutive losses with the same partner.
- Infertility clinics - women being investigated or treated for infertility problems.
- Other – these were women who fell into any of the above categories but had not been routinely seen in the clinical setting and instead had contacted the trial as self-referrals via the trial website. (http://www.birmingham.ac.uk/research/activity/mds/trials/bctu/trials/womens/tablet/index.aspx)

Thyroid function test reference ranges

The reference range for a normal thyroid function for the purpose of the TABLET trial was set according to Roche Diagnostics manufacturer's recommended reference ranges published in 2004, which was derived from 269 healthy non-pregnant females of the reproductive age (20-39years). This range was 0.44mU/L – 3.63mU/L with a Free T4 of 10.0 – 21.0 pmol/L.

For the purpose of the analysis for this study, to be consistent with the generic accepted upper limit for TSH in the UK, we used an upper limit of 4.5mU/L for TSH. Values above this were considered abnormal. A subgroup analysis was done to look at those with TSH 2.5-4.5mU/L. This was done as reproductive medicine clinicians in the UK and endocrinologists in the United States, are moving towards a pre-conception TSH of <2.5mU/L for women undergoing fertility treatment based on the hypothesis that these women could have better success rates, however evidence for this is very limited.

Thyroid function was grouped into mutually exclusive groups with the exception of the euthyroid group, which was further sub-grouped into euthyroid 1a and 1b to explore differences in prevalence using a stricter threshold for normal compared with the traditional upper end of normal. Subclinical hypothyroidism was also split into two groups; moderate (TSH 4.51-10mU/L) and severe (>10mU/L). The rationale for creating two separate groups was because they are treated differently; it is widely accepted to treat subclinical hypothyroidism when the TSH is greater than 10mU/L, however there is disagreement over TSH values between 4.50-10mU/L.

The groups were as follows:

- Euthyroid (1): TSH 0.44-4.5mU/L and fT4 10-21pmol/L
    - Euthyroid (1a): TSH 0.44-2.49mU/L and fT4 10-21pmol/L
    - Euthyroid (1b): TSH 2.50-4.50mU/L and fT4 10-21pmol/L
- Overt hypothyroidism: TSH >4.50mU/L and fT4 <10pmol/L
- Overt hyperthyroidism: TSH <0.44mU/L and fT4 >21pmol/L
- Moderate subclinical hypothyroidism (SCH): TSH 4.51-10mU/L and fT4 10-21pmol/L
- Severe subclinical hypothyroidism (SCH): TSH >10mU/L and fT4 10-21pmol/L
- Subclinical hyperthyroidism: TSH <0.44mU/L and fT4 10-21pmol/L

Demographic data

The following demographic characteristics were recorded for each screened patient: age, body-mass index (BMI), ethnicity and originating population.

Age and BMI were grouped in categories for the purpose of presenting the prevalence data. Ethnicity was selected from a list of 17 options, for the purpose of analysis these were grouped accordingly: “White”; “South Asian” (Indian/Pakistani/Bangladeshi/Other South Asian); “Black” (African/Caribbean/Other Black); “Mixed” (mixed White/Asian, mixed White/Black African, mixed White/Black Caribbean, other mixed background); “Chinese” and “Other” ethnic group. Originating population referred to the clinical setting from where the patients was screened, as previously stated this was; women with a history of one or two

miscarriages (i.e. EPAU setting), women with history of recurrent miscarriage, women seen in the fertility setting or other.

Sample size

A sample size of 5000 would be large enough to have small uncertainty around our point estimate; our 95% confidence interval would be of width 1.5% (in absolute terms). Increase to the sample would reduce uncertainty further but 5000 was the minimum we aimed for in this study. We anticipated we would achieve this minimum number over a 14month period based on individual centre capacities within the relevant clinical settings, research staff numbers and our experience of recruitment figures into the TABLET trial thus far.

Statistical analyses

Using the binomial exact method (http://statpages.org/confint.html), crude proportions and percentages with 95% confidence intervals were calculated to show the prevalence of each thyroid dysfunction group. The prevalence's for each thyroid function group were further explored by examining the prevalence within each demographic subgroup: age, BMI, ethnicity and originating population. Age and BMI were created as categorical variables.

Further analysis was performed, presented in graphical format, looking at the relationship between each demographic outcome and TSH concentration to assess for any trends. The R^2 value was calculated to determine how well each demographic factor was at predicting TSH concentration; with a value near to 1

representing a perfect fit, and a value near to 0 as near to no fit. Age and BMI were treated as continuous variables, for ethnicity and population box plots were created showing the median TSH concentration, along with the interquartile range and range, for each ethnic group and originating population.

Ethical approval

Ethical approval was obtained from Berkshire B Research Ethics Committee (REC) in December 2013. The study protocol and letter of approval can be found in Appendix 4 and Appendix 5 respectively.

Results

A total of 7022 women had thyroid function testing across the UK between December 2013 and the end of February 2015.

Geographical distribution of population

A list of the 42 centres is presented geographically in Figure 26. The distribution of patients in the study, classified into groups of thyroid dysfunction, is shown in Figure 27 and the numbers of women screened at each site is presented in Figure 28.

The pre-screening logs did not show any obvious disparities in age, BMI or ethnicity between those who gave consent and those who did not. The most common reason for declining consent was that the patient stated they would prefer not to know; this

contributed to less than 0.5% of all women approached, we therefore feel satisfied that the cohort is representative of women seen in the corresponding clinical settings.

Figure 26. Geographical distribution of recruiting centres

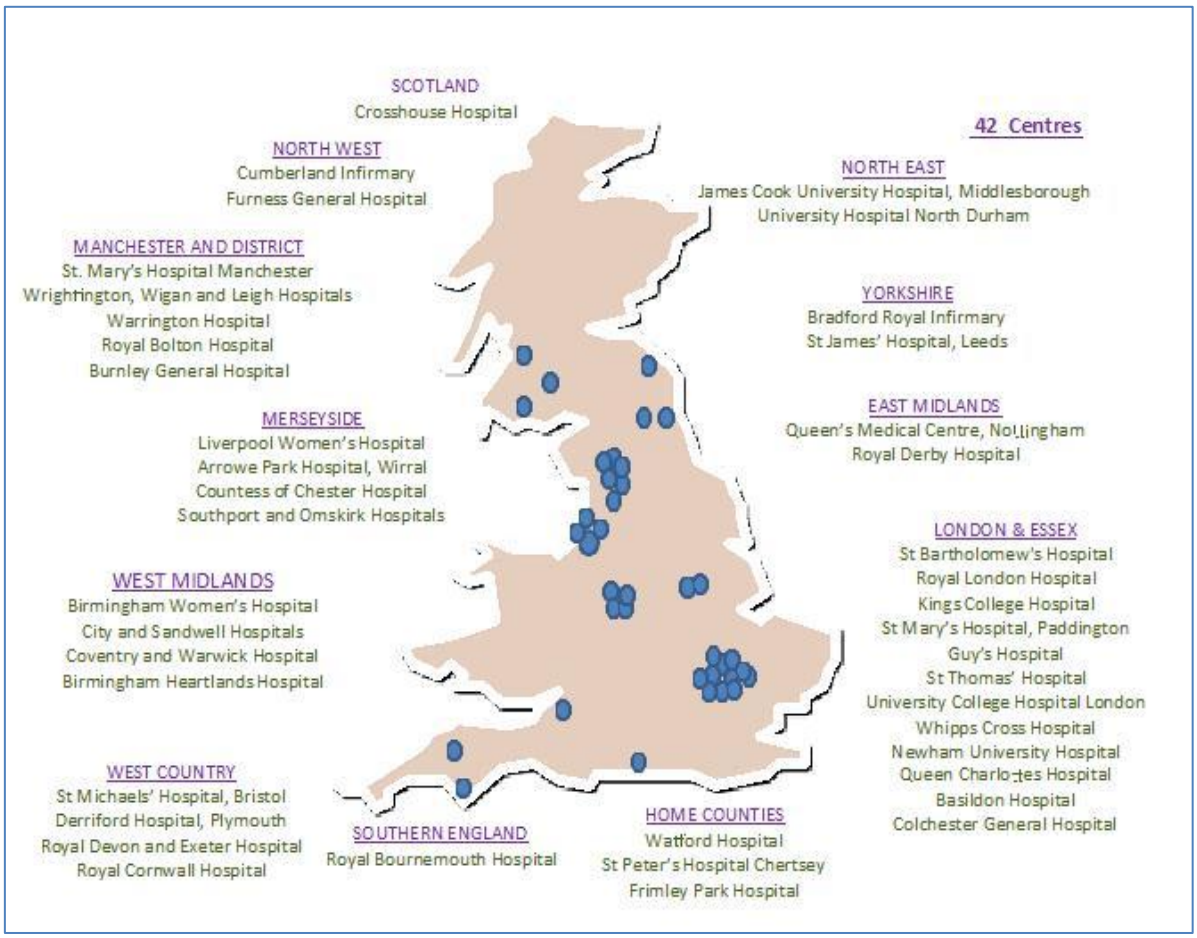


Figure 27. Flowchart of patients in prevalence study

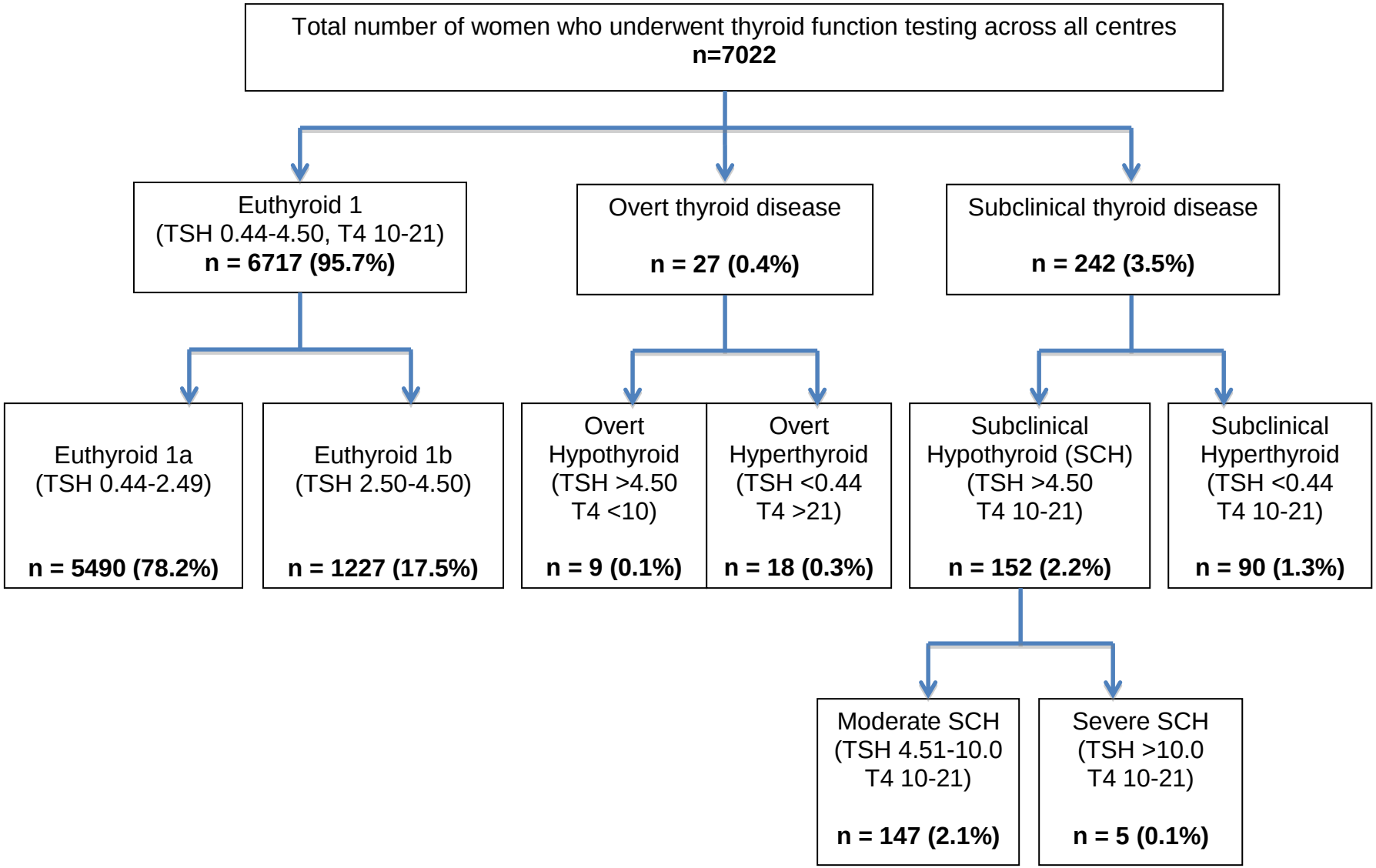
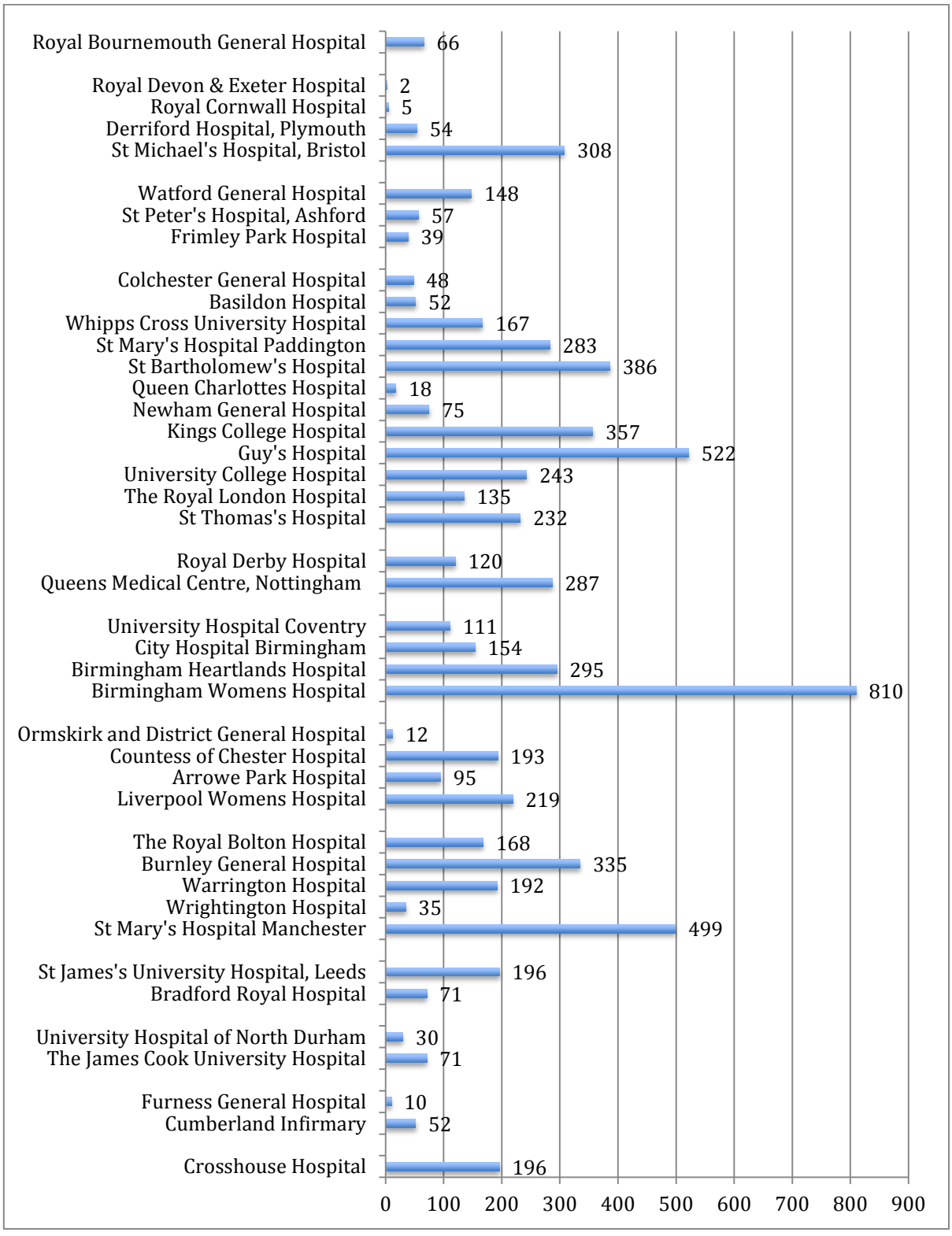


Figure 28. Numbers screened by each participating centre; grouped by geographical region



Overall description of data for each demographic subgroup

Table 20. Overall description of dataset (n=7022) - delete

Demographic Subgroup	Number and % of whole population
Age (n=7022)	
16-21	181 (2.58%)
22-26	888 (12.64%)
27-31	2016 (28.71%)
32-36	2515 (35.82%)
37-41	1422 (20.25%)
BMI (n=6325)	
<18.5	132 (1.88%)
18.5-24.9	3117 (44.39%)
25-29.9	1862 (26.52%)
30-34.9	769 (10.95%)
≥35	445 (6.34%)
Ethnicity (n=7022)	
White	4863 (69.25%)
Black	506 (7.21%)
South Asian	1236 (17.60%)
Chinese	96 (1.37%)
Mixed	146 (2.08%)
Other	175 (2.49%)
Population (n=6898)	
History of 1 or 2 miscarriages (EPAU)	2231 (31.77%)
Infertility	3171 (45.16%)
Recurrent miscarriage	1419 (20.21%)
Other	77 (1.10%)

Overall prevalence of each thyroid function group for whole population

The raw numbers are presented along with percentages and 95% confidence intervals in Table 21 below.

Table 21. Overall prevalence's for whole dataset

Thyroid function	Proportion	Prevalence	95% CI	
			Lower	Upper
Euthyroid				
Euthyroid 1 (TSH 0.44-4.50)	6717/7022	95.66%	95.15%	96.12%
Euthyroid 1a (TSH 0.44-2.49)	5490/7022	78.18%	77.20%	79.14%
Euthyroid 1b (TSH 2.50-4.50)	1227/7022	17.47%	16.59%	18.38%
Overt thyroid disease	27/7022	0.38%	0.25%	0.56%
Hypothyroid	9/7022	0.13%	0.06%	0.24%
Hyperthyroid	18/7022	0.26%	0.15%	0.41%
Subclinical hypothyroid				
Mod. SCH (TSH 4.51-10.0)	147/7022	2.09%	1.77%	2.46%
Severe SCH (TSH >10.0)	5/7022	0.07%	0.02%	0.17%
TSH >4.50	152/7022	2.16%	1.84%	2.53%
TSH >2.50	1379/7022	19.64%	18.72%	20.59%
Subclinical hyperthyroid	90/7022	1.28%	1.03%	1.57%

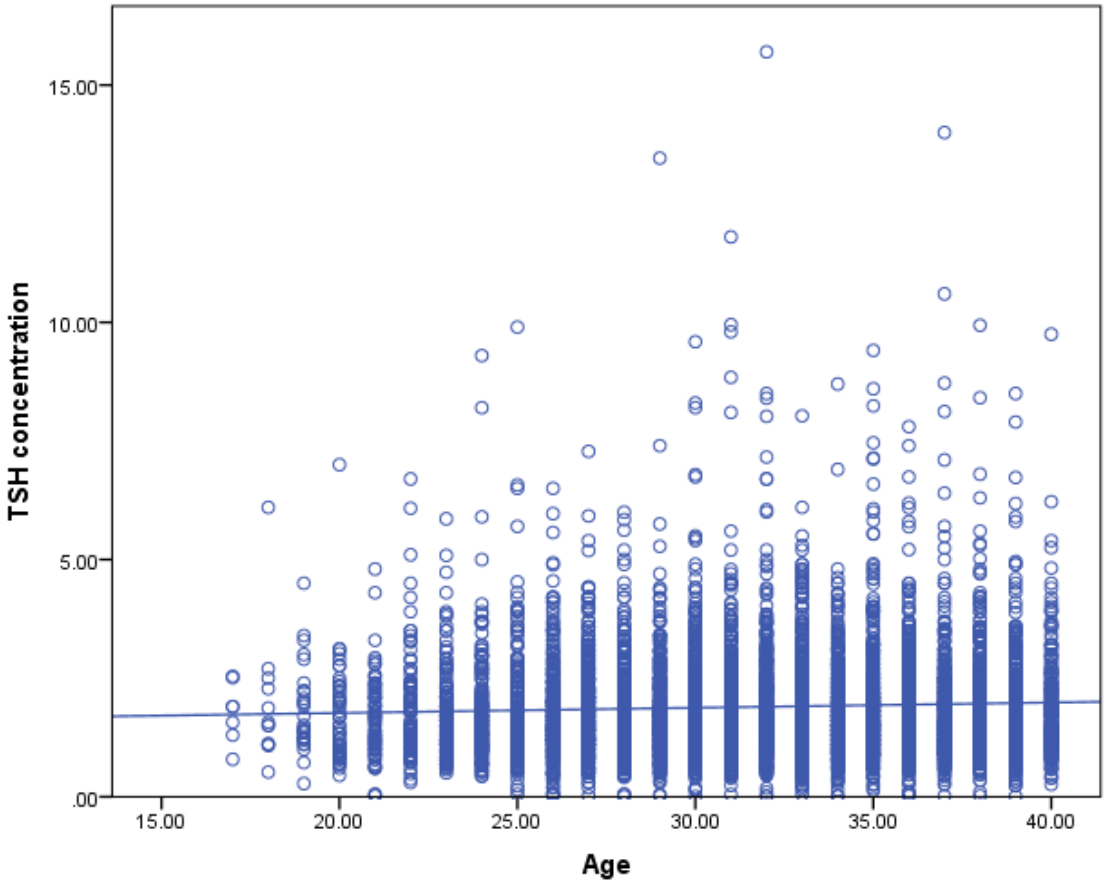
Prevalence of each thyroid dysfunction group subgrouped into age, BMI, ethnicity and originating population

For each main group of thyroid function listed in Table 21, the prevalence's were explored further and broken down into demographic subgroups, each presented as raw proportions, percentages and corresponding 95% confidence intervals. For BMI, ethnicity and originating population these fields were not compulsory to be completed within the study electronic database and so there are missing cases; the total cases used to present the prevalence for each demographic characteristic is shown with "n" values in each table. The prevalences are displayed in tables in appendices 6-13.

Relationship between TSH concentration and each demographic subgroup

These are displayed in graphical format in Figures 29-32.

Figure 29. The relationship between age and TSH concentration

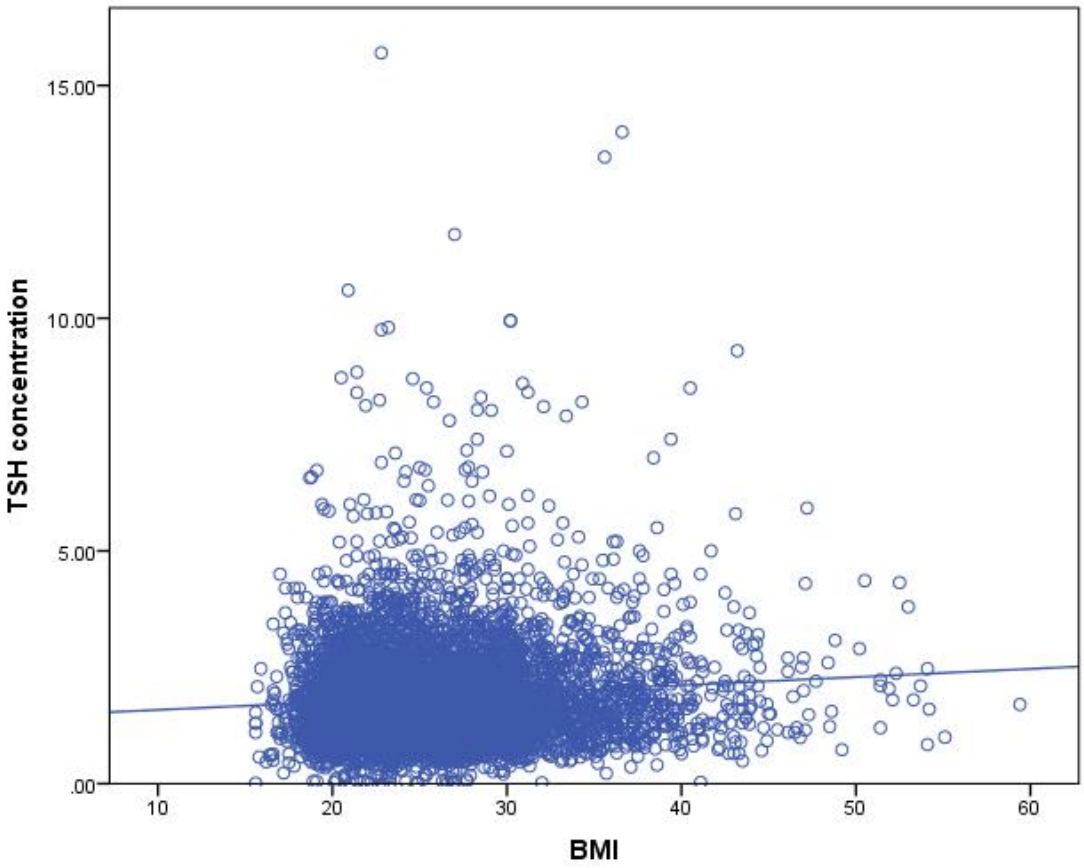


Equation for line of best fit: y = 1.55 + 0.01*x

R^2 linear = 0.0006636

The graph shows there appears to be no significant relationship between age and TSH concentration.

Figure 30. The relationship between BMI and TSH concentration



Equation for line of best fit: y = 1.41 + 0.02*x

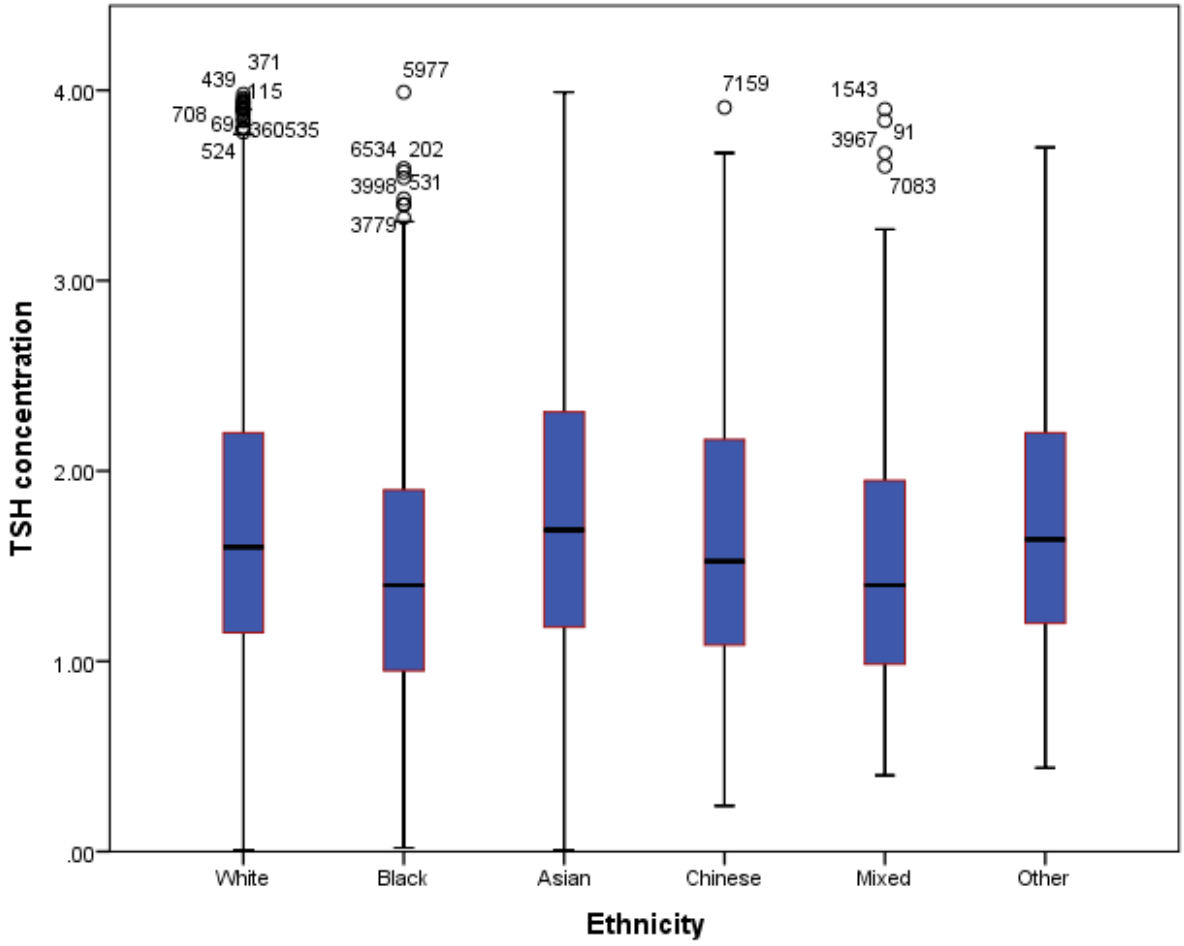
R^2 linear = 0.004

The graph shows there appears to be a relationship between increasing BMI and higher TSH concentration. This is consistent with the analyses done thus far. However the R^2 value is still low suggesting that BMI is not a strong predictor for TSH concentration.

For women who had moderate SCH (as seen in Appendix 11), 1.51% of them had a normal BMI (18.5-24.9), there was a significantly higher prevalence of women with

BMI 25-29.9 (2.31%; p=0.04), 30-34.9 (2.73%; p=0.02) and ≥35 (2.92%; p=0.03). The same pattern was found for severe SCH. In addition the mean BMI for overt hypothyroidism (28.0 SD ±7.7) was significantly higher than the mean BMI for the corresponding euthyroid women (25.9 SD ±5.4) (p<0.001). The reverse findings were found for hyperthyroidism (subclinical and overt).

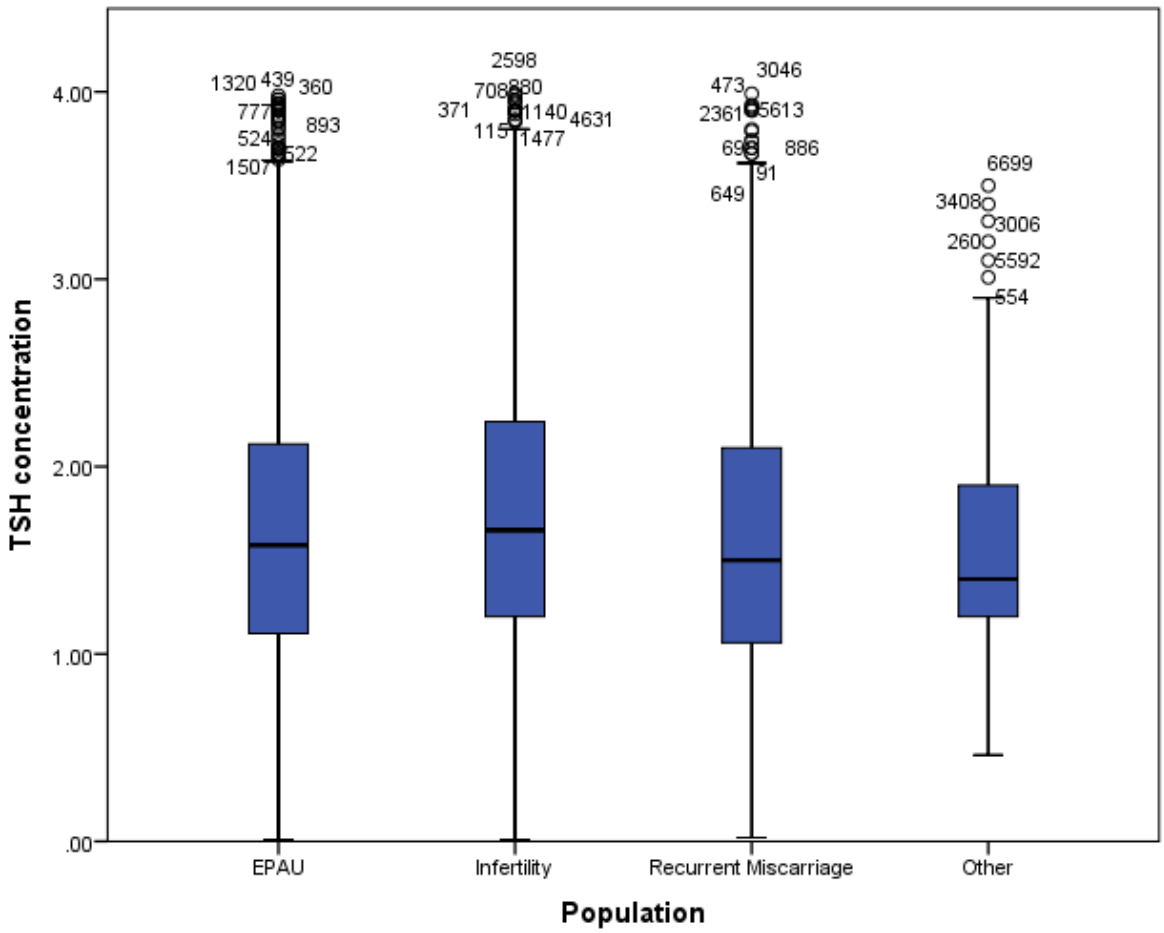
Figure 31. The relationship between ethnicity and TSH concentration



The numbers on the graph represent the case numbers for outliers, as far as possible the extreme outliers have been removed, purely for graphical presentation

purposes, all values were included in the actual analysis. The boxplot shows the median, interquartile range and range; South Asian women appear to have the highest levels of TSH for all parameters and Black women have the lowest.

Figure 32. The relationship between originating population and TSH concentration



The infertility population appear to have the highest median TSH concentration and higher limits for the interquartile range and range (upper end and lower end).

Discussion

Main findings

The aim of this study was to present the prevalence of thyroid function (in specific categories) across the UK in pre-conception women not known to have thyroid disease who are actively trying for a pregnancy.

The overall prevalence of the generically accepted reference range for biochemical euthyroidism (TSH 0.44-4.50, T4 10-21) was 95.66% (95% CI 95.15-96.12%), for overt thyroid disease it was 0.38% (95% CI 0.25-0.56%) and for subclinical thyroid disease it was 3.45% (95% CI 3.03-3.90%). This study shows that there are a small proportion of women who have undiagnosed thyroid disease, and prompts the question of whether women should routinely have screening performed pre-conception to avoid missing these women and improving pregnancy outcomes.

Strengths and limitations

One of the strengths of this study is the large sample size. By having a large sample size it has allowed us to report the prevalence's with greater precision. Another strength is the fact that the population is a good geographic representation of the whole of the UK, rather than being limited to one area.

A further strength was the implementation of pre-screening logs. These ensured that entries were made for all patients approached for screening, including those who declined. When looking at the baseline demographics for those women who declined

there were no significant differences between the demographics for those who gave consent; consequently we can be reassured that the sample of women screened was representative of the women seen in the selected clinical settings.

By collecting demographic data on age, body-mass index, ethnicity and originating population it has allowed us to explore in detail which group of women may be more likely to have thyroid disease. However, the reality is that all the women who were screened belonged to a selected population; whether it was history of miscarriage or infertility. Therefore these women would not necessarily represent the thyroid function of true unselected “low risk” women with no gynaecological or obstetric risk factors. It would be interesting to see comparisons between the women screened from our study and the “normal” or “low risk” population of women trying for a pregnancy with no known history of gynaecological/obstetric or medical problems.

A further limitation of the study is that there are no outcome data to allow us to investigate the relationship between thyroid dysfunction and pregnancy outcomes using a large sample size. This would be particularly useful in examining any potential differences in outcomes for women with TSH 0.44-2.50mU/L and those with TSH 2.50-4.50mU/L. We obtained ethical approval to use the outcome data from one centre (Birmingham Womens Hospital); a cohort study looking at the effect of subclinical hypothyroidism on IVF outcome is presented in chapter 7.

Comparisons to existing literature

The overall prevalence of thyroid disease in our study was very similar to the reported prevalence in the National Health and Nutrition Examination Survey (NHANES III) in the US; which studied an unselected population of over 17,000 people aged 12 or above, between 1988 and 1994[91]. The US study used the upper limit of normal for TSH as 4.5mU/L, the reported prevalence's were 4.3% for subclinical thyroid disease and 0.3% for overt[91] (our study reported 3.5% and 0.4% respectively).

Iodine deficiency is the leading cause of hypothyroidism in developing countries, due to insufficient iodine dietary intake. A cross-sectional survey published in the Lancet in 2011 looked at the iodine status in 810 schoolgirls aged 14–15 years attending secondary school in nine UK centres[93]. The findings of this study suggested that the UK is iodine deficient; 51% of participants were found to have mild iodine deficiency, 16% were classified as moderate and 1% had a severe deficiency[93]. These numbers are not consistent with the small percentage of women diagnosed with hypothyroidism in our prevalence study (0.13%). It may be that mild and moderate iodine deficiencies do not contribute to the development of hypothyroid disease in the same way as severe deficiencies. Or it may also be that the dietary intake of older women of reproductive age is more iodine sufficient than that of younger girls. It is important to note that the findings of our study are not transferrable to iodine insufficient populations, where the prevalence of thyroid disease will be much higher.

One of the core topics of debate at present is the cut off value for TSH to define euthyroidism. As mentioned in the introduction, there is a trend amongst reproductive medicine clinicians in the UK and endocrinologists in the US towards using the definition of abnormal TSH as values above 2.5mU/L. The prevalence overall in our study of women with a TSH 2.5-4.5mU/L was 19.64% (this was as high as 21.60% in South Asian women). There is an on-going debate regarding treatment of these women, particularly in the fertility setting. The National Association of Clinical Biochemistry (NACB) have recently suggested that the upper limit of serum TSH in adults should be reduced to 2.5mU/L, this was based on the serum TSH findings from 95% of healthy euthyroid volunteers which ranged between 0.4mU/L and 2.5mU/L[13]. However, the findings of our study are not consistent with this figure and in fact it is much lower; the prevalence of a TSH between 0.44-2.49mU/L in our study was 78.18% (95% CI 77.20-79.14%). Using the figures from our study this allows us to challenge policy makers and clinicians who are leaning towards routinely commencing LT4 treatment for around 20% of their asymptomatic women, based on no conclusive evidence that the treatment has benefit.

When examining the relationship between increasing age and thyroid dysfunction, specifically exploring TSH concentration, there appeared to be no association. With an increase in age, marked changes in thyroid hormone production, metabolism and action also occur; furthermore there is substantial evidence that increasing age above 60 is associated with higher prevalence of thyroid disease. The prevalence of overt hyperthyroidism in the elderly is increased in populations older than 60 years of age[91,95,92,96] and the frequency of overt hypothyroidism has an increased prevalence

of up to 5% in subjects over 60 years of age[95,97]. Furthermore, the largest epidemiological study in the UK assessing thyroid function, the Whickham survey, showed markedly increased levels of TSH in women over 45[92]. However as the upper limit for age in our study was 41 (limited to women of reproductive age only) this increased prevalence with advancing age was not seen; thus it appears that for women of reproductive age, age itself is not a risk factor for thyroid disease.

Regarding the relationship between increasing BMI and thyroid dysfunction our study did find an association. There have been similar findings in other studies[98–100] and so it is thought that TSH could represent a marker of altered energy balance in obese women[101]. Interestingly, the most recent ATA guidelines suggest that routine screening should be considered in morbidly obese women (≥40); this was based on results of two cohort studies which found prevalence's of SCH and overt hypothyroidism of 13.7% and 19.5% respectively[99,100]. However our study has shown that women with BMI >25 are also at higher risk of subclinical hypothyroidism compared with women of normal BMI.

When looking at the association between ethnicity and thyroid dysfunction it appears that for the "controversial" group of TSH 2.5-4.5mU/L the prevalence across all ethnicities is high (ranging from 11.6% in the mixed group to 21.6% in South Asians). When comparing the ethnicities to a reference group of White women, results showed that South Asian women have a higher prevalence of TSH 2.5-4.5mU/L ($p=0.006$) and moderate SCH, TSH 4.51-10mU/L ($p=0.008$). The reverse was true for Black women; they showed a lower prevalence of TSH 2.5-4.5mU/L ($p<0.001$)

and a higher prevalence of TSH 0.44-2.5mU/L compared to White women. This is in keeping with results from the National Health and Nutrition Examination Survey (NHANES III) where they found serum TSH concentrations were higher in White women than in Black women, independent of serum anti-thyroid antibody concentrations[91]. A more recent U.S. study also identified that Graves disease is more common in Black women compared with White women[102]. In addition to this, our study also showed Black women appear to have statistically significant higher rates of subclinical hyperthyroidism compared with White women. Subclinical hyperthyroidism can be graded as mild or severe, depending on the level of TSH: Grade 1 (mild) TSH 0.1–0.4 μU/mL and suppressed TSH concentration < 0.1 μU/mL would be Grade II (severe)[103]. Subclinical hyperthyroidism is not known to have serious adverse effects on pregnancy and the chances of progression to overt hyperthyroidism, particularly for mild hyperthyroidism is very low; therefore further research is needed to explore reasons for why Black women have lower TSH concentrations and the exact clinical implications.

The box-plot in Figure 31 shows that the median, interquartile range and range values for TSH concentration for Black women are lower compared to White women, and higher for South Asian women compared to White women. It could be in fact that these differences in TSH values are variants of normal; i.e. what may be considered “abnormal” for a White woman may be “normal” for a South Asian or Black woman. Unfortunately, as we do not have any outcome data we cannot confirm whether the variation in TSH concentrations between the ethnic groups has any clinical implications or if they are simply ethnic variations of normal. A Dutch study looked at ethnic differences in maternal thyroid function during pregnancy, the

study population consisted of 2765 Dutch, 308 Moroccan, 421 Turkish and 609 Surinam/Antillean women[104]. This was not representative of a UK population, however their findings were interesting. Results showed that 19% of women who were initially diagnosed as having an abnormal thyroid were found to have a normal thyroid when the reference range applied was ethnicity specific rather than population based[104]. Furthermore, of all women who were considered euthyroid using population-based reference ranges, 1.3% had an abnormal thyroid function test when ethnicity-specific reference ranges were used[104]. Their results showed that ethnic differences in serum TSH and T4 within one population from one geographical area resulted in considerable misclassification of thyroid disease[104]. The results from existing studies and our study highlight the need for population-based reference ranges to account for the relevant ethnic groups.

Finally, when looking at the relationship between originating population and thyroid dysfunction the results shown in the boxplot (Figure 32) shows the median TSH value for women with infertility is higher than women from the other originating populations. The results of our study suggest that TSH concentrations are arguably higher in women with infertility compared to other populations. This finding will be of particular interest to clinicians working in assisted reproduction in the UK given the current trend towards offering Levothyroxine treatment for women with TSH values >2.5mU/L. However, as stated in the introduction, this is yet to be proven as beneficial in improving fertility outcomes.

Conclusions and future work

As discussed throughout this chapter, the lack of clarity over whether thyroid function testing should be performed routinely pre-conception is largely based on the lack of evidence to suggest benefit from treating undiagnosed subclinical disease; for most women with overt disease they would be symptomatic and so this would be detected in primary care. The definition of screening is: *"the systematic and active search of a health problem by applying a test on a large scale in otherwise healthy people"*[105]. There are a series of criteria that should be met in order for a screening programme to be deemed appropriate, in a recent paper by Vila et al, discussing the controversy regarding universal thyroid screening in pregnancy, three broad categories were used; "*1) the disease or health problem should be serious, highly prevalent and have a detectable preclinical stage 2) the screening test should be sensitive and specific, simple and inexpensive, safe, acceptable, reliable, easy to perform and ideally cause minimal discomfort and lastly 3) the diagnosis of the health problem requires facilities with the equity of access and availability to effective, acceptable and harmless treatment"* [105].

It is important to note that screening should not be implemented if early treatment is ineffective and does not have any effect on the natural progression of the disease. Based on the findings of this prevalence study, in conjunction with the criteria for a screening programme and the absence of conclusive evidence to suggest benefit from treating subclinical thyroid disease in pregnancy; we would recommend that screening for thyroid disease should be performed as "case finding" rather than universal.

If conclusive evidence emerges to suggest that pre-conception treatment of thyroid disease (particularly subclinical) has beneficial effects on pregnancy outcomes, then it may be worth considering, as a starting point, routine screening for South Asian women and women with BMI above normal. The findings of this study show that around 4% of women have undiagnosed thyroid disease; of which 85% have subclinical disease. If evidence shows benefit in treating pre-conception women with a TSH of >2.5mU/L then the prevalence of subclinical thyroid disease increases to >20% for all women; resulting in a significant proportion of the population requiring treatment and monitoring of thyroid disease pre-conception and in pregnancy.

There needs to be careful calculation by all stake holders, including health economists, to decide if a burden of 4% (potentially as high as 20%) of the population potentially having pregnancy complications as a result of undiagnosed and untreated thyroid disease outweighs the potential benefit of early detection and treatment through universal screening. Large randomised controlled trials are needed to help answer the question of whether pre-conception treatment of subclinical thyroid disease (in particular using an upper limit TSH value of 2.5mU/L) improves pregnancy outcomes when compared with no treatment. The findings of such studies will then guide decision-making on whether pre-conception thyroid screening should become more widely available.

As mentioned previously, this prevalence study is directly linked to the national TABLET trial. The TABLET trial is continuing recruitment until December 2015. We intend to use all the data collected within the study, this is anticipated to be >10,000

cases, to give provide a greater sample size and examine any differences in more detail.

CHAPTER 6

National study of prevalence of thyroid autoimmunity in women of reproductive age

Introduction

Anti-thyroid antibodies (collectively known as thyroid autoimmunity, TAI) are autoantibodies that are targeted against the thyroid; thyroid peroxidase antibodies (TPOAb), thyroglobulin antibodies (TgAb) and thyrotropin receptor antibodies (TRAb) are the three most clinically important. Thyroid peroxidase (TPO) is the primary enzyme involved in the production of thyroid hormones and is stimulated by TSH[106]. TPOAb work against the function of the TPO enzymes and as a result cause thyroid inflammation[106]. The majority of TPOAb are produced by lymphocytes that infiltrate the thyroid, along with small contributions from the bone marrow and regional lymph nodes[106]. TPOAb cause thyroid cell damage through activation of the complement system and cell cytotoxicity[107]. However, it isn't thought that anti-TPO antibodies contribute significantly to thyroid destruction[108]. TPOAb are the most common anti-thyroid autoantibody; for Hashimoto's thyroiditis TPOAb are present in 90% cases, for those with Graves' disease the prevalence is 75% and with nodular goitre or thyroid carcinoma around 10-20%[107,109]. Also, around 10-15% of biochemically euthyroid individuals can have high TPOAb titres[107,109].

Current evidence shows that TAI is an important risk factor for poor obstetric outcomes, such as miscarriage and pre-term birth; even in women with biochemically normal thyroid function[110,111]. A systematic review, published in the British Medical Journal, showed that the presence of thyroid autoantibodies leads to a significantly increased odds of miscarriage for women from all populations compared to women without autoantibodies, the meta-analysed results showed: subfertility population (OR

3.15 [95% CI 2.23-4.44] p<0.001); recurrent miscarriage population (OR 4.22 [95% CI 0.97-18.44] p=0.06) and “unselected” or other (OR 4.28 [95% CI 2.06-8.92] p<0.001)[19]. There were similar findings for pre-term birth, with increased odds for women with thyroid autoantibodies compared to women without (OR 2.07 [95% CI 1.17-3.68])[19]. Evidence has also shown an association between thyroid autoantibodies, specifically thyroid peroxidase antibodies (TPOAb), with subfertility (OR 1.5 [95% CI 1.1-2.0]) but there was no association with clinical pregnancy rates following IVF treatment[17]. TPO antibody (TPOAb) positivity is also associated with a significantly increased risk of post-partum thyroiditis (OR 11.5 [95% CI 5.6-24])[17].

The presence of thyroid autoantibodies is thought to be relatively common in women of reproductive age; it is particularly high in women with a history of subfertility (prevalence ranging between 10-31%)[112–114] and those with a history of recurrent miscarriage (prevalence ranging between 17-33%)[20,21]. For the general population the prevalence ranges from 6-20%[115,116]. Table 22 shows the pooled prevalences of TPOab across women of differing originating populations taken from existing studies. The prevalences were pooled with a random effects model, with Wolfs method for confidence intervals; given the substantial clinical and statistical heterogeneity as well as threshold variations, the pooled prevalences should be interpreted with caution.

The average prevalence across all studies was 19%, indicating that thyroid autoimmunity is common in women of reproductive age.

Table 22. Pooled prevalence of thyroid antibodies across various populations

Population	Study and year	Thyroid antibodies tested	TPO threshold for test positivity	Prevalence	Pooled prevalence (95% CI)
Unselected women	Stagnaro-Green 1990	TPO, TG	Not stated	108/552 (19.6%)	**11.3%** (7.7%, 17.0%)
	Glinoer 1991	TPO, TG	>100u/ML	45/726 (6.2%)	
	Iijima 1997	TPO, TG	≥1:100 titre	125/1179 (10.6%)	
	Bagis 2001	TPO, TG	>35IU/ml	108/876 (12.3%)	
History of miscarriage	Rushworth 2000	TPO, TG	>1:400 titre	163/870 (19%)	**19%**
History of recurrent miscarriage	Pratt 1993	TPO, TG	Not stated	13/42 (31%)	**25.9%** (19.5%, 34.3%)
	Bussen 1995	TPO, TG	>100IU/ml	11/66 (17%)	
	Esplin 1998	TPO, TG	Not stated	49/149 (33%)	
	Kutteh 1999	TPO, TG	≥65IU/ml	187/900 (20.8%)	
	Dendrinos 2000	TPO, TG	Not stated	15/45 (32.5%)	
History of infertility	Singh 1995	TPO, TG	Not stated	107/487 (22%)	**18.1%** (13.9%, 23.5%)
	Kim 1998	TPO, TG	>100U/ml	23/79 (29.1%)	
	Muller 1999	TPO	>80U/ml	24/173 (14%)	
	Bussen 2000	TPO, TG	>100IU/ml	15/48 (30.6%)	
	Poppe 2003	TPO	>100IU/ml	33/234 (14%)	
	Poppe 2004	TPO	>100IU/ml	9/35 (25.7%)	
	Negro 2005	TPO	>100IU/ml	73/484 (15%)	
	Negro 2007	TPO	>100IU/ml	42/416 (10.1%)	

TPO – Thyroid peroxidase antibody; TG – Thyroglobulin antibody

Despite this apparent clear association between TPOAb positivitiy and miscarriage and pre-term birth, finding an association does not automatically imply a causal relationship. In fact, the aetiology of increased pregnancy loss and pre-term birth in women with TAI remains unknown.

Several "working hypotheses" have been proposed[117–119]. The first hypothesis holds the view that the presence of circulating thyroid antibodies is not directly related to the pregnancy loss, but instead represents a marker of an underlying generalised autoimmune imbalance. An imbalance which, in turn, could explain a greater rejection rate of the foetus[117]. The second hypothesis proposes that despite apparent euthyroidism, the presence of TAI could be associated with a reduced ability of the thyroid gland to adapt adequately to the necessary changes associated with pregnancy[118]; due to the reduced functional reserve characteristic of chronic thyroiditis[111]. The third hypothesis suggests that as increasing age has been associated with increasing titres of TPOAb; that age itself is the risk factor rather than the TPOAb[119]. Although this third hypothesis seems the least plausible, it is most likely that the increased risk of pregnancy loss associated with TAI is of multifactorial origin.

Studies have shown that euthyroid women, who are positive for TPOAb, are more likely to develop impaired thyroid function during pregnancy; in particular subclinical hypothyroidism[114,120]. It is uncertain whether it is the TPOAb or TPO-specific T cells which are the primary cause of thyroid inflammation; which then leads to thyroid failure and hypothyroidism in select individuals[106]. Consequently, the Endocrine

Society Clinical Practice Guidelines (ESCPG) and the American Thyroid Association (ATA) guidelines recommend thyroid function monitoring for women with known thyroid autoimmunity during pregnancy because of the risk of developing hypothyroidism[22]. However, there is no consensus on whether Levothyroxine treatment of euthyroid women with TAI improves pregnancy outcomes. To date there have only been two small randomised controlled trials (n=86 and n=115) which have looked at this research question; the results did not provide sufficient evidence to support treatment with Levothyroxine in a euthyroid women[114,120]. As discussed in Chapter 5, a large multi-centre randomised controlled trial (TABLET trial) is currently on going in the UK and will help to provide more robust findings to help answer this question.

Owing to the lack of evidence suggesting any benefit from Levothyroxine treatment solely for the presence of TPOAb, there is an on-going debate regarding whether women should be routinely screened for TPOAb prior to- or during early pregnancy. There is an argument that states if we know TPOAb positive women are at higher risk of developing thyroid impairment and/or adverse pregnancy outcomes that we should be aiming to identify these women pre-conception with a view to at least offering thyroid monitoring during pregnancy, with or without Levothyroxine treatment. At present the ATA and ESCPG do not recommend routine screening for thyroid antibodies in euthyroid women during pregnancy or prior to undergoing assisted reproduction[121].

To date there has been no large epidemiological study to identify an accurate prevalence of thyroid autoimmunity in women trying actively for a pregnancy in the UK. Aside from the suggestion that there may be higher rates of TPOAb in women with infertility and recurrent miscarriage, there has been no specific identification of "high risk" women using demographic characteristics such as age and body-mass index (BMI). If we can determine a subset of women who may be at higher risk of having TPOAb this would allow for "case finding" of these women pre-conception; this information would then be useful in helping to determine which women may benefit from thyroid monitoring during pregnancy (if treatment with Levothyroxine proves beneficial in the TABLET trial).

The aim of this study was to determine an accurate prevalence of TPOAb in women aiming to conceive across the UK; furthermore we aimed to identify if there are "high risk" women who may benefit from routine pre-conception thyroid autoantibody screening.

Aims and objectives

1. To study the prevalence of thyroid peroxidase antibodies (TPOAb) in pre-conception women of reproductive age in the UK overall.
2. To study the prevalence of TPOAb in combination with normal and abnormal thyroid function, in pre-conception women of reproductive age.
3. To study the prevalence of TPOAb in demographic subgroups; age, body-mass index, ethnicity and originating population.

4. To investigate the prevalence of TPOAb positivity with increasing TSH concentrations.

Methods

This was a large prospective epidemiological prevalence study conducted at 42 hospitals across the UK between December 2013 and February 2015. TPOAb testing was performed in conjunction with thyroid function testing and the work was directly linked to the TABLET trial. Ethical approval was obtained from Berkshire B Research Ethics Committee (letter of approval in Appendix 5).

Eligibility criteria and recruitment setting

The eligibility criteria and clinical settings in which patients were recruited are as stated in the methods for chapter 5.

Thresholds for thyroid antibody tests

There are various assays for TPO antibodies available, each assay has different specified detection limits and differing numerical thresholds for test positivity; these are pre-determined by the assay manufacturer and the sensitivity and specificity of the assays are comparable. For example a result of 35iu in assay A may be positive for that assay, but for assay B the threshold for positivity may be >50iu; the result of 35iu on assay A would be equivalent to >50iu on assay B making the results comparable, and so it is not that assay B only becomes positive at a higher TPOAb level. These variations are an accepted part of normal practice in the UK. Quality

assurance for assays in the laboratories for all the participating centres is provided by UK NEQAS, which shows over 99% concordance in the classification of samples as either positive or negative for TPO antibodies across all assays. Levels that were considered as “indeterminate” by the assay were taken to be positive for the purpose of this study.

Thyroid function grouping and demographic data

The TPOAb status was looked at overall and in individual thyroid dysfunction subgroups as classified in chapter 5; euthyroid 1a and 1b, overt hypo/hyperthyroidism, subclinical hyperthyroidism and subclinical hypothyroidism further split into moderate and severe. The demographic data collected was the same as for the thyroid function study in chapter 5; i.e. age, BMI, ethnicity and originating population.

Statistical analyses

Crude proportions and percentages with 95% confidence intervals were calculated, using the binomial exact method, to show the prevalence of thyroid autoimmunity overall and within each thyroid function group (both normal and abnormal). For the overall TPOAb positive group prevalence’s were examined within each demographic subgroup: age, BMI, ethnicity and originating population (where age and BMI were created as categorical variables).

Finally, an analysis was performed to look at the relationship between TSH concentration and TPOAb positivity. TSH values were split into 9 clinically relevant

groups (<1.00; 1.00-1.49; 1.50-1.99; 2.00-2.49; 2.50-2.99; 3.00-3.49; 3.50-3.99; 4.00-4.49; >4.50). A 2x2 table was created with TSH concentration against TPOAb positivity as the outcome. For each of the 9 groups of TSH concentration, prevalence of TPOAb positivity was calculated and presented in tabular and graphical format.

Calculation of the 95% confidence intervals of prevalence's was performed manually using the binomial exact method, while all other statistical analyses were performed using SPSS (ver. 21.0; SPSS Inc., Chicago).

Results

A total of 6974 women had TPO and thyroid function test results available across the UK between December 2013 and February 2015. This figure is slightly less than the 7022 women who had thyroid function results alone, as seen in chapter 5. This was due to incomplete data entry for all TPOAb results onto the database; 0.68% of cases did not have a TPOAb result entered. The data manager for the TABLET trial chases incomplete data entry on a monthly basis and requests reasons for missing data from each centre; this 0.68% constitutes the women who had insufficient blood samples taken in the first instance and declined repeat testing, consequently they had no result for TPOAb.

Distribution of patients in study and regional prevalence's

Figure 33 shows the distribution of patients in the study, classified into relevant groups of TPO status and thyroid dysfunction; percentages are expressed for the whole cohort. Table 23 shows the percentage prevalence of TPOAb by geographical region.

Figure 33. Flowchart of patients in study

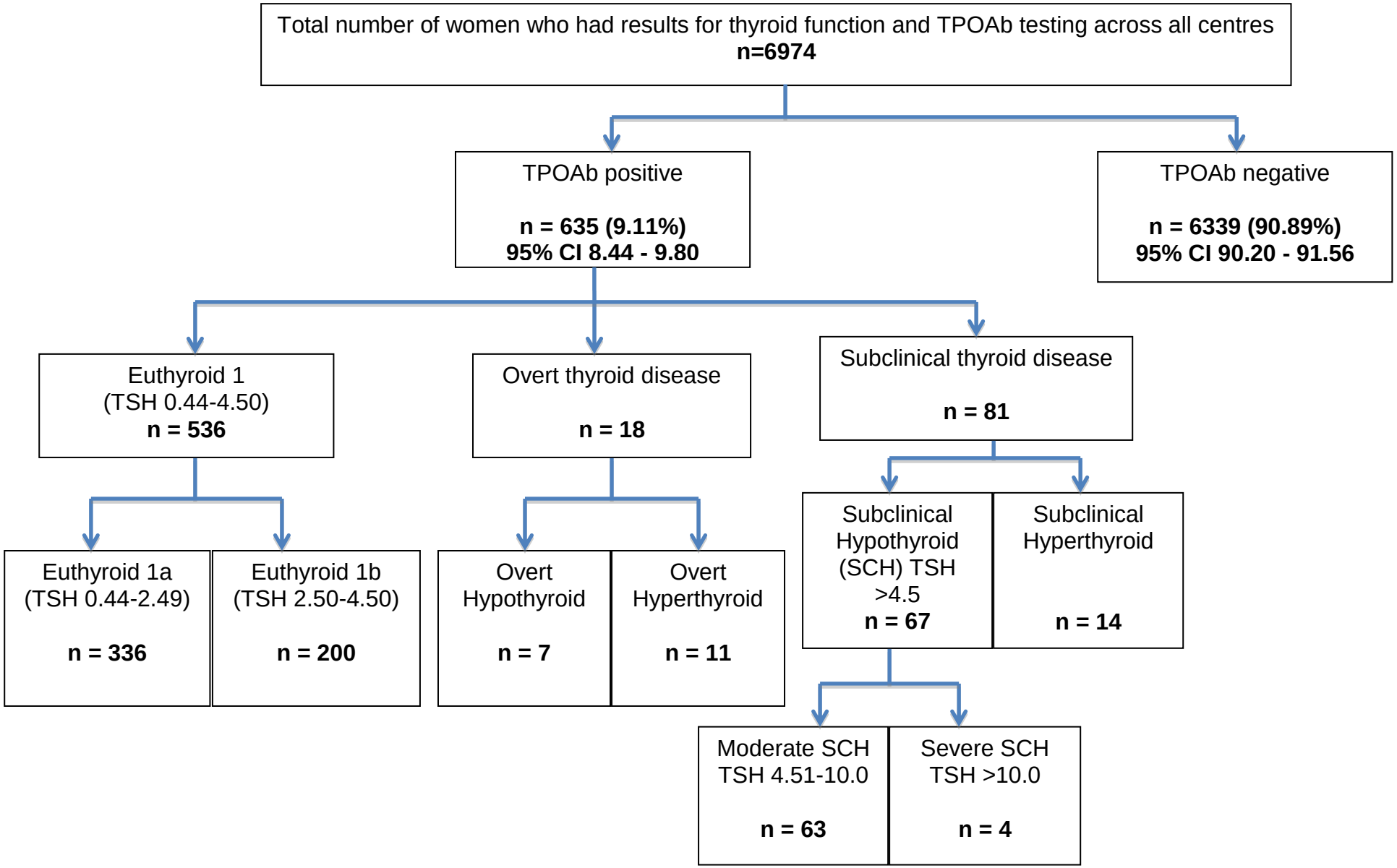


Table 23. TPOAb positivity by geographical region

Region	Numbers screened	Number of TPOAb positive	Regional % TPOAb positivity
Scotland	196	28	14.3%
North West	62	4	6.5%
North East	101	17	16.8%
Yorkshire	267	20	7.5%
Manchester and District	1105	111	10.0%
Merseyside	519	40	7.7%
West Midlands	1246	90	7.2%
East Midlands	407	46	11.3%
London and Essex	2394	207	8.6%
Home Counties	244	20	8.2%
West Country	367	46	12.5%
Southern England	66	6	9.1%
Overall	**6974**	**635**	**9.1%**

The overall prevalence for TPOAb positivity for all screened patients was 9.1%.

There does appear to be some geographical variation in TPOAb positivity; Scotland, North East, West Country and East Midlands appear to have a higher prevalence of TPOAb positivity compared to the rest of the UK. The West Midlands and North West have the lowest prevalence of TPOAb positivity.

Overall description of data and prevalences of TPOAb grouped by thyroid function

The data presented in Tables 24 and 25 show an overall description of the data and prevalence's of TPOAb grouped by thyroid function. The prevalence of TPOAb in women with a normal thyroid function (according to the generically accepted TSH range of 0.44-4.50mU/L) was 7.98% (95%CI 7.34-8.65%). When the euthyroid group were split into 1a (TSH 0.44-2.5mU/L) and 1b (TSH 2.5-4.5mU/L), the prevalence of TPOAb was statistically significantly higher in the euthyroid 1b group; 16.30% vs. 6.12% ($p<0.001$). There was also clear indication of higher prevalence of TPOAb positivity with increasing TSH concentration; prevalence in the moderate SCH group was 42.86% and for severe SCH it was 80.00%. The prevalence of TPOAb was high in both overt disease groups; 61.11% for overt hyperthyroidism and 77.78% in overt hypothyroidism.

Table 24. Overall description of dataset

Subgroup	Number and %
Age (n=6974)	
16-21	180 (2.56%)
22-26	882 (12.56%)
27-31	2006 (28.57%)
32-36	2502 (35.63%)
37-41	1404 (19.99%)
BMI (n=6325)	
<18.5	131 (2.07%)
18.5-24.9	3092 (48.89%)
25-29.9	1851 (29.26%)
30-34.9	763 (12.06%)
≥35	443 (7.00%
Ethnicity (n=6974)	
White	4834 (68.84%)
Black	503 (7.16%)
South Asian	1222 (17.30%
Chinese	96 (1.37%)
Mixed	144 (2.05%)
Other	175 (2.49%)
Population (n=6898)	
History of 1 or 2 miscarriages (EPAU)	2210 (32.04%)
Infertility	3164 (45.87%)
Recurrent miscarriage	1403 (20.34%)
Other	76 (1.10%)

Table 25. Prevalence of TPOAb positivity within each thyroid function group

Thyroid autoimmunity status and function	Proportions	Prevalence	95% CI Lower	95% CI Upper
Euthyroid 1 (TSH 0.44-4.5)	536/6717	7.98%	7.34%	8.65%
Euthyroid 1a (TSH 0.44-2.49)	336/5490	6.12%	5.50%	6.79%
Euthyroid 1b (TSH 2.5-4.5)	200/1227	16.30%	14.27%	18.49%
Mod. SCH (TSH 4.51-10)	63/147	42.86%	34.74%	51.27%
Severe SCH (TSH 10.0<)	4/5	80.00%	28.36%	99.49%
Overt Hypothyroid	7/9	77.78%	39.99%	97.19%
Subclinical hyperthyroid	14/90	15.56%	8.77%	24.72%
Overt hyperthyroid	11/18	61.11%	35.75%	82.70%

Prevalence of TPOAb within demographic subgroups

The data presented in Table 26 shows the prevalence of TPOAb within the demographic subgroups.

Table 26. TPOAb positive prevalence within demographic subgroups

			95% CI	
Subgroup	**Proportions**	**% Prevalence**	**Lower**	**Upper**
Age (n=633)				
16-21	14/180	7.78%	4.32%	12.71%
22-26	69/882	7.82%	6.14%	9.80%
27-31	177/2006	8.82%	7.62%	10.15%
32-36	227/2502	9.07%	7.98%	10.27%
37-41	146/1404	10.40%	8.85%	12.12%
BMI (n=580)				
<18.5	11/131	8.40%	4.27%	14.53%
18.5-24.9	275/3092	8.89%	7.91%	9.95%
25-29.9	160/1851	8.64%	7.40%	10.02%
30-34.9	73/763	9.57%	7.57%	11.88%
≥35	61/443	13.77%	10.70%	17.33%
Ethnicity (n=633)				
White	446/4834	9.23%	8.42%	10.08%
Black	20/503	3.98%	2.45%	6.07%
South Asian	131/1222	10.72%	9.04%	12.59%
Chinese	10/96	10.42%	5.11%	18.32%
Mixed	5/144	3.47%	1.14%	7.92%
Other	21/175	12.00%	7.58%	17.76%
Population (n=622)				
EPAU	213/2210	9.64%	8.44%	10.94%
Infertility	273/3164	8.63%	7.67%	9.66%
Recurrent misc.	127/1403	9.05%	7.60%	10.68%
Other	9/76	11.84%	5.56%	21.29%

Results show that increasing age appears to be associated with higher prevalence of TPOAb. When making comparisons between women with TPO antibodies and those without (not accounting for thyroid function) there was a statistically significant higher mean age for women with TPOAb, however this was small; mean difference 0.5years, p=0.02.

When looking at BMI; the results show that the prevalence of TPOAb was significantly higher in women with BMI ≥35 (13.8%) compared to women with normal BMI (8.9%) ($p<0.001$); furthermore the mean BMI for TPOAb positive women (26.4 (±5.8)) was significantly higher than the mean BMI for TPOAb negative women (25.9 (±5.4)) (p=0.04).

When assessing the relationship between ethnicity and TPOAb status the prevalence ranged from 3.5% in the mixed population to 12% in "other". When comparing the prevalence for each ethnic group to a reference White population (9.23%), there was a statistically significant lower prevalence of TPOAb in Black women (3.98%, $p<0.001$) and mixed race women (3.47%, p=0.02).

Finally there appeared to be no significant relationship between originating population and TPOAb positivity.

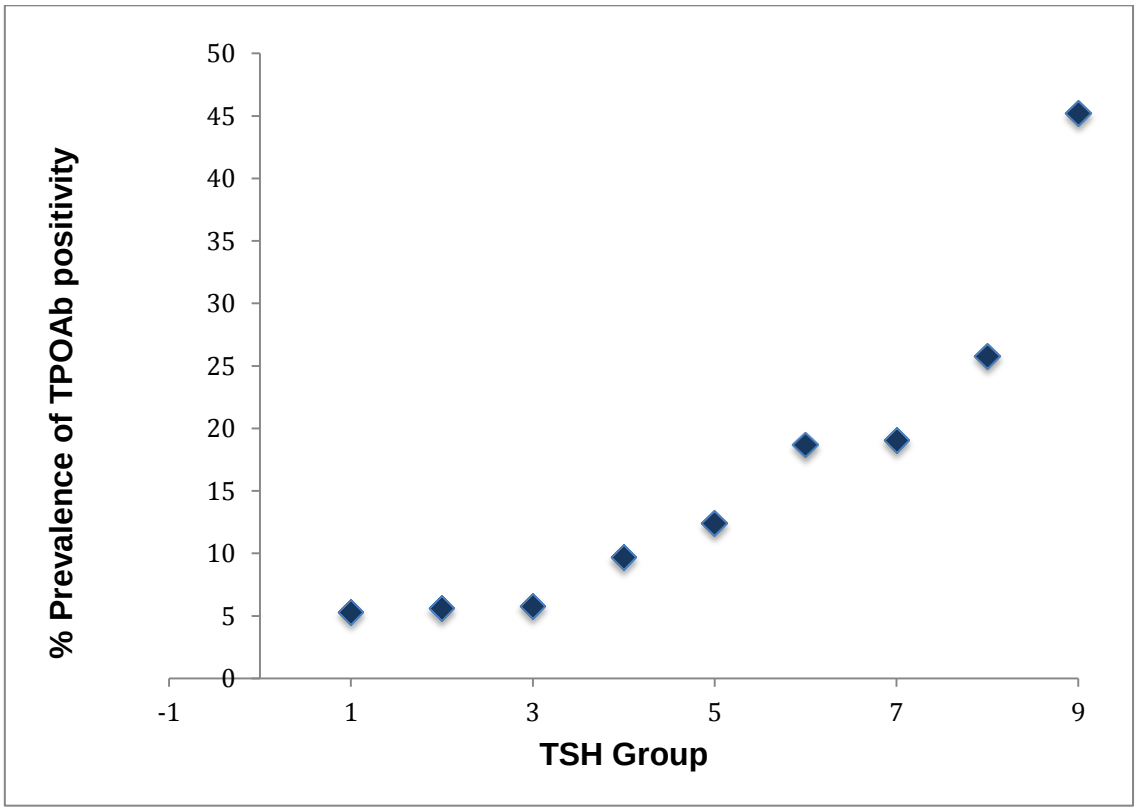
Relationship between TSH concentration and prevalence of TPOAb positivity

Finally, an analysis was performed to look at the relationship between TSH concentration and the prevalence of TPOAb positivity. Table 27 and Figure 34 show that increasing TSH is strongly associated with increased prevalence of TPOAb positivity, particularly TSH values above 2.50mU/l.

Table 27. TSH concentration and probability of being TPOAb positive

	TSH concentration (mU/L)	% Prevalence of TPOAb
1	<1.00	5.3
2	1.00-1.49	5.6
3	1.50-1.99	5.8
4	2.00-2.49	9.7
5	2.50-2.99	12.4
6	3.00-3.49	18.7
7	3.50-3.99	19.1
8	4.00-4.49	25.8
9	≥4.50	45.2

Figure 34. Graphical representation of TSH and prevalence of TPOAb positivity



Discussion

Main findings

The aim of this study was to present the prevalence of thyroid autoimmunity across the UK in pre-conception women (without and without thyroid dysfunction) who are actively trying for a pregnancy. The overall prevalence of TPOAb positivity was 9.11% (95%CI 8.44-9.80). Increasing age, body mass index and TSH concentrations are all associated with higher rates of TPOAb positivity.

Strengths and limitations

The strengths and limitations for this study are in line with those discussed in chapter 5. The large sample size, good geographical distribution of recruiting centres across the UK along with the use of pre-screening logs to ensure a representative population was recruited were all study strengths.

As mentioned in chapter 5 one of the limitations in not having true “low risk” women means that we are unable to accurately compare the prevalence in “high risk” women vs. “low risk”. We have instead compared the higher risk women, as stated in existing studies these are women from the infertility and recurrent miscarriage populations, against women with a history of only one or two miscarriages as the “low risk” control. It would be interesting to see comparisons of TPOAb prevalence between the women screened from our study and the “normal” or true “low risk” population of women trying for a pregnancy with no known history of gynaecological/obstetric or medical problems.

A further limitation of the study is that there is no outcome data to allow us to investigate the relationship between thyroid autoimmunity and dysfunction and pregnancy outcomes using a large sample size. However, this study is directly linked to the TABLET trial; this trial will provide evidence for or against the hypothesis that treatment using 50mcg of Levothyroxine for euthyroid women with TPOAb improves live birth rates by 10% compared to placebo.

Comparisons with existing literature

When looking at the relationship between age and TPOAb positivity we found there was a statistically significant higher mean age for women with TPOAb. This finding is in keeping with results from a meta-analysis of 22 studies looking at thyroid autoimmunity and miscarriage[122]. The results of the meta-analysis showed that women with TAI were found to have slightly higher mean age [age difference, 1·29 years] (95% CI 0·43–2·16, *p*=0.003) compared with those without TAI[122]. However, a different systematic review and meta-analysis published in the British Medical Journal found no significant difference between the groups (weighted mean difference 0.87 years, −0.06 to 1.80 years; p=0.07)[19]. As the results of our study suggest only a marginal increase in prevalence of TPOAb with increasing age, it is difficult to determine a cut off age at which the prevalence of TPOAb becomes higher to help identify which women may benefit from screening. In our study the prevalence for women in the highest age group (37-41) was 10.40% compared to 8.82% in the reference age group (27-31). In practical terms a difference of 1.58% in prevalence of TPOAb between older women and younger women is of questionable relevance. This finding in conjunction with conflicting evidence from existing meta-analyses looking at age and TPOAb prevalence provides little support for routine screening of “older” women.

Existing literature looking at the relationship between thyroid autoimmunity and body-mass index is limited. A recent large Danish study of over 70,000 participants found that BMI is positively correlated with autoimmune diseases; in particular Type 1 Diabetes[123]. However they found no significant relationship between raised BMI and

thyroid autoimmune diseases such as Hashimoto's or Graves disease[123]. The results of our study have shown that the prevalence of TPOAb was significantly higher in women with BMI ≥35 compared to women with normal BMI; furthermore the mean BMI for TPOAb positive women was significantly higher than the mean BMI for TPOAb negative women. If evidence from the awaited TABLET trial shows benefit in treating euthyroid women with TPOAb then based on the findings of our study it would be worth considering offering routine pre-conception TPOAb screening for women with BMI ≥35.

When looking at the relationship between ethnicity and TPOAb status the prevalence ranged from 3.5% in the mixed population to 12% in "other". Both Black and "Other" ethnic groups were found to have a statistically significant lower prevalence of TPOAb positivity compared to White women. This finding was consistent with the results of the NHANES III study; where a lower prevalence of TPOAb was found in African American women compared with White women. These differences in thyroid disease prevalence's between ethnic groups may be due to different environmental exposures, genetics, or a combination of both. Future work needs to explore potential reasons to explain these differences and the corresponding clinical implications. Regarding the implication of universal screening for TPOAb, on the basis of the findings of our study and existing literature we can advise against routine screening for TPOAb in Black women or mixed race women. As for South Asian women, given the higher prevalence of subclinical thyroid disease in this group (as shown in chapter 5), it may be worth testing for TPOAb pre-conception as a surrogate marker of potential thyroid disease progression in pregnancy.

The final demographic explored was originating population. Existing literature has suggested that the prevalence of TPOAb in women of reproductive age is relatively common. The prevalence in an "unselected" population ranges from 6% to 20%[115,116]. The prevalence is considered to be even higher in women with a history of recurrent miscarriage (17-33%)[20,21,124] and in women with a history of infertility (10-31%)[112–114]. In contrast to the existing studies the findings of our study suggest no significant differences in the prevalence of TPOAb between "lower risk" women i.e. those with history of 1 or 2 miscarriages compared with "higher risk" women i.e. with infertility or history of recurrent miscarriage. Furthermore the prevalence of TPOAb in all of the main populations (excluding "other") was much lower than the prevalence's stated in the literature. The existing literature and table of pooled prevalence's across different populations for TPOAb (as shown in Table 22 in the introduction) have shown the prevalence to be around 19%; this is much higher than the 9.11% observed in our study. Furthermore we looked at the prevalence's for select centres that recruited from specific populations. Guys Hospital in London solely recruit from their fertility centre, while St Marys Hospital in Manchester and St Bartholomews Hospital in London recruit >90% of their patients from the fertility setting; the corresponding prevalence of TPOAb in these centres was 7.5%, 8.4% and 9.6% respectively. These prevalence's are much lower than the pooled prevalence from existing studies showing the prevalence of TPOAb for the infertility population to be 18.1% (Table 22). St Marys Hospital in Paddington is a tertiary referral unit and one of the leading centres for recurrent miscarriage, they solely recruit from their recurrent miscarriage clinics; the prevalence of TPOAb in this population was 7.4%. Again this is much lower than the data from existing studies

suggesting a pooled prevalence of 25.9% for women with recurrent miscarriage. Surprisingly, this figure is also much lower than the 19% thyroid antibody prevalence found by the team at St Mary's Paddington themselves in a study they conducted in 2000[125]; however that study was focused on thyroglobulin antibodies and thyroid microsomal antibodies rather than TPOAb. It could also be that if the screening was not being offered routinely to all patients that there may be some selection bias, resulting in a lower than expected prevalence.

As discussed in Chapter 5 one of the central challenges with defining thyroid function at present is agreeing on the cut off TSH value to define euthyroidism. If we compare the prevalence of TPOAb in those with TSH 0.44-2.49 (euthyroid 1a) to those with TSH 2.5-4.5mU/L (euthyroid 1b) there is a statistically significant greater odds of being TPO positive for those with TSH 2.5-4.5mU/L; 6.12% vs. 16.30% (OR 2.99 (2.48-3.60) $p<0.001$). A prevalence of 16.30% indicates that TPOAb are common amongst women who have a TSH 2.5-4.5mU/L i.e. an otherwise normal thyroid function. Given the association between TPOAb positivity and increased risk of developing thyroid disease in pregnancy, this result supports the notion that women with a TSH of >2.5mU/L should potentially be offered routine testing for TPOAb and subsequent monitoring of their thyroid function in pregnancy.

Finally we looked at the relationship between TSH concentration and the prevalence of TPOAb positivity. Results showed that overall increasing TSH concentration was strongly correlated with a higher prevalence of TPOAb positivity. This was also reflected by the higher prevalences of TPOAb positivity seen in women with

moderate SCH and severe SCH. The graph presented in Figure 34 shows that there appears to be a large increase in probability of being TPOAb positive beyond a TSH concentration of 2.50mU/L; this increase is even more marked above 4.50mU/L. This relationship between increasing TSH concentrations and increased likelihood of TPOAb positivity is consistent with what has been shown in existing studies[91]. As mentioned previously, it may be useful to consider pre-conception TPOAb testing in women with TSH values above 2.5mU/L (or even 4.5mU/L as a starting point), especially if treatment is not being offered. Knowing the TPOAb status may then help to decide whether the patients thyroid function should be monitored through their pregnancy.

Conclusions and future work

The lack of clarity over whether thyroid antibody testing should be performed routinely pre-conception is largely based on the lack of evidence for a treatment with known benefit (should the patient test positive).

For a screening test to be offered universally, as mentioned in Chapter 5, a detailed list of strict pre-requisites must be met. Realistically, with the prevalence of TPOAb in women of reproductive age shown to be much lower than existing studies suggested, there is less of a case for universal screening. As with thyroid function testing, there would have to be careful calculation to decide if the burden of 9.11% of the population potentially having pregnancy complications, as a result of undiagnosed thyroid autoimmunity, warrants the potential benefit of early detection and treatment through implementation of universal screening. The results of the

TABLET trial will inevitably guide decision-making on whether screening should be offered more widely; dependent on whether Levothyroxine is proven to be beneficial or not. If Levothyroxine is proven to have benefit in women with TPOAb, based on the findings of this study and collated findings from existing studies, we would recommend routine TPOAb screening for women with BMI ≥35 and South Asian women with TSH concentrations >2.5mU/L. Given the low prevalence of TPOAb in Black women and mixed race women there is no case for universal screening in these populations. Finally, there appears to be no benefit to targeting "high risk" populations such as women with infertility or history of recurrent miscarriage in favour of women with a history of just one or two miscarriages.

The work from this chapter is directly linked to the national TABLET trial. The TABLET trial is continuing recruitment until December 2015. We intend to use all the data collected within the study, this is anticipated to be >10,000 cases, to give provide a greater sample size and examine any differences in more detail.

CHAPTER 7

The effect of subclinical hypothyroidism on IVF outcome

Introduction

Subclinical hypothyroidism (SCH) is defined as an elevated level of thyroid stimulating hormone (TSH) accompanied by a normal level of free thyroxine (FT4) in the circulation. The reported incidence of SCH, in the infertility population has ranged between 1-43% with a mean of around 13%[15,84]; it is dependent upon the thresholds used for diagnosis and the iodine status of population. Observational studies have indicated that pregnant women with subclinical hypothyroidism have an increased risk of adverse pregnancy outcomes such as miscarriage, perinatal loss, preterm birth, pre-eclampsia and low IQ in the offspring[17,126].

Debates regarding recommendations to screen for SCH and/or thyroid autoimmunity (TAI), and whether abnormal results should be treated pre-conceptually or in pregnancy, have been on going for the last two decades. The work of chapters 5 and 6 have focused on this issue. The central challenging question has been whether Levothyroxine (LT4) treatment will alter fertility, obstetric or neonatal outcomes. To date there are two randomised controlled trials which have looked at infertile women with subclinical hypothyroidism undergoing IVF treatment[127,128]. Both studies suggested an improvement in birth rates, improvement in implantation of embryos and decrease in miscarriage for those women supplemented with LT4 compared with untreated women [127,128]. The threshold for TSH treatment was 4.0mU/L in the study by Abdel-Rahman et al and 4.5mU/L in the study by Kim et al. However, both studies were very small with only 70 women in Abdel-Rahman et al[127] and 64 women in Kim et al[128]. A recent systematic review by Velkeniers et al[18] meta-

analysed the results of the two studies by Kim et al[128] and Abdel-Rahman et al[127] and also included data from a third study, by Negro et al[129]. The study by Negro et al randomised women if they had TSH values above 4.2mU/L and tested positive for thyroid peroxidase antibodies (TPOAb); this was also a small study of only 86 women[129]. The meta-analysis showed that LT4 supplementation versus no treatment (or placebo) resulted in a significant increase in delivery rate and implantation rate, and a decrease in miscarriage.

Levothyroxine, however, is not without its side effects. If doses are too high patients can experience symptoms of hyperthyroidism, therefore it is imperative that should patients be commenced on LT4 that they have regular thyroid function test monitoring, pre-conception and during pregnancy. Furthermore, a study by Browne et al of over 14,000 cases found that periconceptual thyroxine medication was significantly associated with several congenital birth defects; left ventricular outflow tract obstruction heart defects, hydrocephaly, hypospadias, and isolated anorectal atresia[130]. However, because of evidence for adverse pregnancy outcomes associated with subclinical hypothyroidism, a subset of clinicians, particularly fertility specialists, prefer to routinely treat SCH.

The real debate lies in defining SCH. Given this potential benefit from LT4 treatment in treating SCH some clinicians are beginning to move towards treating subfertile women using even stricter reference ranges for “normal”; i.e. aiming to achieve a pre-conception TSH of <2.5mU/L, however there is no conclusive evidence to support this. At present the British Thyroid Association have no clear pre-conception

recommendations specific to subfertile women. As discussed in chapter 5, the current guidance is to routinely offer treatment if non-pregnant women have a TSH >10mU/L. Following this, the recommendation then jumps straight to pregnancy and advises that the target TSH in early pregnancy (i.e. before 12weeks gestation) for women who are already on Levothyroxine treatment (for overt thyroid disease) should be <2.5mU/L; with close monitoring of thyroid function throughout the pregnancy. This is in-keeping with guidance from the American Thyroid Association guidelines for first trimester serum TSH concentration[16]. The most commonly accepted upper limit of TSH, as per guidance from the United Kingdom National External Quality Assessment Service (UK NEQAS), for normal non-pregnant women is around 4.5mU/L. However, there is no clear guidance regarding the treatment of subfertile women who are asymptomatic and have a TSH between 4.5-10mU/L, although it is stated in the "UK Guidelines for the Use of Thyroid Function Tests" published in July 2006 that LT4 treatment *may be indicated* in women who have TSH concentrations in this range and are trying for a pregnancy. In contrast to the British Thyroid Association guidelines, the recent ATA and AACE (American Association of Clinical Endocrinologists) guideline advises that treatment with Levothyroxine should be considered in women of childbearing age with serum TSH levels between 2.5mU/l and the upper limit of normal (or a given laboratory's reference range) if they are planning a pregnancy, including assisted reproduction in the immediate future[89]. However, the American Congress of Obstetricians and Gynaecologists have not supported these recommendations. As there is no consensus in the UK over when or how to treat pre-conception subfertile women who have a result between 4.5-10mU/L (or even 2.5-10mU/L), this decision on whether to treat with LT4

replacement or not comes down to the discretion of the clinician and patient preference.

In the United States there is an indication from the National Association of Clinical Biochemistry (NACB) that it is likely in the future that the upper limit of the TSH euthyroid reference range will be reduced to 2.5mU/L for all adults, even without pregnancy. This is based on evidence that more than 95% of rigorously screened normal euthyroid volunteers have serum TSH values between 0.4 and 2.5mU/L[94]. Although individuals who have a TSH between 2.5-4.5mU/L may then be classified as having subclinical hypothyroidism there is no clear evidence to suggest there are any adverse outcomes in this group[84]. Studies looking at a pre-conception threshold of 2.5mU/L for TSH, in the subfertility population, have shown mixed findings regarding the use of LT4 treatment. A study of 1055 infertility patients, by Reh et al, found no differences in clinical pregnancy rates, live birth or miscarriage rates for patients undergoing their first IVF cycle when comparing those with a TSH <2.5mU/L and those between 2.5 and 4.5mU/L[131]. Similarly, a recent study by Chai et al, of 627 women, also found no differences in miscarriage or live birth rates for those with TSH <2.5mU/L compared with those with TSH ≥2.5mU/L; furthermore the thyroid autoantibody level did also not affect these IVF outcomes[132]. In contrast, a study by Fumarola in 2013, of 164 women, found that TSH >2.5mU/L was associated with reduced clinical pregnancy rates (22.3% in TSH ≤ 2.5mU/L group versus 8.9% in TSH > 2.5mU/L group; p=0.045); although there were no significant differences for any other IVF outcome measures[133].

In view of the lack of clear guidance and inconsistent findings in existing evidence regarding when to treat subclinical hypothyroidism preconception in the infertile population, this led to the development of our study. We aimed to investigate the effect of treated and untreated subclinical hypothyroidism, of varying thresholds of TSH concentration, in women undergoing treatment with assisted reproductive technologies. The purpose of this study was to determine if it is safe practice to allow a higher threshold of TSH concentration, than is traditionally used, before commencing LT4 treatment pre-conceptually in infertile women with SCH, thus challenging the trend of moving towards stricter upper limit TSH thresholds.

Aims and objectives

1. To compare IVF outcomes for women who are euthyroid (TSH 0.44-3.63mU/L, fT4 10-21) and women with SCH (TSH >3.63mU/L, fT4 10-21).
2. To compare IVF outcomes for euthyroid women (TSH 0.44-3.63mU/L, fT4 10-21) and untreated SCH (TSH 3.64-5.99mU/L, fT4 10-21)
3. To compare IVF outcomes for women with TSH (<2.5mU/L) and women with untreated SCH (2.50-5.99mU/L).

Methods

Study design

This was a prospective cohort study conducted at the Birmingham Women's Hospital Assisted Conception Unit (ACU) between June 2012 and December 2013. The cohort consisted of women undergoing fresh cycle in-vitro fertilisation (IVF) or intracytoplasmic sperm injection (ICSI) treatment at the unit during this time period. Since June 2012, thyroid function testing and thyroid peroxidase antibody (TPOab) testing has been offered routinely to all women being seen in the ACU as part of a national multi-centre randomised controlled trial called TABLET. All women who agreed to have their thyroid function and TPO antibody tested gave written consent to allow their data to be included in research studies linked to their thyroid function blood test, prior to their blood samples being taken. Furthermore, the Birmingham Women's Hospital research governance department granted ethical approval allowing access to patient notes for data extraction.

Eligibility criteria

The eligibility criteria for this study followed that of the TABLET trial (as discussed in chapters 5 and 6). Women were offered the screening blood test if they were aged between 16-41 years, not known to be on treatment for thyroid dysfunction in the past or present, not known to have any cardiac disorders and not currently taking amiodarone or lithium.

Classification of SCH

The reference range for biochemical euthyroidism was set according to Roche Diagnostics manufacturer's recommended reference ranges published in 2004, which was derived from 269 healthy non-pregnant females of the reproductive age (20-39years). This range was 0.44mU/L – 3.63mU/L with a Free T4 of 10.0 – 21.0 pmol/L (2.5^{th} to 97.5^{th} centile for both). This is the accepted reference range used for the TABLET trial.

Selective allocation of LT4 treatment

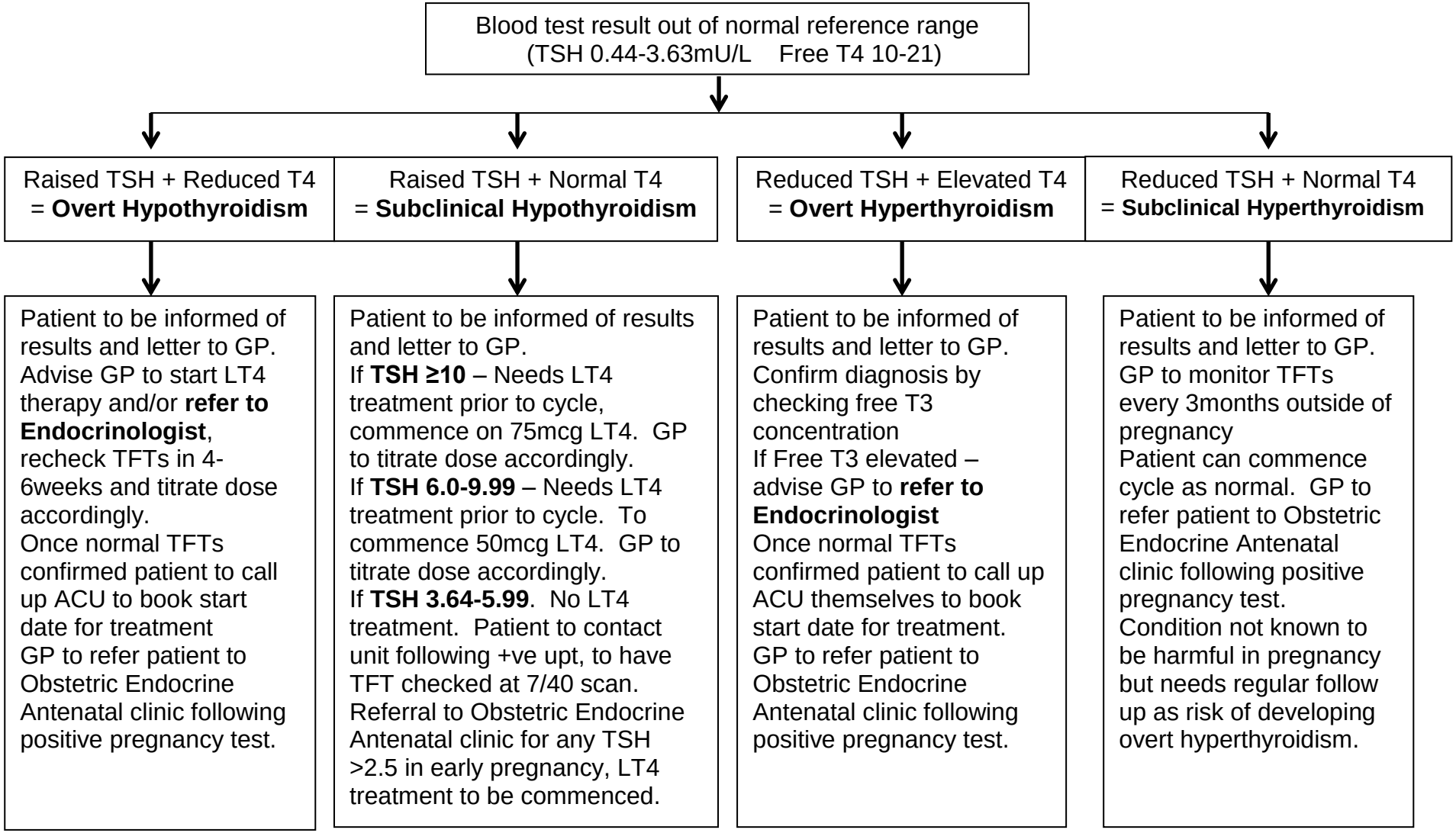
Prior to the commencement of this study, screening for thyroid dysfunction and/or thyroid antibody testing was not routine practice within the Birmingham Women's Hospital fertility centre. Consequently, a local guideline for the fertility unit was created regarding the management of any abnormal thyroid function results detected incidentally. The decision was made to classify differing severities of SCH arbitrarily based on the concentrations of TSH into "mild", "moderate" and "severe". In order to be categorised as having SCH, the free T4 level had to be within the normal range. Mild SCH was defined as a TSH between 3.63-5.99mU/L; moderate 6.00-9.99mU/L and severe >10mU/L. It was agreed that for any patient with a TSH above the commonly accepted upper limit of 4.5mU/L, who presented with symptoms of hypothyroidism, that they should be started on LT4 treatment. TPO antibody testing was also performed and reported as: negative ≤59iu; indeterminate 60-99iu; positive ≥100iu. Treatment of LT4 was dependent on TSH concentrations alone, with or without the presence of TPO antibodies. This was based on the current lack of evidence for whether LT4 therapy improves outcomes for women with thyroid

autoimmunity. The results of the on-going large TABLET trial are eagerly awaited to help answer this important clinical question. Women with “mild” SCH who would not be receiving supplementary LT4 treatment prior to their fertility treatment, would have their TFT rechecked at their routine 7 week viability scan and would be commenced on LT4 treatment, if necessary, to maintain the TSH at <2.5mU/L in the first trimester.

Figure 35 displays the management guideline implemented into the fertility unit from June 2012.

NB: All women commenced on LT4 treatment pre-conception were not to undergo any IVF treatment until their TFTs were shown to be within the normal reference range. All women taking LT4 pre-conception were advised to double their dose on two days of the week (Monday & Friday) following a positive pregnancy test and then the TFT would be checked at the routine 7 week viability scan (empirical dose increase). This was in line with American Thyroid Association recommendations based on the findings of the THERPY” trial[134].

Figure 35. Management Pathway for abnormal thyroid function in BWH Assisted Conception Unit



Statistical analysis

The baseline patient characteristics, cycle characteristics and outcome data were described giving frequencies with percentages (to one decimal place) for categorical outcomes with chi-squared test analysis to test for differences. For continuous variables, means with standard deviations were calculated and differences were analysed using students t-test. To estimate the contribution of thyroid function to live birth rate (defined as the birth of one of more living infants) and clinical pregnancy (defined as the presence of a gestational sac on ultrasound), univariate and multiple logistic regression analyses were performed to calculate odds ratios and corresponding 95% confidence intervals. An enter technique was used for multiple logistic regression. Covariates were preselected when they had a known effect on IVF outcome. Variables added to the model were: age, body mass index, duration and cause of infertility, basal FSH, previous history of IVF, previous live birth and previous miscarriage. Subgroup analyses were performed comparing the euthyroid women (TSH 0.44-3.63mU/L) and untreated SCH women (TSH 3.64-5.99), as well as comparing a stricter euthyroid cut off (TSH 0.44-2.5mU/L) against a larger untreated group (TSH 2.5-5.99mU/L).

Data were analysed using SPSS (ver. 21.0; SPSS Inc., Chicago).

Results

Overall study population

A total of 944 women seen in the infertility setting provided written consent to thyroid function and TPO antibody (TPOAb) testing between June 2012 and December 2013. Of this 944, a total of 560 underwent IVF or ICSI treatment by December 2013. Of this 560, 36 women were eligible for randomisation into the TABLET trial (euthyroid and TPOAb positive). These 36 women were removed from the analysis as we could not un-blind their treatment allocation due to the trial being still in the recruitment phase. A summary flow diagram outlining the patient selection in the study is shown in Figure 36. Amongst women who underwent IVF treatment, 11.7% had subclinical hypothyroidism. Of these, 68.3% had a TSH of 3.63-5.99mU/L and the remainder had a TSH equal to or greater than 6.0mU/L. The latter group were those who received LT4 treatment prior to commencing their IVF treatment.

SCH vs. euthyroid

The baseline characteristics of the cohort are displayed in Table 28, and the cycle characteristics and outcome data are displayed in Tables 29 and 30.

Figure 36. Summary of patients in the study

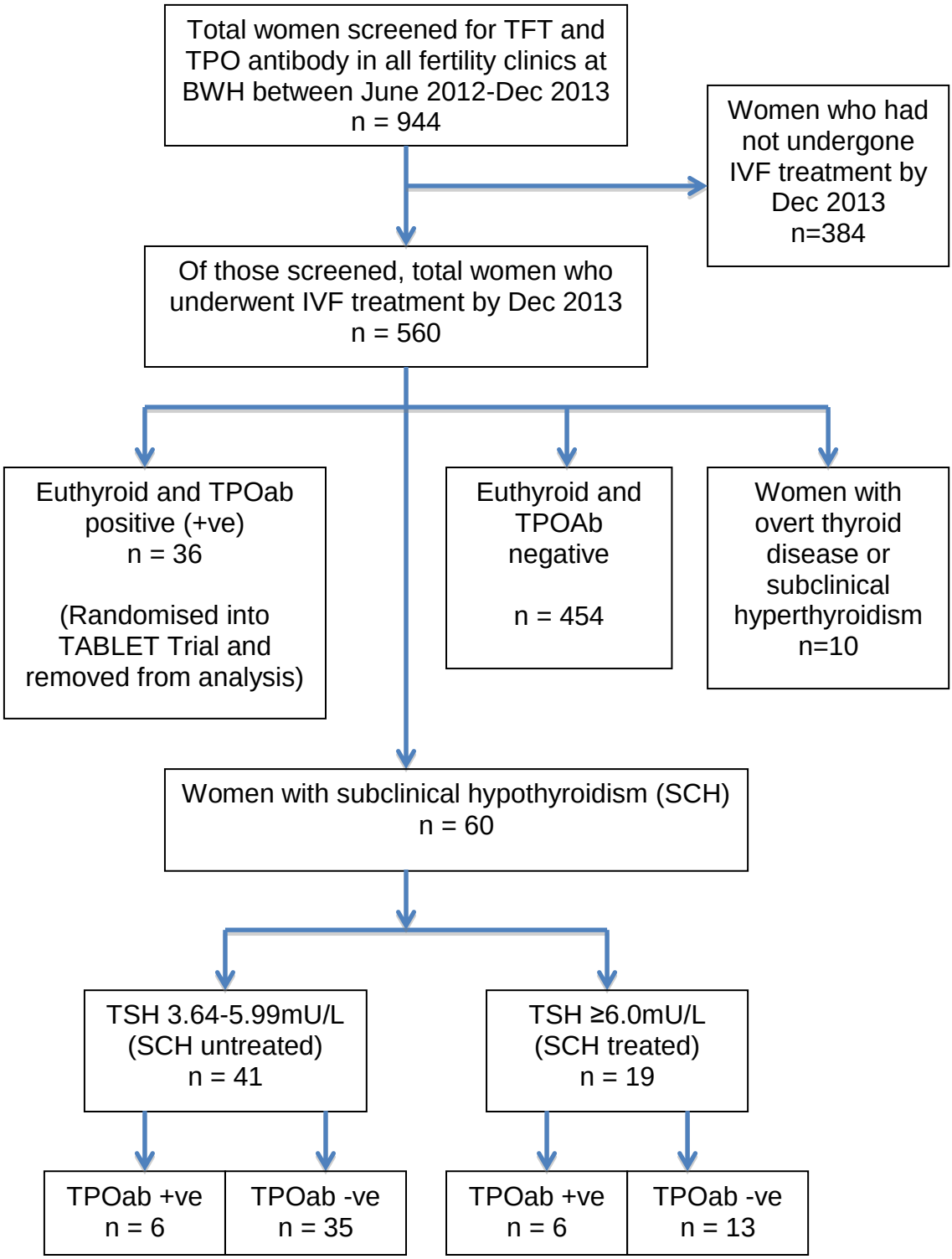


Table 28. Baseline characteristics of the cohort

	n (%) *or* Mean (±SD)		
	Women with SCH (n=60)	**Euthyroid women (n=454)**	**P value**
Age	31.9 (4.5)	32.5 (4.2)	0.3
BMI	25.8 (4.3)	25.1 (4.2)	0.2
Ethnicity			
White	27 (45.0%)	221 (48.7%)	0.7
Asian	28 (46.7%)	183 (40.3%)	0.5
Black	2 (3.3%)	24 (5.3%)	0.5
Chinese	2 (3.3%)	7 (1.5%)	0.3
Mixed	0 (0)	5 (1.1%)	0.8
Other	1 (1.7%)	14 (3.1%)	0.6
Mean TSH concentration (mU/L)	5.83 (3.56)	1.78 (0.70)	n/a
Mean T4 concentration (pmol/L)	14.3 (2.5)	15.5 (1.8)	**0.001**
Average cycle length (if regular)	25.2 (9.5)	24.8 (10.3)	0.8
Duration of infertility (in completed years)	3.9 (2.7)	4.3 (3.0)	0.2
Day 2 FSH	7.4 (2.2)	7.5 (2.6)	0.7
Fibroids	1 (1.7%)	20 (4.4%)	0.3
History of previous IVF treatment	14 (23.3%)	122 (26.9%)	0.7
Parity	0.2 (0.4)	0.2 (0.5)	--

Table 28. continued

	n (%) *or* Mean (±SD)		
	Women with SCH (n=60)	**Euthyroid women (n=454)**	**P value**
History of previous miscarriage	16 (26.7%)	118 (26.0%)	0.9
History of previous live birth	8 (13.3%)	86 (18.9%)	0.4
Cause of infertility			
Male factor	32 (53.3%)	213 (46.9%)	0.6
Tubal factor	5 (8.3%)	88 (19.4%)	0.08
Endometriosis	6 (10.0%)	44 (9.7%)	0.9
Anovulation	4 (6.7%)	25 (5.5%)	0.7
Diminished ovarian reserve	2 (3.3%)	27 (5.9%)	0.4
Unexplained	9 (15.0%)	79 (17.4%)	0.7
Other	0 (0%)	2 (0.4%)	0.8
Type of treatment			
IVF	25 (41.7%)	219 (49.2%)	0.6
ICSI	33 (58.3%)	226 (50.8%)	0.7

Table 29. Cycle characteristics of the study cohort

	Women with SCH (n=60)	Euthyroid women (n=454)	P value
Cycle data			
No. of oocytes retrieved	9.0 (4.2)	10.4 (5.4)	**0.02**
No. of mature oocytes inseminated	7.4 (3.5)	8.4 (4.5)	**0.05**
Total no. of embryos	1.2 (0.4)	1.3 (0.5)	0.1
Fertilisation rate*	0.65 (0.19)	0.63 (0.23)	0.5
No. of embryos transferred			
0	1 (1.7%)	17 (3.7%)	0.4
1	45 (75.0%)	321 (70.7%)	0.8
2	14 (23.3%)	116 (25.6%)	0.8
Day of embryo transfer	3.6 (1.4)	3.5 (1.4)	0.6
Blastocyst transfer	29 (48.3%)	194 (42.7%)	0.6
No. of embryos frozen	1.6 (2.1)	1.2 (1.8)	0.2

Table 30. Outcome data for study cohort

Outcome data	SCH (n=60) (Treated and untreated)	Euthyroid (n=454)	Odds ratio	P value
Implantation rate**	0.41 (0.5)	0.38 (0.5)	1.08 (0.63-1.86)	0.6
Biochem preg.	29 (48.3%)	210 (46.3%)	1.04 (0.65-1.68)	0.9
Clinical preg.	23 (38.3%)	189 (41.6%)	0.92 (0.55-1.53)	0.8
Miscarriage[a]	4 (17.4%)	24 (12.8%)	1.26 (0.42-3.76)	0.7
Live birth[b]	19 (31.7%)	165 (36.3%)	0.87 (0.50-1.50)	0.6
Singleton[b]	17 (28.3%)	148 (32.6%)	0.87 (0.49-1.53)	0.6
Twins[b]	2 (3.3%)	17 (3.7%)	0.89 (0.20-3.95)	0.9
Gestation (Mean±SD)	38.8 (1.7)	39.0 (1.3)		**0.01**

*Fertilisation rate = no. of embryos / no. of eggs inseminated

** Implantation rate = no. of foetal hearts on scan / no. of embryos transferred

[a] Expressed as a percentage of all clinical pregnancies

[b] Expressed as a percentage of all cycles

The data presented in Table 30 shows that there are no significant differences in the outcomes for women undergoing IVF treatment who are either euthyroid or have subclinical hypothyroidism diagnosed preconception (regardless of treatment received). The only statistically significant difference noted from the baseline characteristics was found in the mean T4 concentration, which showed that women with SCH have a lower mean T4 than women who are euthyroid; in-keeping with what would be expected physiologically. The other results of statistical significance were for number of oocytes retrieved and number of mature oocytes inseminated, both appeared to be significantly lower in women with SCH than in euthyroid women (p=0.02 and p=0.048 respectively). Despite this, there were no differences in fertilisation rate, implantation rate, or live births. Gestation at which the baby was delivered was statistically significantly higher for euthyroid women compared with women with SCH, however there were no births below 37weeks for any women who delivered in the study. When conducting a multivariate analysis adjusting for age, BMI, ethnicity, day2 FSH, cause of infertility, duration of infertility, previous history of IVF, previous miscarriage and previous live birth the results remained unchanged, there was no statistically significant difference noted between the two groups for the primary outcome of live birth (OR 0.94 (0.71-2.24) p=0.6).

Untreated SCH vs. euthyroid (TSH 0.44-3.64mU/L)

To answer the question of whether untreated SCH pre-conceptually could result in adverse outcomes compared with euthyroid women, we compared those with a TSH 0.44-3.64 to the untreated SCH group (TSH 3.64-5.99), the results are shown in Table 31.

Table 31. Outcome data comparing euthyroid (TSH<3.64) and untreated SCH (3.64-5.99)

	Mean (±SD) Or n (%)			
	Untreated SCH TSH 3.64-5.99 (n=41)	**Euthyroid TSH 0.44-3.64 (n=454)**	**Odds ratio**	**P value**
Outcome data				
Implantation rate	0.44 (0.5)	0.38 (0.5)		0.5
Biochemical preg	22 (53.7%)	210 (46.3%)	1.16 (0.67-2.00)	0.6
Clinical preg[a]	18 (43.9%)	189 (41.6%)	1.05 (0.59-1.88)	0.9
Miscarriage[b]	3 (16.7%)	24 (12.8%)	1.38 (0.40-4.79)	0.6
Live birth[b]	15 (36.6%)	165 (36.3%)	1.01 (0.54-1.87)	1.0
Gestation	39.0 (1.2)	39.0 (1.3)		--

[a] Expressed as a percentage of all clinical pregnancies

[b] Expressed as a percentage of all cycles

There were no statistically significant differences in fertility outcomes for those with untreated SCH pre-conception and the euthyroid women with TSH 0.44-3.64mU/L.

Untreated SCH vs. euthyroid TSH (0.44-2.5mU/L)

To further explore the data and help answer the question of whether a stricter cut off TSH concentration of 2.5mU/L shows any adverse outcomes in those untreated; we compared the euthyroid women TSH 0.44-2.5mU/L with women with a TSH concentration between 2.5-5.99mU/L, the results are shown in Table 32.

Table 32. Outcome data comparing TSH <2.5 and untreated SCH (2.50-5.99)

	Mean (SD) Or n (%)			
	Untreated SCH TSH 2.50-5.99 (n=119)	**Euthyroid TSH <2.50 (n=376)**	**Odds ratio**	**P value**
Outcome data				
Implantation rate	0.44 (0.5)	0.38 (0.5)		0.3
Biochemical preg	57 (47.8%)	175 (46.5%)	1.03 (0.72-1.48)	0.9
Clinical preg	52 (43.7%)	163 (43.4%)	1.01 (0.69-1.47)	1.0
Miscarriage[a]	8 (6.8%)	18 (4.9%)	1.40 (0.60-3.31)	0.4
Live birth[b]	44 (36.9%)	145 (37.8%)	0.96 (0.65-1.42)	0.8
Gestation	38.8 (1.8)	39.0 (1.2)		0.3

[a] Expressed as a percentage of all clinical pregnancies

[b] Expressed as a percentage of all cycles

The results show no statistically significant differences in fertility outcomes between the stricter threshold euthyroid group and the broader “untreated SCH” group.

Discussion

Main findings

The aim of this study was to determine if fertility outcomes, primarily live birth, were different for euthyroid women compared with women with subclinical hypothyroidism (SCH). It is important to note that the euthyroid women were all TPOAb negative, as those who were TPOAb positive were randomised into the TABLET trial. Those with subclinical hypothyroidism were composed of TPOAb positive and negative women. The group of women with subclinical hypothyroidism comprised of those who were treated and those who were untreated; this was dependent on the TSH concentration. The purpose of this distinction was to explore the safety of allowing a higher threshold of TSH concentration, challenging the widely accepted upper limits, before treating subclinical hypothyroidism in pre-conception women. The results of the study showed there were no significant differences in fertility outcomes for women with SCH (both treated and untreated) compared with euthyroid women; there were also no significant differences when using a stricter reference range for euthyroid (TSH 0.44-2.5mU/L).

The results showed that having a raised TSH appeared to reduce the total number of oocytes collected as well as the number of mature oocytes inseminated, despite no difference in ovarian reserve (basal FSH) or rates of anovulation to potentially explain this. This reduced response to ovarian stimulation and potentially reduced oocyte quality seen in women with SCH may suggest there is a pre-conception defect compared to euthyroid women. If this were the case then only choosing to treat SCH in pregnancy would miss this pre-conception window, presuming there is underlying

thyroxine related aetiology. Despite the reduced number of oocytes and mature oocytes inseminated seen in women with SCH this did not impact on any of the other important fertility outcomes; fertilisation rate, clinical pregnancy rate, miscarriage rate or live birth compared to euthyroid women.

Of the 41 women who had SCH but were untreated pre-conceptually, 15 women went on to have successful live births, giving a live birth rate of 37%, which is just above the national average for women undergoing fertility treatment and similar to the euthyroid TPO negative women in the study. Seven (47%) of those who had a successful pregnancy had a persistently raised TSH greater than 2.5mU/L when checked at 7weeks gestation. A potential explanation of why more than half the group had a TSH fall to <2.5mU/L in early pregnancy is probably due to a combination of factors. Firstly, we used a different analyser for the TFTs performed after the initial screening. This was because the local trust (Birmingham Women's Hospital) uses a Beckman analyser, however, for the purpose of the TABLET Trial (i.e. for the screening test) a Roche analyser had to be used and so these samples were sent to a neighbouring hospital (Queen Elizabeth Hospital). The Beckman analyser has no positive bias for TSH unlike the Roche. From our data on the same first trimester samples analysed on both Beckman and Roche it appears as though a TSH of 2.5mU/L on Beckman is about equivalent to 4.0mU/L on a Roche in the first trimester (personal communication from Dr Shiao Chan). Secondly, another reason to help explain why 8/15 women had lower TSH levels in early pregnancy compared to pre-conception, could be due to the physiological fall in TSH that has been observed in previous studies in the first 10weeks of pregnancy (with rising levels of

hCG). Cross-sectional studies by Dashe et al[135] and Cotzias et al[136] showed that the upper limit of TSH (defined as the 97.5th population centile) at the start of pregnancy is between 4.94-5.09mU/L and then drops to 3.0mU/L by about 10weeks with no intervention.

All seven women who were untreated pre-conception (TSH 3.64-5.99mU/L), but received treatment from 7weeks onwards, went on to have successful live births and they all maintained euthyroidism on LT4 treatment. The baseline characteristics (i.e. age, BMI, basal FSH etc.) and cycle characteristics (i.e. no. of oocytes retrieved, no. of mature oocytes inseminated etc.) for these seven women were compared to the other eight women who did not require treatment in their pregnancies, to look for any potential predictors of who may need treatment in early pregnancy. No statistically significant differences were detected; however this was most likely due to the small numbers. Of these seven women who required treatment in their pregnancy, two patients were also thyroid peroxidase antibody positive. In total there were six women who were TPOAb positive and had TSH values between 3.64-5.99mU/L (i.e. untreated SCH); four of these women did not conceive. Given that the two women who were TPO positive and did conceive both required LT4 treatment in pregnancy, this could suggest that being TPOAb positive with SCH does make it more likely for the TSH to go up in pregnancy, however we cannot make any strong conclusions due to the very small numbers.

Just over half the women who became pregnant, and did not receive LT4 treatment prior to conception, had TSH levels <2.5mU/L in early pregnancy. In the context of

the reference ranges used in this study, our findings show that there is a plausible argument for not routinely commencing LT4 treatment pre-conception for all women with a TSH 3.64-5.99mU/L and to recheck and treat in early pregnancy (if necessary) instead. Furthermore there is evidence that has shown there are higher risks of congenital defects associated with thyroxine medication. As mentioned in the introduction, a large study by Browne et al of over 14,000 cases found that periconceptual thyroxine medication was significantly associated with several congenital birth defects[130]. The findings of our study, and the evidence to suggest thyroxine medication may cause congenital defects, supports the notion of not empirically treating all cases of SCH.

Given that the seven women with untreated SCH who did require LT4 treatment in early pregnancy all went on to have successful live births, this provides some support for the case of not treating SCH pre-conception, but rather monitoring and managing from early pregnancy instead.

Strengths and limitations

One of the advantages of this study is the prospective design and rigorous data collection; there were no missing entries for any variables. As thyroid function testing was previously not routinely performed in the Birmingham Women's assisted conception unit it lead to the development of a novel guideline, allowing us to challenge the growing consensus on treatment of pre-conception subclinical hypothyroidism in subfertile women.

The main limitation of our cohort study is the small sample size, this may account for the non-significant differences found in outcomes between those with SCH and those without, resulting in a Type 2 error. Due to these small numbers, reflected by the wide confidence intervals associated with the odds ratios seen in the analyses, we cannot draw strong conclusions from this study. It may be that no differences are seen owing to the small numbers, and that using a larger sample size may reveal significant differences.

One of the concerns regarding subclinical hypothyroidism is impaired neurological development in the foetus[137]. However, the long-term impact on the child, including neurodevelopmental effects, was not determined by our study. A landmark trial, published in the New England Journal of Medicine, studied thyroid function and IQ scores of children at three years of age[90]. A total of 794 women with Thyrotropin levels above the 97.5th percentile and free T_4 levels below the 2.5th percentile were randomised from around 12-13weeks in pregnancy[90]. This study showed no benefit in the use of Levothyroxine in improving cognitive function in children born to hypothyroid women (as detected in pregnancy) compared to the children of untreated women[90]. Consequently this leads us to question what value Levothyroxine treatment would have for improving childhood neurodevelopmental outcomes for children born from women with subclinical thyroid disease if no apparent benefit is seen with children born from women treated for overt disease.

A final limitation is that women who were euthyroid pre-conception (TSH 0.44-3.63mU/L) and those who had an optimal TSH in the 1st trimester range (i.e. <2.5mU/L) did not have their thyroid function repeated later in pregnancy to see if it

remained in euthyroid range. The reason for this was that no clinical benefit was thought to be gained from re-testing known euthyroid individuals throughout their pregnancy and the consensus amongst the clinicians when creating the initial guideline was to only repeat blood test for abnormal results, as per normal practice.

Comparison to existing literature

The current NICE guidelines state that measurement of thyroid function in asymptomatic women with infertility, as the only isolated risk factor, should not be offered routinely[16,138]. This is due to the lack of strong evidence regarding adverse IVF and pregnancy outcomes for women with subclinical thyroid disease and also the lack of evidence proving any benefit from treating such women pre-conception.

The findings of our cohort study suggest that it *may* be safe to allow a higher threshold of subclinical hypothyroidism before considering treatment for pre-conception subfertile women; however this conclusion must be interpreted with caution given the small sample size of the study. This "higher threshold" is of course an arbitrary value but based on our work we can say that a TSH concentration up to 6.0mU/L may be safe to leave untreated as long as the thyroid function test is checked again in early pregnancy; or if a woman does not become pregnant it should be re-checked in around 3months time. This is based on what we know of the natural progression of SCH with studies reporting progression to overt disease in 2-5% and reversal to normal in as many as 62% after 5 years follow up[139,140].

Only three trials to date have looked at the effect of LT4 treatment in subfertile women with subclinical hypothyroidism[127–129]. The upper limits of TSH used in the studies by Negro et al, Abdel-Rahman et al and Kim et al were 4.2mU/L, 4.2mU/L and 4.0mU/L respectively[127–129]. When the results of these studies were meta-analysed it showed that LT4 supplementation in women with SCH resulted in a significant increase in delivery rates and a decrease in miscarriage compared to women who did not receive treatment. This finding is in contrast to our cohort study, which showed no difference in live birth or miscarriage rates between the treated and untreated women; furthermore we used a higher threshold for treatment (TSH ≥6.0mU/L). In addition, our study found that having a raised (and untreated) TSH appeared to reduce the total number of oocytes collected as well as the number of mature oocytes inseminated, however the meta-analysed data of the three trials showed no difference between women treated and women untreated for these outcomes. One of the reasons for the discrepancies in our results compared with those of the meta-analysis is that our study is limited by sample size and so may be underpowered to show any differences; 60 women with SCH in our cohort vs. 220 women in the meta-analysed data. It is evident that larger numbers are required to accurately determine whether pre-conception LT4 treatment of subclinical hypothyroidism can improve outcomes for women undergoing assisted reproduction.

Conclusions and future work

This cohort study utilised thyroid function tests taken from patients being screened for the national TABLET trial. This trial is continuing recruitment until December

2015. The aim will be to collect all the outcome data for the women screened in the fertility setting and add to this analysis to give greater power and examine any differences in more detail. Ethical approval has been obtained for this. We anticipate that a further 1500 women will have been screened in the fertility setting in the Birmingham Womens Hospital by December 2015. By the end of December 2015 we would anticipate a total of 1700 euthyroid women and 240 with SCH; providing us with the largest cohort study of this kind. Using live birth rate as the primary outcome, to see a minimally important difference of 5% between euthyroid women and women with SCH (37% vs. 32%) this sample size will have 57% power with an alpha of 0.05. In order to reach 80% power the sample size would have to increase by 1.8 times to 2890 euthyroid women and 432 women with SCH to detect the 5% difference between 32% and 37% for an alpha of 0.05. This will be possible through collaboration with the other large fertility units participating in the TABLET trial (i.e. Guy's Hospital London, St Bartholomews Hospital London and St Marys Hospital Manchester). These centres would have all prospectively collected thyroid function tests on several thousand pre-conception subfertile women, we have permission from the corresponding principal investigators for each site and the local research and development units to have access to these results and the necessary fertility outcome data. Although this collaboration will provide a greater sample size, due to the individual units having differing management pathways for treating SCH, collating the results will be problematic. What is needed is a large, appropriately powered multi-centre randomised controlled trial, including women from the fertility setting, to investigate if pre-conception Levothyroxine supplementation in women with subclinical hypothyroidism improves fertility and pregnancy outcomes.

Furthermore, the decision on whether to use the strict upper limit for TSH of 2.5mU/L or the currently recognised 4.5mU/L would also need to be considered carefully.

It is important for us to continue work in this area to understand the implications of subclinical hypothyroidism in subfertile women so that we can make better informed decisions on whether to treat or not.

CHAPTER 8

Discussion

There are many factors that can affect IVF outcome, the work of this thesis has focused solely on pre-treatment factors and the impact of certain important under-investigated factors on IVF outcome. The three key factors explored in this thesis were ethnicity, body-mass index (BMI) and thyroid function. All three factors are common, easily available and contribute important information that can impact on decision making for both patients and clinicians before achieving a pregnancy.

Ethnicity and IVF outcome; summary of findings and future work

The relationship between ethnicity and assisted reproduction has been disputed for some time with various studies showing conflicting results. The results of the systematic review, cohort study and updated meta-analysis presented in chapters 2 and 3 provide robust evidence for the hypothesis that there is an association between ethnic background and IVF success. The work of this thesis shows that Black and South Asian women have a statistically significant reduced chance of live birth following fresh IVF treatment (when compared with White women) and the commonly known confounders do not explain this. For Black women, the odds of having a clinical pregnancy are also reduced when compared to White women. However, for South Asian women it appears that there are no differences in clinical pregnancy rates compared to White women. This suggests that South Asian women are more likely to suffer miscarriage compared to White women. Interestingly, for the Black population, when the frozen cycles were analysed separately from the fresh cycles the results showed no difference in live birth or clinical pregnancy rates compared with White women. This poses the question of whether Black women would perform better if an “elective freeze” approach was used; i.e. all embryos to be

routinely frozen and implanted at a later date. This suggestion of elective freeze for all women is currently proving popular amongst reproductive medicine clinicians and was discussed at this year's European Society for Human Reproduction and Embryology (ESHRE) conference in Lisbon. Consequently, the findings of the work in this thesis should certainly prompt clinicians to consider trialling this change in practice, certainly in the first instance with Black women. Furthermore, the findings of this study should prompt investigation into the mechanisms underpinning the disparities seen between ethnic groups; thus allowing for modification of laboratory and or clinical practice to improve IVF outcome for all ethnicities. Finally, there also needs to be careful consideration of whether information regarding ethnicity and its potential affect on IVF outcome should be routinely provided to patients as part of pre-treatment counselling. Although this is not a factor that women are able to change, it may still have implications on their decision-making. One must be cautious when providing this information to patients as there is an argument to state that if NHS funding criteria are dictated by predictors that are associated with poorer chances of success (i.e. age >38, BMI >30 and smoking) then why should ethnicity not be included? Given the potential controversy that may surround counselling on the basis of ethnicity, we would advise that this should be a decision to be made by individual clinicians when counselling their patients.

BMI and IVF outcome; summary of findings and future work

The relationship between raised BMI and adverse pregnancy outcomes is well known[13], however there remains debate over the association between BMI and IVF outcomes. The aim of the work in this thesis was to determine to what degree BMI

affects IVF outcome, and the interplay of this with other factors. This was investigated by the creation of a prediction model to estimate the chances of live birth following IVF, incorporating BMI as a predictor. To date, successful prediction of live birth after assisted reproductive technology (ART) has been limited and so far no model has accounted for BMI. We developed a novel model, which encompasses key prognostic factors that have not previously been used; such as ovarian reserve and ethnicity. The association between BMI and live birth following IVF, was strong when looked at in the univariate analysis; however this association became non-significant when other important confounders were adjusted for in the final model. Despite this, we have demonstrated that there is still value in counseling women to lose weight, as shown in chapter 4 there was around a 5% reduction in chance of live birth (as calculated by our model) for women with a BMI >30 compared to women with a normal BMI. We believe this model, once converted into a user-friendly mobile application and or web-page, will hold an important role in the counseling and decision making for women at the critical decision-making point in their journey, i.e. *before* they embark on their first treatment cycle. Furthermore, this model provides a personalised approach to counselling and estimates chances of success based on easily measurable variables that are specific to the individual woman; rather than using success rates based on age-related national HFEA data. Future work will involve further external validation in the form of geographical validation and then conversion of the model into a mobile application and/or web-page for utility by clinicians and patients. Once this is undertaken, the next step would be to assess impact analysis; establish whether the prediction model improves decisions, in terms of quality or cost-effectiveness of patient care[69,70]. No existing IVF prediction models

have reached the impact analysis stage to date. As this model is limited to pre-treatment variables only, there is a need for future work to be conducted to explore models that incorporate treatment variables; thus allowing for adjusted calculations as the patient progresses through their treatment.

Thyroid function and autoimmunity; summary of findings and future work

The final aspect of the PhD focused on the prevalence of thyroid dysfunction and thyroid autoimmunity and also the effect of subclinical thyroid disease on IVF outcome. The work in these chapters also explored the relationship between thyroid dysfunction/autoimmunity and BMI and ethnicity, linking all three key factors within the thesis together. This prevalence work looked at a broader group of women actively trying for a pregnancy, including women with history of one or two miscarriages and women with recurrent miscarriage; as well as those with infertility.

At present there is debate over whether screening pre-conception women for thyroid disease should become universal. This is in view of the adverse pregnancy outcomes seen for women with thyroid disease (including both overt and subclinical). The argument for the case of universal screening is that these women may be missed if not routinely screened, and therefore would not receive the appropriate treatment required to potentially avoid or reduce negative pregnancy outcomes. This argument holds more strength for the cases of undiagnosed overt thyroid disease (as these are the group of women over whom there is no debate regarding treatment), however our study has shown this to be prevalent in only 0.38% of women. Does the financial and practical burden of screening all women to pick up 0.38% outweigh the

benefits from potentially reducing adverse obstetric outcomes for these women? The bigger question relates to the screening and or treatment of subclinical disease. Using an upper limit cut off of 2.5mU/L for subclinical hypothyroidism would result in around 20% of women potentially requiring monitoring and/or treatment of their thyroid in pregnancy, which could constitute a significant burden to the NHS. Subclinical hypothyroidism is linked to adverse pregnancy outcomes, as discussed in chapters 5 and 7; however the findings of our cohort study in chapter 7 did not show any differences in fertility or pregnancy outcomes between women with treated SCH (using Levothyroxine) and the untreated women. The main limitation of this cohort study was the sample size. The study will have a greater sample size when the whole dataset is analysed and will hopefully provide greater insight into the question; recruitment will continue until the end of December 2015.

Based on the findings of the prevalence work in chapters 5 and 6 our findings suggest that screening for thyroid disease would be more effective as selective “case finding” screening rather than routine. The “high risk” groups in particular depend on the thyroid disease. In general it appears that South Asian woman and women with a BMI above normal have the highest prevalence’s of subclinical hypothyroidism, and Black women have the highest rates of hyperthyroid disease (both subclinical and overt); therefore these groups should be considered for TFT testing prior to conception. Having said this, it may well be that South Asian women having higher TSH concentrations and Black women having lower TSH concentrations are in fact variants of normal. Further work needs to be conducted to explore this and to look at

developing population-based reference ranges to account for the relevant ethnic groups.

Finally, linked to thyroid disease is thyroid autoimmunity. The presence of TPOAb have strongly been linked to increased rates of miscarriage and pre-term birth[19]. However, as seen for subclinical thyroid disease there is limited conclusive evidence to suggest that there is any benefit in reducing these adverse outcomes using Levothyroxine treatment. Regarding pre-conception screening for TPOAb, there appeared to be a very subtle link between TPOAb positivity and increasing age, however no real cut off could be defined and the association was weak so we would conclude that age alone is not useful in determining which women should be screened. Regarding BMI, the results of our study have shown that the prevalence of TPOAb was significantly higher in women with BMI ≥35 compared to women with normal BMI; furthermore the mean BMI for TPOAb positive women was significantly higher than the mean BMI for TPOAb negative women. If evidence shows benefit in treating euthyroid women with TPOAb then based on the findings of our study it may be worth considering offering routine pre-conception TPOAb screening for women with BMI ≥35. Interestingly Black and Mixed race women appear to have the lowest prevalence of TPO antibodies and so routine TPOAb screening for these women would not be advised. There did not appear to be an association between TPOAb positivity and clinical originating population (i.e. whether the woman had a history of recurrent miscarriage or infertility) and so we would not advise specifically targeting these women in favour of women with a history of 1 or 2 miscarriages. Lastly, increasing TSH concentration appeared to be associated with increased probability;

in particular the probability of being TPOAb positive seems to increase with a TSH level above 2.5mU/L. Therefore, it may be worth considering TPOAb screening women who have TSH values >2.5mU/L; in particular it may be worth targeting South Asian women who are known to have higher TSH concentrations.

Ultimately, large randomised controlled trials (like the awaited TABLET trial) are needed to help answer the question of whether pre-conception treatment of subclinical thyroid disease and thyroid antibodies improves pregnancy outcomes when compared with no treatment. The findings of such studies will then guide decision-making on whether pre-conception thyroid screening should become more widely available.

Conclusion

Couples undergoing assisted reproductive technologies often experience great emotional and financial burden and the decision to undergo IVF can be challenging. It is therefore crucial that women are well informed about their chances of success and that they are appropriately stratified, investigated and managed before commencing their fertility treatment. The work presented in this thesis has adopted a mixed methodological approach to provide new information, as well as challenge existing evidence, regarding the impact of ethnicity and BMI on IVF outcome. Furthermore, it has lead to the development of a novel IVF counselling tool for implementation in clinical practice. And finally this work has provided new information (and disproved existing information) regarding the prevalence of thyroid dysfunction and autoimmunity in women of reproductive age in the UK, including

those undergoing assisted reproduction, to help guide screening programmes. The findings of this compilation of work will hopefully further educate patients and clinicians and aid effective counselling and clinical decision-making *prior* to achieving a pregnancy.

APPENDICES

Appendix 1. Complete case multivariate logistic regression model for live birth (n=2911)

	Parameter Estimates	Standard Error	P value	Odds Ratio	95% CI	
					Lower	Upper
Age						
≤36 years	-0.0234409	0.014	0.08	0.98	0.95	1.00
>36 years	-0.1860316	0.029	<0.001	0.83	0.78	0.88
Body mass index	-0.0094761	0.011	0.4	0.99	0.97	1.01
Cause of infertility:						
Male factor	0.1142310	0.108	0.3	1.12	0.91	1.38
Tubal factor	-0.1280070	0.123	0.3	0.88	0.69	1.12
Anovulation	-0.0852442	0.145	0.6	0.92	0.69	1.22
Unexplained	0.1503198	0.131	0.3	1.16	0.90	1.50
Other (e.g. Endo, fibroids)	0.0430051	0.105	0.7	1.04	0.85	1.28
Ethnicity:						
White	0			Reference		
Asian	-0.0818836	0.159	0.6	0.92	0.67	1.26
Black	-0.9654467	0.418	0.02	0.38	0.17	0.86
Chinese	0.7115788	0.569	0.2	2.04	0.67	6.22
Other	-0.7086007	0.398	0.08	0.49	0.23	1.07
Not stated	0.1056809	0.419	0.8	1.11	0.49	2.53
Mixed	-0.0308901	0.207	0.9	0.97	0.65	1.46

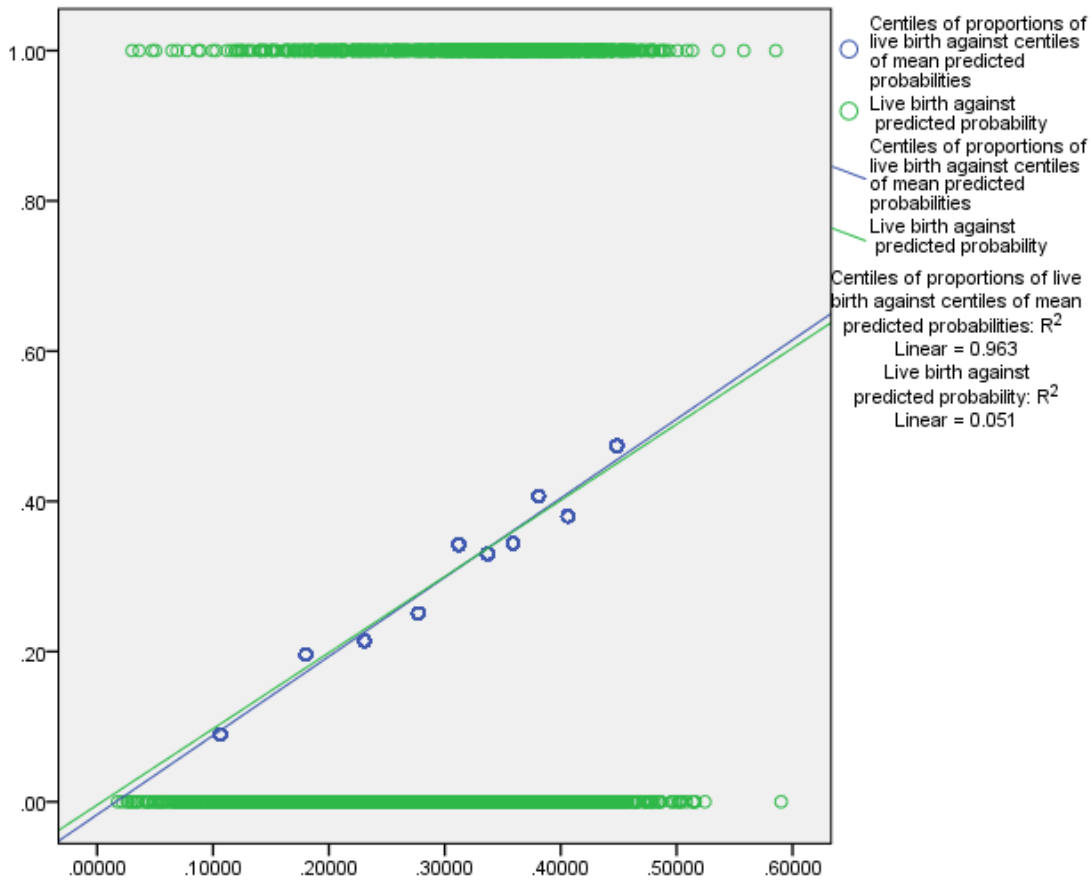
Appendix 1 continued

	Parameter Estimates	Standard Error	P value	Odds Ratio	95% CI	
					Lower	Upper
Previous Live Birth						
No	0			Reference		
Yes	0.3073074	0.123	0.01	1.36	1.07	1.73
Previous Miscarriage						
No	0			Reference		
Yes	-0.0329350	0.106	0.8	0.97	0.79	1.19
AFC	0.0366093	0.009	<0.001	1.04	1.02	1.06
AFC (squared)	-0.0004876	0.000	<0.001	1.00	1.00	1.00
Duration of infertility:						
0-4 years	0			Reference		
5≤ years	-0.0653509	0.031	0.04	0.94	0.88	1.00
Constant	-0.1601944	0.547	0.8	0.85		

Appendix 2. Calibration and internal validation for complete dataset prediction model (n=2911)

Predicted probability of live birth

Observed probability of live birth plotted against predicted probability of Live Birth for each tenth of predicted probability.



This model was internally validated using a bootstrapping technique. The optimism adjusted c-statistic (AUROC curve) for this was 0.62 and the optimum adjusted calibration slope was 1.18.

Appendix 3. Prevalence study consent form

Thyroid AntiBodies and
LEvoThyroxine Study

Blood Screening Consent Form

Please initial boxes

I confirm that I have read and understand the participant screening information sheet dated 26/3/2012 version 4.0 for the above study. I have had the opportunity to consider the information, ask questions and these have been answered satisfactorily. ☐

I agree to provide a blood sample for thyroid antibody and thyroid function testing. ☐

I understand that the thyroid test results and data collected at screening will be anonymised, and looked at by researchers at The University of Birmingham, and I give my permission for these individuals to have access to my anonymised information. ☐

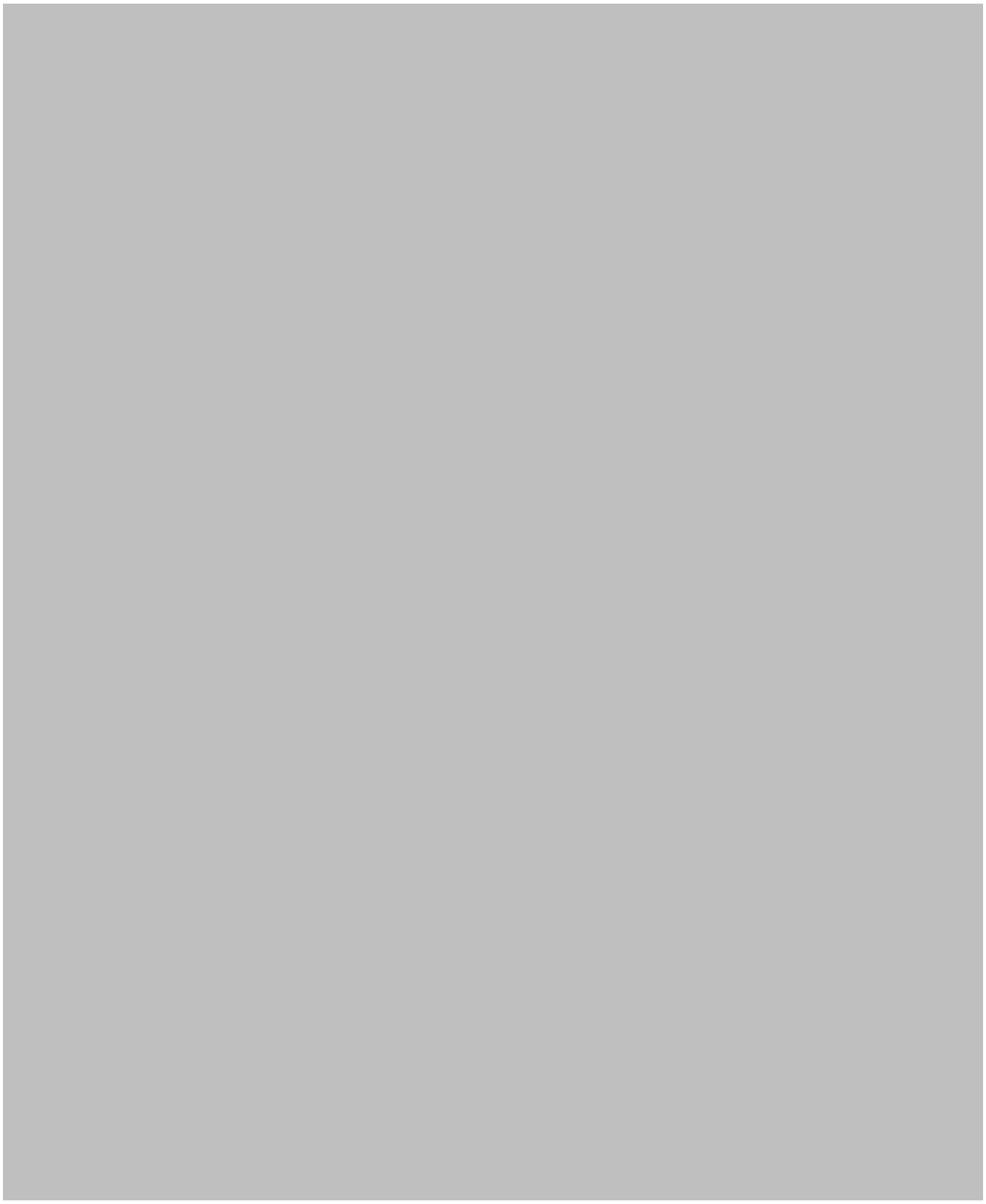
I understand that sections of any of my medical notes may be looked at by responsible individuals from the research team, regulatory authorities or from the NHS Trust where it is relevant to my taking part in research. I give permission for these individuals to have access to my records. ☐

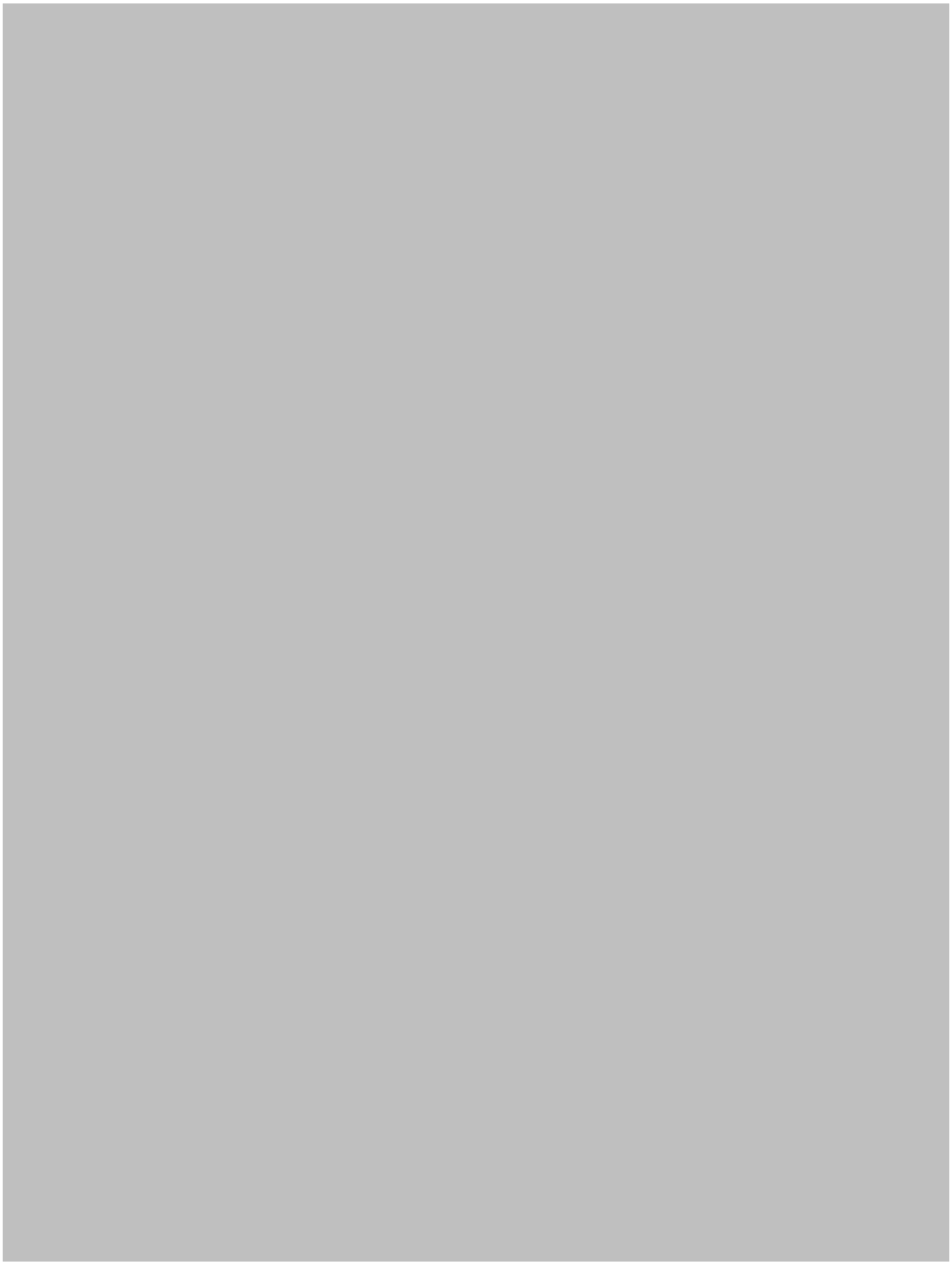
I understand that my participation is voluntary and that I am not obliged to take part in the subsequent trial, and that my medical care or legal rights will not be affected. ☐

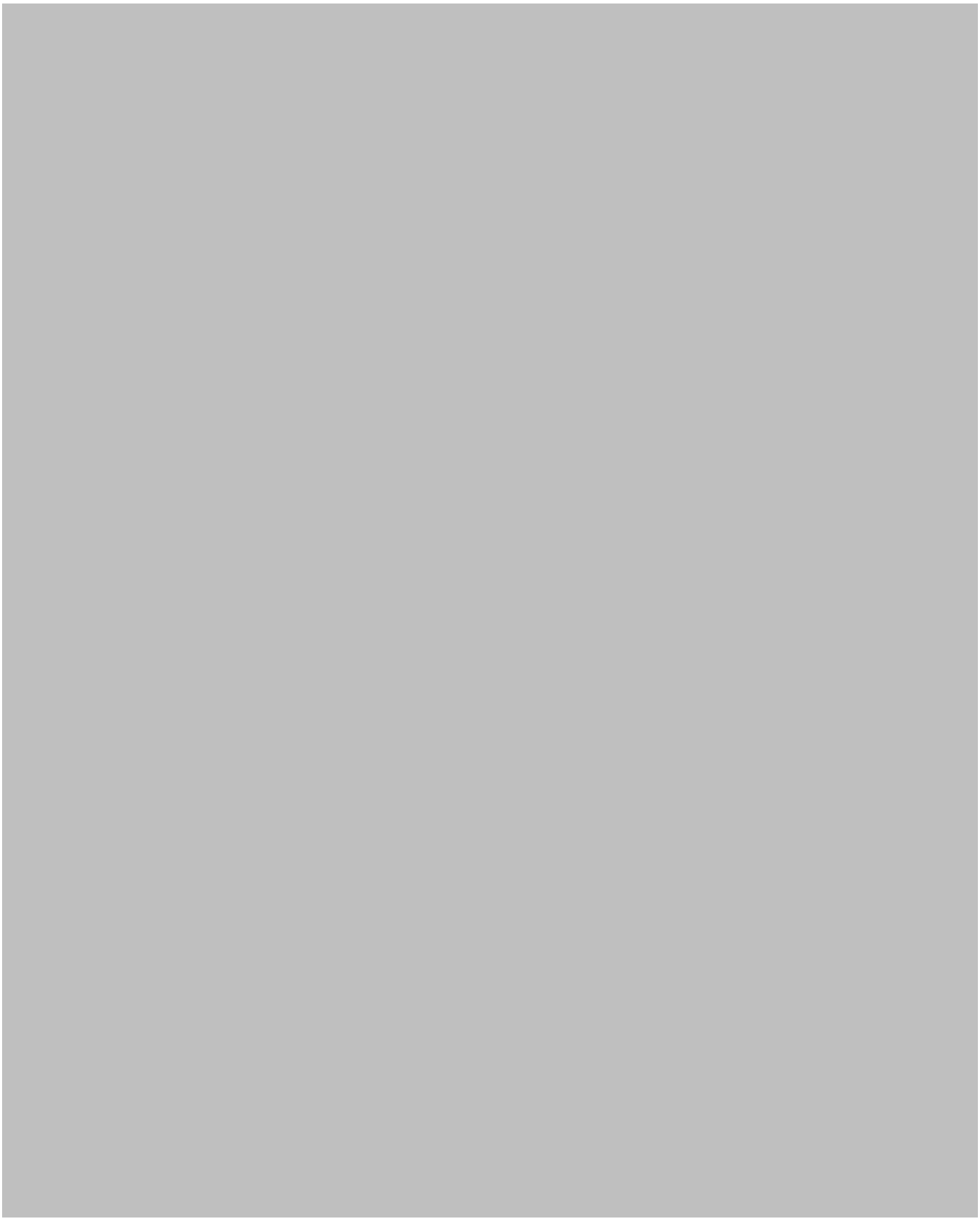
----------	_____	_____
Name of Patient	Date	Signature
----------	_____	_____
Name of Researcher	Date	Signature

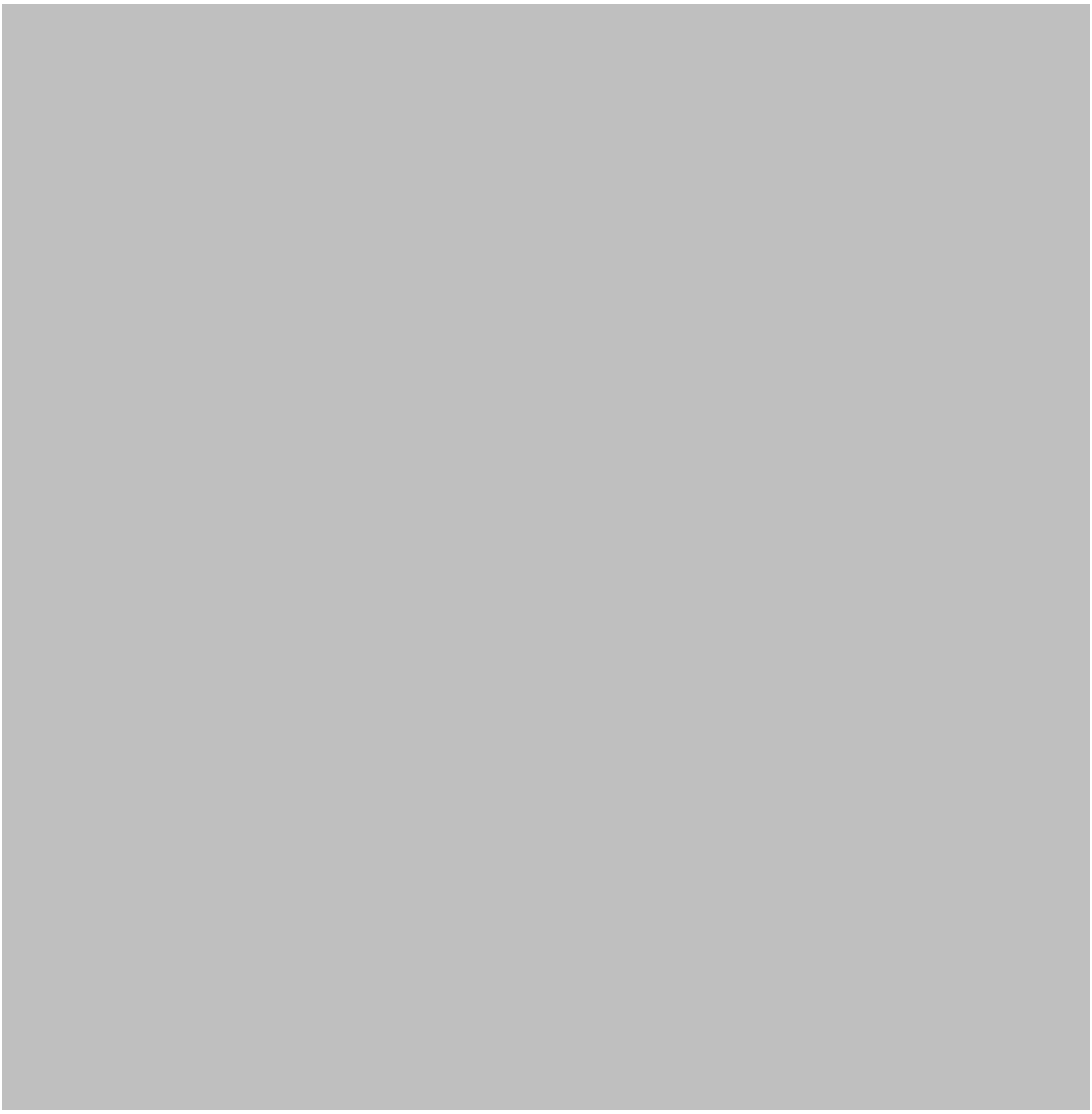
3 Copies of consent Forms: 1 copy for patient, 1 copy for site file, 1 copy to be kept in patient's hospital notes

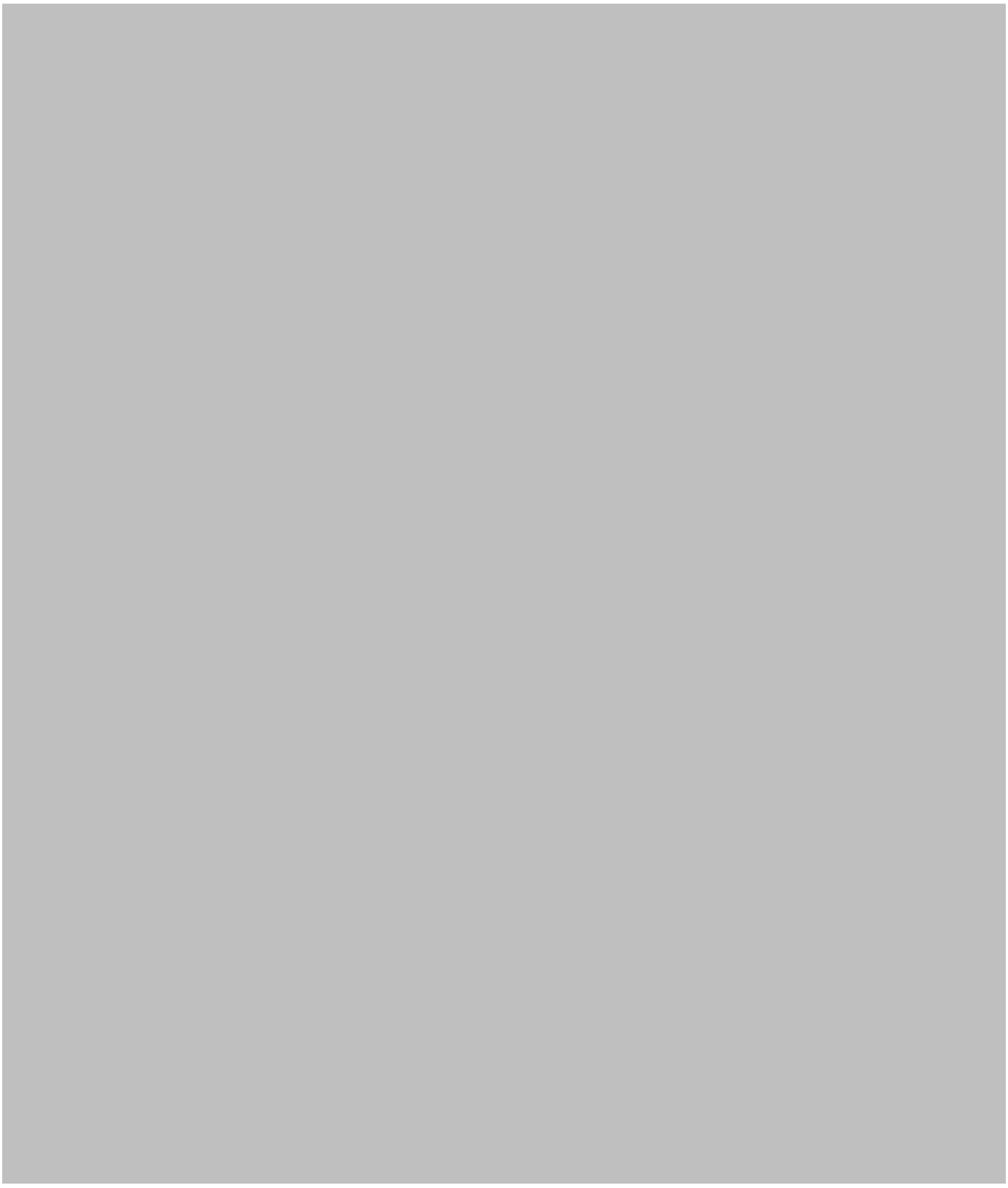
Appendix 4. Prevalence study protocol

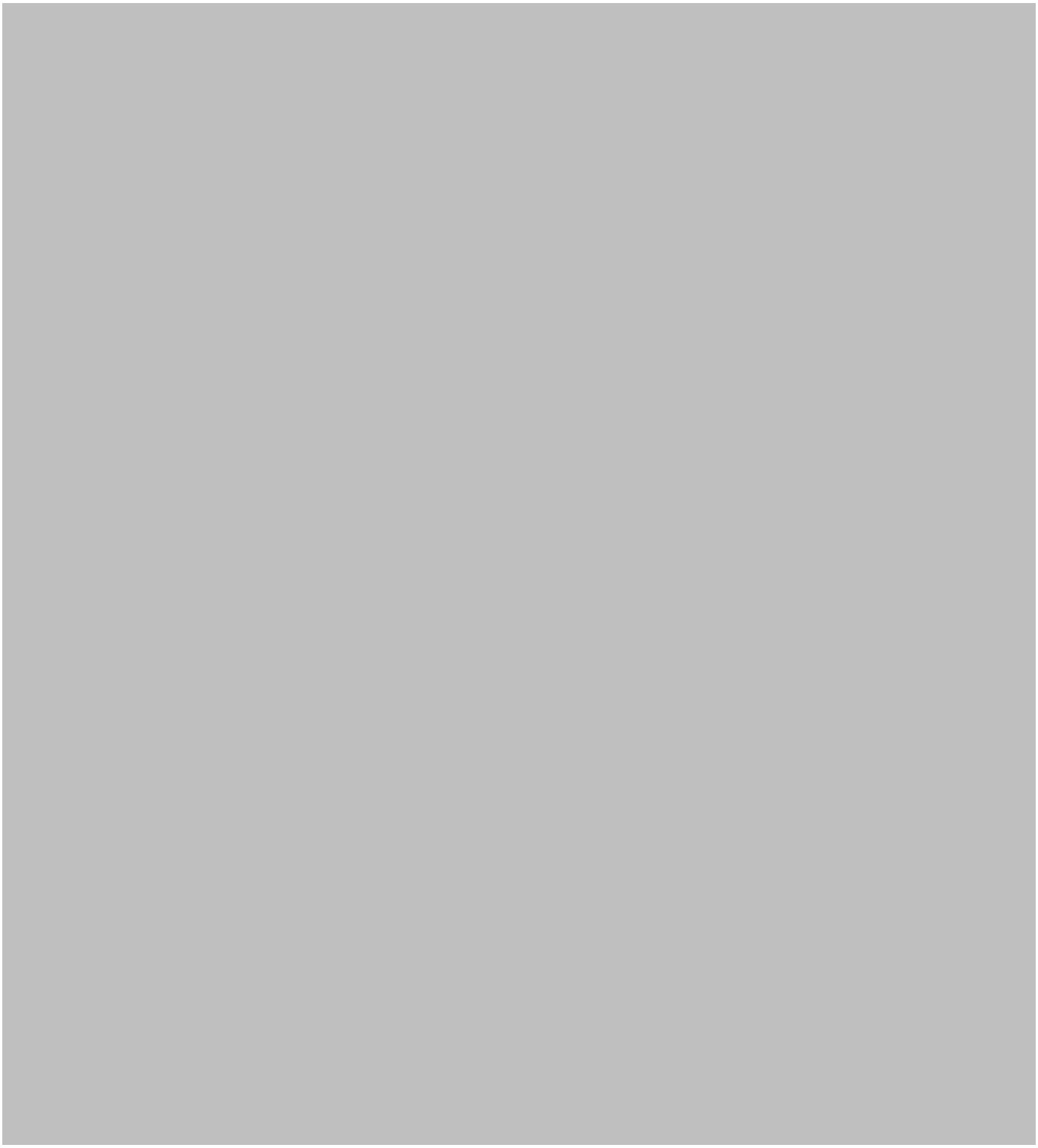


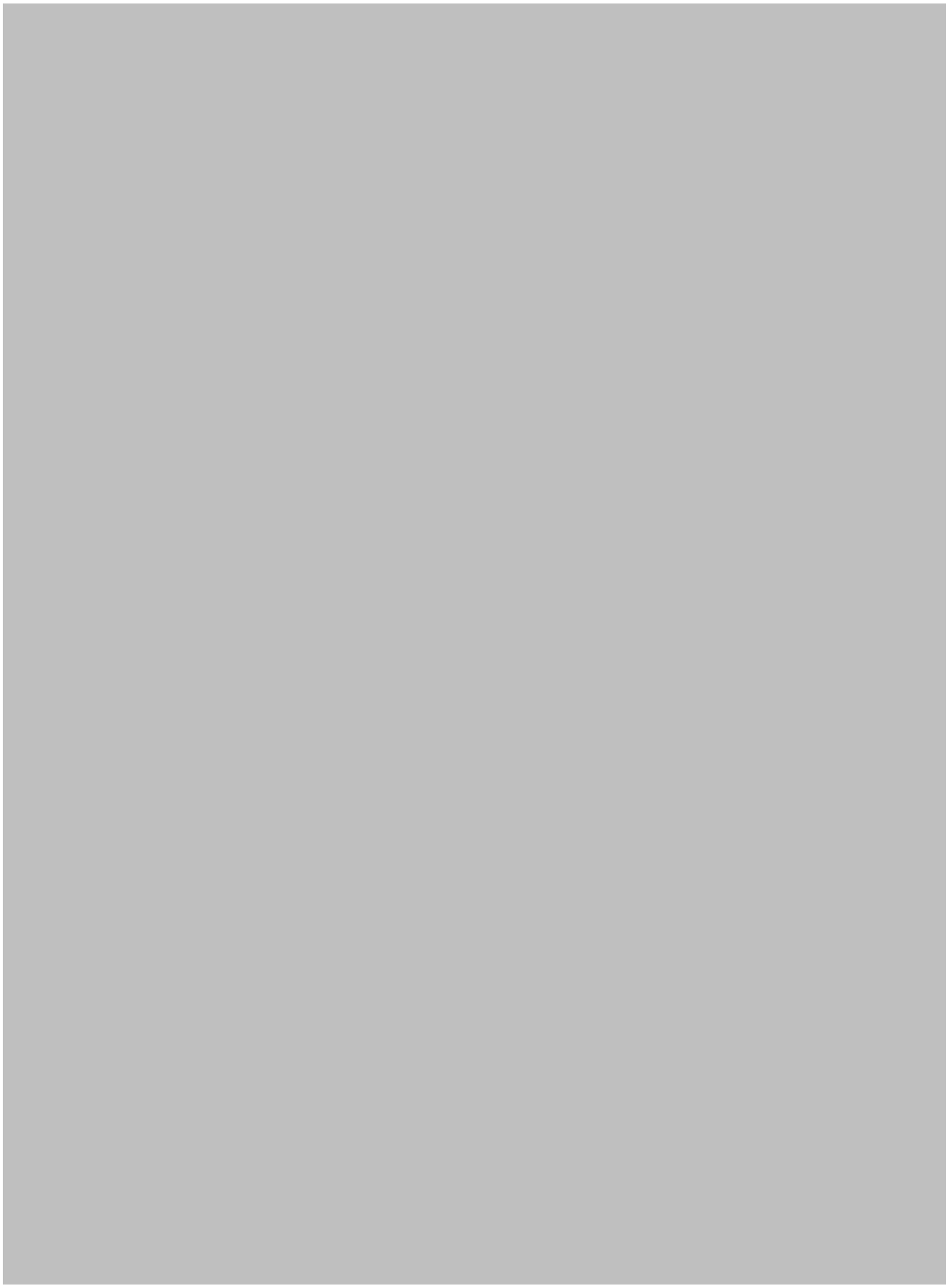


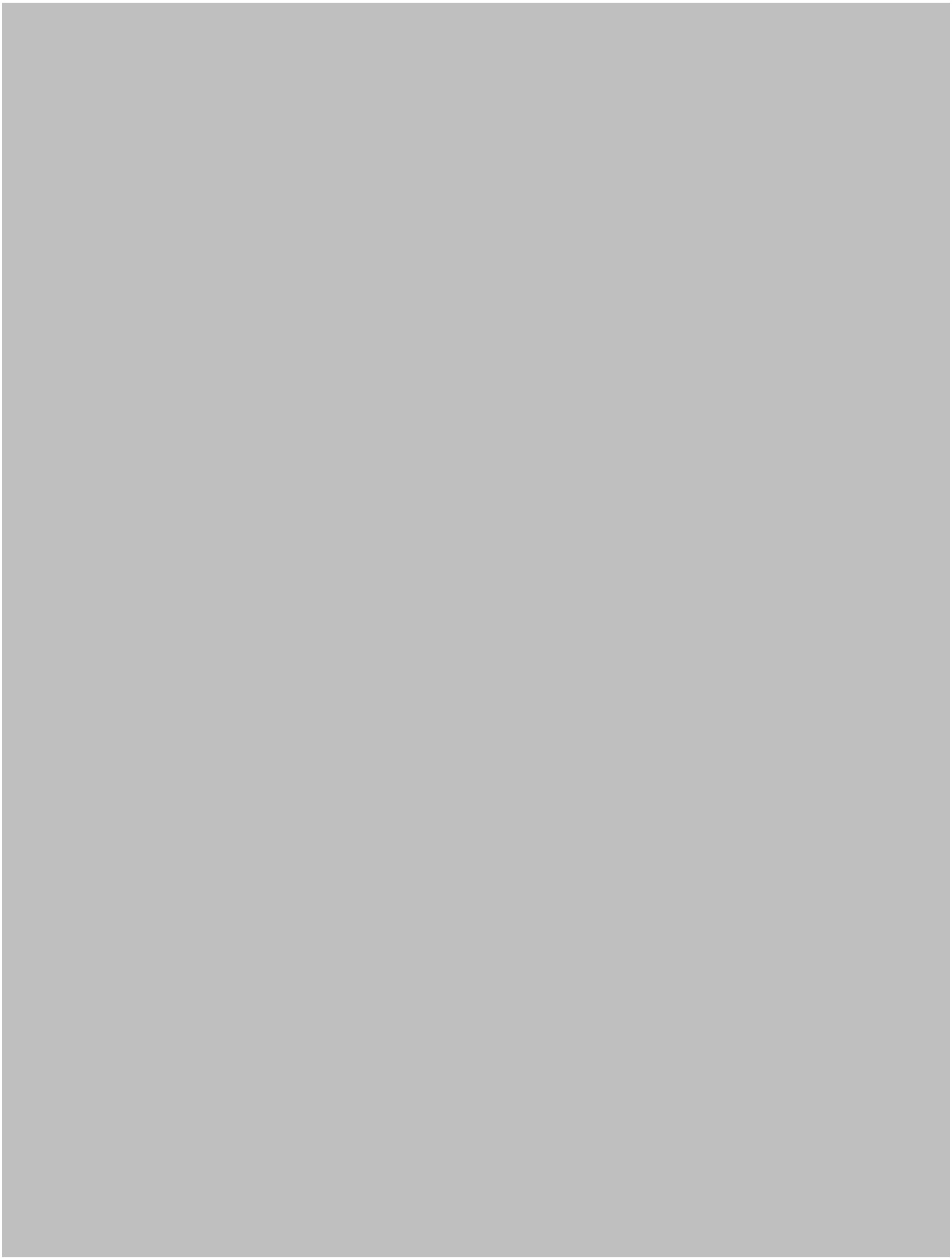


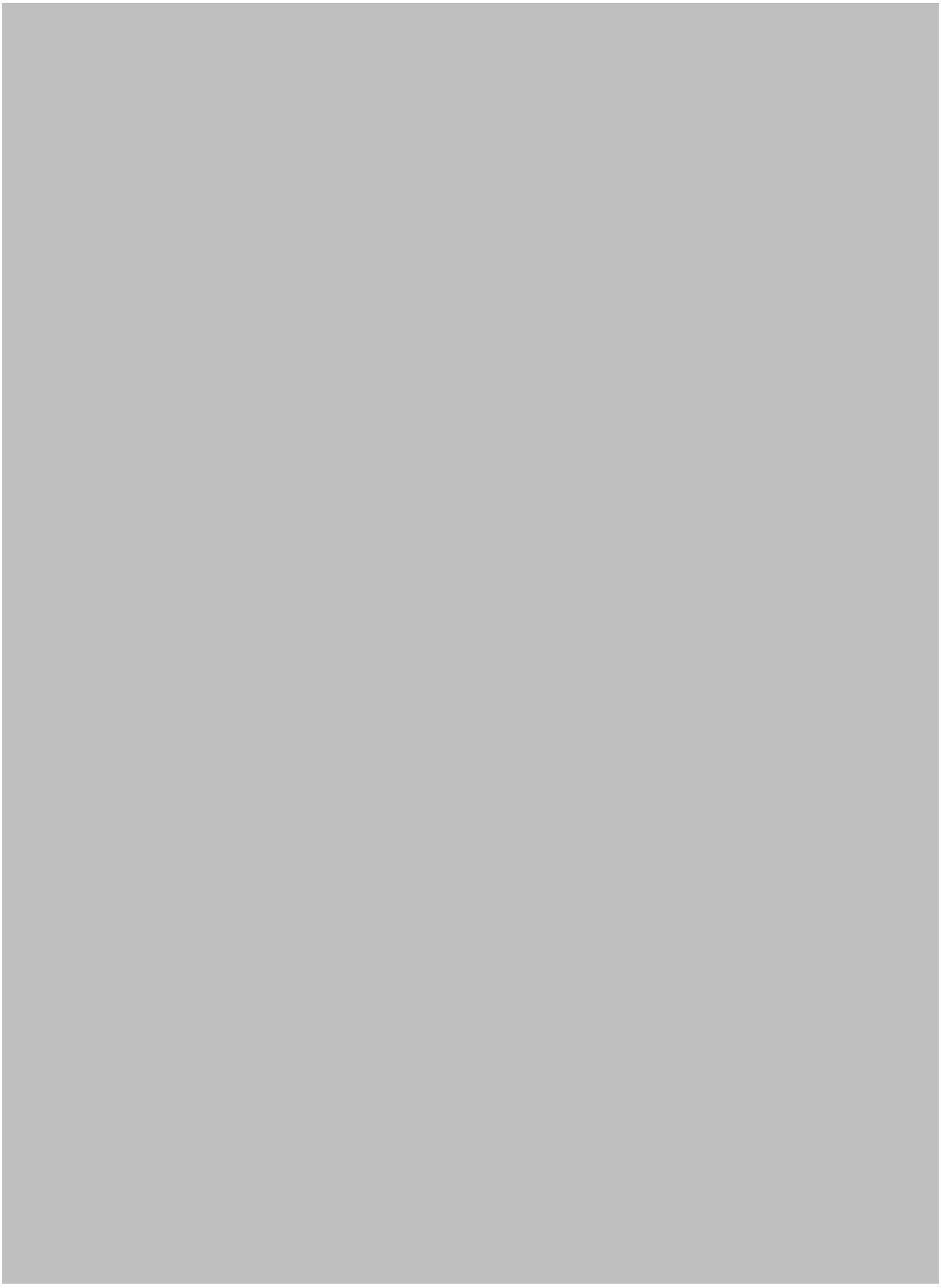


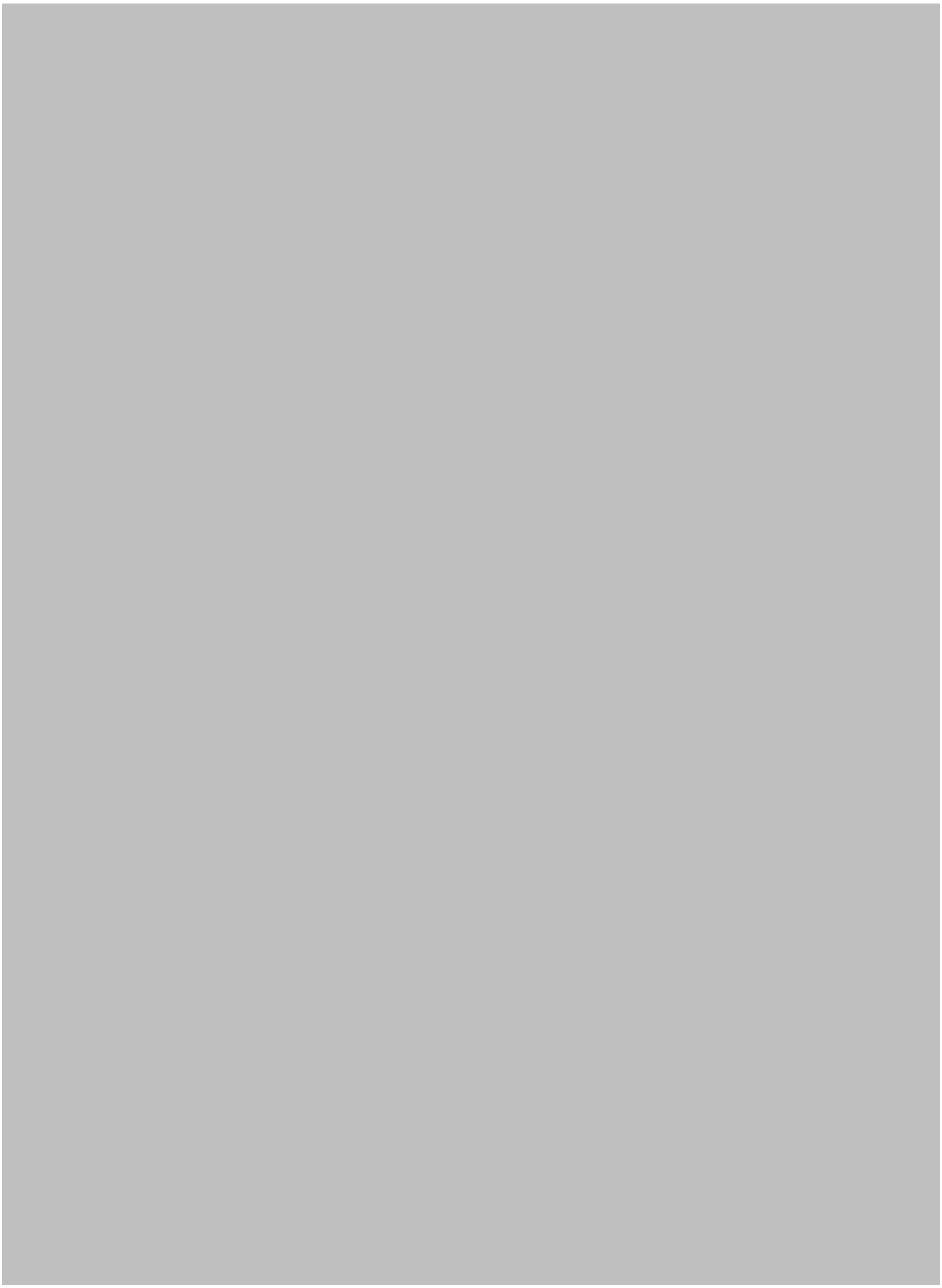


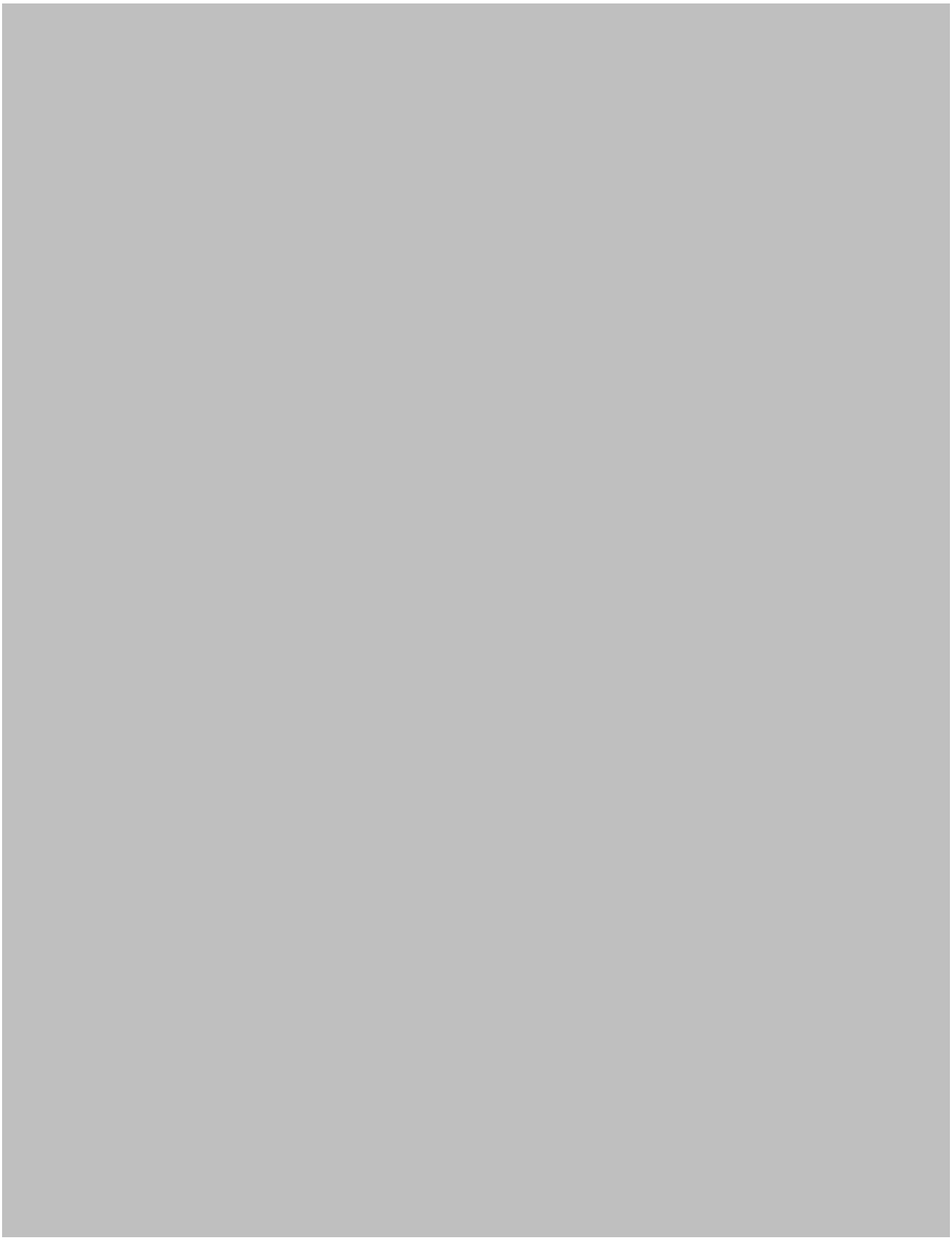


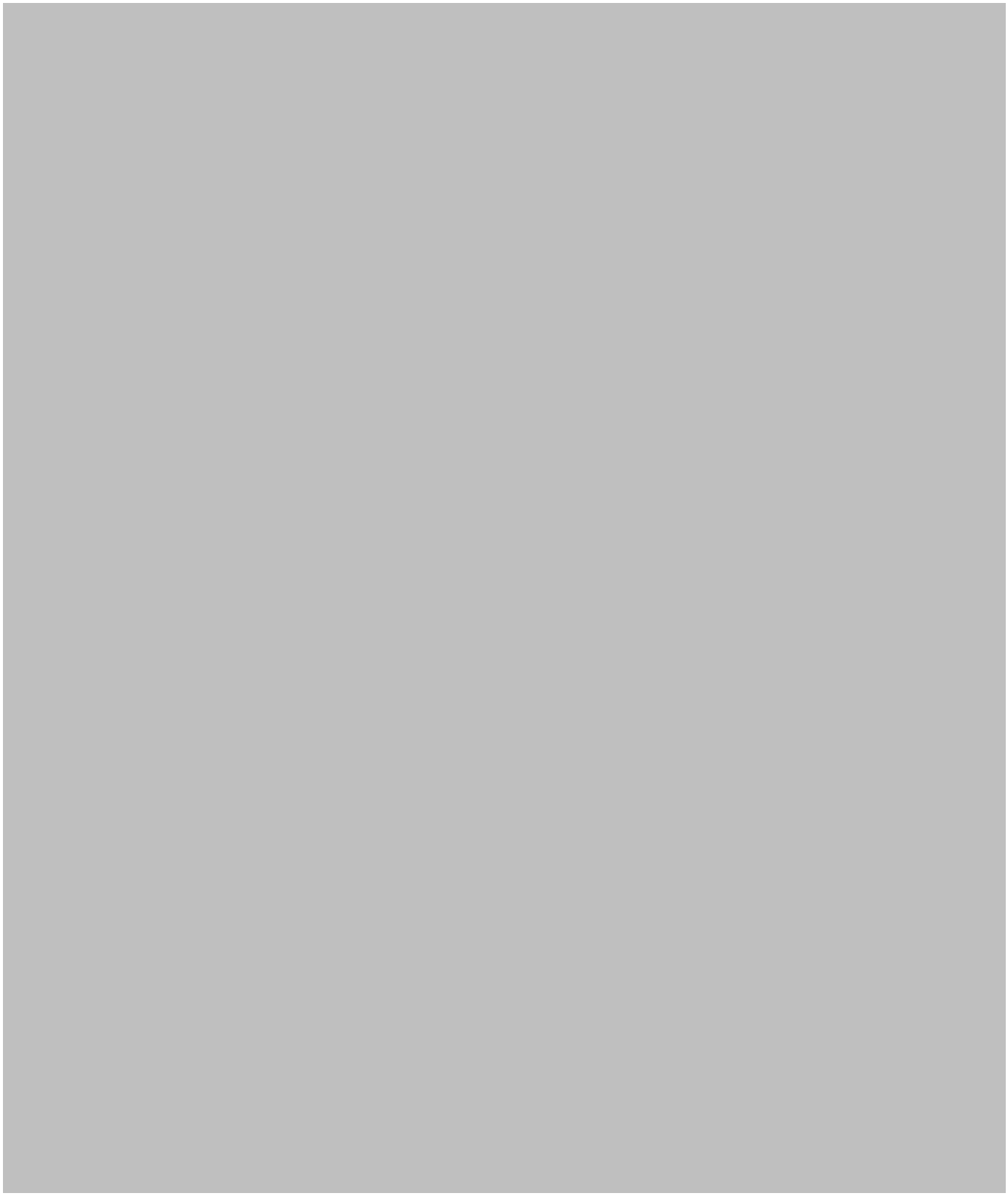


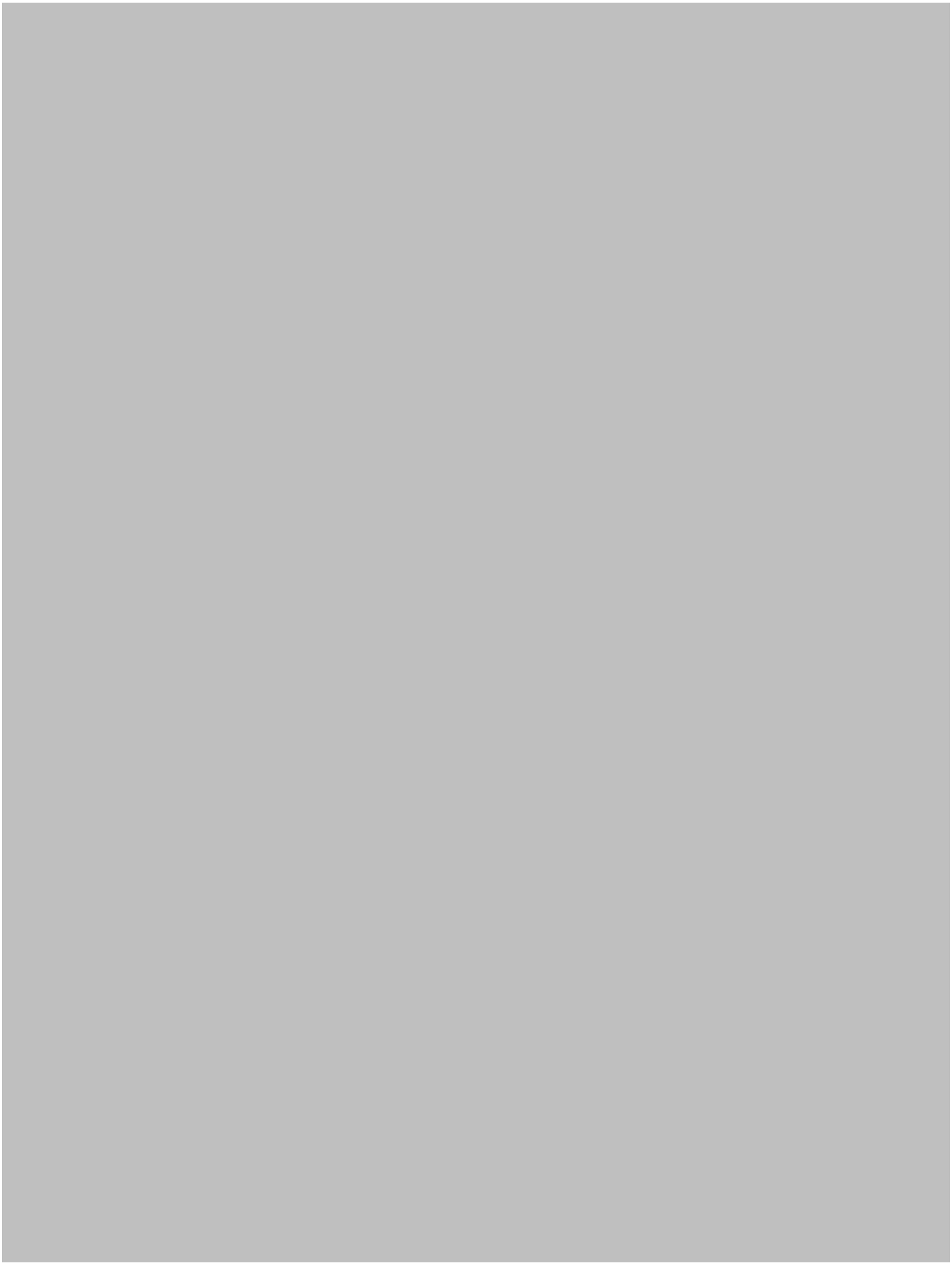




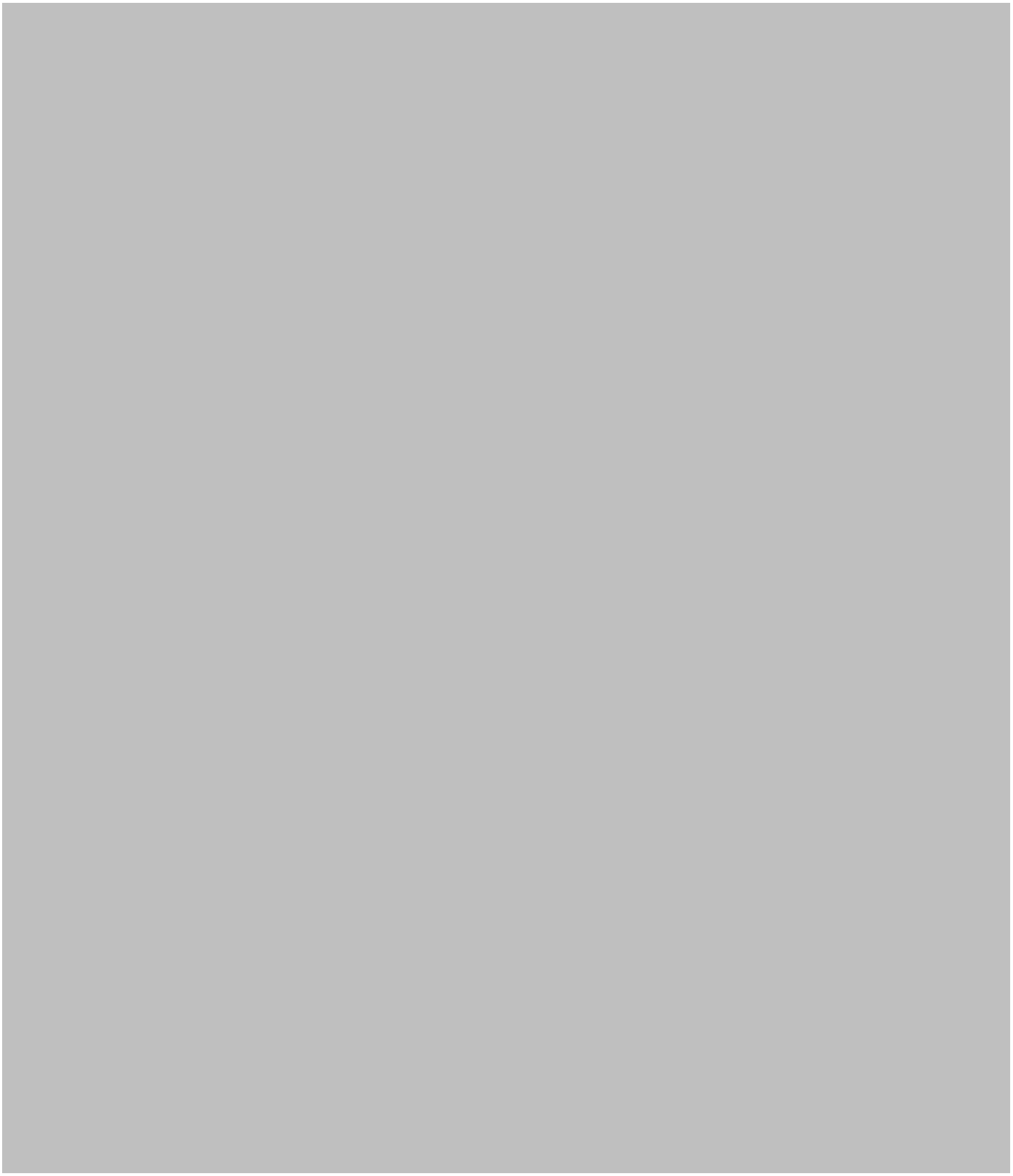


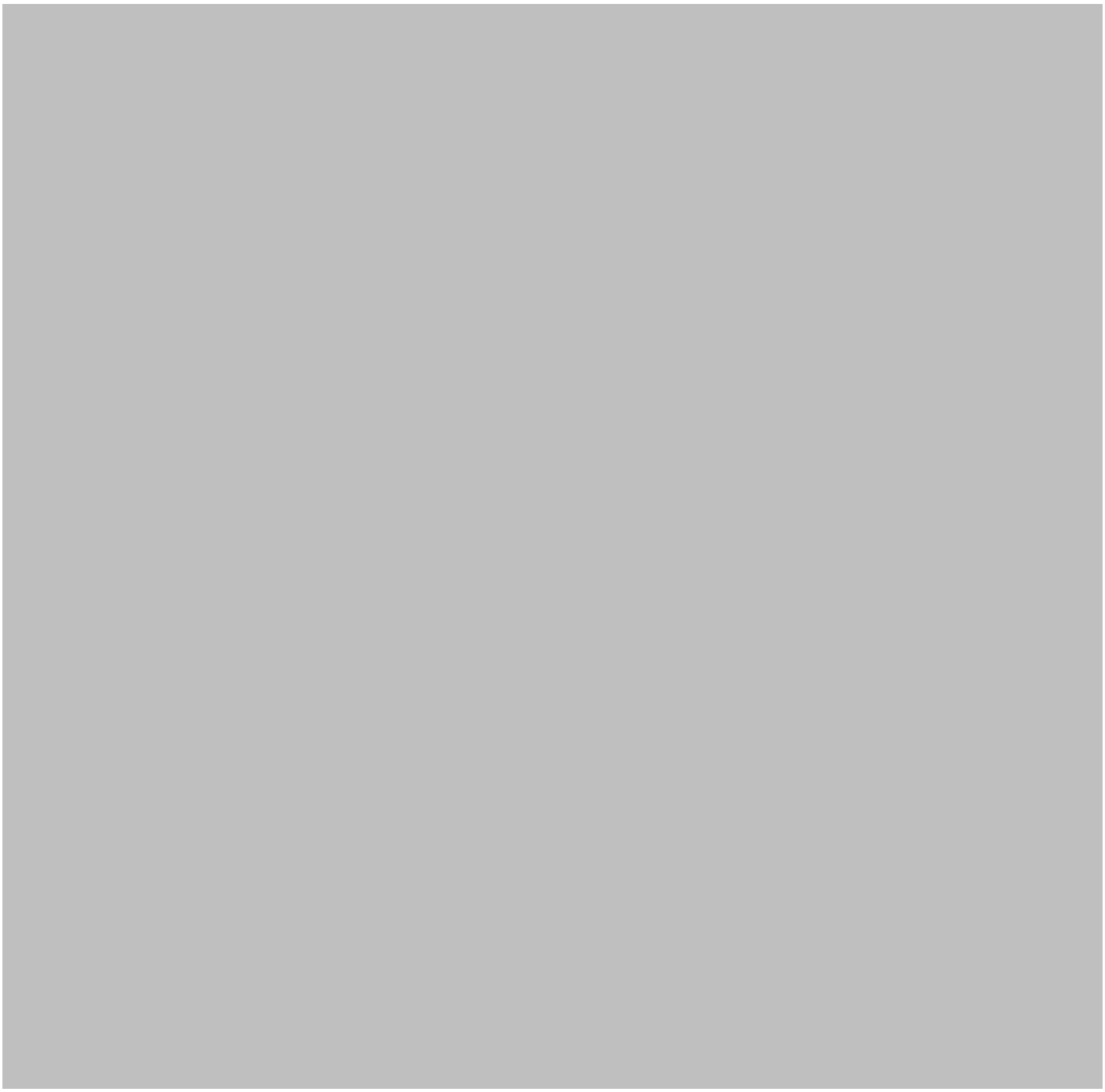






Appendix 5. Ethical approval for prevalence study from Berkshire B REC





Appendix 6. Prevalence of euthyroid 1 (TSH 0.44-4.50 and T4 10-21)

Subgroup	Proportion	% Prevalence	95% CI Lower	95% CI Upper
Age (n=6717)				
16-21	172/181	95.03%	90.77%	97.70%
22-26	847/888	95.38%	93.79%	96.67%
27-31	1937/2016	96.08%	95.14%	96.89%
32-36	2402/2515	95.51%	94.62%	96.28%
37-41	1359/1422	95.57%	94.37%	96.58%
BMI (n=6060)				
<18.5	129/132	97.73%	93.50%	99.53%
18.5-24.9	2996/3117	96.12%	95.38%	96.77%
25-29.9	1775/1862	95.33%	94.27%	96.24%
30-34.9	739/769	96.10%	94.48%	97.35%
≥35	421/445	94.61%	92.08%	96.51%
Ethnicity (n=6717)				
White	4665/4863	95.93%	95.33%	96.47%
Black	478/506	94.47%	92.10%	96.29%
South Asian	1165/1236	94.26%	92.81%	95.49%
Chinese	93/96	96.88%	91.14%	99.35%
Mixed	143/146	97.95%	94.11%	99.57%
Other	173/175	98.86%	95.93%	99.86%
Population (n=6599)				
EPAU	2117/2231	94.89%	93.89%	95.77%
Infertility	3049/3171	96.15%	95.42%	96.79%
Recurrent misc.	1357/1419	95.63%	94.43%	96.63%
Other	76/77	98.70%	92.98%	99.97%

Appendix 7. Prevalence of euthyroid 1a (TSH 0.44-2.49 and T4 10-21)

Subgroup	Proportion	% Prevalence	95% CI Lower	95% CI Upper
Age (n=5490)				
16-21	147/181	81.22%	74.75%	86.63%
22-26	699/888	78.72%	75.87%	81.37%
27-31	1575/2016	78.13%	76.26%	79.91%
32-36	1951/2515	77.57%	75.89%	79.19%
37-41	1118/1422	78.62%	76.40%	80.73%
BMI (n=4961)				
<18.5	103/132	78.03%	70.00%	84.77%
18.5-24.9	2478/3117	79.50%	78.04%	80.90%
25-29.9	1469/1862	78.89%	76.97%	80.73%
30-34.9	579/769	75.29%	72.09%	78.30%
≥35	332/445	74.61%	70.29%	78.59%
Ethnicity (n=5490)				
White	3836/4863	78.88%	77.71%	80.02%
Black	416/506	82.21%	78.60%	85.45%
South Asian	898/1236	72.65%	70.08%	75.12%
Chinese	75/96	78.13%	68.53%	85.92%
Mixed	126/146	86.30%	79.64%	91.43%
Other	139/175	79.43%	72.68%	85.16%
Population (n=5399)				
EPAU	1749/2231	78.40%	76.63%	80.09%
Infertility	2454/3171	77.39%	75.89%	78.83%
Recurrent misc.	1133/1419	79.84%	77.66%	81.90%
Other	63/77	81.82%	71.38%	89.69%

Appendix 8. Prevalence of euthyroid 1b (TSH 2.5-4.5 and T4 10-21)

Subgroup	Proportion	% Prevalence	95% CI Lower	Upper
Age (n=1227)				
16-21	25/181	13.81%	9.14%	19.71%
22-26	148/888	16.67%	14.27%	19.28%
27-31	362/2016	17.96%	16.30%	19.70%
32-36	451/2515	17.93%	16.45%	19.49%
37-41	241/1422	16.95%	15.03%	19.00%
BMI (n=1099)				
<18.5	26/132	19.70%	13.29%	27.51%
18.5-24.9	518/3117	16.62%	15.33%	17.97%
25-29.9	306/1862	16.43%	14.78%	18.20%
30-34.9	160/769	20.81%	17.99%	23.85%
≥35	89/445	20.00%	16.38%	24.02%
Ethnicity (n=1227)				
White	829/4863	17.05%	16.00%	18.13%
Black	62/506	12.25%	9.52%	15.43%
South Asian	267/1236	21.60%	19.34%	24.00%
Chinese	18/96	18.75%	11.51%	28.00%
Mixed	17/146	11.64%	6.93%	17.99%
Other	34/175	19.43%	13.85%	26.08%
Population (n=1221)				
EPAU	368/2231	16.49%	14.98%	18.10%
Infertility	595/3171	18.76%	17.42%	20.17%
Recurrent misc.	224/1419	15.79%	13.93%	17.79%
Other	34/77	44.16%	32.84%	55.93%

Appendix 9. Prevalence of overt hypothyroidism (TSH >4.50 and T4 <10)

			95% CI	
Subgroup	**Proportion**	**% Prevalence**	**Lower**	**Upper**
Age (n=9)				
22-26	2/888	0.23%	0.03%	0.81%
27-31	2/2016	0.10%	0.01%	0.36%
32-36	4/2515	0.16%	0.04%	0.41%
37-41	1/1422	0.07%	0.00%	0.39%
No cases in 16-21				
BMI (n=7)				
18.5-24.9	2/3117	0.06%	0.01%	0.23%
25-29.9	4/1862	0.21%	0.06%	0.55%
≥35	1/445	0.22%	0.01%	1.25%
No cases in <18.5 or 30-34.9				
Ethnicity (n=9)				
White	7/4863	0.14%	0.06%	0.30%
Black	2/506	0.40%	0.05%	1.42%
No cases in any other ethnic groups				
Population (n=9)				
EPAU	3/2231	0.13%	0.03%	0.39%
Infertility	4/3171	0.13%	0.03%	0.32%
Recurrent misc.	2/1419	0.14%	0.02%	0.51%
No cases in "Other"				

Appendix 10. Prevalence of overt hyperthyroidism (TSH <0.44 and T4 >21)

Subgroup	Proportion	% Prevalence	95% CI Lower	95% CI Upper
Age (n=18)				
16-21	2/181	1.10%	0.13%	3.93%
22-26	3/888	0.34%	0.07%	0.98%
27-31	5/2016	0.25%	0.08%	0.58%
32-36	6/2515	0.24%	0.09%	0.52%
37-41	2/1422	0.14%	0.02%	0.51%
BMI (n=15)				
<18.5	1/132	0.76%	0.02%	4.15%
18.5-24.9	11/3117	0.35%	0.18%	0.63%
25-29.9	3/1862	0.16%	0.03%	0.47%
30-34.9	1/769	0.13%	0.00%	0.72%
No cases in ≥35				
Ethnicity (n=18)				
White	10/4863	0.21%	0.10%	0.38%
Black	2/506	0.40%	0.05%	1.42%
South Asian	6/1236	0.49%	0.18%	1.05%
No cases in any other ethnic groups				
Population (n=9)				
EPAU	3/2231	0.13%	0.03%	0.39%
Infertility	4/3171	0.13%	0.03%	0.32%
Recurrent misc.	2/1419	0.14%	0.02%	0.51%
No cases in "Other"				

Appendix 11. Prevalence of moderate SCH (TSH 4.51-10 and T4 10-21)

Subgroup	Proportions	% Prevalence	95% CI Lower	95% CI Upper
Age (n=147)				
16-21	3/181	1.66%	0.34%	4.77%
22-26	19/888	2.14%	1.29%	3.32%
27-31	33/2016	1.64%	1.13%	2.29%
32-36	57/2515	2.27%	1.72%	2.93%
37-41	35/1422	2.46%	1.72%	3.41%
BMI (n=134)				
18.5-24.9	47/3117	1.51%	1.11%	2.00%
25-29.9	43/1862	2.31%	1.68%	3.10%
30-34.9	21/769	2.73%	1.70%	4.14%
≥35	13/445	2.92%	1.56%	4.94%
No cases in <18.5				
Ethnicity (n=147)				
White	96/4863	1.97%	1.60%	2.41%
Black	7/506	1.38%	0.56%	2.83%
South Asian	40/1236	3.24%	2.32%	4.38%
Chinese	1/96	1.04%	0.03%	5.67%
Mixed	2/146	1.37%	0.17%	4.86%
Other	2/175	1.14%	0.14%	4.07%
Population (n=144)				
EPAU	45/2231	2.02%	1.47%	2.69%
Infertility	67/3171	2.11%	1.64%	2.68%
Recurrent misc.	31/1419	2.18%	1.49%	3.09%
Other	1/77	1.30%	0.03%	7.02%

Appendix 12. Prevalence of severe SCH (TSH >10.0 and T4 10-21)

Subgroup	Proportions	% Prevalence	95% CI Lower	95% CI Upper
Age (n=5)				
22-26	8/888	0.90%	0.39%	1.77%
27-31	2/2016	0.10%	0.01%	0.36%
37-41	3/1422	0.21%	0.04%	0.62%
No cases in 16-21 or 32-36				
BMI (n=5)				
18.5-24.9	1/3117	0.03%	0.00%	0.18%
25-29.9	1/1862	0.05%	0.00%	0.30%
≥35	3/445	0.67%	0.14%	1.96%
No cases in <18.5 or 30-34.9				
Ethnicity (n=5)				
White	3/4863	0.06%	0.01%	0.18%
South Asian	2/1236	0.16%	0.02%	0.58%
No cases in any other ethnic groups				
Population (n=5)				
EPAU	3/2231	0.13%	0.03%	0.39%
Infertility	2/3171	0.06%	0.01%	0.23%
No cases in "Recurrent misc." or "Other"				

Appendix 13. Prevalence of subclinical hyperthyroidism (TSH <0.44 & T4 10-21)

Subgroup	Proportions	% Prevalence	95% CI Lower	95% CI Upper
Age (n=90)				
16-21	2/181	1.10%	0.13%	3.93%
22-26	14/888	1.58%	0.86%	2.63%
27-31	21/2016	1.04%	0.65%	1.59%
32-36	36/2515	1.43%	1.00%	1.98%
37-41	17/1422	1.20%	0.70%	1.91%
BMI (n=79)				
<18.5	2/132	1.52%	0.18%	5.37%
18.5-24.9	42/3117	1.35%	0.97%	1.82%
25-29.9	28/1862	1.50%	1.00%	2.17%
30-34.9	2/769	0.26%	0.03%	0.94%
≥35	5/445	1.12%	0.37%	2.60%
Ethnicity (n=81)				
White	55/4863	1.13%	0.85%	1.47%
Black	12/506	2.37%	1.23%	4.11%
South Asian	11/1236	0.89%	0.45%	1.59%
Chinese	2/96	2.08%	0.25%	7.32%
Mixed	1/146	0.68%	0.02%	3.76%
No cases for "Other"				
Population (n=80)				
EPAU	39/2231	1.75%	1.25%	2.38%
Infertility	21/3171	0.66%	0.41%	1.01%
Recurrent misc.	20/1419	1.41%	0.86%	2.17%
No cases for "Other"				

References

1. Fertility treatment in 2013. Trends and figures. HFEA. [Internet]. [cited 2015 Jan 1]. Available from: http://www.hfea.gov.uk/docs/HFEA_Fertility_Trends_and_Figures_2013.pdf

2. McCarthy-Keith DM, Schisterman EF, Robinson RD, O'Leary K, Lucidi RS, Armstrong AY. Will decreasing assisted reproduction technology costs improve utilization and outcomes among minority women? Fertil Steril. 2010 Dec;94(7):2587–9.

3. Purcell K, Schembri M, Frazier LM, Rall MJ, Shen S, Croughan M, et al. Asian ethnicity is associated with reduced pregnancy outcomes after assisted reproductive technology. Fertil Steril. 2007 Feb;87(2):297–302.

4. Seifer DB, Zackula R, Grainger DA, Society for Assisted Reproductive Technology Writing Group Report. Trends of racial disparities in assisted reproductive technology outcomes in black women compared with white women: Society for Assisted Reproductive Technology 1999 and 2000 vs. 2004-2006. Fertil Steril. 2010 Feb;93(2):626–35.

5. Seifer DB, Frazier LM, Grainger DA. Disparity in assisted reproductive technologies outcomes in black women compared with white women. Fertil Steril. 2008 Nov;90(5):1701–10.

6. Fujimoto VY, Luke B, Brown MB, Jain T, Armstrong A, Grainger DA, et al., Society for Assisted Reproductive Technology Writing Group. Racial and ethnic disparities in assisted reproductive technology outcomes in the United States. Fertil Steril. 2010 Feb;93(2):382–90.

7. Centre for Maternal and Child Enquiries & Royal College of Obstetricians and Gynaecologists. CMACE & RCOG Joint Guideline. Management of women with obesity in pregnancy . London: CMACE & RCOG, 2010.

8. Cedergren MI. Maternal morbid obesity and the risk of adverse pregnancy outcome. Obstet Gynecol. 2004 Feb;103(2):219–24.

9. Legge A, Bouzayen R, Hamilton L, Young D. The impact of maternal body mass index on in vitro fertilization outcomes. J Obstet Gynaecol Can JOGC J Obstétrique Gynécologie Can JOGC. 2014 Jul;36(7):613–9.

10. Schliep KC, Mumford SL, Ahrens KA, Hotaling JM, Carrell DT, Link M, et al. Effect of male and female body mass index on pregnancy and live birth success after in vitro fertilization. Fertil Steril. 2015 Feb;103(2):388–95.

11. Zhang JJ, Feret M, Chang L, Yang M, Merhi Z. Obesity adversely impacts the number and maturity of oocytes in conventional IVF not in minimal stimulation

IVF. Gynecol Endocrinol Off J Int Soc Gynecol Endocrinol. 2015 May;31(5):409–13.

12. Caillon H, Fréour T, Bach-Ngohou K, Colombel A, Denis MG, Barrière P, et al. Effects of female increased body mass index on in vitro fertilization cycles outcome. Obes Res Clin Pract. 2015 Mar 10;

13. Rittenberg V, Seshadri S, Sunkara SK, Sobaleva S, Oteng-Ntim E, El-Toukhy T. Effect of body mass index on IVF treatment outcome: an updated systematic review and meta-analysis. Reprod Biomed Online. 2011 Oct;23(4):421–39.

14. Brenta G. Diabetes and thyroid disorders. Br J Diabetes Vasc Dis. 2010 Jul 1;10(4):172–7.

15. Abalovich M, Amino N, Barbour LA, Cobin RH, De Groot LJ, Glinoer D, et al. Management of thyroid dysfunction during pregnancy and postpartum: an Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab. 2007 Aug;92(8 Suppl):S1–47.

16. Stagnaro-Green A, Abalovich M, Alexander E, Azizi F, Mestman J, Negro R, et al., American Thyroid Association Taskforce on Thyroid Disease During Pregnancy and Postpartum. Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and postpartum. Thyroid Off J Am Thyroid Assoc. 2011 Oct;21(10):1081–125.

17. Van den Boogaard E, Vissenberg R, Land JA, van Wely M, van der Post JAM, Goddijn M, et al. Significance of (sub)clinical thyroid dysfunction and thyroid autoimmunity before conception and in early pregnancy: a systematic review. Hum Reprod Update. 2011 Oct;17(5):605–19.

18. Velkeniers B, Van Meerhaeghe A, Poppe K, Unuane D, Tournaye H, Haentjens P. Levothyroxine treatment and pregnancy outcome in women with subclinical hypothyroidism undergoing assisted reproduction technologies: systematic review and meta-analysis of RCTs. Hum Reprod Update. 2013 Jun;19(3):251–8.

19. Thangaratinam S, Tan A, Knox E, Kilby MD, Franklyn J, Coomarasamy A. Association between thyroid autoantibodies and miscarriage and preterm birth: meta-analysis of evidence. BMJ. 2011;342:d2616.

20. Bussen S, Steck T. Thyroid autoantibodies in euthyroid non-pregnant women with recurrent spontaneous abortions. Hum Reprod Oxf Engl. 1995 Nov;10(11):2938–40.

21. Kutteh WH, Yetman DL, Carr AC, Beck LA, Scott RT. Increased prevalence of antithyroid antibodies identified in women with recurrent pregnancy loss but not in women undergoing assisted reproduction. Fertil Steril. 1999 May;71(5):843–8.

22. Vissenberg R, van den Boogaard E, van Wely M, van der Post JA, Fliers E, Bisschop PH, et al. Treatment of thyroid disorders before conception and in early pregnancy: a systematic review. Hum Reprod Update. 2012 Jul;18(4):360–73.

23. Dayal MB, Gindoff P, Dubey A, Spitzer TLB, Bergin A, Peak D, et al. Does ethnicity influence in vitro fertilization (IVF) birth outcomes? Fertil Steril. 2009 Jun;91(6):2414–8.

24. Bendikson K, Cramer DW, Vitonis A, Hornstein MD. Ethnic background and in vitro fertilization outcomes. Int J Gynaecol Obstet Off Organ Int Fed Gynaecol Obstet. 2005 Mar;88(3):342–6.

25. Baker VL, Luke B, Brown MB, Alvero R, Frattarelli JL, Usadi R, et al. Multivariate analysis of factors affecting probability of pregnancy and live birth with in vitro fertilization: an analysis of the Society for Assisted Reproductive Technology Clinic Outcomes Reporting System. Fertil Steril. 2010 Sep;94(4):1410–6.

26. Csokmay JM, Hill MJ, Maguire M, Payson MD, Fujimoto VY, Armstrong AY. Are there ethnic differences in pregnancy rates in African-American versus white women undergoing frozen blastocyst transfers? Fertil Steril. 2011 Jan;95(1):89–93.

27. Langen ES, Shahine LK, Lamb JD, Lathi RB, Milki AA, Fujimoto VY, et al. Asian ethnicity and poor outcomes after in vitro fertilization blastocyst transfer. Obstet Gynecol. 2010 Mar;115(3):591–6.

28. Shahine LK, Lamb JD, Lathi RB, Milki AA, Langen E, Westphal LM. Poor prognosis with in vitro fertilization in Indian women compared to Caucasian women despite similar embryo quality. PloS One. 2009;4(10):e7599.

29. Shuler A, Rodgers AK, Budrys NM, Holden A, Schenken RS, Brzyski RG. In vitro fertilization outcomes in Hispanics versus non-Hispanic whites. Fertil Steril. 2011 Jun;95(8):2735–7.

30. Wellons MF, Fujimoto VY, Baker VL, Barrington DS, Broomfield D, Catherino WH, et al. Race matters: a systematic review of racial/ethnic disparity in Society for Assisted Reproductive Technology reported outcomes. Fertil Steril. 2012 Aug;98(2):406–9.

31. Jayaprakasan K, Pandian D, Hopkisson J, Campbell BK, Maalouf WE. Effect of ethnicity on live birth rates after in vitro fertilisation or intracytoplasmic sperm injection treatment. BJOG Int J Obstet Gynaecol. 2014 Feb;121(3):300–6.

32. Khan K, Ter Riet G, Glanville J, Sowden A, Kleijnen J. Undertaking Systematic Reviews of Research on Effectiveness: CRD's Guidance for Carrying out or Commissioning Reviews. 2nd ed. York: NHS Centre for Reviews and Dissemination, University of York; 2001.

33. McKibbon A, Eady A, Marks S. Identifying and selecting studies for inclusion. In PDQ Evidence based Principles and Practice. Hamilton, Canada: BC Decker Incorporation; 1999. p. 125–7.

34. Glasziou P, Irwig L, Bain C, Colditz G. In Systematic Reviews in Health Care: A practical guide. 2nd ed. Cambridge, UK: Cambridge University Press; 2001. p. 67–73.

35. Feinberg EC, Larsen FW, Catherino WH, Zhang J, Armstrong AY. Comparison of assisted reproductive technology utilization and outcomes between Caucasian and African American patients in an equal-access-to-care setting. Fertil Steril. 2006 Apr;85(4):888–94.

36. Luke B, Brown MB, Stern JE, Missmer SA, Fujimoto VY, Leach R. Racial and ethnic disparities in assisted reproductive technology pregnancy and live birth rates within body mass index categories. Fertil Steril. 2011 Apr;95(5):1661–6.

37. Higgins JPT, Green S (Eds): Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011] In http://www.cochrane-handbook.org.

38. Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P: The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2013. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp.

39. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. PLoS Med. 2007;4(10):e296.

40. Lashen H, Afnan M, Sharif K. A controlled comparison of ovarian response to controlled stimulation in first generation Asian women compared with white Caucasians undergoing in vitro fertilisation. Br J Obstet Gynaecol. 1999 May;106(5):407–9.

41. Nichols JE, Higdon HL, Crane MM, Boone WR. Comparison of implantation and pregnancy rates in African American and white women in an assisted reproductive technology practice. Fertil Steril. 2001 Jul;76(1):80–4.

42. Sharara F, Fouany M, Sharara Y, Abdo G. Racial differences in ART outcome between white and South Asian women. Middle East Fertil Soc J. 2012;17(2):89–92.

43. Sharara FI, McClamrock HD. Differences in in vitro fertilization (IVF) outcome between white and black women in an inner-city, university-based IVF program. Fertil Steril. 2000 Jun;73(6):1170–3.

44. Mahmud G, López Bernal A, Yudkin P, Ledger W, Barlow DH. A controlled assessment of the in vitro fertilization performance of British women of Indian origin compared with white women. Fertil Steril. 1995 Jul;64(1):103–6.

45. Stroup D, Berlin J, Morton S, Olkin I, Williamson G, Rennie D. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA. 2000;283:2008–12.

46. Census 2011. 2011 Census: KS201UK Ethnic group, local authorities in the United Kingdom, Accessed 21 February 2014.

47. Harb HM, Al-rshoud F., Dhillon R, Harb M, Coomarasamy A. Ethnicity and miscarriage: a large prospective observational study and meta-analysis. Fertil Steril. 2014 Sep 1;102(3):e81.

48. Loendersloot LL van, Wely M van, Limpens J, Bossuyt PMM, Repping S, Veen F van der. Predictive factors in in vitro fertilization (IVF): a systematic review and meta-analysis. Hum Reprod Update. 2010 Nov 1;16(6):577–89.

49. Roque M, Lattes K, Serra S, Solà I, Geber S, Carreras R, et al. Fresh embryo transfer versus frozen embryo transfer in in vitro fertilization cycles: a systematic review and meta-analysis. Fertil Steril. 2013 Jan;99(1):156–62.

50. Malizia BA, Hacker MR, Penzias AS. Cumulative live-birth rates after in vitro fertilization. N Engl J Med. 2009 Jan 15;360(3):236–43.

51. Van Der Steeg J, Steures P, Eijkemans M, Habbema J, Bossuyt P, Hompes P, et al. Do clinical prediction models improve concordance of treatment decisions in reproductive medicine? BJOG Int J Obstet Gynaecol. 2006 Jul 1;113(7):825–31.

52. Wiegerinck MAHM, Bongers MY, Mol BWJ, Heineman M-J. How concordant are the estimated rates of natural conception and in-vitro fertilization/embryo transfer success?*. Hum Reprod. 1999 Mar 1;14(3):689–93.

53. Andersen AN, Goossens V, Bhattacharya S, Ferraretti AP, Kupka MS, Mouzon J de, et al. Assisted reproductive technology and intrauterine inseminations in Europe, 2005: results generated from European registers by ESHRE ESHRE. The European IVF Monitoring Programme (EIM), for the European Society of Human Reproduction and Embryology (ESHRE). Hum Reprod [Internet]. 2009 Feb 18 [cited 2014 Nov 16]; Available from: http://humrep.oxfordjournals.org/content/early/2009/02/18/humrep.dep035

54. Australian Government Department of Health and Ageing (2006). Report of the independent review of assisted reproductive tehcnologies. Available: http://www.health.gov.au/internet/main/publishing.nsf/Content/ART-Report. Accessed September 2014.

55. Centers for Disease Control and Prevention ASRM, Society for Assisted Reproductive Technology (2008) 2006 Assisted reproductive technology success rates: National summary and fertility clinic reports. Altanta: CDC. Available: http://www.cdc.gov.art/art2006/index.htm. Accessed September 2014.

56. Stolwijk AM, Zielhuis GA, Hamilton CJ, Straatman H, Hollanders JM, Goverde HJ, et al. Prognostic models for the probability of achieving an ongoing pregnancy after in-vitro fertilization and the importance of testing their predictive value. Hum Reprod Oxf Engl. 1996 Oct;11(10):2298–303.

57. Stolwijk AM, Straatman H, Zielhuis GA, Jansen CA, Braat DD, van Dop PA, et al. External validation of prognostic models for ongoing pregnancy after in-vitro fertilization. Hum Reprod Oxf Engl. 1998 Dec;13(12):3542–9.

58. Templeton A, Morris JK, Parslow W. Factors that affect outcome of in-vitro fertilisation treatment. Lancet. 1996 Nov 23;348(9039):1402–6.

59. Hunault CC, Eijkemans MJC, Pieters MHEC, te Velde ER, Habbema JDF, Fauser BCJM, et al. A prediction model for selecting patients undergoing in vitro fertilization for elective single embryo transfer. Fertil Steril. 2002 Apr;77(4):725–32.

60. Hunault CC, te Velde ER, Weima SM, Macklon NS, Eijkemans MJC, Klinkert ER, et al. A case study of the applicability of a prediction model for the selection of patients undergoing in vitro fertilization for single embryo transfer in another center. Fertil Steril. 2007 Jun;87(6):1314–21.

61. Nelson SM, Lawlor DA. Predicting live birth, preterm delivery, and low birth weight in infants born from in vitro fertilisation: a prospective study of 144,018 treatment cycles. PLoS Med. 2011;8(1):e1000386.

62. Minaretzis D, Harris D, Alper MM, Mortola JF, Berger MJ, Power D. Multivariate analysis of factors predictive of successful live births in in vitro fertilization (IVF) suggests strategies to improve IVF outcome. J Assist Reprod Genet. 1998 Jul;15(6):365–71.

63. Te Velde ER, Nieboer D, Lintsen AM, Braat DDM, Eijkemans MJC, Habbema JDF, et al. Comparison of two models predicting IVF success; the effect of time trends on model performance. Hum Reprod Oxf Engl. 2014 Jan;29(1):57–64.

64. Smith ADAC, Tilling K, Lawlor DA, Nelson SM. External validation and calibration of IVFpredict: a national prospective cohort study of 130,960 in vitro fertilisation cycles. PloS One. 2015;10(4):e0121357.

65. Jirge PR. Ovarian reserve tests. J Hum Reprod Sci. 2011;4(3):108–13.

66. Ulug U, Ben-Shlomo I, Turan E, Erden HF, Akman MA, Bahceci M. Conception rates following assisted reproduction in poor responder patients: a

retrospective study in 300 consecutive cycles. Reprod Biomed Online. 2003 Jan 1;6(4):439–43.

67. Shaban MM, Abdel Moety GAF. Role of ultrasonographic markers of ovarian reserve in prediction of IVF and ICSI outcome. Gynecol Endocrinol Off J Int Soc Gynecol Endocrinol. 2014 Apr;30(4):290–3.

68. Broer SL, Dólleman M, van Disseldorp J, Broeze KA, Opmeer BC, Bossuyt PMM, et al., IPD-EXPORT Study Group. Prediction of an excessive response in in vitro fertilization from patient characteristics and ovarian reserve tests and comparison in subgroups: an individual patient data meta-analysis. Fertil Steril. 2013 Aug;100(2):420–9.e7.

69. Steyerberg EW. Clinical prediction models. A practical approach to development, validation and updating. New York, USA: Springer Science + Business Media, LCC; 2009.

70. Moons KGM, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart Br Card Soc. 2012 May;98(9):683–90.

71. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J. 2014 Aug 1;35(29):1925–31.

72. Coppus SFPJ, van der Veen F, Opmeer BC, Mol BWJ, Bossuyt PMM. Evaluating prediction models in reproductive medicine. Hum Reprod Oxf Engl. 2009 Aug;24(8):1774–8.

73. Tremellen K, Savulescu J. Ovarian reserve screening: a scientific and ethical analysis. Hum Reprod Oxf Engl. 2014 Dec;29(12):2606–14.

74. Bonilla-Musoles F, Castillo JC, Caballero O, Pérez-Panades J, Bonilla F, Dolz M, et al. Predicting ovarian reserve and reproductive outcome using antimüllerian hormone (AMH) and antral follicle count (AFC) in patients with previous assisted reproduction technique (ART) failure. Clin Exp Obstet Gynecol. 2012;39(1):13–8.

75. Bahn Chair RS, Burch HB, Cooper DS, Garber JR, Greenlee MC, Klein I, et al., American Thyroid Association, American Association of Clinical Endocrinologists. Hyperthyroidism and other causes of thyrotoxicosis: management guidelines of the American Thyroid Association and American Association of Clinical Endocrinologists. Thyroid Off J Am Thyroid Assoc. 2011 Jun;21(6):593–646.

76. Zimmerman D. Fetal and neonatal hyperthyroidism. Thyroid Off J Am Thyroid Assoc. 1999 Jul;9(7):727–33.

77. Earl R, Crowther CA, Middleton P. Interventions for preventing and treating hyperthyroidism in pregnancy. Cochrane Database Syst Rev. 2010;(9):CD008633.

78. Bjoro T, Holmen J, Krüger O, Midthjell K, Hunstad K, Schreiner T, et al. Prevalence of thyroid disease, thyroid dysfunction and thyroid peroxidase antibodies in a large, unselected population. The Health Study of Nord-Trondelag (HUNT). Eur J Endocrinol Eur Fed Endocr Soc. 2000 Nov;143(5):639–47.

79. Col NF, Surks MI, Daniels GH. Subclinical thyroid disease: clinical applications. JAMA. 2004 Jan 14;291(2):239–43.

80. Samuels MH. Subclinical thyroid disease in the elderly. Thyroid Off J Am Thyroid Assoc. 1998 Sep;8(9):803–13.

81. Sawin CT, Geller A, Wolf PA, Belanger AJ, Baker E, Bacharach P, et al. Low serum thyrotropin concentrations as a risk factor for atrial fibrillation in older persons. N Engl J Med. 1994 Nov 10;331(19):1249–52.

82. Bauer DC, Ettinger B, Nevitt MC, Stone KL, Study of Osteoporotic Fractures Research Group. Risk for fracture in women with low serum levels of thyroid-stimulating hormone. Ann Intern Med. 2001 Apr 3;134(7):561–8.

83. Faber J, Galløe AM. Changes in bone mass during prolonged subclinical hyperthyroidism due to L-thyroxine treatment: a meta-analysis. Eur J Endocrinol Eur Fed Endocr Soc. 1994 Apr;130(4):350–6.

84. Surks MI, Ortiz E, Daniels GH, Sawin CT, Col NF, Cobin RH, et al. Subclinical thyroid disease: scientific review and guidelines for diagnosis and management. JAMA. 2004 Jan 14;291(2):228–38.

85. Casey BM, Dashe JS, Wells CE, McIntire DD, Leveno KJ, Cunningham FG. Subclinical hyperthyroidism and pregnancy outcomes. Obstet Gynecol. 2006 Feb;107(2 Pt 1):337–41.

86. Stagnaro-Green A. Screening pregnant women for overt thyroid disease. JAMA. 2015 Feb 10;313(6):565–6.

87. De Groot L, Abalovich M, Alexander EK, Amino N, Barbour L, Cobin RH, et al. Management of thyroid dysfunction during pregnancy and postpartum: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab. 2012 Aug;97(8):2543–65.

88. Association of Clinical Biochemists (ACB), British Thyroid Association (BTA), British Thyroid Foundation (BTF) (2006) UK guidelines for the use of thyroid function tests. Available from: http://www.british-thyroid-association.org/info-for-patients/Docs/TFT_guideline_final_version_July_2006.pdf (accessed 5th April 2015), 39–43.

89. Garber JR, Cobin RH, Gharib H, Hennessey JV, Klein I, Mechanick JI, et al., American Association of Clinical Endocrinologists and American Thyroid Association Taskforce on Hypothyroidism in Adults. Clinical practice guidelines for hypothyroidism in adults: cosponsored by the American Association of Clinical Endocrinologists and the American Thyroid Association. Endocr Pract Off J Am Coll Endocrinol Am Assoc Clin Endocrinol. 2012 Dec;18(6):988–1028.

90. Lazarus JH, Bestwick JP, Channon S, Paradice R, Maina A, Rees R, et al. Antenatal Thyroid Screening and Childhood Cognitive Function. N Engl J Med. 2012 Feb 9;366(6):493–501.

91. Hollowell JG, Staehling NW, Flanders WD, Hannon WH, Gunter EW, Spencer CA, et al. Serum TSH, T(4), and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). J Clin Endocrinol Metab. 2002 Feb;87(2):489–99.

92. Tunbridge WM, Evered DC, Hall R, Appleton D, Brewis M, Clark F, et al. The spectrum of thyroid disease in a community: the Whickham survey. Clin Endocrinol (Oxf). 1977 Dec;7(6):481–93.

93. Vanderpump MPJ, Lazarus JH, Smyth PP, Laurberg P, Holder RL, Boelaert K, et al., British Thyroid Association UK Iodine Survey Group. Iodine status of UK schoolgirls: a cross-sectional survey. Lancet Lond Engl. 2011 Jun 11;377(9782):2007–12.

94. Baloch Z, Carayon P, Conte-Devolx B, Demers LM, Feldt-Rasmussen U, Henry J-F, et al., Guidelines Committee, National Academy of Clinical Biochemistry. Laboratory medicine practice guidelines. Laboratory support for the diagnosis and monitoring of thyroid disease. Thyroid Off J Am Thyroid Assoc. 2003 Jan;13(1):3–126.

95. Mariotti S, Franceschi C, Cossarizza A, Pinchera A. The aging thyroid. Endocr Rev. 1995 Dec;16(6):686–715.

96. Díez JJ. Hyperthyroidism in patients older than 55 years: an analysis of the etiology and management. Gerontology. 2003 Oct;49(5):316–23.

97. Vanderpump MP, Tunbridge WM, French JM, Appleton D, Bates D, Rodgers H, et al. The incidence of diabetes mellitus in an English community: a 20-year follow-up of the Whickham Survey. Diabet Med J Br Diabet Assoc. 1996 Aug;13(8):741–7.

98. Knudsen N, Laurberg P, Rasmussen LB, Bülow I, Perrild H, Ovesen L, et al. Small Differences in Thyroid Function May Be Important for Body Mass Index and the Occurrence of Obesity in the Population. J Clin Endocrinol Metab. 2005 Jul 1;90(7):4019–24.

99. Michalaki MA, Vagenakis AG, Leonardou AS, Argentou MN, Habeos IG, Makri MG, et al. Thyroid function in humans with morbid obesity. Thyroid Off J Am Thyroid Assoc. 2006 Jan;16(1):73–8.

100. Rotondi M, Leporati P, La Manna A, Pirali B, Mondello T, Fonte R, et al. Raised serum TSH levels in patients with morbid obesity: is it enough to diagnose subclinical hypothyroidism? Eur J Endocrinol Eur Fed Endocr Soc. 2009 Mar;160(3):403–8.

101. Iacobellis G, Ribaudo MC, Zappaterreno A, Iannucci CV, Leonetti F. Relationship of thyroid function with body mass index, leptin, insulin sensitivity and adiponectin in euthyroid obese women. Clin Endocrinol (Oxf). 2005 Apr;62(4):487–91.

102. McLeod DA, Caturegli P, Cooper DS, Matos PG, Hutfless S. VAriation in rates of autoimmune thyroid disease by race/ethnicity in us military personnel. JAMA. 2014 Apr 16;311(15):1563–5.

103. Mitchell AL, Pearce SHS. How should we treat patients with low serum thyrotropin concentrations? Clin Endocrinol (Oxf). 2010 Mar;72(3):292–6.

104. Tim Korevaar, Marco Medici, Yolande de Rijke, Willy Visser, Sabine de Muinck Keizer-Schrama, Vincent Jaddoe, et al. Ethnic differences in maternal thyroid parameters during pregnancy: the generation r study. Non-neoplastic Thyroid Disorders [Internet]. The Endocrine Society; 2013 [cited 2015 Jul 3]. p. MON – 456 – MON – 456. Available from: http://press.endocrine.org/doi/abs/10.1210/endo-meetings.2013.THPTA.4.MON-456

105. Vila L, Velasco I, González S, Morales F, Sánchez E, Torrejón S, et al. Controversies in endocrinology: On the need for universal thyroid screening in pregnant women. Eur J Endocrinol Eur Fed Endocr Soc. 2014 Jan;170(1):R17–30.

106. Trbojević B, Djurica S. [Diagnosis of autoimmune thyroid disease]. Srp Arh Celok Lek. 2005 Oct;133 Suppl 1:25–33.

107. Chardès T, Chapal N, Bresson D, Bès C, Giudicelli V, Lefranc M-P, et al. The human anti-thyroid peroxidase autoantibody repertoire in Graves' and Hashimoto's autoimmune thyroid diseases. Immunogenetics. 2002 Jun;54(3):141–57.

108. Melmed, Shlomo. Williams Textbook of Endocrinology (12th ed.). Philadelphia: Elsevier/Saunders. p. 355.

109. Saravanan P, Dayan CM. Thyroid autoantibodies. Endocrinol Metab Clin North Am. 2001 Jun;30(2):315–37, viii.

110. Glinoer D, Soto MF, Bourdoux P, Lejeune B, Delange F, Lemone M, et al. Pregnancy in patients with mild thyroid abnormalities: maternal and neonatal repercussions. J Clin Endocrinol Metab. 1991 Aug;73(2):421–7.

111. Glinoer D, Delange F. The potential repercussions of maternal, fetal, and neonatal hypothyroxinemia on the progeny. Thyroid Off J Am Thyroid Assoc. 2000 Oct;10(10):871–87.

112. Kim CH, Chae HD, Kang BM, Chang YS. Influence of antithyroid antibodies in euthyroid women on in vitro fertilization-embryo transfer outcome. Am J Reprod Immunol N Y N 1989. 1998 Jul;40(1):2–8.

113. Muller AF, Verhoeff A, Mantel MJ, Berghout A. Thyroid autoimmunity and abortion: a prospective study in women undergoing in vitro fertilization. Fertil Steril. 1999 Jan;71(1):30–4.

114. Negro R, Formoso G, Coppola L, Presicce G, Mangieri T, Pezzarossa A, et al. Euthyroid women with autoimmune disease undergoing assisted reproduction technologies: the role of autoimmunity and thyroid function. J Endocrinol Invest. 2007 Jan;30(1):3–8.

115. Stagnaro-Green A, Glinoer D. Thyroid autoimmunity and the risk of miscarriage. Best Pract Res Clin Endocrinol Metab. 2004 Jun;18(2):167–81.

116. Poppe K, Velkeniers B, Glinoer D. The role of thyroid autoimmunity in fertility and pregnancy. Nat Clin Pract Endocrinol Metab. 2008 Jul;4(7):394–405.

117. Stuart, A.E. (1994) Autoantibodies and pregnancy loss. Lancet. 343, 747-748.

118. Glinoer D. The regulation of thyroid function in pregnancy: pathways of endocrine adaptation from physiology to pathology. Endocr Rev. 1997 Jun;18(3):404–33.

119. Sinclair D. Thyroid antibodies: which, why, when and who? Expert Rev Clin Immunol. 2006 Sep;2(5):665–9.

120. Negro R, Formoso G, Mangieri T, Pezzarossa A, Dazzi D, Hassan H. Levothyroxine treatment in euthyroid pregnant women with autoimmune thyroid disease: effects on obstetrical complications. J Clin Endocrinol Metab. 2006 Jul;91(7):2587–91.

121. Mehran L, Tohidi M, Sarvghadi F, Delshad H, Amouzegar A, Soldin OP, et al. Management of Thyroid Peroxidase Antibody Euthyroid Women in Pregnancy: Comparison of the American Thyroid Association and the Endocrine Society Guidelines. J Thyroid Res. 2013 May 12;2013:e542692.

122. Chen L, Hu R. Thyroid autoimmunity and miscarriage: a meta-analysis. Clin Endocrinol (Oxf). 2011 Apr 1;74(4):513–9.

123. Harpsøe MC, Basit S, Andersson M, Nielsen NM, Frisch M, Wohlfahrt J, et al. Body mass index and risk of autoimmune diseases: a study within the Danish National Birth Cohort. Int J Epidemiol. 2014 Mar 7;dyu045.

124. Pratt D, Novotny M, Kaberlein G, Dudkiewicz A, Gleicher N. Antithyroid antibodies and the association with non-organ-specific antibodies in recurrent pregnancy loss. Am J Obstet Gynecol. 1993 Mar;168(3 Pt 1):837–41.

125. Rushworth FH, Backos M, Rai R, Chilcott IT, Baxter N, Regan L. Prospective pregnancy outcome in untreated recurrent miscarriers with thyroid autoantibodies. Hum Reprod. 2000 Jul 1;15(7):1637–9.

126. Toulis KA, Goulis DG, Venetis CA, Kolibianakis EM, Negro R, Tarlatzis BC, et al. Risk of spontaneous miscarriage in euthyroid women with thyroid autoimmunity undergoing IVF: a meta-analysis. Eur J Endocrinol Eur Fed Endocr Soc. 2010 Apr;162(4):643–52.

127. Abdel Rahman AH, Aly Abbassy H, Abbassy AAE. Improved in vitro fertilization outcomes after treatment of subclinical hypothyroidism in infertile women. Endocr Pract Off J Am Coll Endocrinol Am Assoc Clin Endocrinol. 2010 Oct;16(5):792–7.

128. Kim C-H, Ahn J-W, Kang SP, Kim S-H, Chae H-D, Kang B-M. Effect of levothyroxine treatment on in vitro fertilization and pregnancy outcome in infertile women with subclinical hypothyroidism undergoing in vitro fertilization/intracytoplasmic sperm injection. Fertil Steril. 2011 Apr;95(5):1650–4.

129. Negro R, Mangieri T, Coppola L, Presicce G, Casavola EC, Gismondi R, et al. Levothyroxine treatment in thyroid peroxidase antibody-positive women undergoing assisted reproduction technologies: a prospective study. Hum Reprod Oxf Engl. 2005 Jun;20(6):1529–33.

130. Browne ML, Rasmussen SA, Hoyt AT, Waller DK, Druschel CM, Caton AR, et al., National Birth Defects Prevention Study. Maternal thyroid disease, thyroid medication use, and selected birth defects in the National Birth Defects Prevention Study. Birt Defects Res A Clin Mol Teratol. 2009 Jul;85(7):621–8.

131. Reh A, Grifo J, Danoff A. What is a normal thyroid-stimulating hormone (TSH) level? Effects of stricter TSH thresholds on pregnancy outcomes after in vitro fertilization. Fertil Steril. 2010 Dec;94(7):2920–2.

132. Chai J, Yeung W-YT, Lee C-YV, Li H-WR, Ho P-C, Ng H-YE. Live birth rates following in vitro fertilization in women with thyroid autoimmunity and/or subclinical hypothyroidism. Clin Endocrinol (Oxf). 2014 Jan;80(1):122–7.

133. Fumarola A, Grani G, Romanzi D, Del Sordo M, Bianchini M, Aragona A, et al. Thyroid function in infertile patients undergoing assisted reproduction. Am J Reprod Immunol N Y N 1989. 2013 Oct;70(4):336–41.

134. Yassa L, Marqusee E, Fawcett R, Alexander EK. Thyroid hormone early adjustment in pregnancy (the THERAPY) trial. J Clin Endocrinol Metab. 2010 Jul;95(7):3234–41.

135. Dashe JS, Casey BM, Wells CE, McIntire DD, Byrd EW, Leveno KJ, et al. Thyroid-stimulating hormone in singleton and twin pregnancy: importance of gestational age-specific reference ranges. Obstet Gynecol. 2005 Oct;106(4):753–7.

136. Cotzias C, Wong S-J, Taylor E, Seed P, Girling J. A study to establish gestation-specific reference intervals for thyroid function tests in normal singleton pregnancy. Eur J Obstet Gynecol Reprod Biol. 2008 Mar;137(1):61–6.

137. Haddow JE, Palomaki GE, Allan WC, Williams JR, Knight GJ, Gagnon J, et al. Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. N Engl J Med. 1999 Aug 19;341(8):549–55.

138. National Institute for Health and Care Excellence. Assessment and Treatment for Peopl with Fertility Problems. NICE clinical guideline 156. London: NICE; 2013.

139. Meyerovitch J, Rotman-Pikielny P, Sherf M, Battat E, Levy Y, Surks MI. Serum thyrotropin measurements in the community: five-year follow-up in a large network of primary care physicians. Arch Intern Med. 2007 Jul 23;167(14):1533–8.

140. Díez JJ, Iglesias P. Spontaneous subclinical hypothyroidism in patients older than 55 years: an analysis of natural course and risk factors for the development of overt thyroid failure. J Clin Endocrinol Metab. 2004 Oct;89(10):4890–7.