Gender differences in the professional performance of doctors practising in the UK

Emily Armandine Unwin

UCL

A thesis submitted for the degree of Doctor of Philosophy
Declaration

I, Emily Armandine Unwin, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature:

Date: 4th September 2019
Abstract

Throughout this thesis I aim to address the research question “is there a gender difference in doctors’ professional performance, and if so why?”

Professionalism is at the core of delivering good medical care. Most doctors pursue their medical careers without complaints that question their professionalism and even fewer have action taken against their medical registration; however, previous research suggests certain groups of doctors – including men – are referred more often to the medical regulator.

I explored the research question through six quantitative studies:

1. Multivariate analyses of administrative data from the UK medical register to investigate the relationship between gender and having sanctions imposed.
2. Systematic review and meta-analysis to quantify the association between gender and medico-legal action.
3. Multivariate analyses of results of a clinical postgraduate medical examination to investigate the association between gender and pass rates.
4. Systematic review and meta-analysis of postgraduate examination performance to investigate the association with gender.
5. Administrative data linkage study of General Practitioners’ performance at an examination designed to assess doctors under investigation for fitness to practise concerns, by gender.
6. Path analysis of longitudinal survey data of doctors to investigate whether empathy, communication style and personality mediate the association between gender and sanctions.

Female doctors were less likely to have experienced medico-legal action and performed better on average than male doctors at clinical postgraduate medical examinations, suggesting that the skills highly valued in clinical
examinations may be protective. Empathy, communication style or personality were not significantly correlated with sanctions, however, gender differences were present.

As pressure continues to grow on our health system it is important to support our medical workforce. Gaining a clearer understanding of the doctors who have fitness to practise concerns will better enable the medical community to support and enhance the professional performance of all doctors.
Impact statement

“There is no power for change greater than a community discovering what it cares about.”

Margaret Wheatley, Turning to one another, 2002 [1]

As a collective doctors are one of the most trusted professions [2], and this trust is underpinned by “a set of values, behaviours and relationships” that are collectively known as medical professionalism [3]. Professionalism is an important component of doctors’ daily practice as they aim to promote and support the health of the community they serve. A lack of professionalism can be detrimental to patient care and can undermine the trust in the medical profession.

Exploring the professional performance of doctors from a gender perspective is timely and relevant. For over a decade concerns have been raised about the impact on the medical profession caused by the increasing number of women choosing to study and practice medicine [4]. Published findings indicate that professional performance may be influenced by gender [5, 6]. Gender is currently high on the political and societal agenda, with an increase in focus on the difference between how members of each gender are treated and valued, both within medicine and within society more generally. Confirming whether the professional performance of doctors is influenced by gender is important to ensure patients continue to receive optimal care.

This doctoral thesis contributes to the emerging and growing practice of linking and analysing large secondary datasets in medical education research to uncover patterns and trends, and demonstrates that there is a gender difference in the professional performance of doctors.

To promote and disseminate the findings of this doctoral research throughout the academic community and beyond, the findings have been published in peer-reviewed journals [7-10] and presented at academic conferences. The
impact of this research and the public interest in these findings has been demonstrated by the news coverage received from both the medical and general press [11-18] and the high Altmetric Attention Score compared with outputs of the same age that each article received, with the 2015 meta-analytic review being the top article of 2015 for BMC Medicine in terms of page views.

The findings of this thesis will benefit both the UK medical regulator and the medical Royal Colleges, by providing insights into the performance measures they use. This research is also highly relevant to the medical education community - it not only demonstrates the benefits of using large secondary datasets in medical education research, but also provides a basis upon which further research can be developed. The overall aim of exploring doctors' professional performance is to enhance patient care and therefore patients too will be beneficiaries of this research.

Finally, this thesis provides a formal acknowledgement of the existence of gender differences in the professional performance of doctors and has raised the awareness of this important and pertinent issue. It has begun to create an understanding of how and why professional performance differs between the genders and has opened a forum of thought and discussion for those involved in medical education to consider how professional performance of doctors is assessed and the areas where training and further improvement in performance may be required.
Acknowledgements

I would like to express my deepest gratitude to my supervisors; Professor Dame Jane Dacre, Dr Katherine Woolf and Dr Henry Potts, for their selfless time, care, support and reassurance throughout this PhD experience. I have learnt great deal from you all – thank you.

I am grateful to colleagues at the General Medical Council for the data provided, queries answered and for the constructive recommendations and valuable comments on this thesis. I would also like to thank colleagues at the Royal College of Physicians for the data provided, queries answered and for the useful comments on Chapter 5.

Appreciation is due to all my colleagues at RDME, UCL who have supported me throughout this doctoral journey with words of encouragement, laughter and friendship – it has been an exciting and dynamic place to work. I would especially like to acknowledge the contribution of Dr Clare Wadlow (CW), Dr Rowena Viney (RV) and Dr Leila Mehdizadeh (LM) in Chapters 3 and 4, Chapter 6, and Chapter 7, respectively. I would also like to extend a special thanks to the GMC team at RDME, UCL for their help in facilitating access to the data used in Chapter 7 and answering my (many!) queries.

I am indebted to Professor Chris McManus for providing me with data and for his time and insightful comments.

I would also like to thank all my friends who have always shown an interest in what I was doing and how I was progressing.

Finally, I am hugely grateful to my family. Without your patience, understanding, support and unfailing belief in me, this thesis would not have been possible.
# Table of contents

Declaration.................................................................................................................. 2

Abstract .......................................................................................................................... 3

Impact statement ............................................................................................................. 5

Acknowledgements ......................................................................................................... 7

Peer reviewed publications ............................................................................................ 22

Chapter 1 - Introduction ................................................................................................. 24

1.1 Chapter summary ..................................................................................................... 24

1.2 Introduction .............................................................................................................. 25

1.3 Thesis structure ....................................................................................................... 25

1.4 Thesis aims and questions ....................................................................................... 28

1.5 Defining the problem ............................................................................................... 29

1.6 Conclusion ............................................................................................................... 64

Chapter 2 – Study populations, context and methodology ......................................... 65

2.1 Chapter summary ..................................................................................................... 65

2.2 Introduction .............................................................................................................. 66

2.3 Context ..................................................................................................................... 68

2.4 Datasets ................................................................................................................... 71
Chapter 3 - A cross-sectional study examining the association between a doctor’s gender and receiving sanctions against their medical registration in the UK

3.1 Chapter summary ................................................................. 89

3.2 Introduction ........................................................................... 91

3.3 Methods .............................................................................. 94

3.4 Results ................................................................................ 100

3.5 Discussion .......................................................................... 109

Chapter 4 - Gender differences in medico-legal action against doctors: a systematic review and meta-analysis ................................................................. 119
4.1 Chapter summary ................................................................. 119
4.2 Introduction........................................................................... 121
4.3 Methods.............................................................................. 121
4.4 Results................................................................................ 128
4.5 Discussion .......................................................................... 150

Chapter 5 - Passing MRCP(UK) PACES: a cross-sectional study examining the performance of doctors by gender and country............................................. 158

5.1 Chapter summary ................................................................. 158
5.2 Introduction........................................................................... 160
5.3 Methods.............................................................................. 161
5.4 Results................................................................................ 168
5.5 Discussion .......................................................................... 173

Chapter 6 – Gender differences in postgraduate medical examination pass rates: a systematic review and meta-analysis.................................................. 180

6.1 Chapter summary ................................................................. 180
6.2 Introduction........................................................................... 182
6.3 Methods.............................................................................. 182
6.4 Results................................................................................ 188
6.5 Discussion .......................................................................... 212
Chapter 7 – Gender differences in fitness to practise test scores – a cohort study of GPs ........................................................................................................... 219

7.1 Chapter summary ...................................................................................... 219

7.2 Introduction.................................................................................................. 221

7.3 Methods...................................................................................................... 222

7.4 Results........................................................................................................ 227

7.5 Discussion .................................................................................................. 239

Chapter 8 – Disciplinary action against doctors by gender - associations between empathy, communication style and personality traits: a cohort study ........................................................................................................... 246

8.1 Chapter summary ...................................................................................... 246

8.2 Introduction.................................................................................................. 248

8.3 Methods...................................................................................................... 250

8.4 Results........................................................................................................ 263

8.5 Discussion .................................................................................................. 280

Chapter 9 - Discussion.................................................................................... 289

9.1 Chapter summary ...................................................................................... 289

9.2 Introduction.................................................................................................. 290

9.3 Interpretation of findings ........................................................................... 291
9.4 Strengths and limitations of the research .................................. 300

9.5 Unanswered question and further research ................................ 301

9.6 Conclusion .................................................................................. 303

References ....................................................................................... 305

Appendices ...................................................................................... 349
List of tables

Table 2.1 Summary of the datasets used throughout this thesis. .................. 72

Table 3.1 Distribution of variables by gender of doctors ............................. 101

Table 3.2 The distribution of warnings and sanctions for each variable and the association of individual factors with warnings or sanctions ........ 103

Table 3.3 The adjusted OR for having warnings or sanctions against registration for each variable compared to its baseline having adjusted for all other variables ......................................................... 105

Table 3.4 Stratum-specific ORs for having warnings or sanctions imposed on registration if the doctor is female for each specialty divided into four categories ................................................................. 106

Table 3.5 The number and proportion of male and female doctors by outcome type ................................................................. 107

Table 3.6 ORs of a female doctor having each outcome type, while controlling for number of years since PMQ, region of PMQ, and specialty (as binary variables) ......................................................... 108

Table 3.7 Stratum-specific ORs for receiving undertakings is the doctor is female (with specialty as a binary variable) ................................. 108

Table 3.8 Stratum-specific ORs for erasure from the medical register if a doctor is female (with region of PMQ as a binary variable) .......... 109

Table 4.1 Search terms used for electronic database search ......................... 123

Table 4.2 Names of journals and conference abstracts searched electronically ......................................................................................... 124
Table 4.3 Characteristics of eligible studies..................................................... 131

Table 4.4 Methodological quality assessment using the Newcastle-Ottawa scale ........................................................................................................................ 141

Table 5.1 Distribution of variables by gender of the candidates (N=7670). 169

Table 5.2 Comparison of outcome between candidates with no missing data and candidates who were missing data for at least one variable ....... 170

Table 5.3 The adjusted odds ratio (OR) for passing PACES at first attempt for female candidates compared to male candidates, and ethnic minority candidates compared to white candidates, after adjusting for all other variables ........................................................................................................... 172

Table 6.1 Search terms used for electronic database search .................... 184

Table 6.2 Characteristics of eligible studies..................................................... 193

Table 6.3 Methodological assessment of cohort studies using the Newcastle-Ottawa Scale (NOS) ........................................................................................................... 201

Table 6.4 Methodological assessment of cross-sectional studies using a modified version of the Newcastle-Ottawa Scale (NOS) ................. 204

Table 6.5 Characteristics of the eligible studies included in the meta-analysis ................................................................................................................................. 207

Table 7.1 Definition of outcome types............................................................ 225

Table 7.2 Demographics of the 120 GPs who completed a KT and SS between 2008 and 2014 as part of a GMC investigation process, by outcome type .................................................................................................................. 229
Table 7.3 The distribution of each variable by gender of the 120 GPs under investigation by the GMC, and the association of individual variables with gender ........................................................................................................ 231

Table 7.4 Mean and standard deviation of the standardised KT scores (z-scores) by GP group and gender ........................................................................................................ 233

Table 7.5 Mean and standard deviation of the standardised SS scores (%) by GP group and gender ........................................................................................................ 236

Table 7.6 Correlation matrix of the five clinical assessment domains .... 236

Table 7.7 Univariate associations between investigation outcome, gender, KT score, and SS score ........................................................................................................ 237

Table 8.1 Description of the 3 datasets used to create the overall study dataset................................................................................................................................. 253

Table 8.2 Correlation matrix of the communication factors .................... 258

Table 8.3 Table showing the factor items, factor loadings and uniqueness of variance for the exploratory factor analysis on communication style .. 260

Table 8.4 Demographics of the 2763 cohort members who are or have been registered with the GMC ....................................................................................................... 263

Table 8.5 Number and percentage of participants who responded to the questions for each co-variate, by gender and sanction status ............... 265

Table 8.6 Correlation matrix of all variables included in the path model .... 267

Table 8.7 The association between gender and active and/or historical sanctions ................................................................................................. 268
Table 8.8 Mean scores, standard deviations and 95% confidence intervals for empathy, communication and personality measures by gender .......... 271

Table 8.9 Mean scores, standard deviations and 95% confidence intervals for empathy, communication and personality measures by sanction status .................................................................................................................................................. 272

Table 8.10 Mean scores, standard deviations and 95% confidence intervals for empathy, communication and personality measures by gender and sanction status .................................................................................................................................................. 273

Table 8.11 Comparison of participants with no data missing for the co-variates of interest and those participants with data missing, by gender and sanction status .................................................................................................................................................. 275

Table 8.12 Comparison of data of participants with complete data and those with incomplete data for the co-variates of interest ........................................... 275

Table 8.13 Beta coefficients, 95% confidence intervals and p values of the relationships between gender, outcome, personality measures, communication styles and empathy measures used to construct path diagram (using multiple imputation) ........................................................................................................ 277

Table 8.14 Beta coefficients, 95% confidence intervals and p values of the relationships between gender, outcome, personality measures and empathy measures (not using multiple imputation) ............................... 279
List of figures

Figure 1.1 Flow diagram of the thesis chapters ........................................... 27

Figure 1.2 Good Medical Practice describes what it means to be a good doctor [19]................................................................. 31

Figure 1.3 RCP's working party on medical professionalism description of medical professionalism [3].................................................. 32

Figure 1.4 Diagram of the medical career pathways in the UK and the methods used to routinely assess professional performance .......... 39

Figure 1.5 Logic model displaying the potential factors influencing the gender difference in doctors' performance and demonstrating how this doctoral research explored a selection of these factors............................... 63

Figure 2.1 Description of the structure of the chapter ............................... 67

Figure 2.2 Key dates in the history of the GMC from 1858 to 2019 ............ 70

Figure 3.1 Summary of GMC investigation pathway of complaints to the GMC that have triggered a GMC investigation................................. 92

Figure 4.1 Flow chart showing reports retrieved, and articles excluded and included in the review based on the PRISMA statement .............. 128

Figure 4.2 Results of meta-analysis of 19 reports on the association between doctors' gender and experience of medico-legal action using both fixed-effects and random-effects modelling.................................................. 147

Figure 4.3 Funnel plot of effect size by inverse standard error ............... 150

Figure 5.1 Carousel of PACES stations (source: MRCP(UK), permission obtained to reproduce figure in September 2016)...................... 164
Figure 5.2 The seven core skills assessed during PACES (source: MRCP(UK), permission obtained to reproduce figure in September 2016) ............................................................ 165

Figure 6.1 Flow chart showing reports retrieved, excluded and articles included in the review based on the PRISMA statement ................. 188

Figure 6.2 Results of meta-analysis of 16 findings on the association between doctors' gender and pass rates on postgraduate medical examinations using random-effect modelling........................................ 210

Figure 6.3 Funnel plot of effect size by inverse standard error .............. 212

Figure 7.1 Flow diagram of how the population of GPs under investigation who completed KT and who completed SS ........................................... 228

Figure 7.2 Flow diagrams of the volunteer GPs who completed KT and volunteer GPs who completed SS ...................................................... 230

Figure 7.3 Distribution of standardised knowledge test (KT: written assessment) and Simulated Surgery (SS: clinical assessment) scores by GP group and gender .................................................. 234

Figure 7.4 Line graph showing the interaction between gender and GP group in KT scores ........................................................................ 235

Figure 7.5 Written and clinical assessment scores of GPs under investigation and GMC investigation outcome, by GP gender ....................... 238

Figure 7.6 Path diagram showing the relationship between gender, KT score, SS score and outcome. ..................................................... 239
Figure 8.1 Timeline of when the five postal questionnaires were administered and when the sanction data was collected (See [343] for further details).
                                                                                   .................................................................................................................. 252

Figure 8.2 Flow chart showing describing information flows and data linkage
                                                                                   .................................................................................................................. 254

Figure 8.3 Scree plot of the eigenvalues of factors ........................................... 259

Figure 8.4 Timeline showing when data of each co-variate were collected.
                                                                                   .................................................................................................................. 261

Figure 8.5 Flow diagram showing when data in each measure was collected and the response rate .................................................. 264

Figure 8.6 Flow diagram showing the number of cohort members with complete and missing data ........................................... 274

Figure 8.7 Path diagram (using multiple imputation) showing the relationship between gender, empathy measures, communication styles, personality measures and outcome. .................................................. 278

Figure 8.8 Path diagram (not using multiple imputation) showing the relationships between gender, empathy measures, communication style, personality measures and outcome. .................................................. 280
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>95% CI</td>
<td>95% confidence interval</td>
</tr>
<tr>
<td>A level</td>
<td>Advanced level</td>
</tr>
<tr>
<td>ABIM</td>
<td>American Board of Internal Medicine</td>
</tr>
<tr>
<td>ABPN</td>
<td>American Board of Psychiatry and Neurology</td>
</tr>
<tr>
<td>ABS</td>
<td>American Board of Surgery</td>
</tr>
<tr>
<td>ACER</td>
<td>Australian Council for Educational Research</td>
</tr>
<tr>
<td>ACP-ASIM</td>
<td>American College of Physicians-American Society of Internal Medicine</td>
</tr>
<tr>
<td>AKT</td>
<td>Applied Knowledge Test</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>ARCP</td>
<td>Annual Review of Competence Progression</td>
</tr>
<tr>
<td>BMAT</td>
<td>BioMedical Admissions Test</td>
</tr>
<tr>
<td>CCT</td>
<td>Certificate of Completion of Training</td>
</tr>
<tr>
<td>CEGPR</td>
<td>Certificate of Eligibility for GP Registration</td>
</tr>
<tr>
<td>CESR</td>
<td>Certificate of Eligibility for Specialist Registration</td>
</tr>
<tr>
<td>CMT</td>
<td>Core Medical Training</td>
</tr>
<tr>
<td>COMLEX-USA</td>
<td>Comprehensive Osteopathic Medical Licensing Examination</td>
</tr>
<tr>
<td>CPD</td>
<td>Continuing Professional Development</td>
</tr>
<tr>
<td>CSA</td>
<td>Clinical Skills Assessment</td>
</tr>
<tr>
<td>CSM</td>
<td>Communicator Style Measure</td>
</tr>
<tr>
<td>d.f.</td>
<td>Degrees of freedom</td>
</tr>
<tr>
<td>ECFMG</td>
<td>Educational Commission for Foreign Medical Graduates</td>
</tr>
<tr>
<td>EEA</td>
<td>European Economic Area</td>
</tr>
<tr>
<td>F1</td>
<td>1st year of Foundation Programme</td>
</tr>
<tr>
<td>F2</td>
<td>2nd year of Foundation Programme</td>
</tr>
<tr>
<td>F1CC</td>
<td>Foundation year 1 Certificate of Completion</td>
</tr>
<tr>
<td>FMGEMS</td>
<td>Foreign Medical Graduate Examination in the Medical Sciences</td>
</tr>
<tr>
<td>FPCC</td>
<td>Foundation Programme Certificate of Completion</td>
</tr>
<tr>
<td>FRCA</td>
<td>Fellowship of the Royal College of Anaesthetists</td>
</tr>
<tr>
<td>FRCR</td>
<td>Fellowship of the Royal College of Radiologists</td>
</tr>
<tr>
<td>FIP</td>
<td>Fitness to practice</td>
</tr>
<tr>
<td>GAMSAT</td>
<td>Graduate Medical School Admissions Test</td>
</tr>
<tr>
<td>GCSE</td>
<td>General Certificate of Secondary Education</td>
</tr>
<tr>
<td>GDPR</td>
<td>General Data Protection Regulation</td>
</tr>
<tr>
<td>GMC</td>
<td>General Medical Council</td>
</tr>
<tr>
<td>GP</td>
<td>General Practitioner</td>
</tr>
<tr>
<td>HEE</td>
<td>Health Education England</td>
</tr>
<tr>
<td>IMG</td>
<td>International medical graduates</td>
</tr>
<tr>
<td>ITE</td>
<td>In-Training Examination</td>
</tr>
<tr>
<td>KT</td>
<td>Knowledge test</td>
</tr>
<tr>
<td>LETB</td>
<td>Local Education Training Boards</td>
</tr>
<tr>
<td>LRMP</td>
<td>List of Registered Medical Practitioners</td>
</tr>
<tr>
<td>MCAT</td>
<td>Medical College Admission Test</td>
</tr>
<tr>
<td>MFPHM</td>
<td>Membership of the Faculty of Public Health Medicine</td>
</tr>
</tbody>
</table>
Peer reviewed publications

The citations of the peer reviewed publications that have arisen from work relating to this thesis and the chapters of the thesis that present this published work are presented below.

Chapter 3:


Chapter 4:


Chapter 5:


Chapter 7:


Other publication:

The citation of a peer reviewed publication that has arisen from work related to this thesis and is tangential to the thesis is presented below.
Chapter 1 - Introduction

“Patients must be able to trust doctors with their lives and health…[doctors] must show respect for human life and make sure [their] practice meets the standard expected…”

GMC, Good medical practice, 2013 [19]

1.1 Chapter summary

I begin this chapter by setting out the structure of this thesis before stating the research question, aims and objectives. I discuss the concept of medical professionalism and the methods used to measure and monitor doctors’ performance throughout their careers. I move on to describe how the gender ratios in medicine have changed over the course of history and the association between gender and professional performance in doctors. I close this chapter with a description of the theoretical and methodological stance I have taken throughout this thesis.
1.2 Introduction

It is widely acknowledged that professionalism is at the core of good medical practice [3, 20], however there is no simple definition of medical professionalism. Following the research question, aims and objectives of this thesis, I then aim to briefly summarise the history of medical professionalism and review the definitions published by several medical organisations, whilst also discussing the importance of medical professionalism in today’s society. Partly due to the of a lack of a single definition of medical professionalism there are multiple methods used to measure and monitor the professional performance of doctors in the United Kingdom (UK), and I will describe these further.

The gender composition of UK medical workforce has changed dramatically over the last century, with an increasing number of women choosing to study and practise medicine. Within this chapter, I will describe how the proportion of women studying and practising medicine through the course of history to the modern day has varied, and the perceived impact on the medical profession. I will also briefly describe the relationship between gender and the performance measures used through medical careers, along with the other major factors that are known to affect the professional performance of doctors. I will present a justification for exploring the gender difference in the professional performance of doctors practising in the UK. Finally, I will reflect on my academic background and skills, and how these have influenced my choice of research question and the methods I have chosen to use to explore and address the aims of this thesis.

1.3 Thesis structure

The thesis is composed of nine chapters; in this section I will explain what the reader can expect from them (see Figure 1.1).

I have stated the contents of this first and opening chapter above.
In the second chapter I set the context of the research by describing the medical Royal Colleges and Faculties, and the UK medical regulator (the General Medical Council, GMC). I outline the different datasets explored to complete this research and the study populations that were examined. I continue by defining the key terms used throughout this thesis and then state the main outcome measures and their validity and reliability. I then report the statistical methods employed throughout this thesis, and then follow with a rationale for using large secondary datasets to explore the research question. I report the study designs used throughout this thesis, before finally describing the current interest in the professional performance of doctors.

Chapters 3-8 each present one of the six research studies completed to inform the answer to the thesis’ main research question. Each six of these studies had their own specific aims, which contributed to the overall thesis’ aims. Each chapter has its own methods section, results and discussion. The studies are presented in chronological order, and each study attempts to build upon the preceding studies and findings.

The final chapter summarises and discusses the overall findings of this thesis. The overall strengths and limitations of this doctoral research are explored before considering the potential for further research opportunities in this area.
Figure 1.1 Flow diagram of the thesis chapters

**Chapter 1: Introductory chapter**
States the thesis aims and research question. Defines the problem the thesis is addressing, and justifies why it is a problem that requires further exploration.

**Chapter 2: Methods chapter**
Explains and justifies the methodological approaches used to address the research question.

**Chapter 3: Research study 1**
Large cross-sectional study establishing the magnitude of the association between doctors’ gender and a measure of extreme poor professional performance (having sanctions or warnings imposed on their medical registration) in doctors practising in the UK.

**Chapter 4: Research study 2**
Systematic review and meta-analysis of the gender differences in the medico-legal action against doctors to establish the robustness of the association between gender and a measure of poor professional performance.

**Chapter 5: Research study 3**
Cross-sectional study exploring the gender difference in the performance of doctors who have completed the clinical component of the Membership of the Royal College of Physicians (United Kingdom) diploma.

**Chapter 6: Research study 4**
Systematic review and meta-analysis of the gender difference in the performance at postgraduate medical examinations.

**Chapter 7: Research study 5**
Cohort study examining the performance of General Practitioners who have fitness to practise concerns and General Practitioners who volunteer to complete a high-stakes examination by gender.

**Chapter 8: Research study 6**
Cohort study examining how a selection of psychological factors may mediate the relationship between doctors’ gender and having sanctions imposed on their medical registration.

**Chapter 9: Discussion chapter**
Explains and interprets the thesis findings. Considers how the findings may inform medical education and the future of medical education research.
1.4 Thesis aims and questions

1.4.1 Research question

Throughout this thesis I seek to address the following research question:

*Is there a gender difference in doctors’ professional performance, and if so why?*

1.4.2 Aims

1. To investigate the relationship between gender and professional performance in doctors practising in the UK.
2. Facilitate a greater understanding of the ways gender may influence professional performance in medical doctors.

1.4.3 Objectives

1. Data from the UK List of Registered Medical Practitioners (LRMP) will be analysed to quantify the association between gender and odds of action against medical registration.
2. A meta-analysis of the published peer-reviewed literature will measure the association between gender and medico-legal action against medical registration globally and over time.
3. Clinical examination data from the oldest medical college in England will be used to investigate the relationship between gender and pass rates.
4. A meta-analysis of the published peer-reviewed literature will quantify the association between gender and postgraduate medical examination performance from multiple specialties in Canada, Ireland, UK and United States of America (USA).
5. Data will be extracted from General Practitioners (GPs) who have completed a written and clinical examination used by the UK medical regulator to assess fitness to practise. The data will be synthesised to establish if there is a gender difference in examination performance, and whether this gender difference varies between a group of GPs.
who were undergoing a formal investigation by the medical regulator and a group of GPs who volunteered to complete the examinations.

6. Data from a representative cohort of doctors practising in the UK will be used to explore whether empathy, communication style and personality differ between gender, and whether these variables mediate the association between gender and risk of action against doctors’ medical registration.

1.5 Defining the problem

Within this section I aim to enable the reader to gain a better understanding of my decision to focus on the relationship between gender and professional performance in doctors practising in the UK. I will define and highlight the importance of professional performance in doctors, and describe how professional performance can be measured in the medical workforce, before justifying my choices of performance measures used throughout this thesis. I will explain my decision to focus on doctors’ gender by defining gender and the role of gender in modern medicine. I will also describe the literature on gender and professional performance in medicine, before finally justifying my choice to explore doctors’ professional performance from a gender perspective. I will close this section with an explanation of the theoretical and methodological underpinnings of this thesis.

1.5.1 Professional performance in doctors

1.5.1.1 Defining medical professionalism

Medicine is a profession. To gain a greater understanding of medical professionalism I feel it is important to firstly consider what a profession is. A profession can be defined as an occupation that requires the “acquisition and application of […] knowledge and technical skills” [21]. Membership to a profession is controlled through a demonstration of the required knowledge and skills usually by examination and certification. Members of a profession are expected to share and adhere to codes of moral conduct and ethical obligations. The professions themselves must also regulate the practice of
their members through professional associations. In the UK the GMC and the medical Royal Colleges and Faculties regulate the medical profession.

The relationship between the medical profession and society has been described as a social contract [22]. This relationship was viewed as reciprocal with doctors being granted status, respect, autonomy, self-regulation and financial rewards and in return they were expected to be competent, altruistic, moral and attend to the health care needs of the members of society [22]. Medical professionalism is seen to be fundamental to this relationship between doctors and society, and forms the basis of the obligations and expectations of the medical profession [22]. This relationship, however, is dynamic and evolves as the needs and demands of society changes and to the changes in healthcare. An example of this is when society began to challenge the medical profession’s authority in the late 20th century. The medical profession was criticised for poor self-regulation and for protecting poorly performing colleagues; it was perceived as increasingly self-serving and slow to address new diseases or healthcare issues [22, 23]. In response to, and to address society’s concerns, the medical profession has had to reconsider the definition of medical professionalism in line with the expectations of a contemporary society and demonstrate that medical professionalism is still held in high regard by practising doctors [22, 24].

In today’s society there is no simple definition of medical professionalism. Medical professionalism is complex and must reflect the diverse and heterogeneous nature of medicine and medical careers. As elegantly argued by Swick “the concept of medical professionalism must account for the nature of the medical profession and must be grounded in what physicians actually do and how they act, individually and collectively” [25]. Many of North America and Europe’s major medical organisations have provided a definition of medical professionalism to guide their members, medical education and medical practice, thereby reflecting the importance of medical professionalism today. However, each of these definitions varies slightly, but
in general I propose there is an overall consensus between these organisations. I will proceed by describing some of the guidance published.

For doctors practising in the UK, the UK medical regulator (GMC) has published a guidance document entitled “Good medical practice” [19], which sets out the core guidance of professional values and behaviours expected of all doctors practising in the UK – see Figure 1.2.

**Figure 1.2 Good Medical Practice describes what it means to be a good doctor [19]**

It says as a good doctor you will:

- Make the care of your patient your first concern
- Be competent and keep your professional knowledge and skills up to date
- Take prompt action if you think patient safety is being compromised
- Establish and maintain good partnerships with your patients and colleagues
- Maintain trust in you and the profession by being open, honest and acting with integrity

Though “Good medical practice” does not provide an explicit definition of medical professionalism, it does describe how the GMC expects doctors to behave and sets the expectation that doctors are responsible for the development and maintenance of their individual professional performance. The GMC states that doctors must be competent in all aspects of their work and that they must participate in activities that maintain and develop their competence, thereby keeping up to date. Doctors are also expected to monitor and improve their work continuously throughout their medical careers. “Good medical practice” makes it clear that it is the individual doctor’s responsibility to ensure that their professional performance meets the standards expected and set out by the GMC. The GMC’s role in ensuring that doctors who practice in the UK are demonstrating and upholding medical professionalism is extended through the regular process of revalidation (discussed further in section 1.5.1.3.3.5) in which all practising doctors must participate, and through the fitness to practise procedures (discussed further in section 3.2) in which doctors whose fitness to practise
has raised concerns are reviewed and assessed. (Further details on the role of the GMC are described in section 2.3.2).

The Royal College of Physicians (RCP) is a medical professional organisation in the UK whose mission is to influence the design and delivery of healthcare, promote good healthcare, and support doctors to fulfil their potential [26]. In 2005, the RCP’s Working Party on medical professionalism published a report that defined medical professionalism as “a set of values, behaviours and relationships that underpin the trust the public has in doctors” [3] (see Figure 1.3 for further details). More recently, the RCP have published a report with the aim of facilitating a greater practical understanding of medical professionalism for practising doctors, by breaking down medical professionalism into seven roles: doctor as healer; doctor as patient partner; doctor as team worker; doctor as manager and leader; doctor as advocate; doctor as learner and teacher; and doctor as innovator [20].

![Figure 1.3 RCP’s working party on medical professionalism description of medical professionalism [3]](image)

"Medicine is a vocation in which doctor’s knowledge, clinical skills, and judgement are put in the service of protecting and restoring human well-being. This purpose is realised through a partnership between patient and doctor, one based on mutual respect, individual responsibility, and appropriate accountability.

In their day-to-day practice, doctors are committed to:

- Integrity
- Compassion
- Altruism
- Continuous improvement
- Excellence
- Working in partnership with members of the wider healthcare team

These values, which underpin the science and practice of medicine, form the basis for a moral contract between the medical profession and society. Each party has a duty to work to strengthen the system of healthcare on which our collective human dignity depends."

In the USA, the American Board of Internal Medicine’s (ABIM) Project Professionalism defined six components of medical professionalism:
altruism; accountability; excellence; duty; honour and integrity; and respect for others [27]. Finally, the international collaboration between the ABIM Foundation, the American College of Physicians-American Society of Internal Medicine (ACP-ASIM) Foundation, and the European Federation of Internal Medicine delivered the Physicians’ Charter [28]. The resultant Physicians’ Charter states three fundamental principles (primacy of patient welfare; patient autonomy; social justice) and ten professional responsibilities (professional competence; honesty with patients; patient confidentiality; maintaining appropriate relationships with patients; improving quality of care; improving access to care; just distribution of finite resources; scientific knowledge; maintaining trust by managing conflicts of interest; professional responsibilities) [28].

The above demonstrates that practising doctors are expected to consistently demonstrate and uphold medical professionalism, and though there is no universally agreed definition of medical professionalism, the overlap between the different international organisations on the role of a doctor in maintaining professionalism demonstrates an overall consensus of what medical professionalism is. It is my belief and understanding that medical professionalism is a combination of behaviours, characteristics, relationships and responsibilities expected of a competent doctor and that the professional performance of a doctor is the demonstration of this medical professionalism in their everyday medical practice.

The renewed focus on medical professionalism demonstrates how medical professionalism is also influenced by the expectations and needs of the society they serve and it is a concept that is continuously evolving with time [29].
1.5.1.2 The importance of medical professionalism

“[Medical professionalism] encompasses who doctors are, how they work and what they value”

Tweedie, Hordern & Dacre, Advancing medical professionalism, 2018 [20]

There has been a resurgence of the importance of medical professionalism, with international medical organisations publishing on the topic [30, 28, 3] (see 1.5.1.1). Medical professionalism is important because it defines what is expected of a doctor in today’s modern society [20]. It is at the heart of good medical care, with its purpose being to set the standard and support good medical practice. It also underpins the public’s confidence in the profession and improves the productivity of health systems [3, 20].

The GMC’s publication “Good medical practice”, with its explicit statements about the standards expected of a doctor practising in the UK, demonstrates not only how important the professional performance of doctors is in maintaining the trust of the public, but also how highly valued professionalism is by the medical regulator.

1.5.1.3 Measuring professional performance in doctors

Professionalism is complex and encompasses many different and varied components. It is related to individual characteristics of doctors, such as their skills, knowledge, personality and health but is also influenced by organisational factors [31, 32]. The multi-dimensional aspects of professionalism can make it challenging to measure and evaluate objectively [33].

Instruments and tools have been developed to specifically assess and measure professionalism in medical students and doctors. These instruments are diverse and can include: self-administered ratings; multi-source feedback; simulations; patients’ opinions; direct observations; and role model evaluation [34, 35]. A systematic review examining these tools concluded that only a limited number are methodologically sound with a
number having validity and reliability issues [34]. The findings of this systematic review are not alone in highlighting that the measurement of professionalism in medicine is problematic [36, 37].

Despite this, doctors continually have specific aspects of their professional performance measured and assessed through a variety of methods throughout their medical careers (see 1.5.1.3.1 to 1.5.1.3.3). The purpose of this continual assessment is to ensure that all practising doctors are up to date and fit to practice. It also provides greater assurance to their patients, the public, their employers and other healthcare professionals. The overarching aim of measuring the professional performance of doctors is to maintain and improve the quality of care delivered.

I will provide brief description of the most commonly used methods to assess doctors’ performance throughout the continuum of their medical careers here.

1.5.1.3.1 Medical school selection methods

The assessment of how a doctor may perform professionally starts when a prospective doctor applies to medical school. Medicine is a competitive career choice and doctors fulfil an important and demanding role in society. Training a future doctor is expensive and time-consuming. Thus medical schools are careful when selecting individuals to join their medicine courses. Obtaining a place to study medicine at a UK university is competitive – in 2013 the competition ratio for medicine was 11.2, this was more than double the average competition ratio for all other higher education courses [38]. A successful applicant to an undergraduate medical course needs to perform well on a variety of selection processes to secure a place at their chosen medical school. Medical school selection processes are complex and vary between medical schools [38, 39]. However, they all tend to use measures of academic attainment and measures of the core skills, values and attributes that are felt to be important for future doctors and will be further developed during medical training. The selection measures used by medical schools
are guided by Health Education England’s (HEE) values based recruitment (VBR), which aims to attract and recruit on the basis of individual values and behaviours that align with the values of the National Health Service (NHS) Constitution [40]. I will provide a brief description of the most common UK medical school selection methods below.

1.5.1.3.1.1 Academic attainment
All UK medical schools select prospective medical students based on their academic attainment [38]. Academic attainment at school has been shown to be a significant predictor of future academic performance at medical school [41-45]. Medicine is viewed as an academically demanding degree and therefore medical schools are keen to accept students who will be able to meet the demands of their course. Most UK medical schools require high grades at both A levels and GCSEs, although there is variability between medical schools in the actual grade requirements [46].

1.5.1.3.1.2 Aptitude tests
Intellectual ability has been demonstrated to be predictive of future academic ability [44], but intellectual ability alone does not make a good doctor, other personal qualities are required [47]. In response to a lack of reliable and valid selection methods, aptitude tests were introduced by some medical schools as a further method used in the selection process of prospective medical students. They were introduced as part of medical schools’ selection process in response to concerns about the increasing proportion of A level candidates achieving top grades (and therefore making it difficult for admission departments to differentiate between applicants) and for the call to increase diversity (as research suggested that candidates from state schools were under-represented) [47].

There are several different aptitude tests available to UK medical schools and I will discuss the three most commonly used aptitude tests here.

The University Clinical Admissions Test (UCAT), formerly known as the UK Clinical Aptitude Test (UKCAT), is currently used by twenty-five UK medical
schools as part of their selection procedures [46] (for more information on UCAT see [48]). Introduced in 2006, UCAT measures different aspects of the prospective medical student’s cognitive ability through verbal reasoning, quantitative reasoning, abstract reasoning and decision-making [49].

The BioMedical Admissions Test (BMAT) was introduced in 2003 [50] and is currently used by seven UK medical schools for UK applicants [51]. It consists of three sections. The first section assesses aptitude and skills, the second section assesses scientific knowledge and applications, and the final section is a writing task. It is intended to be challenging and aims to assess candidates’ academic ability.

The Australian Council for Educational Research (ACER) developed the Graduate Medical School Admissions Tests (GAMSAT) [52]. GAMSAT is an aptitude test used specifically for those candidates who have completed an undergraduate degree. Nine medical schools in the UK currently require it as a prerequisite for entry into their graduate programmes [53]. It consists of three sections: the first is reasoning in humanities and social sciences; the second is a written communication; and the third is reasoning in biological and physical sciences [54]. Its purpose is to assess candidates’ critical thinking and ability to present thoughts in a clear and logical manner.

1.5.1.3.2 Performance in undergraduate medical education

In the UK, medical students will spent between 4-7 years at medical school. Medical students’ performance is continually evaluated throughout undergraduate medical courses. Though the methods used vary between medical schools and may include hurdle assessments, workplace assessments and high stakes examinations [55], all UK medical schools are expected to meet the standards for medical education and training as set by the GMC [56].

A further measure of performance during undergraduate medical training are any self-declared Fitness to Practice declarations made to the GMC. The
GMC provides guidance on how medical schools manage fitness to practise and professionalism concerns in medical students [57]. Before being granted provisional medical registration (and being able to practice as a medical doctor), medical students in the UK are required to report any significant behavioural and health concerns that may impact their fitness to practise.

Influential research from North America has demonstrated that documented concerns raised about medical students’ academic performance or professional behaviour during their undergraduate medical training were associated with future disciplinary action by the state medical board once they qualified and were practising practitioners [58-60]. The conclusions of these studies have been challenged, with arguments stating that the link between behaviour as a medical student and future unprofessional conduct as a physician is weak and lacks the predictive power required to develop policies of either remediation or exclusion for medical students with problems [61, 62]. However, it remains that unprofessional behaviour in medical school has been found to be a risk factor for future concerns in clinical practice.

**1.5.1.3.3 Performance in postgraduate medicine**

As mentioned earlier, doctors’ performance is continuously monitored throughout their medical careers through a variety of measures. Figure 1.4 illustrates the methods used to assess professional performance in doctors practising in the UK throughout the continuum of their careers and the most commonly used measures of performance in postgraduate medical training that will be discussed further in this section.
**Figure 1.4 Diagram of the medical career pathways in the UK and the methods used to routinely assess professional performance**

<table>
<thead>
<tr>
<th>Foundation programme</th>
<th>Postgraduate medical training</th>
<th>Specialist and GP registers</th>
</tr>
</thead>
</table>

1.5.1.3.3.1 *Foundation Completion Certificates*

Upon graduation from medical school, newly qualified doctors enter the UK Foundation Programme. The UK Foundation Programme is a two-year structured programme during which medical graduates rotate between a variety of specialties and healthcare settings with the aim of building upon their medical school education and providing generic training in medicine. Medical graduates are required to successfully complete this two-year workplace based learning before applying for specialty or general practice training. For the first 12 months, newly qualified doctors are provisionally registered with the GMC with a licence to practise. Following satisfactory completion of the first year of the Foundation Programme (known as F1), a F1 Certificate of Completion (F1CC) is awarded and the foundation doctor is able to apply for full registration with the GMC and to progress to the second year of the Foundation Programme (known as F2). A Foundation Programme Certificate of Completion (FPCC) is awarded to foundation doctors who complete F2. The FPCC confirms that the doctor has reached
the level of competence expected for a specialty training appointment and the doctor can apply for specialty or general practice training. (See [63] for further details on the UK Foundation Programme).

1.5.1.3.3.2 Annual Review of Competence Progression (ARCP)
Doctors in training who began their training after August 2007 have their progress in their training programme assessed through the Annual Review of Competence Progression (ARCP). The ARCP replaces the Review of in Training Assessment (RITA) and is a formal process, which occurs at least annually. The ARCP process is a review of the competencies achieved by a doctor in training. These competencies are set down in the curriculum of the Royal College or Faculty of the specialty in which the doctor is training and are focused on the doctor’s education and attainment. During an ARCP, the trainee and the trainer of the trainee present evidence of the trainee’s progression to a panel of senior doctors. This panel decides, based on the evidence presented, whether the trainee is competent in the required skills and has the necessary knowledge to progress to the next level of training. Broadly speaking the outcome from an ARCP panel is either ‘satisfactory’ or ‘unsatisfactory’ [64]. There are several possible reasons for an ‘unsatisfactory’ outcome at an ARCP including examination failure, non-engagement with educational supervisor, issues with record keeping and evidence, issues with professional personal skills, and trainer absence [64].

The ARCP process is used for revalidation purposes for doctors in training. Revalidation is a process that all doctors, regardless of training status, must undergo and it is focused on the doctor’s fitness to practise and maintaining a doctor’s licence to practise. However, it is important to note that, although the ARCP has been enhanced to ensure the requirements of revalidation are met, the two processes are different (see 1.5.1.3.5 for further details about revalidation).

1.5.1.3.3.3 Postgraduate membership examinations
Each Royal College or Faculty has its own examinations, often called ‘membership examinations’. Doctors training in the specialty of the college
or faculty are required to complete and pass these examinations to complete their specialty training and obtain a certificate of completion of training (CCT) enabling them to apply for positions as a hospital consultant or GP (for more details on CCT see 1.5.1.3.3.4). Failing a postgraduate medical examination is one cause of not progressing satisfactorily at the ARCP [65].

Poor performance at postgraduate examinations has been linked with future disciplinary action, with those doctors obtaining higher scores at these examinations being at reduced risk of future disciplinary action [66, 67].

1.5.1.3.3.4 Certificate of completion of training (CCT)
Completion of specialty training can take between 3-8 years to complete, depending on the specialty [68]. A CCT is awarded to doctors who have successfully completed an approved training programme and have gained the competencies required by their Royal College or Faculty, including passing membership examinations for certain specialties [69]. The award of a CCT means that a doctor will become eligible to be included on the GMC’s Specialist or GP register. The Specialist and GP registers are a list of doctors who are eligible to take up an appointment as a consultant or a GP in the UK.

1.5.1.3.3.5 Revalidation
"Revalidation is based on the affirmation of good practice rather than the identification of bad apples."

Sir Donald H. Irvine, former President of the GMC, 2003 [70]

In 2004 the 5th report of the Shipman Inquiry, Dame Janet Smith, the chairman of the inquiry, made recommendations for revalidation [71]. Following her recommendations revalidation was introduced in 2012. Revalidation is a process for doctors to demonstrate that they are up to date and fit to practise, and thereby provide greater assurance to patients, the public and employers [72]. All doctors who have a licence to practice in the UK are required to participate in the revalidation process (usually every five years) to maintain their medical licence. There are three possible outcomes
of the revalidation process: revalidation of medical licence; defer the submission due to insufficient evidence or on-going local process; or non-engagement which can lead to the medical licence being withdrawn.

1.5.1.3.3.6 Action against doctors’ medical registration
It could be argued that the most definitive measure of poor professional performance in a doctor is action against their medical registration, especially if that action results in a restriction or barring of the doctor’s medical practice. A record of unprofessional behaviour at medical school has been found to be associated with future action against a doctor’s medical registration [58, 59]. Poor academic performance, at both an undergraduate and postgraduate level, has been shown to be associated with future action against a doctor’s medical registration [73, 66, 59, 74]. There is also evidence to suggest that psychological measures, including irresponsibility, lack of self-improvement and poor initiative, are associated with unprofessional behaviour in medical students [75]. (see section 1.5.2.1 for further details on the GMC’s role in assessing doctors’ fitness to practise).

1.5.2 Justification for the professional performance measures selected in this thesis
As a doctor, I have experienced first-hand the privilege to treat, care and advise the public, often when they are feeling at their most vulnerable. I have seen how valued the role of a clinician is, but also the importance of striving to maintain and improve the quality and standard of care delivered by clinicians.

Doctors hold a privileged position of trust in our society [2] and the public’s confidence in the standards and quality of care provided by the medical profession are held in high regard by both the profession and the regulatory body. This trust and the confidence in the medical profession can be compromised when doctors perform poorly and the consequences of doctors’ poor performance can be devastating [76].
I have chosen to explore two measures of professionalism in doctors: action against doctors’ medical registration, and performance at postgraduate medical examinations.

1.5.2.1 Action against doctors’ medical registration

Being investigated by the GMC is arguably one of the most serious and stressful events in a doctor’s professional life in the UK [77]. At the time of this research, if concerns about a doctor’s fitness to practise are raised with the GMC, the GMC may choose to open a formal investigation into the doctor’s fitness to practise. (The GMC also has other processes to investigate other concerns [78].) The investigation process aims to gain a complete overview of the doctor’s practice and varies depending on the evidence required. GMC case examiners will consider the evidence collected during the investigation. If the case examiners decide the doctor’s fitness is not impaired, they will either close the case or issue a warning. However, if the case examiners decide there is evidence to suggest the doctor’s fitness to practise is impaired, they may agree undertakings with the doctor or refer the case for a medical practitioners tribunal hearing. Medical practitioners tribunals are the final stage of a GMC investigation. At the end of the hearing, the tribunal may decide the doctor’s fitness is not impaired and will either close the case or issue a warning. When a tribunal find evidence that a doctor’s fitness has been impaired, they may do one of the following:

- Agree undertakings with the doctor;
- Put conditions on the doctor’s registration;
- Suspend the doctor’s name from the medical register;
- Erase the doctor from the medical register.

Having conditions imposed allows the doctor to continue practising, but will restrict the medical work the doctor is able to do. If a doctor is suspended, they will be unable to practise during the period of the suspension. If a doctor is erased, the individual will not be able to work as a doctor in the UK.
I have chosen to focus on any action against a doctor’s medical registration (erasure, suspension, conditions, undertakings and warnings). I have chosen to explore this measure of extreme poor professional performance because it identifies doctors whose professionalism has been compromised, and in the most severe cases their performance can be considered a potential serious risk to patient safety. I also chose to investigate this measure of poor professionalism because of the impact the investigation process and its outcomes can have on both the doctors themselves (in terms of impact on mental health, personal loss of earnings and impact on career [77]), on the health service (in terms of loss of workforce) and in terms of patient care and patient safety.

1.5.2.2 Postgraduate medical examination performance

Postgraduate medical examinations are important because passing these examinations is required in order to progress in a medical career. As a result, all doctors practising in the UK will be required to pass postgraduate medical examinations (compared to the very small proportion of doctors who have sanctions imposed against their medical registration). Using postgraduate medical examination performance will allow me to explore a much larger and more representative sample of doctors practising in the UK. Exploring postgraduate medical examinations also enables me to research different aspects of professional performance – written examinations focus on scientific knowledge, whereas clinical examinations assess clinical knowledge and skills, interpersonal relationships with patients and communication skills. Doctors invest time and money to prepare and sit postgraduate medical examinations and the result of these examinations can influence the progression of their careers. Failure to pass an examination has a financial implication, hinders career progression and can lower morale and confidence. A further reason for my choice of this professional performance measure is the established association between postgraduate examination performance and action against a doctor’s medical registration [66, 67], with clinical assessments in particular relating to action against medical registration [67].
1.5.2.3 Summary of argument for using selected performance measures

It could be argued that the measures I will use to explore professional performance in doctors are not measures that have been developed specifically for this purpose. However, I argue that performance at these measures reflect doctors’ “real-life” performance and can impact their medical careers and therefore it is important to explore these further. I also propose that the general consensus is that, although there are multiple methods for specifically assessing performance in doctors, the current evidence suggests these tools are not robust [34]. It is therefore reasonable to proceed to assess doctors’ professional performance using the measures I have selected.

Facilitating a greater understanding of the professional performance of doctors will enhance patient care, maintain the public’s confidence in the medical profession, and enable the medical profession and the medical education community to better support medical students and doctors.

1.5.3 Gender and medicine

1.5.3.1 Gender vs. sex – using the term gender

There is much debate about when to use the term “sex” and when to use the term “gender” in research. Sex is defined as the biological sex assigned to a person at birth, whereas gender is defined as a personal identification of one’s own gender. As a result, a person’s sex may not necessarily be the same as their gender. In other words, sex and gender are not synonymous.

Sex can be biologically defined by the difference in karyotype: 46XX for females and 46XY for males. The different karyotypes are usually reflected in phenotype differences between males and females, including differences in internal reproductive organs, external genitalia and hormones.

Sex is also one of the protected characteristics under the Equality Act of 2010 [79], meaning that it is against UK law to discriminate against a person either at the workplace or in wider society on the basis of their sex.
It would be reasonable to justify the use of the term “sex” in the context of this thesis. Performance is a complex composite of behavioural traits and it could be argued that these behaviours may be the result of ancestral intra-sexual competition or other genetic differences that predispose one group (i.e. males) to exhibit certain behaviours (i.e. risk taking) or to choose specialties in which these behaviours can be expressed (i.e. surgery). The argument that doctors’ behavioural traits are related to their biological sex, and that these behavioural traits influence their professional performance would therefore justify the use of the term “sex” throughout this thesis. There are also examples of experienced researchers in medical education using the term “sex” in the presentation of their research [80, 81].

Gender, on the other hand, is not as clear-cut. Unlike sex, at present there is no agreed upon validated tool for assessing gender [82]. Gender is self-identified and is influenced by social, cultural, environmental and behavioural factors [82]. It would therefore be reasonable to argue that doctors’ performance is influenced by their behaviour, and that behaviour is not solely influenced by genetics, but also by environmental exposures. It is likely that doctors’ performance and hence behaviour is influenced by a combination of both genetic and environmental influences. It has been suggested that data should be collected on both “sex” and “gender”, and that the findings are presented for both these variables separately. However, if this is not possible than it is recommended that the term “gender” is used, especially if the data on this variable is collected by self-report [82].

The research that forms the basis of this thesis has been performed using secondary datasets, and as such I had to work with the variables collected. Of the datasets used for the completion of the quantitative research studies in this thesis, six of them contained data on the gender or sex of the study population. The five large administrative datasets collected data on self-reported gender. The sixth dataset (survey data from a cohort of doctors who applied medical school in 1990) asked the study participants to self-report their sex. I have therefore decided to use the term “gender”
throughout this thesis. This decision is based on the fact that the majority of the datasets used for this research specifically collected data on “gender” and because in all the datasets the variable of gender/sex was self-reported. I also believe that performance measures analysed in this thesis are not solely influenced by genetics, but by the doctors’ social, cultural and environmental influences. I feel it is important to stress that this thesis is not an argument justifying the use of “sex” or “gender” in medical education research. It is a description of doctors’ performance categorised by their own self-report of belonging to one of two groups. Whether a doctor’s self-reported gender differs to their biologically assigned sex at birth is not in discussion here. Patterns of performance differ between men and women, and I aim to explore and discuss this throughout this thesis.

It is relevant and important to note that since I started working on this thesis the wider understandings of gender has evolved. In 2015 a poll of 1,000 individuals aged 18 to 34, half of the respondents viewed gender as a spectrum, rather than binary [83]. The adoption of gender-neutral marketing for children’s toys by some retailers reflects this shift away from binary gender [84]. However, for the purposes of this thesis gender is binary.

In conclusion, throughout this thesis I examine the difference in performance between those doctors who report themselves as being men and those who report themselves as being women and as such the term “gender” will be used throughout this thesis when describing the two groups.

1.5.3.2 The importance of exploring gender trends in medicine

Headlines about gender in medicine always grab attention [85-87, 11, 15]. This is likely to be due, in part, that this previously male dominated profession is attracting an increasing and high proportion of women [88-90, 4]. It was not until the late nineteenth century that women were able study at UK medical schools and practise as medical doctors, though the number of women doctors remained small with only 25 women registered as doctors in England and Wales by 1881 [91]. However, the political and societal
changes from the 1960s onwards (including The Sex Discrimination Act in 1975 [92]) resulted in an increasing number of women doctors, with a steady increase from the late 1960s up to the early 1990s when women became the majority intake into UK medical schools [93, 94, 91]. Though, at the time of writing, men continue to represent the majority of all UK licensed doctors (though only just at 53%) [95], this majority has been slowly reducing since the mid-1990s [94]. At the time of writing, women now form the majority of all UK doctors in training [95], and women are the majority in several medical specialties including General Practice [95], Obstetrics and Gynaecology, Clinical Oncology, Paediatrics, Pathology and Psychiatry [96]. However, across all hospital specialties men continue to hold the majority of consultant posts [97]. The impact of the increasing numbers of women choosing to practise medicine is a topic of great interest and discussion [98, 4]. A recent literature review summarised the main competing arguments regarding the implications of the increasing number of women in medicine [99]. It found that any gender differences in motivations, attitudes or academic performance were either narrowing or so small as to be of little practical importance. The authors found conflicting evidence for a gender difference in communication style, concluding that there was a growing body of literature demonstrating that there may not be gender-specific communication styles. However, they did find that medical encounters are shaped by gendered expectations and stereotypes and that these gendered expectations may influence the reported difference between male and female doctors. Finally, they concluded that much of the research conducted has presented conflicting findings and that further research is required to better understand the complexity of how gender may influence medical education.

Women in medicine were recently in the headlines, as new attention is paid to a gender pay gap amongst doctors working full-time in the UK. Women have been found to earn a third less than their male colleagues, a larger pay gap than is seen in many other professions across all employment sectors and much higher than the national average of a fifth [100]. A large review of the gender pay gap in the NHS has recently been commissioned by the
Department of Health and Social Care with the aim of understanding the causes of the gender pay gap in medicine and developing strategies to narrow the pay gap [101, 102]. This research is likely to shed light on any horizontal and vertical occupational segregation by gender in medical careers.

1.5.3.3 Gender and professional performance in medical careers

Since the introduction of revalidation in 2012, the performance of doctors practising in the UK is continuously monitored and assessed, beginning at application to medical school and continuing throughout their medical school and professional careers. To provide further context to the research within this thesis, I will briefly describe the gender differences in the different methods used to measure performance at each stage of the UK medical education continuum.

1.5.3.3.1 Gender and performance at medical school selection methods

Research analysing UK medical school admission statistics has demonstrated that women are more likely to apply and be accepted to medicine, regardless of whether the course is the traditional 5/6-year programme or the accelerated 4-year graduate programme [103]. Interestingly, women applicants have not been found to perform better than men at the academic requirements of medical schools, with no significant difference in the overall Universities and Colleges Admissions Service (UCAS) tariff points (which assigns a numerical score to the grades of each type of school qualification) between the genders [41]. Nor do they perform better than their male counterparts at the aptitude tests used by UK medical schools: in fact it is the men who perform better at the three most commonly used aptitude tests by UK medical schools (UCAT, BMAT and GAMSAT) [104, 105, 50, 106]. The finding that women are more likely to be offered a place at medical school suggests that women are either outperforming male candidates at the non-academic selection measures, including personal statements, interviews (these can take different forms) and work experience
or that selection is biased in favour of women. Further research is required to explore this in more depth.

1.5.3.3.2 Gender and performance in undergraduate medical education

Evidence from the literature has consistently shown a gender difference in the performance of medical students during the clinical years of their undergraduate training, with men not performing as well at clinical examinations (Objective Structured Clinical Examinations, OSCEs) [107-111]; being at greater risk of failing the final medical examinations [112, 113]; and having a reduced likelihood of achieving an honours result [113].

As I mentioned in section 1.5.1.3.2, self-declared Fitness to Practice declarations made to the GMC can be used as a performance measure during undergraduate medical training. Before being granted provisional medical registration (and being able to practice as a medical doctor), medical students in the UK are required to report any significant behavioural and health concerns that may impact their fitness to practise. It has been shown that men are more likely to have a Fitness to Practice declaration overall, and that this declaration is much more likely to be due to a conduct-related issue [114].

1.5.3.3.3 Gender and performance in postgraduate medicine

The gender difference in professional performance observed at the undergraduate level in medicine continues in postgraduate medicine. Men are more likely to obtain a less satisfactory outcome at ARCP and are therefore are at greater risk of not being able to progress satisfactorily in their training [65].

Gender differences have also been reported in the pass rates of postgraduate medical examinations. For example, the proportion of men who passed the postgraduate medical examinations for internal medicine specialty training (MRCP(UK)) in 2017 was lower than women at both the clinical examination (PACES) and the Specialty Certificate Examination
(SCE). However, the proportion of men who passed the written components of the MRCP(UK) examinations was either higher, or equivalent to the proportion of women who passed the examinations [115]. In General Practice training, a lower proportion of men passed both the written and clinical elements of the postgraduate medical examination for GP training (MRCGP) at first attempt in 2017 [116].

Throughout this thesis I will explore in greater depth the gender difference in performance at the MRCP(UK) Part 2 Clinical examination (PACES), investigating how any gender difference in examination performance varies between UK medical graduates, international medical graduates practising in the UK, and international graduates not currently practising in the UK (see Chapter 5). I will also investigate the gender differences in postgraduate medical examinations, exploring the different examination formats, different specialties and different countries (see Chapter 6). My aim in Chapter 5 and Chapter 6 is to provide a better understanding of whether there is a gender difference in the examination performance of doctors and whether these differences differ in terms of whether the doctor completed their undergraduate training in the country where they are sitting their postgraduate medical examinations, and whether examination format or specialty influence the association between gender and examination performance.

With regards to the regular revalidation a doctor practising in the UK is required to complete to maintain a license to practise, a gender difference in outcomes has been reported. Between 2012-2017, a higher proportion of men were revalidated (78% of men compared with 73% of women), but a higher proportion of men obtained the outcome of non-engagement during this period (0.4% of men compared with 0.1% of women). A slightly higher proportion of women have had their revalidation deferred (27% of women compared with 21% of men), but this gender discrepancy in deferral rates may be due to women doctors taking maternity leave [117, 118].
Finally, male doctors have been found to be at increased risk of fitness to practise issues [66, 74, 95], although this finding is not consistent [58, 59]. I will further explore the gender difference and other predictor variables in those doctors who have action taken against their medical registration and other similar measures of extreme poor professional performance in Chapter 3 and Chapter 4. There is also evidence to suggest that other factors may influence the risk of action against doctors’ medical registration, with evidence that academic performance influences the risk of future action against medical registration and an association between psychological measures and unprofessional behaviour in medical students. I will further delve into how knowledge and clinical skills differ between doctors who have had concerns raised about their fitness to practise and those who have had no concerns raised, whilst also examining for any gender difference in the examinations measuring knowledge and clinical skills (see Chapter 7). I will then explore whether there is a difference in a selection of psychological variables between male and female doctors, and doctors who have had actions against their medical registration compared to those who have not (see Chapter 8).

1.5.3.3.4 Gender and medical career trajectories

The proportion of men and women doctors on the UK medical register is approaching an even split, with men representing 53% of all doctors in 2017 [95]. This near even split is also achieved on the General Practice register (a register of doctors who have completed their postgraduate training in General Practice and are or were eligible to practice as GPs in the UK), with men representing 47% of this population. However, on the Specialist register (a register of doctors who have completed their postgraduate training in any specialty other than General Practice and are or were eligible to practice as consultants in the UK), men dominate with 65% representation. However, it is reasonable to hypothesise that an even split in terms of gender ratios may be achieved on the Specialist register in the future because in 2017 women formed the majority of medical students graduating from a UK university and they were the majority of doctors on the Foundation Training Programme (all
newly qualified doctors are required to complete this two year training programme before being able to apply for specialty or General Practice training). However, there is historical evidence to suggest that women do not progress to hospital consultant posts at the rate that would be expected in certain specialties [93, 119]. It would be appropriate to monitor the gender split in hospital specialties over the coming years and investigate the reasons behind any observed shortfall.

A recent cohort study related to this thesis, that I co-authored, found women were no more likely to apply for specialty training, but are more likely to receive an offer and accept an offer [120]. Following the career progression of this cohort would be of interest.

1.5.3.4 Gender differences on other variables known to affect medical performance

There are multiple factors that may influence the performance of medical students and of doctors at examinations. It is important to be aware of these factors and their influence on performance when researching how gender may affect professional performance, and so a brief overview of these is given below. I will not be exploring these factors in any great depth, but some of the factors are included as confounders in the research studies conducted as part of this thesis.

1.5.3.4.1 Socio-economic status

Socio-economic status has been found to bear no relationship to performance at medical school, once educational attainment has been taken into account [121, 122]. However, once practising as doctors, socio-economic status has been shown to be associated with lower pass rates for all specialty examinations and with a higher rate of unsatisfactory outcomes at the ARCP (excluding those unsatisfactory outcomes associated with examination failure) [123].
1.5.3.4.2 Ethnicity and country of graduation

It is a consistent finding that ethnicity does affect attainment at both undergraduate and postgraduate medical examination performance, with white candidates performing better than their colleagues from ethnic minorities [124-126]. This pattern is seen both in those doctors who have graduated from UK medical schools, and those doctors who have graduated from medical schools outside of the UK [127]. An interaction between ethnicity and gender has also been found when exploring the academic performance of medical students and doctors [108, 125]. A differential level of attainment at postgraduate medical examinations has also been reported between UK medical graduates and those doctors who graduated from a medical school outside of the UK, with UK graduates performing better at these assessments [124, 126].

When examining the small population of doctors who are referred to the UK medical regulator and are required to complete a performance assessment as part of an investigation into their fitness to practise, doctors who report being from ethnic minorities are over represented compared with the medical register [128]. It was also found that doctors who had qualified outside of the UK, with the exception of South African medical graduates, were more likely to have to undergo a performance assessment when compared to their UK qualified counterparts [128].

1.5.3.4.3 Number of years in practice

The number of years a doctor has been practising has also been shown to affect performance, with the performance of doctors decreasing with increasing years of practice [129-131], though the evidence is conflicting with a recent study of surgeons and patient mortality finding older surgeons had lower patient mortality [132]. There is also evidence to suggest that doctors who have practised for longer are at greater risk of having sanctions imposed against their medical registration [67].
1.5.3.5 Why is it important to examine performance from a gender perspective

I have chosen to examine the professional performance of doctors from a gender perspective. The recent shift in the gender distribution of UK medical students and doctors and the concern raised by Dame Carol Black in 2004 that the increase in women doctors may endanger the medical profession has triggered on-going debate [88, 89]. Dame Carol’s concerns centred on the career choices of women doctors and the resulting impact on the profession. She also raised fears about how the influence of the medical profession may be dampened by a workforce where women are the majority. The concerns raised by Dame Carol have been furthered by others, with suggestions that a predominantly female workforce will have implications for workforce planning, education and research [4]. These views suggest that there is a perceived difference in how men and women doctors practise medicine. I sought to explore whether professional performance was influenced by gender and whether any difference in professional performance was measurable, and if a difference is present, unpacking what the medical profession can learn from this difference to improve the professional performance of doctors as whole.

My choice to focus on the gender perspective of doctors’ professional performance is also grounded from my own personal experiences. I am interested in the role gender plays in medical careers for several reasons: firstly as a woman and as a qualified doctor, but also as an alumnus of UCL Medical School. The Royal Free Hospital, where I did the majority of my clinical training and where I later returned to study for this thesis, was the first teaching hospital in London to admit women for medical training [133]. It was during my first year of clinical training when Dame Carol Black triggered the debate about women in medicine [87]. The combination of training in a hospital with a history in supporting women to enter medical careers and to be training at a time when the role of women in medicine was being hotly debated have had an influence in my interest in how gender may influence performance in doctors. It is also no coincidence that I chose to work
towards a doctorate under the supervision of Professor Dame Jane Dacre, who was the chair of a research steering group at the RCP on Women in Medicine (which published their findings in 2009 [94]), and is the third woman to be elected president of the RCP in its 500-year history.

1.5.4 Theoretical and methodological underpinnings

This thesis is a piece of medical education research. Medical education research explores and critically evaluates the teaching, learning and professional practice of knowledge, skills and attitudes among healthcare providers throughout the continuum of their training and professional careers [134]. The aim of medical education research is to improve the quality and promote standards of the education and training of students and practitioners in the healthcare professions, including medicine. Medical education research guides and informs policies and educational decision-making and may take many forms: evaluating the transfer and acquisition of knowledge, attitudes and skills; developing, implementing and evaluating innovations in practice; and identifying patterns and trends in medical students and doctors learning and performance. The field of medical education research is broad and interdisciplinary, encompassing research techniques from medicine, education, sociology, psychology and linguistics, to name but a few [134, 135]. As a relatively young field of research, it is an exciting time to be involved in medical education. It is not a tightly defined discipline and as it continues to grow and develop, this expansion encourages rich and lively debates about the different theoretical and methodological paths researchers in this domain can develop [136].

Prior to commencing any research process, it is vital that a researcher reflects on the theoretical framework they will follow. The paradigm the researcher holds will influence how research questions are framed and the methodology that will be used.

I have chosen to use a positivist approach for my doctoral research. In positivism, the researcher is independent from the research and their role is
to collect data and interpret it in an objective way [137]. Positivism uses a systematic and scientific approach to research and tends to give rise to the use of quantitative methodology. Quantitative research methodology is structured and aims to measure and quantify research outcomes. The advantage of using a positivist approach is that the research findings are intended to be generalisable to a wider population and the research process itself should be replicable. However, positivism can be narrow in what it explores, often examining only one variable at a time while controlling for the effects of other variables. As such it can fail to deliver a deeper understanding of an issue, especially when examining complex phenomena such as professionalism. It can also be argued that it is not possible for a researcher to be truly detached and objective about the research they are conducting.

My choice to follow a positivist theoretical framework is grounded from my own professional and academic background as a medical doctor and holder of a master’s degree in epidemiology.

Epidemiology is a public health discipline with a focus on methodology. It is the study of the distribution and determinants of health outcomes and diseases in populations [138]. Epidemiology is data-driven and relies on a systematic unbiased approach to the collection, analysis and interpretation of data. Evidence based medicine introduces these epidemiological methods to the field of medical education, through promoting practitioners to use the current best evidence to inform individual level patient decision-making, rather than using beliefs and opinions. Evidence-based medicine is defined as “conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients” [139]. It promotes the use of evidence from well-designed and well-conducted research to inform and guide decision-making. It stems from the thinking that decisions based on beliefs or opinions may be limited by knowledge gaps and biases. Evidence-based medicine is regarded by many as the gold standard of clinical practice and has championed research study designs such as
systematic reviews, meta-analyses and randomised controlled trials (RCTs). There are, however, limitations to using evidence-based medicine methodology to guide decisions. Firstly, the study designs championed by evidence-based medicine can be time-consuming and expensive and therefore the methods may not always be feasible or appropriate. Due to the time-consuming nature of some of the methods, there may be a time lag from when a research question is posed to when the results are available and disseminated, this could result in the findings no longer being relevant. It is, however, a common misconception that evidence-based medicine requires RCTs to deliver the best evidence: evidence-based medicine is about using the correct methodology to answer a specific research question. A further limitation is that the research methods used in both the discipline of epidemiology and in evidence-based medicine analyse and interpret findings on populations. Individuals may vary from population norms and therefore the findings may not be applicable to each individual. Another limitation is that certain groups may be under-represented in the population examined and therefore the findings may not be generalisable to that group. Population size and selection methods need to be carefully considered to overcome this limitation. Despite these limitations, I have chosen to follow the epistemological stance from epidemiology and evidence-based medicine to guide my research methodology throughout this thesis.

When I first embarked on this doctoral journey, little was known about doctors practising in the UK whose professional performance fell below the expected standard. It therefore felt appropriate to examine performance measures in large populations of doctors with the objective of identifying trends and patterns in the professional performance of doctors. Quantitative methodology would enable me to achieve this objective. The findings of this doctoral research could then be used to inform future studies that may wish to delve deeper and explore specific aspects of professional performance, perhaps using alternative methodologies.
As mentioned earlier, professionalism is a highly complex area. The epistemological approach I have decided to follow allowed me to measure and quantify the aspects of professionalism I have chosen to specifically explore, however it may have failed to allow me to deliver a deeper understanding of how the other aspects that I have not explored and that may work together to influence professional performance in doctors. I have chosen to focus on poor professional performance as measured by action against a doctor’s medical registration or the failing of professional examinations. I have elected to study these measures quantitatively with the aim of identifying patterns or trends that may aid a better understanding of when doctors’ professional performance falls below the expected standard. My choice to focus on doctors’ poor professional performance was influenced by my university department’s strong links with the GMC. At the time of writing, the Research Department of Medical Education (RDME) at UCL had a contract with the GMC to design and set up specific assessments that reflect the practice of doctors who are being investigated by the GMC because concerns have been raised about their fitness to practise. The contract also included the RDME conducting research on doctors’ fitness to practise. Large amounts of quantitative data regarding the assessments and the performance of doctors who complete these assessments are potentially available to the RDME researchers working with the GMC unit. The opportunity to work with large datasets improves the chances of detecting true effects and exposing meaningful relationships between variables, especially when the outcome is rare, as in the case of action against medical registration.

Along with my department’s strong connection with the GMC, two of my doctoral supervisors have strong links with the RCP: Professor Dame Jane Dacre was both President of the RCP (2014-2018) and Medical Director of Membership of the Royal College of Physicians (UK) (MRCP(UK)) examination (2010-2013) and Dr Katherine Woolf is an educational advisor to MRCP(UK) examination. These strong links with the MRCP(UK) examination team provided me with the opportunity to access the large
quantitative data regarding the MRCP(UK) examinations. Using large datasets, such as the MRCP(UK) examination database, ensure that my research findings are generalisable to a wider population.

As mentioned earlier, professionalism is a large and complex area and therefore different research methodologies could have been used to explore it further. I could have used qualitative research methods to gain a different understanding of the doctors who do not perform as well as their peers. This form of research methodology may have enabled me to better understand how these doctors perceive their professional performance and how actions against their medical registration or failing an examination has impacted on their career and health. However, I was concerned that these doctors may not wish to engage in research exploring their performance. Failing a professional examination or having sanctions imposed against one’s medical registration can have negative consequences on a doctor’s career and can adversely impact their mental health through stress [77]. These doctors may also feel resentment towards the medical profession and regulator. It is for these reasons I thought that these doctors might not wish to participate actively in research exploring poor professional performance.

I remain aware of the limitations of my choice of methodology, but believe that it is still important research to conduct. The RCP Working Party report from 2005 highlighted the need for more research on professionalism in the medical profession [3]. This thesis goes towards addressing that need by measuring and quantifying how doctors’ professional performance differs between the genders. The research I have presented in this thesis can then be built upon by future research, which can explore more deeply why these differences exist and how the medical community can better support all doctors.

The quantitative methodological approaches I have chosen to use to address my research question include: meta-analyses to combine the results of multiple studies to robustly confirm trends and quantify the size of effects; and multivariate statistical methods to analyse large secondary datasets and
establish patterns in performance. Secondary data analysis has both benefits and drawbacks, and these are discussed in greater detail in Section 2.10.

As previously mentioned, the field of medical education is diverse and as such my approach is not unusual; experienced researchers in this field have presented and published their research using quantitative methodology and multivariate statistical techniques to explore large datasets. There are published examples of large longitudinal datasets, with both primary and secondary data being analysed using quantitative statistical methods including multivariate logistic regression and path analysis to address diverse medical education research topics such as: exploring how prior academic achievement are predictive of outcomes in medical school or postgraduate medical education [140, 44]; assessing the impact of the different uses of a medical school admission test [141]; and investigating the predictors of conduct or health-related issues that could impair medical students’ fitness to practise [114].

One further point that needs to be addressed is my decision for this research to have a UK focus. My choice to focus on the professional performance of doctors practising in the UK is influenced by the fact that the datasets available and accessible to me during this doctoral journey contained data that concerned UK practising doctors in the majority. Another reason for the UK focus was that countries around the world differ in how they train, examine and regulate doctors. This added the potential complexity of merging data from different countries. However, I was keen to explore how my UK focused research compared with similar research internationally. I therefore supplemented the research I performed using UK data with published international data through systematic reviews and meta-analyses. These systematic reviews and meta-analyses enabled me to discover and compare patterns and trends in doctors’ professional performance both in the UK and internationally.
Finally, before drawing this section to a close I would like to present a logic model to illustrate the potential factors influencing the gender differences in doctors’ professional performance and to highlight the factors I have chosen to focus upon throughout this thesis. The model demonstrates the factors I have chosen to explore and clearly shows which studies within this thesis explore each of these selected factors (see Figure 1.5). The logic model also allows me to present the desired outputs, outcomes and impact of this doctoral research.
Figure 1.5 Logic model displaying the potential factors influencing the gender difference in doctors’ performance and demonstrating how this doctoral research explored a selection of these factors

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Activities</th>
<th>Outputs</th>
<th>Outcomes</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristics</td>
<td>Research studies conducted to address research question</td>
<td>Confirm a gender difference in medico-legal action experience against doctors&lt;sup&gt;1,2,4&lt;/sup&gt;</td>
<td>Raise awareness about the gender difference in the professional performance of doctors</td>
<td>To contribute towards a greater understanding of the gender difference in the professional performance of doctors</td>
</tr>
<tr>
<td>Ethnicity&lt;sup&gt;5&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country of primary medical qualification&lt;sup&gt;1,2&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Socioeconomic status insights</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practice</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years qualified&lt;sup&gt;1&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specialty&lt;sup&gt;3&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hours worked</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient encounters</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge&lt;sup&gt;2,3&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skills&lt;sup&gt;2,4,6&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication style&lt;sup&gt;4&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personality&lt;sup&gt;6&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empathy&lt;sup&gt;2&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organisation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organisational structures</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local cultures</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Society</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gendered expectations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Possible gender biases</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entry into medical school</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Application to postgraduate examination</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Referral to regulatory body</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 1: Establishes association between doctors’ gender and extreme poor professional performance (disciplinary action) in doctors practicing in the UK</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 2: Establishes the robustness of the association between doctors’ gender and extreme poor professional performance (medico-legal action)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 3: Explores the association between doctors’ gender and performance at a clinical postgraduate examination</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 4: Explores association between doctors’ gender and performance at postgraduate examinations across examination type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 5: Compares and examines the performance of GPs with FIP concerns and volunteer GPs who complete FIP assessments, by gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 6: Examines how personality, communication style and empathy may mediate the relationship between gender and disciplinary action</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The numbering in superscript identifies the study/ies used to explore the inputs and outputs proposed.
1.6 Conclusion

Recent years have seen an increase in focus on medical professionalism, with medical organisations across the globe publishing guidance for their members. This interest in medical professionalism has been driven in part by widely publicised health scandals [142-145] and the need for the profession to maintain the public’s trust. There has also been increased attention on the role of gender in medicine, with concerns being raised about the increase in the number of women choosing to practise medicine and the resulting impact this may have on the profession [87, 88, 98, 4]. The practice of modern medicine is changing with patients presenting with more complex health care needs, changes in doctors’ working conditions with the advent of multi-disciplinary working, and growing expectations from patients to be more involved in their health care decisions [146]. Within this thesis I will explore the professional performance, specifically poor performance, in doctors from a gender perspective in today’s society with the aim to better understand, and ultimately contribute towards maintaining and improving doctors’ professional performance.
Chapter 2 – Study populations, context and methodology

2.1 Chapter summary

Within this chapter I describe the study populations examined throughout this thesis, with a particular focus on the career trajectory of doctors training and working in the UK. I provide a brief history of the GMC and RCP for context. Finally, I describe and justify the methodology used to explore the main research question “Is there a gender difference in doctors’ professional performance, and if so why?”
2.2 Introduction

My research aims were explored using a quantitative approach including meta-analyses and multivariate statistical analyses of large secondary datasets.

I will begin by examining the gender difference in a population of doctors whose professional performance has raised concerns about their fitness to practise. Very few doctors have disciplinary action taken against their medical registration: 0.6% of all doctors registered to practise in the UK received a sanction or a warning between 2012 to 2016 [95]. When this group of doctors is further divided into subgroups, for example by gender, the statistical power to detect true differences between the subgroups is reduced. This is the primary reason behind my decision to explore large datasets. Large datasets enable the researcher to capture as much data as possible and therefore improves the chances of detecting true effects and exposing meaningful relationships between variables.

Within this chapter I describe the context in which this research was set, together with the populations used to explore the research question and aims of this thesis. Figure 2.1 outlines the content of this chapter.
**Figure 2.1 Description of the structure of the chapter**

<table>
<thead>
<tr>
<th>Context</th>
<th>This section detail a brief history and role of the Royal College of Physicians and of the General Medical Council.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Datasets</td>
<td>A brief description of the different datasets used during the course of the thesis and their source.</td>
</tr>
<tr>
<td>Study populations</td>
<td>A description of the different study populations examined throughout the thesis.</td>
</tr>
<tr>
<td>Definition of key terms</td>
<td>Provides a brief definition and explanation of the recurring terms in this thesis, along with a justification for their use.</td>
</tr>
<tr>
<td>Main outcome measures</td>
<td>This section will describe the two main performance measures explored throughout this thesis: a disciplinary action against a doctor’s medical registration; and pass rates at postgraduate medical examinations.</td>
</tr>
<tr>
<td>Validity and reliability</td>
<td>The terms validity and reliability will be described and the importance of these two measures will be discussed.</td>
</tr>
<tr>
<td>Statistical methods</td>
<td>The descriptive and inferential statistical tests used throughout this thesis will be explained.</td>
</tr>
<tr>
<td>Rationale for using secondary datasets and ethical and data protection considerations</td>
<td>A discussion of the advantages and disadvantages of using secondary datasets, followed by a discussion of the ethical and data protection considerations.</td>
</tr>
<tr>
<td>Study designs used to address the research questions</td>
<td>A description of the study designs used throughout the thesis and a discussion of their advantages and disadvantages.</td>
</tr>
<tr>
<td>Interest in research exploring doctors’ performance</td>
<td>A brief justification for the exploring doctors’ professional performance.</td>
</tr>
</tbody>
</table>
2.3 Context

2.3.1 The medical Royal Colleges and Faculties

There are 24 medical Royal Colleges and Faculties in the UK and Ireland. Their role is to promote patient care and safety. They achieve this by setting the standards in their specialty for doctors’ education and training. One method by which the medical Royal Colleges and Faculties support and regulate doctors’ training is through the delivery of medical postgraduate examinations for their specialty. Doctors training in the UK are required to complete the relevant examinations to complete training and to become a specialist in their chosen medical or surgical specialty, or to become a GP.

In this thesis I focus on the performance of doctors who complete the Membership of the Royal College of Physicians (UK), (MRCP(UK)), Part 2 Clinical Examination (Practical Assessment of Clinical Examination Skills - PACES). Those doctors who wish to specialise in internal medicine complete this examination. The Royal College responsible for this examination is the RCP. The RCP is the oldest medical college in England and was founded in 1518 by a Royal Charter from King Henry VIII. It is a professional membership body for physicians, with 34,000 members and fellows worldwide [147]. It aims to improve the practise of medicine through influencing the way healthcare is designed and delivered; promoting good health and disease prevention; and supporting physicians to fulfil their potential. As I described in section 1.5.4, I focus in part on the RCP’s membership examinations because of the opportunity to access their examination data.
2.3.2 General Medical Council (GMC)

2.3.2.1 History of the GMC

Established following the Medical Act of 1858, the GMC (formerly the General Council of Medical Education and Registration) is the UK medical regulator. All doctors practising within the UK are required to be registered with the GMC. The role of the GMC has developed over its 160 years of existence.

In 1858, the purpose of the Medical Act was to distinguish “qualified” practitioners from “unqualified” ones, thereby bringing order and clarification to the growing number of practitioners offering their services to the public. It was to be the Council’s role to implement the Act, and as a result the medical register was formed and the Council became responsible for the registration, basic medical education and professional discipline of medical practitioners. In 1886, the Medical Act strengthened the Council’s educational powers with the introduction of passing examinations for registration. In 1950, the Council introduced a pre-registration year for all newly qualified doctors.

Today the GMC’s role, set out by the Medical Act 1983, is to protect patients and improve medical education and practice across the UK. It does this by: deciding which doctors are qualified and fit to practise in the UK; overseeing medical education and training in the UK; setting the standards doctors working in the UK are expected to adhere to (the standards are set out in the publication Good Medical Practice); and, where necessary, take action to prevent a doctor putting the safety of patients, or the public’s confidence in doctors at risk (see 2.7.1 for more details).

\[\text{In 1951, the General Council of Medical Education and Registration was abbreviated to the General Medical Council (GMC).}\]
**Figure 2.2 Key dates in the history of the GMC from 1858 to 2019**

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1858</td>
<td>Introduction of the Medical Act</td>
</tr>
<tr>
<td>1886</td>
<td>Medical Act strengthens the GMC's educational powers and the introduction of passing exams prior to registration</td>
</tr>
<tr>
<td>1926</td>
<td>Lay membership to the Council introduced</td>
</tr>
<tr>
<td>1950</td>
<td>Medical Act introduces a pre-registration year for all newly qualified doctors</td>
</tr>
<tr>
<td>1995</td>
<td>Good Medical Practice is published and sets out the generic medical standards of good practice</td>
</tr>
<tr>
<td>1999</td>
<td>Decision that all actively practising doctors should have their practice evaluated regularly to ensure they are up to date and fit to practise</td>
</tr>
<tr>
<td>2003</td>
<td>Lay membership increased to 40%</td>
</tr>
<tr>
<td>2012</td>
<td>Revalidation introduced</td>
</tr>
<tr>
<td>2013</td>
<td>Current version of Good Medical Practice published</td>
</tr>
<tr>
<td>2019</td>
<td>Present day</td>
</tr>
</tbody>
</table>
2.3.2.2 The medical register

Since its conception in 1858, the GMC has complied and held registers of all qualified doctors able to practise in the UK. Initially the medical register was published as a book; today the medical register, known as the List of Registered Medical Practitioners or LRMP, is an electronic database accessible to the public [150]. It contains details of all doctors on the GMC’s register, and hence all doctors who are eligible, or have been eligible, to practise medicine in the UK since October 2005i. The LRMP is updated daily with doctors joining the register and the registration status of doctors changing. The LRMP provides details of doctors’ GMC reference numbers, gender, year and place of primary medical qualification (PMQ), current registration status, date of registration, date of entry on the GP and Specialist registers iii, and publicly available fitness to practise history iv since October 2005, including any active sanctions against a doctor’s medical registration.

The purpose of the medical register is to give confidence to the public that the doctors practising medicine in the UK have the training, skills and experience to meet the standards expected of the profession and the standards expected by patients.

2.4 Datasets

I found the opportunity to apply big data research techniques to medical education research exciting, however, one of the greatest challenges of this thesis was discovering the existence of, sourcing, and obtaining the permission to use the appropriate secondary datasets to answer the

---

i The 20th October 2005 was the date when the GMC first began to publish full details of a doctor’s registration on the LRMP online.

ii For inclusion on the GP and/or Specialist registers, a doctor must have completed their Specialty Training and be a fully qualified consultant and/or GP.

iv Confidential information relating solely to the doctor’s physical or mental health will not be publicly available.
research question. I obtained multiple datasets from a variety of sources (see Table 2.1). The advantages and disadvantages of using large secondary datasets in research are discussed in section 2.10.

<table>
<thead>
<tr>
<th>Study chapter</th>
<th>Data details</th>
<th>Data provider</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 3</td>
<td>The List of Registered Medical Practitioners (LRMP) from May 2013 data</td>
<td>GMC</td>
</tr>
<tr>
<td>Chapter 5</td>
<td>Practical Assessment of Clinical Examination Skills (PACES) first time attempts between 2010 and 2013</td>
<td>MRCP(UK)</td>
</tr>
<tr>
<td>Chapter 7</td>
<td>Doctors under investigation by the GMC, who were required to complete tests of competence (ToC) between 2008 and 2013</td>
<td>GMC &amp; RDME</td>
</tr>
<tr>
<td></td>
<td>Reports containing ToC knowledge test (KT) and simulated surgery scores (SS) for the General Practitioners (GPs) who were investigated by the GMC between 2008 and 2013</td>
<td>RDME</td>
</tr>
<tr>
<td></td>
<td>Updated case outcome data for investigated GPs</td>
<td>GMC</td>
</tr>
<tr>
<td></td>
<td>Demographic data of the doctors who volunteered to completed the ToC</td>
<td>RDME</td>
</tr>
<tr>
<td></td>
<td>Multiple datasets for the General Practice pilots between 2008 and 2013 containing the ToC KT scores of the volunteer GPs</td>
<td>RDME</td>
</tr>
<tr>
<td></td>
<td>SS scores of the GPs who volunteered to complete a ToC</td>
<td>RDME</td>
</tr>
<tr>
<td>Chapter 8</td>
<td>Survey data from doctors who applied and were accepted to one of at least five English medical schools in 1990</td>
<td>Professor McManus, RDME</td>
</tr>
<tr>
<td></td>
<td>Registration status of the GMC registered cohort members as of August 2015</td>
<td>GMC</td>
</tr>
<tr>
<td></td>
<td>Historical and active sanctions of the GMC registered cohort members as of August 2015</td>
<td>GMC</td>
</tr>
</tbody>
</table>

GMC: General Medical Council

MRCP(UK): Membership of the Royal Colleges of Physicians (United Kingdom)

RDME: UCL Research Department for Medical Education

2.5 Study populations

The study populations I explored throughout this thesis are qualified medical practitioners who are or were registered to practise medicine in the UK. A total of five study populations were examined:

i) Doctors on the GMC medical register (LRMP) on the 29th May 2013 (Chapter 3).
ii) Doctors attempting the MRCP(UK) Part 2 clinical assessment (PACES) for the first time, at any official MRCP exam centre in the UK or internationally (Chapter 5).

iii) Doctors who were working as GPs at the time that a complaint about their fitness to practise was raised with the GMC and who - following a triage and selection process - were required to complete tests of competence (ToC) as part of a GMC performance assessment (Chapter 7).

iv) Doctors who were working as GPs, with no known fitness to practise concerns, who volunteered to complete either the written assessment and/or clinical assessment of the GMC’s ToC (Chapter 7).

v) Doctors who had applied to study at one of five medical schools in 1990 and had agreed to be part of a research cohort study (Chapter 8).

There is some overlap between the study populations of this thesis, with some doctors appearing in more than one of the study populations. However, for the purposes of this thesis, the study populations are treated independently.

2.6 Definition of key terms

2.6.1 Ethnic group

The categorisation of people into ethnic groups is complex. Assigning oneself to an ethnic group may be based on physical characteristics, cultural differences, and country of birth. The UK census in 2001 asked “What is your ethnic group?” and provided the following categories: White; Mixed; Asian or Asian British; Black or Black British; Chinese or other ethnic group [151]. The debate on how to categorise a population into ethnic groups is beyond the scope of this thesis (see [152, 153] for further details). However, levels of attainment have been found to differ between different ethnic groups [125]. It is therefore, important to consider the effect of ethnic groups in research examining performance and attainment. That said, the level at
which ethnic groups are described can vary from detailed and specific to general groupings. As mentioned earlier, dividing the study population into multiple subgroups can result in a lack of statistical power because of the small numbers of study participants in each of the subgroups. This is one of the reasons that previous researchers have dichotomised ethnicity into: “White” (those who report themselves as being “White British”, “White Irish” and “White other”); and “ethnic minorities” to categorise all other ethnic groups, with the exception of those participants who declined or omitted to provide an ethnic group [154]. The use of the term “ethnic minorities” is also in line with the Office for National Statistics (ONS) style guide [155]. This thesis uses the terms “white” and “ethnic minorities” when describing the ethnicity of the study population.

2.6.2 World region of primary medical qualification

To legally practise medicine in the UK all doctors must be registered with the GMC and have a licence to practise, regardless of the sector of employment, the type of employment and the level of registration.

Doctors who have obtained their primary medical qualification outside of the UK may be asked to provide evidence of their English language capability and evidence of their fitness to practise. If the doctor has obtained their primary medical qualification outside of the European Economic Area (EEA) or Switzerland they may also be required to demonstrate the knowledge and skills necessary to practise medicine (see [156] for further details).

It has been shown that doctors who have obtained their primary medical degree outside of the country where they are practising do differ in terms of professional performance [127, 157, 128]. It is therefore important to consider the influence of the country where a doctor obtained their primary medical qualification when examining the professional performance in doctors. Though collapsing the different countries in which a doctor obtained their primary medical qualification into a variable with two levels (UK graduate and international medical graduate (IMG)) or three levels (UK
graduate, EEA graduate and IMG) reduces the detail in which this variable can be explored, it should maintain a sufficient level of statistical power to enable meaningful analyses.

2.7 Main outcome measures

The main outcomes are measures of professional performance. In Chapter 3, Chapter 7 and Chapter 8 I explored the outcome of a warning or sanction imposed on a doctor’s medical registration. In Chapter 5 and Chapter 7 I explored the performance at two postgraduate medical examinations: both were high-stakes examinations, but each had different objectives.

2.7.1 GMC and assessing doctors’ fitness to practise

The role of the GMC is to ensure proper standards in the practice of medicine in the UK, thereby protecting, promoting and maintaining the health and safety of the patient population and the community as a whole [158]. Under the Medical Act 1983, the GMC holds the powers to take action against a doctor’s registration if the doctor’s fitness to practise is impaired including deficient professional performance [149, 57]. A doctor’s fitness to practise can be found to be impaired for the reason of: misconduct; deficient professional performance; conviction or caution for a criminal offence; adverse mental or physical health; not having the necessary knowledge of English; and a determination/decision by a regulatory body responsible for the regulation of a health or social care profession, either in the UK or overseas [149, 57]. The GMC triage information received identifying complaints that require investigation [78]. The pathway of an investigation depends on the nature of the concern. The final outcome of an investigation process by the GMC can be the following:

- **Case closed**: No evidence of impaired fitness to practise.
- **Warning**: Issued when a doctor’s performance has not been in keeping with the principles set by the GMC for doctors, but a restriction on the doctor’s registration is not necessary.
• **Undertakings:** An agreement between the GMC and the doctor about the doctor’s future practice. The doctor must adhere to these undertakings to maintain their registration.

• **Conditions:** Set out by the GMC and restrict a doctor’s practice. The doctor must comply with these conditions to maintain their registration.

• **Suspension or erasure:** The doctor’s license is withdrawn by the GMC and they are no longer able to practise.

Undertakings, conditions, suspension or erasure are defined as sanctions throughout this thesis. Both active and historical sanctions are recorded on the LRMP. Warnings will also be recorded on the LMRP, but only for a limited period of time, while the warning is active.

2.7.1.1 Performance assessments and tests of competence

If the GMC receives a concern regarding a doctor’s fitness to practise, they may require the doctor complete a performance assessment to assess their fitness to practise. One component of the performance assessment are the tests of competence (ToC). The ToC are individually tailored to the doctor’s specialty, experience and work environment. They usually consist of at least a written assessment and a clinical component. Further details about the ToC are described in Section 7.3.1.1.1.

2.7.2 MRCP(UK) PACES

The MRCP(UK) Part 2 Clinical PACES examination was first introduced in 2001 and is the clinical assessment component of the MRCP(UK) Diploma. It is a structured standardised assessment and was developed with the aim of providing a valid and reliable assessment of physical examination and communication skills [159]. The format of PACES was revised considerably in 2009 and a newer format of PACES was introduced in 2010 [160]. MRCP(UK) is currently reviewing the PACES examination and changes will be implemented in 2020 (see [161] for further details about PACES 2020). Greater details about PACES are given in Section 5.3.2.
2.8 Validity and reliability of the performance measures used

The performance measures used throughout this thesis are high-stakes measures, meaning that their outcome could significantly impact a doctor’s medical career. It is for this reason that these measures must function as intended and must be both reliable and valid.

2.8.1 Validity and reliability of the GMC Fitness to Practise procedures

In Chapter 3 I explore the association between gender and the outcome of GMC Fitness to Practise (FtP) procedures. Undergoing a GMC FtP procedure is stressful for the doctor involved and can result in a doctor’s medical practice being restricted or even terminated. There is also a substantial financial impact of FtP procedures [162]. It is vital that these procedures function as intended and that they are reliable and valid to maintain the confidence of both the public and of the medical profession. Following evidence that doctors with certain demographics – non UK-trained, ethnic minorities, men, and older doctors – were overrepresented in FtP procedures and were at increased risk of receiving a higher impact outcome [163, 164], the GMC commissioned a report into the reliability and validity of the FtP procedures [165]. The report concluded that there was no evidence of bias or discriminatory practices identified in the FtP procedures, but did provide suggestions on how the GMC could further demonstrate that the decision-making during FtP procedures were fair and valid [166].

2.8.2 Validity and reliability of postgraduate medical examinations

In Chapter 5 and Chapter 7 I explore the performance of the genders at two types of postgraduate medical examinations. In Chapter 5, the examination is the clinical component of a set of examinations designed to test the clinical knowledge and skills of trainee doctors who wish to enter into higher specialist training, following on from core medical training (CMT). In Chapter 7, the assessment has the aim of testing the doctor’s knowledge and clinical skills as part of an investigation into the doctor’s fitness to practise.
Both of these examinations assess the doctor’s knowledge and skills, however, the purpose of these examinations are different but equally important. The MRCP(UK) Part 2 Clinical (PACES) is designed to test the clinical knowledge and skills of doctors who hope to enter into higher specialist training. It aims to ensure that those who successfully complete the examination have demonstrated that they are competent across a range of skills and have met the high level of standard expected of a doctor in a higher specialty training post. The GMC ToC are written and clinical assessments that assess a doctor’s knowledge, clinical and consultation skills. They assess whether the doctor is performing to the minimum standard expected of a F2 doctor (two years post-graduation). There is no ‘pass mark’ for the ToC and the performance of the doctor at the ToC is taken into consideration along with other assessment instruments used to assess the doctor’s fitness to practise. Despite their different purposes, it is vital that these examinations are both reliable and valid to ensure that they are correctly identifying the doctors who perform well and those who do not.

An examination is said to be valid if it measures what it aims to measure. In the context of postgraduate medical examinations, the content of a valid test should be highly related to the medical specialty that the doctor is working in. The reliability of an examination is the extent to which the examination is dependable and repeatable for individual candidates. Otherwise stated, if a candidate completes an examination, and then sat the same examination a few days later, a reliable examination would yield the same result, assuming there had been no change in conditions or knowledge acquisition. In summary, a valid and reliable examination is an examination that tests what it has set out to test and its’ results are reproducible and consistent.

2.8.2.1 The validity and reliability of MRCP(UK) PACES

The integrated brief clinical encounters included within PACES reflect the cases a doctor may encounter on a hospital ward or clinic, and contribute to the validity of the examination [167]. PACES has also be demonstrated to be
a particularly strong predictor of action against doctors’ medical registration, which is further evidence of its validity [67].

PACES is a type of objective structured clinical examination (OSCE), and like most OSCEs, it has stations or encounters, which are organised in a carousel around which candidates rotate. Unique among postgraduate medical examinations, PACES has two examiners per station rather than one: the reliability of PACES is therefore considerably enhanced through this pairing of examiners and the high number of independent judgements made by each of the examiners [167].

2.8.2.2 The validity and reliability of the GMC tests of competence (ToC)

The ToC are written and clinical assessments that form part of the GMC’s performance assessment, which is used to assess doctors when their fitness to practise has raised serious concerns [168] (for further details on ToC see 7.3.1.1.1). The written assessment of the GMC’s ToC were developed in 2005 with the aim of delivering a test that was fair, valid and reliable for single candidates or small groups of candidates [169]. Full details of how these tests were developed and the methodology used to ensure the tests are standardised, valid and reliable are described by Dacre et al. [169], but in summary the team developing the ToC employ trained item writers, experienced test developers and content experts to ensure each test is valid, matched to a blueprint and is standard set. Each item included in a ToC is pilot-tested by a volunteer group who complete the examination under comparable conditions to doctors under investigation.

2.9 Statistical methods

In Chapter 3, Chapter 5, Chapter 7 and Chapter 8 I present and describe the original quantitative research studies of this thesis. In Chapter 4 and Chapter 6 I describe the systematic reviews of the literature and meta-analyses completed as part of this thesis. The data were analysed using STATA v.12/SE. Within each chapter I individually describe the statistical
methodology used. I will briefly describe the statistical tests that recur throughout this thesis in this section.

2.9.1 Descriptive statistics

The aim of descriptive statistics is to summarise and describe the dataset. Measures of central tendency and spread were reported where relevant.

2.9.2 Inferential statistics

Inferential statistical analyses make judgements or inferences about a population. A causal modelling analysis approach was used to facilitate inferences about possible causal relationships from statistical data.

2.9.2.1 Bivariate

2.9.2.1.1 Categorical data

Chi-squared was used to test the relationship between categorical variables. Mantel-Haenszel was used to assess whether odds ratios were homogenous or heterogeneous between strata of categorical variables, and to provide an estimate of the odds ratio of an exposure while controlling for a possible confounder variable.

2.9.2.1.2 Continuous data

The two-sample t-test was used with continuous data to determine if two population means were equal.

2.9.2.2 Multivariate

2.9.2.2.1 Categorical data

Logistic regression was used to predict the relationship between independent (exposure) and dependent (outcome) variables. Binary logistic regression was used when the outcome variable was binary, and multinomial logistic regression was used when the outcome variable had greater than two levels, but there was no hierarchical order to the variables.
2.9.2.2 Continuous data

Two-way ANOVA was used to compare the mean differences between two populations that have been split on two independent variables.

Path analysis was used to represent possible causal relationships and to allow several relationships between variables to be tested simultaneously. In Chapter 7 I describe this method in greater detail.

2.9.2.3 Standardisation

In Chapter 5 and Chapter 7 I standardised the examination scores into z-scores. Standardisation is a measure of how many standard deviations below or above the population mean a raw score is. It allows for the comparison of scores between different populations.

2.9.2.4 Effect sizes

In Chapter 5 and Chapter 7 I calculated effect sizes (Cohen’s $d$) to provide a measure of the sizes of differences. Cohen advised that an effect size of $d = 0.2$ is ‘small’, $d = 0.5$ is ‘medium’, and $d = 0.8$ is ‘large’ [170].

2.10 Rationale for using secondary datasets and ethical and data protection considerations

I used secondary data sources to conduct the research studies in Chapter 3, Chapter 5, Chapter 7 and Chapter 8. The datasets I used in Chapter 3, Chapter 5 and Chapter 7 been collected for administrative purposes and not for the purpose of research. In Chapter 8 I conducted the research using survey data that had been collected for research purposes, however the surveys were designed and collected by Professor McManus at UCL and were not collected for my research study.

In summary, all of the studies included in this thesis have been conducted using secondary data sources.
2.10.1 Advantages and disadvantages of using secondary data

Secondary data analysis can be defined as the analysis of data that has previously been collected and that has not been collected specifically for the research question being addressed [171]. The use of secondary data sources for research has advantages and disadvantages, which will be discussed below. There are also important ethical issues that need to be considered when using secondary data for research purposes.

A huge amount of data is collected as part of routine management or administrative information systems [172] and it has become increasingly popular to use these large datasets in research. The value of providing large amounts of administrative data for researchers to use has become recognised in the medical education community, resulting in the development of the UK Medical Education Database (UKMED) in 2015 - 2016 [173, 174].

One of the greatest advantages of using secondary datasets for research is that the data already exist and as such there is a saving in terms of time, money and resources [171, 172], however secondary data can be costly in terms of acquisition. To use secondary data, knowledge of the existence of the data and the means to access it are required, which can take considerable time and effort [171]. Researchers analysing secondary data may also need to expend time and effort into processing the data into suitable form for their research purposes [175].

A further advantage of using secondary data is that it can be of higher quality and more representative than if a primary survey had been completed [172]. This can be particularly relevant when attempting to research a sensitive topic or a population that may be difficult to reach for primary research [172]. However, the researcher using secondary data will have no control or involvement in data gathering. As a result the researcher will be required to spend time in understanding the nature of the data and how it was assembled [175]. It is also unlikely that the data will address all the topics of interest to the researcher. It is important that the researcher is aware of how
their hypotheses may be limited by the available data and how the data may in turn influence their hypotheses and research questions [175]. Nonetheless, secondary data can encourage and stimulate the development of new research ideas and trajectories [175, 171]. Researchers using secondary data must also be mindful of what is included and omitted from the data they are using and how this can influence research theory [175].

It is my view that secondary data analysis is not a substitute for primary data analysis. I believe that primary and secondary data have both their advantages and disadvantages, and that they can provide useful research material in ways that the other may not be able to. As such, I feel they can complement each other.

I have chosen to only use secondary data in this thesis because, when I chose to examine the performance of doctors, particularly examining poor performance measures, I chose to examine outcomes that were sensitive: passing or failing an examination, and disciplinary action against medical registration. The sensitive nature of the research topic, coupled with the fact that some members of the study population would have gone through a very difficult time in their careers, led me to suppose that any primary data gathering attempts may be hampered by a poor response rate and biases, such as recall or response. Secondary data enabled me to access this sensitive and possibly stigmatising data.

A further reason for my choice to conduct my research using secondary data was the large number of doctors included in the datasets. These large datasets increased the power of any statistical analyses and provided the opportunity to detect true effects and expose meaningful associations between variables; this was especially relevant for the studies that explored rare outcomes and the studies that had relatively small numbers with the exposure of interest. They also enabled me to build upon the skills I had acquired during my master's degree in epidemiology and to use multivariate statistical methods, which are usually employed on large datasets.
Finally, the use of secondary datasets allowed me to explore the research question using multiple datasets. Had I opted to collect the data used in this research project first hand, it is unlikely I would have been able to explore this research question in a variety of study populations. Exploring professional performance in several study populations has enabled me to demonstrate that my research findings are generalisable and robust.

2.10.2 Ethical and data protection considerations when using secondary data

When data is collected for administrative purposes, those who provide the data may not be aware that their data may be used for research and they are certainly unlikely to be aware of the type of research their data will be used for. It is for these reasons that, as researchers, we have an obligation to consider ethical concerns and to adhere to data protection regulations prior to conducting research.

The most significant change to data protection law to occur in 20 years happened during the course of this thesis. General Data Protection Regulation (GDPR) was implemented in May 2018 and replaced the Data Protection Act 1998. GDPR was implemented because of the way new technologies have fundamentally changed the way we use personal data and its objective is to increase transparency and honesty in how data is used and stored. GDPR has changed the way in which personal data can be collected, used, retained and deleted. Those working with data are required to consider the data required and if data is identifiable, the data controller and processor must be able to justify the need for identifiable information and demonstrate how individuals’ privacy and confidentiality will be maintained. Thought must be given as to the safe and secure storage of data and of the length of time the data will need to be stored. It is in the public’s interest to ensure the research conducted using individuals’ data is compliant with ethical and GDPR standards.

For the research conducted in this thesis, I liaised closely with the UCL Research Ethics Committee and UCL Data Protection team. The research
received ethical approval from the University College London Research Ethics Committee (Project ID: 5025/001) and was registered with the UCL Data Protection Team (UCL Data Protection Registration Number Z6364106/2013/08/11) at the beginning of my doctoral journey. Following the introduction of GDPR in 2018, I liaised further with both the UCL Research Ethics Committee and UCL Data Protection team to update both on how the research has evolved and to ensure compliance. The project was then registered under UCL Data Protection Registration Number Z6364106/2019/03/119.

As stated earlier, the research explored secondary datasets. Permission to access and use the data for the research studies was sought from the data controllers for each of the datasets. Information sharing agreements, confidentiality agreements and research agreements were completed and signed prior to receiving or accessing data. The ownership of each dataset has been acknowledged in the relevant study chapters of this thesis.

Once received, soft copies of the datasets were stored on a secure drive with restricted access on the UCL network. When no remote connection to the secure drive was available, copies of the data would be stored on an encrypted Universal Serial Bus (USB) stick and would subsequently be deleted when not in use. No hard copies of the data were stored.

All the data received and accessed had identifiable information. This identifiable information was not required for three out of the four research studies conducted. As such, I permanently deleted the identifiable data from the datasets used for two out of these three research studies and notified the owners of the data. I anonymised the dataset I used for the third study and I used this anonymised version of the dataset for my research. For the remaining study I required identifiers to enable me to merge multiple datasets.

All research outputs from this thesis are unidentifiable with regards to the individual members of the study populations and all data have been
aggregated in an anonymous form when disseminating results. The owners of the data have been given the opportunity to review all research outputs prior to submission for publication.

2.11 Study designs used to address the research question

To address the research question of this thesis, the quantitative research studies were either cross-sectional (Chapter 3 and Chapter 5) or cohort (Chapter 7 and Chapter 8) in design. Cross-sectional and cohort studies are both observational in design: that is, they simply record data and do not intervene on the study population. Using observational study designs to address the research question of this thesis is appropriate because this thesis aims to provide preliminary data regarding the association between gender and professional performance in doctors. It is not possible to control the exposure (gender).

A cross-sectional study aims to obtain a representative sample of a well-defined population and to obtain all measurements at a single point in time [176]. It is often referred to as capturing a ‘snapshot’ in time of the data. The advantages of cross-sectional studies are that they are relatively low cost and simple to perform and there is no risk of loss to follow-up (because there is no follow-up). The results from cross-sectional studies may inform the hypotheses of more complex studies, such as cohorts [176]. However, the drawbacks of cross-sectional studies are primarily the risk of non-response bias if the individuals who consented to be part of the study differ from those individuals who decline to be part of the study [176]. However, there was no risk of non-response bias in the cross-sectional studies conducted as part of this thesis because the entire of the population of interest were included in the study and consent for each individual study participant was not sought (nor was it required). A further limitation of cross-sectional studies is the difficulty to infer temporal association between the variables of interest [176]. Once again, this was not a concern for the cross-sectional studies conducted towards this thesis because the majority of the variables explored were demographic variables and all were non-modifiable. Finally, one of the
greatest drawbacks of a cross-sectional study is only being able to infer an association and not causation [176].

A cohort study can be either retrospective (Chapter 7) or prospective (Chapter 8). A retrospective cohort study is usually constructed from databases of records that have previously been collected and the temporal sequence of independent and dependent variables are explored [177]. A prospective cohort follows a study population over time and records when the dependent variable occurs [178]. The advantages of cohort study designs are that they enable the temporal association between variables to be explored and they allow for the examination of multiple effects of one exposure. However cohort designs are at risk of bias, including selection bias (which in turn can impact the external validity of a study), non-response bias, attrition bias, response bias and confounding [178]. The risk of bias for the cohort studies conducted during this thesis will be discussed in the relevant study chapters. A cohort study design is usually the study design of choice if the exposure variable is rare, it is not the ideal study design for rare outcomes, once again this will be discussed further in the relevant study chapters.

Cross-sectional and cohort studies have been previously used by researchers in the medical education community to explore performance in medical students and doctors [44, 5, 154].

2.12 Interest in research exploring doctors’ performance

Gender is a standard demographic variable usually included in research designed to explore performance in doctors. However, if gender is not the primary focus of a research question, it may not be explored in greater depth or its effect may not have been adjusted for the possible confounding effects of other variables. When work began on this thesis in 2012, there were very few studies examining poor performance in doctors practising in the UK [5, 74, 163]. Over the course of this thesis, the interest in poorly performing doctors has increased, partly due to recent health scandals and public
demand [142, 179, 144], but the number of studies examining these doctors still remains low [128, 131, 114, 9, 10]. The GMC are committed to better understanding poorly performing doctors and dedicate a chapter of their annual medical education publication to these doctors [180-184, 95] and commission research to better understand the doctors who they investigate [185]. This interest and desire for further knowledge about poorly performing doctors demonstrate that this thesis is providing timely research evidence that is of relevance to the medical profession, medical education researchers and the public. The published studies exploring poorly performing doctors have provided an essential evidence base on which I have been able to develop this thesis and complete further interpretative research.

2.13 Conclusion

Throughout this thesis I employ methods used historically in epidemiological studies, which routinely analyse large datasets. These epidemiological techniques have been applied to research in the medical education field by other researchers. The major limitation of the methodologies employed throughout this thesis is the sole use of secondary data sets. However, the use of large secondary datasets does have its advantages and the value of using large administrative datasets in medical education research has been recognised, with the development of UKMED during the course of this thesis. Using large secondary datasets for medical education research when it was still regarded as fairly novel has been exciting. With this thesis I provide timely evidence to develop a greater understanding of the gender difference in the professional performance of doctors.
Chapter 3 - A cross-sectional study examining the association between a doctor’s gender and receiving sanctions against their medical registration in the UK

3.1 Chapter summary

3.1.1 Background

Studies from the UK and abroad have demonstrated that doctors’ gender is associated with disciplinary action against their medical registration, with men being more likely to face disciplinary action. However the applicability and interpretation of these findings are limited due to the differences in the medical and legal systems of foreign countries and the lack of adjustment of the findings for potential confounders.

3.1.2 Aims

To examine the association between doctors’ gender and receiving a warning or sanction against their medical registration, while controlling for other known confounding variables.

3.1.3 Design and setting

This cross-sectional study analyses the GMC’s LRMP database of doctors practising in the UK. The database included all doctors who are or have been registered to practise medicine in the UK since October 2005 up to May 2013.

3.1.4 Methods

Binary logistic regression modelling, controlling for confounders, described the association between a doctor’s gender and warnings or sanctions being imposed against a doctor’s medical registration. The confounding variables included years since primary medical qualification, world region of primary medical qualification and specialty.
3.1.5 Results

Of the 329,542 doctors on the LRMP, 2,697 (0.8%) had a warning or sanction against their medical registration, 516 (19.1%) of whom were women. In the fully adjusted model, women had nearly a third of the odds (OR: 0.37, 95% CI: 0.33 to 0.41) of having a warning or sanction compared to men. There was evidence that the association varies with specialty, with women who had specialised as GPs being the least likely to have warnings or sanctions imposed (OR: 0.26, 95% CI: 0.22 to 0.31).

3.1.6 Conclusions

Women have reduced odds of having a warning or sanction imposed against their medical registration when compared with their male colleagues. This association remained after adjustment for the confounding factors. These results were representative of all doctors registered to practise in the UK.
3.2 Introduction

In the period between 2012 - 2016 the GMC received a complaint about 1 in 10 doctors' fitness to practise [95]. Following a triage process of the complaints received, the GMC may investigate the complaint further. GMC case examiners consider the evidence collected during a GMC investigation and will decide whether the doctor's fitness to practise is impaired. If the case examiners conclude there is no evidence of impairment they may close the case or issue a warning to the doctor. If the case examiners find evidence to suggest the doctor's fitness to practise is impaired they may agree undertakings with the doctor or refer the case to the Medical Practitioners Tribunal Service (MPTS). The MPTS will make an independent decision about whether the doctor is fit to practise medicine. A tribunal can restrict or remove a doctor's right to practise medicine in the UK, known as a sanction (see Figure 3.1).
Between 2012 - 2016 less than 1 in 100 doctors had a warning or sanction imposed following a complaint to the GMC [95]. However, the level of stress and fear associated with the fitness to practise investigation process can have significant detrimental effect on the investigated doctor’s mental health, as highlighted by a report in the BMJ [77].

I aimed to explore the characteristics of doctors who have a warning or sanction imposed on their medical registration in the UK, with a focus on doctors’ gender. The justifications for focusing on doctors’ gender when examining disciplinary actions against doctors is two-fold:

---

* This diagram was developed with support from the GMC and using GMC diagrams.
Firstly, there has been a dramatic change in the demographics of the medical profession in the UK, with a significant increase in the number of women entering this previously male-dominated profession (see 1.5.3). This change in the demographic profile of UK doctors has brought with it a heightened interest in how the increase in female doctors may impact the profession.

Secondly, although data reporting disciplinary action by gender from other professions is limited, it has been reported that though the distribution of male and female solicitors in the general population is roughly equal (in 2012 the solicitor population in the UK was composed of 53% male and 47% female), there was a much lower proportion of female solicitors being referred to the Solicitors Disciplinary Tribunal, when compared to male solicitors (of the 180 solicitors referred to the Solicitors Disciplinary Tribunal in 2012, 27% were female) [186]. The gender difference in disciplinary action is also present in the general population, and starts from childhood with female school-aged students in England being less likely to be permanently excluded from school compared to male school-aged students [187]. If we look at the extreme end of the scale of disciplinary action, females are less likely to be defendants of court proceedings (25% of defendants were female in 2013), and the minority of the prison population is female (5% of all prisoners were female in 2014) [188]. Though easily accessible and publically available data available on disciplinary action is limited, it appears that there is a pattern forming that shows women are less likely to find themselves being disciplined. It would be interesting to explore whether this pattern is replicated in the medical profession, because not only are doctors a highly selected group of individuals (they have undergone several selection procedures to qualify as doctors) and therefore are not representative of the general population, but also because of the shift in the gender-distribution of the profession.

Examining and understanding the predictors of doctors having warnings and/or sanctions imposed may aid the medical profession in identifying doctors whose performance might raise future concerns and therefore enable
better support structures to help prevent GMC referral. Research from Canada, the USA and Australia and New Zealand has suggested that male and female doctors differ in terms of risk of disciplinary action, with female doctors being at reduced risk [189-195]. However, the applicability of the findings from these studies to the UK may be limited due to differences in the medical and legal systems in these countries. In 2011, Wakeford explored the situation in the UK [5]. He examined the factors associated with the severest outcomes of the GMC disciplinary procedures, suspension or erasure from the medical register. In agreement with the worldwide studies, he demonstrated that women doctors were four times less likely to be suspended or erased when compared to men. However, the interpretation of this finding is limited because the measure of effect for doctors’ gender was not adjusted for potential confounders.

With this cross-sectional study I aim to examine the association between doctors’ gender and receiving sanctions against their medical registration, while adjusting for known confounders, allowing for a meaningful comparison of men and women doctors and their experience of disciplinary action in the UK.

3.3 Methods

3.3.1 Study design, setting and source of data

This is a cross-sectional study using UK-wide data from the LMRP held by the GMC.

Under the Medical Act 1983, the GMC is required to keep up-to-date registers of qualified doctors (see section 2.3.2.2). The full register is the LRMP. The LRMP is a list of all doctors registered to practise medicine in the UK, and as such it changes daily. It lists those doctors who are practising (or could practise) medicine, as well as those doctors who have been suspended or whose names have been erased from the medical register. To support this research, the GMC provided a snapshot of doctors registered on the LRMP on 29 May 2013. The list included all doctors who
have been registered\textsuperscript{vi} with the GMC at any point in the period 20 October 2005\textsuperscript{vii} to 28 May 2013. The different categories of registration status included provisionally or fully registered; suspended; not registered – administrative reason, or deceased, or having relinquished registration; and not registered – erased after Fitness to Practise panel hearing. The database provided details of the doctor’s gender; the year, country and institutions of the doctor’s primary medical qualification and the doctor’s current registration status, including whether they currently had any sanctions imposed on their medical registration (see below for details). It classified doctors as GPs (on the GP register) and as hospital specialists (on the Specialist register). For inclusion on the GP or Specialist registers, a doctor must be a fully qualified consultant or GP (the doctor must have successfully completed their specialty training). Doctors who are neither on the GP nor Specialist registers can be primarily divided into two groups: the first being doctors who are currently working within a speciality training programme with the aim of becoming a GP or a consultant in a specialty, and the second group being composed of doctors in non-training posts. Non-training posts are for doctors who are not fully qualified consultants or GPs. Non-training posts are focused to meet the NHS service requirements and the doctors who choose to undertake a non-training post do so for a variety of reasons including difficulty in obtaining a place in a specialty training programme due to high competition, or doctors who prefer the work-life balance that the non-training post can provide [196].

Permission to use the data for research purposes was obtained from the GMC. The GMC provided a copy of the LRMP as of 29 May 2013 for the

\textsuperscript{vi} Not all these doctors had a licence to practise, but remained registered either because they were working abroad or taking a break and wanted to demonstrate that they are in good standing with the GMC.

\textsuperscript{vii} The 20 October 2005 was the date when the GMC first began to publish full details of a doctor’s registration status on the LRMP online.
purposes of this research project, once an information sharing agreement had been signed between UCL and the GMC.

3.3.2 Population and primary outcome

The study population included all doctors who were listed on the LRMP on 29 May 2013.

The outcomes of interest were warnings or sanctions imposed on a doctor’s medical registration that were recorded on the LRMP on 29 May 2013.

The types of warning and sanctions included:

**Warning:** Issued when a doctor’s performance has not been in keeping with the principles set by the GMC for doctors, but a restriction on the doctor’s registration is not necessary.

**Undertakings:** An agreement between the GMC and the doctor about the doctor’s future practice. The doctor must adhere to these undertakings to maintain their registration.

**Conditions:** Imposed by an independent panel and restrict a doctor’s practice. The doctor must comply with these conditions to maintain their registration. Conditions can initially be imposed for a maximum of 3 years and then be renewed in periods of up to 36 months.

**Suspension or erasure:** The doctor is prevented from practising medicine in the UK. Suspension from the register can last 12 months, but may be indefinite in certain circumstances.

A sanction can be imposed if a doctor’s fitness to practise has been proved to be impaired. The impairment can result from misconduct, poor professional performance, physical or mental ill health, or a conviction or fitness to practise determination by another regulatory body either in the UK or overseas [197]. However, it is recognised that the reason for impairment
can cross more than one category (e.g. a doctor with a drug misuse problem could be classified as having mental ill health, yet the effects of the drug abuse could impact their professional performance). The duration of a sanction on a doctor’s registration varies and it is possible for doctors to have more than one sanction against their registration, and this typically represents the outcomes of different complaints. It was not possible to establish the date a sanction was imposed or the reason why a sanction has been imposed from the data provided by the GMC.

The outcome of interest was collapsed into a binary variable: doctors with a warning or sanction against their registration and doctors with no warning or sanction against their registration.

The exposure of interest was doctor's gender, as declared by the doctor to the GMC.

3.3.3 Selection of variables

I selected the variables included in the study before any statistical analysis was performed. Data were available on the year of primary medical qualification, country of primary medical qualification and primary specialty, if on the Specialist or GP registers. I selected these variables as a priori confounders based on findings from earlier research [163, 193, 194, 5], which demonstrated that these factors might influence the risk of having sanctions imposed.

Once I had selected the variables, and prior to any statistical analysis, I performed a variable reduction process, to reduce the number of categories into meaningful categories. I converted the variable ‘year of primary medical
qualification’ into ‘number of years since qualification’ by subtracting the year of primary medical qualification from 2013viii.

I then collapsed this variable into six categories. The first category ‘0-2 years’ represented Foundation Training, the second category ‘3-10 years’ represented the bulk of time a doctors would be likely to spend in Specialty Training. I divided the remaining categories into 10-year blocks. I collapsed the variable ‘country of PMQ’ into three categories: ‘UK’, ‘EEA’ and ‘International’. I obtained the list of countries included in the EEA category from the European Union website [198] and included all countries that were members of the EEA before May 2013. I divided the variable ‘specialty’ into 14 categories. Doctors who were not present on either the Specialist register or the GP register were categorised as having ‘no specialty’ and represented trainee and other non-specialist doctors. Doctors on both the Specialist register and the GP register were recorded as having ‘dual specialty’ and doctors only on the GP register were categorised as ‘GP’. For doctors only on the Specialist register, their primary specialty was taken and recorded into 1 of 11 categories. The eleven categories represented the ten largest specialties: anaesthetics; emergency medicine; medicine; obstetrics and gynaecology; ophthalmology; paediatrics; pathology; psychiatry; radiology; and surgery. The eleventh category grouped all the remaining smaller specialties into ‘other specialties’. To categorise those doctors on the Specialist register, two researchers (myself and CW) independently allocated each primary specialty to a specialty category. The $\kappa$ statistic demonstrated a good level of agreement ($\kappa=0.72$). Any disagreements about the specialty category allocation were resolved through discussion.

viii As mentioned above, a doctor could appear in the LRMP dataset if they had been removed from the medical register. The actual date of those doctors being removed could lie anytime between 20 October 2005 and 28 May 2013. However, as no actual removal dates were given for those doctors, I used 2013 for all doctors.
3.3.4 Secondary outcome and exposure

Following my initial analyses of the data examining the outcome as a binary variable, I chose to explore the association between gender and each of the five different outcome types, because the severity of the outcomes that can be imposed on a doctor’s medical registration varies, with a warning leaving no restrictions on a doctor’s medical practise, through to erasure where a doctor is no longer able to practise medicine in the UK. The number of doctors with each outcome type was likely to be small and therefore I chose to collapse the co-variates into binary variables to maximise the power of the analyses. Number of years since primary medical qualification (PMQ) was collapsed into ‘0-10 years’ and ‘>10 years’, to represent the bulk of time a doctor who spend in training or as a fully qualified GP or specialist; world region where primary medical qualification was received was collapsed into ‘PMQ obtained in UK’ and ‘PMQ obtained outside of the UK’; and finally specialty was collapsed into ‘On the GP and/or Specialist registers’ and ‘On neither the GP nor Specialist registers’.

3.3.5 Statistical methods

I first performed bivariate analyses to look for crude associations in the data, followed by Mantel-Haenszel analyses, before going on to complete multivariate analyses using binary logistic regression modelling. In the initial logistic regression model I included only the exposure (gender) and the outcome (warning or sanction) variables to provide a crude measure of effect. I built the final logistic regression model to include all potential confounder variables, while checking for multicollinearity. The final logistic regression model enabled the calculation of an adjusted measure of effect. I assessed the final model for the presence of effect modifiers following the findings from the Mantel-Haenszel analyses.

I used the STROBE Statement [199] to guide the reporting of this study.
3.4 Results

There were 329,542 doctors registered with the GMC at any point in the period 20 October 2005 to 28 May 2013, of whom 40.3% were women. Table 3.1 shows the distribution of variables by the gender of doctors. The median number of years since qualification was 19 years. The distribution of the number of years since a doctor had qualified was skewed to the right with the largest proportion of doctors qualifying 11-20 years ago (28.1%). The majority of the doctors had received their primary medical qualification from a UK medical school (59.3%).
Table 3.1 Distribution of variables by gender of doctors

<table>
<thead>
<tr>
<th>Variable</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=196 814</td>
<td>N=132 728</td>
<td>N=329 542</td>
</tr>
<tr>
<td>Sanction imposed on registration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>194 633 (98.9%)</td>
<td>132 212 (99.6%)</td>
<td>326 845 (99.2%)</td>
</tr>
<tr>
<td>Yes</td>
<td>2181 (1.1%)</td>
<td>516 (0.4%)</td>
<td>2697 (0.8%)</td>
</tr>
<tr>
<td>Number of years since receipt of PMQ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-2</td>
<td>6332 (3.2%)</td>
<td>8830 (6.7%)</td>
<td>15 162 (4.6%)</td>
</tr>
<tr>
<td>3-10</td>
<td>28 548 (14.5%)</td>
<td>37 220 (28.0%)</td>
<td>65 768 (20.0%)</td>
</tr>
<tr>
<td>11-20</td>
<td>52 437 (26.6%)</td>
<td>40 023 (30.2%)</td>
<td>92 460 (28.1%)</td>
</tr>
<tr>
<td>21-30</td>
<td>39 146 (19.9%)</td>
<td>23 069 (17.4%)</td>
<td>62 215 (18.9%)</td>
</tr>
<tr>
<td>31-40</td>
<td>30 206 (15.4%)</td>
<td>12 136 (9.1%)</td>
<td>42 342 (12.9%)</td>
</tr>
<tr>
<td>≥ 41</td>
<td>40 145 (20.4%)</td>
<td>11 450 (8.6%)</td>
<td>51 595 (15.7%)</td>
</tr>
<tr>
<td>Region where PMQ received</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UK</td>
<td>108 323 (55.0%)</td>
<td>86 989 (65.5%)</td>
<td>195 312 (59.3%)</td>
</tr>
<tr>
<td>EEA</td>
<td>25 333 (12.9%)</td>
<td>15 880 (12.0%)</td>
<td>41 213 (12.5%)</td>
</tr>
<tr>
<td>International</td>
<td>63 158 (32.1%)</td>
<td>29 859 (22.5%)</td>
<td>93 017 (28.2%)</td>
</tr>
<tr>
<td>Specialty</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No specialty</td>
<td>94 815 (48.2%)</td>
<td>73 309 (55.2%)</td>
<td>168 124 (51.0%)</td>
</tr>
<tr>
<td>Anaesthetics</td>
<td>8 710 (4.4%)</td>
<td>3797 (2.9%)</td>
<td>12 507 (3.8%)</td>
</tr>
<tr>
<td>Emergency Medicine</td>
<td>754 (0.4%)</td>
<td>209 (0.2%)</td>
<td>963 (0.3%)</td>
</tr>
<tr>
<td>General Practice</td>
<td>37 959 (19.3%)</td>
<td>32 264 (24.3%)</td>
<td>70 223 (21.3%)</td>
</tr>
<tr>
<td>Medicine</td>
<td>15 076 (7.7%)</td>
<td>6775 (5.1%)</td>
<td>21 851 (6.6%)</td>
</tr>
<tr>
<td>Obstetrics &amp; Gynaecology</td>
<td>2934 (1.5%)</td>
<td>1966 (1.5%)</td>
<td>4900 (1.5%)</td>
</tr>
<tr>
<td>Ophthalmology</td>
<td>2508 (1.3%)</td>
<td>1078 (0.8%)</td>
<td>3586 (1.1%)</td>
</tr>
<tr>
<td>Paediatrics</td>
<td>3906 (2.0%)</td>
<td>3891 (2.9%)</td>
<td>7797 (2.4%)</td>
</tr>
<tr>
<td>Pathology</td>
<td>5589 (2.8%)</td>
<td>2965 (2.2%)</td>
<td>8554 (2.6%)</td>
</tr>
<tr>
<td>Psychiatry</td>
<td>5494 (2.8%)</td>
<td>3077 (2.3%)</td>
<td>8571 (2.6%)</td>
</tr>
<tr>
<td>Radiology</td>
<td>172 (0.1%)</td>
<td>41 (0.03%)</td>
<td>213 (0.1%)</td>
</tr>
<tr>
<td>Surgery</td>
<td>16 452 (8.4%)</td>
<td>1942 (1.5%)</td>
<td>18 394 (5.6%)</td>
</tr>
<tr>
<td>Other specialties</td>
<td>1330 (0.7%)</td>
<td>867 (0.7%)</td>
<td>2197 (0.7%)</td>
</tr>
<tr>
<td>Dual specialty</td>
<td>1115 (0.6%)</td>
<td>547 (0.4%)</td>
<td>1662 (0.5%)</td>
</tr>
</tbody>
</table>

Approximately half of all the doctors were neither on the GP nor the Specialist registers (51%), of which the majority (58%) had received their primary medical qualification greater than 10 years previously. It is interesting to note that half of the doctors registered to practise medicine in the UK in this period were not registered specialists (they were neither on the GP nor Specialist register) and the majority of these doctors had qualified greater than 10 years ago, suggesting that these doctors are not trainee...
 doctors, but doctors who have not completed speciality training and are currently working in non-training posts. Of those doctors who had specialised, General Practice was the most popular specialty (21.3%), followed by Medicine (6.6%). Only 0.5% of the doctors were on both the Specialist and GP registers.

In total, 2,697 (0.8%) doctors had warnings or sanctions imposed against their registration on 29 May 2013. There was a higher proportion of male doctors who had warnings or sanctions against their registration when compared with female doctors (1.1% of all male doctors compared with 0.4% of all female doctors, $\chi^2=505.4$, p<0.001). There was strong evidence for an association between receiving warnings or sanctions and the number of years since receipt of primary medical qualification, with doctors who qualified 31-40 years ago having the highest proportion of warnings or sanctions; world region of primary medical qualification, with doctors who qualified outside the EEA with the highest proportion with warnings or sanctions; and specialty, with doctors on both the Specialist and GP registers having the highest proportion of doctors with warnings or sanctions. These results are presented in Table 3.2.
Table 3.2 The distribution of warnings and sanctions for each variable and the association of individual factors with warnings or sanctions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total number of doctors</th>
<th>Warnings or sanctions (%)</th>
<th>p value ($X^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Male</td>
<td>196 814</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>132 728</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td><strong>Number of years since receipt of PMQ</strong></td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>0-2</td>
<td>15 162</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>3-10</td>
<td>65 768</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>11-20</td>
<td>92 460</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>21-30</td>
<td>62 215</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>31-40</td>
<td>42 342</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>≥41</td>
<td>51 595</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td><strong>Region where PMQ received</strong></td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>UK</td>
<td>195 312</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>EEA</td>
<td>41 213</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>International</td>
<td>93 017</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td><strong>Specialty</strong></td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>No specialty</td>
<td>168 124</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>Anaesthetics</td>
<td>12 507</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>Emergency Medicine</td>
<td>963</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>General Practice</td>
<td>70 223</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>Medicine</td>
<td>21 851</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Obstetrics &amp; Gynaecology</td>
<td>4900</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>Ophthalmology</td>
<td>3586</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Paediatrics</td>
<td>7797</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>Pathology</td>
<td>8554</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>Psychiatry</td>
<td>8571</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>Radiology</td>
<td>213</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Surgery</td>
<td>18 394</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>Other specialties</td>
<td>2197</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>Dual specialty</td>
<td>1662</td>
<td>1.4</td>
<td></td>
</tr>
</tbody>
</table>

Using bivariate analyses, I compared female doctors with male doctors. There was a strong trend between gender and the number of years since the doctor received their primary medical qualification, with female doctors being more likely to have recently qualified and the proportion of female doctors reducing as the number of years since primary medical qualification increased. I also found that female doctors were more likely to have qualified
in the UK (65.5% of all female doctors compared with 55% of all male doctors) and there was a slightly higher proportion of women when compared to men who were not on the Specialist or GP registers. When examining those doctors who were registered in a specialty, a higher proportion of female doctors were on the GP register compared with male doctors (24.3% of female doctors compared with 19.3% of male doctors) and a higher proportion of male doctors were registered with a hospital specialty (32% of male doctors compare with 20.1% of female doctors).

In summary, the number of years since receipt of primary medical qualification, world region where primary medical qualification was received, and registered specialty were associated with the outcome (warnings or sanctions) and the exposure (gender) and as such these co-variates were considered as confounders.

The unadjusted OR for having warnings or sanctions imposed against a doctor's registration comparing female doctors with male doctors was 0.35 (95% CI: 0.32 to 0.38), suggesting that being a female doctor is protective of having warnings or sanctions imposed.

Mantel-Haenszel analyses and tests of homogeneity were conducted to examine the change in strength of the association between warnings or sanctions and gender while controlling for each confounder separately.

There was strong evidence that the true ORs were different between the different specialty categories (p=0.0002); therefore, specialty was considered as an effect modifier when conducting multivariate analyses.

Table 3.3 represents the results from the binary logistic regression model built to adjust for all the variables. After taking into account the number of years since primary medical qualification and world region where a doctor received their primary medical qualification and specialty, female doctors had nearly a third of the odds of having warnings or sanctions imposed on their
registration compared to male doctors (OR: 0.37, 95% CI: 0.33 to 0.41, p<0.0001).

### Table 3.3 The adjusted OR for having warnings or sanctions against registration for each variable compared to its baseline having adjusted for all other variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Adjusted OR</th>
<th>95% CI</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Male</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.37</td>
<td>0.33 to 0.41</td>
<td></td>
</tr>
<tr>
<td><strong>Number of years since receipt of PMQ</strong></td>
<td></td>
<td></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>0-2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-20</td>
<td>3.42</td>
<td>2.20 to 5.32</td>
<td></td>
</tr>
<tr>
<td>11-20</td>
<td>3.85</td>
<td>2.48 to 5.98</td>
<td></td>
</tr>
<tr>
<td>21-30</td>
<td>5.66</td>
<td>3.63 to 8.81</td>
<td></td>
</tr>
<tr>
<td>31-40</td>
<td>6.44</td>
<td>4.13 to 10.05</td>
<td></td>
</tr>
<tr>
<td>≥41</td>
<td>3.12</td>
<td>2.00 to 4.87</td>
<td></td>
</tr>
<tr>
<td><strong>Region where PMQ received</strong></td>
<td></td>
<td></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>UK</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EEA</td>
<td>1.33</td>
<td>1.17 to 1.50</td>
<td></td>
</tr>
<tr>
<td>International</td>
<td>1.64</td>
<td>1.51 to 1.80</td>
<td></td>
</tr>
<tr>
<td><strong>Specialty</strong></td>
<td></td>
<td></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>No specialty</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anaesthetics</td>
<td>0.65</td>
<td>0.52 to 0.82</td>
<td></td>
</tr>
<tr>
<td>Emergency Medicine</td>
<td>0.66</td>
<td>0.31 to 1.39</td>
<td></td>
</tr>
<tr>
<td>General Practice</td>
<td>1.43</td>
<td>1.29 to 1.58</td>
<td></td>
</tr>
<tr>
<td>Medicine</td>
<td>0.49</td>
<td>0.40 to 0.60</td>
<td></td>
</tr>
<tr>
<td>Obstetrics &amp; Gynaecology</td>
<td>1.22</td>
<td>0.93 to 1.59</td>
<td></td>
</tr>
<tr>
<td>Ophthalmology</td>
<td>0.53</td>
<td>0.33 to 0.83</td>
<td></td>
</tr>
<tr>
<td>Paediatrics</td>
<td>0.64</td>
<td>0.47 to 0.87</td>
<td></td>
</tr>
<tr>
<td>Pathology</td>
<td>0.65</td>
<td>0.50 to 0.86</td>
<td></td>
</tr>
<tr>
<td>Psychiatry</td>
<td>0.81</td>
<td>0.63 to 1.04</td>
<td></td>
</tr>
<tr>
<td>Radiology</td>
<td>0.47</td>
<td>0.07 to 3.38</td>
<td></td>
</tr>
<tr>
<td>Surgery</td>
<td>0.78</td>
<td>0.66 to 0.93</td>
<td></td>
</tr>
<tr>
<td>Other specialties</td>
<td>0.36</td>
<td>0.17 to 0.77</td>
<td></td>
</tr>
<tr>
<td>Dual specialty</td>
<td>1.37</td>
<td>0.90 to 2.09</td>
<td></td>
</tr>
</tbody>
</table>

I felt all of the *a priori* confounders were confounders because the adjusted OR changed when each variable was added to the model. There was no evidence of multicollinearity.
The Mantel-Haenszel analyses suggested that specialty may be an effect modifier; a statistical test for effect modification was therefore performed by first collapsing the specialty variable into four groups to increase the power of the test. Table 3.4 demonstrates that specialty was felt to be an important effect modifier with female doctors being less likely to have warnings or sanctions imposed when compared to male doctors, but the effect was greater for GPs than for doctors with no specialty or those had specialised in a hospital specialty.

Table 3.4 Stratum-specific ORs for having warnings or sanctions imposed on registration if the doctor is female for each specialty divided into four categories

<table>
<thead>
<tr>
<th>Variable</th>
<th>Stratum-specific OR</th>
<th>95% CI</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specialty category</td>
<td></td>
<td></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>No specialty</td>
<td>0.43</td>
<td>0.38 to 0.49</td>
<td></td>
</tr>
<tr>
<td>GP specialty</td>
<td>0.26</td>
<td>0.22 to 0.31</td>
<td></td>
</tr>
<tr>
<td>Hospital specialty</td>
<td>0.44</td>
<td>0.36 to 0.56</td>
<td></td>
</tr>
<tr>
<td>Dual specialty</td>
<td>0.09</td>
<td>0.13 to 0.70</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.5 details the number of male and female doctors with each outcome type. Female doctors were less likely to have each type of outcome compared with male doctors.
Table 3.5 The number and proportion of male and female doctors by outcome type

<table>
<thead>
<tr>
<th>Variable</th>
<th>Male</th>
<th>Female</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=196,814</td>
<td>N=132,728</td>
<td>(X^2)</td>
</tr>
<tr>
<td>Warning</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>No</td>
<td>196,060 (99.6%)</td>
<td>132,580 (99.9%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>754 (0.4%)</td>
<td>148 (0.1%)</td>
<td></td>
</tr>
<tr>
<td>Undertakings</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>No</td>
<td>196,506 (99.8%)</td>
<td>132,596 (99.9%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>308 (0.2%)</td>
<td>132 (0.1%)</td>
<td></td>
</tr>
<tr>
<td>Conditions</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>No</td>
<td>196,463 (99.8%)</td>
<td>132,614 (99.9%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>351 (0.2%)</td>
<td>114 (0.1%)</td>
<td></td>
</tr>
<tr>
<td>Suspension</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>No</td>
<td>196,461 (99.8%)</td>
<td>132,662 (99.9%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>353 (0.2%)</td>
<td>66 (0.1%)</td>
<td></td>
</tr>
<tr>
<td>Erasure</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>No</td>
<td>196,324 (99.8%)</td>
<td>132,658 (99.9%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>490 (0.2%)</td>
<td>70 (0.1%)</td>
<td></td>
</tr>
</tbody>
</table>

Note that a doctor may have more than one type of sanction against their medical registration.

Bivariate analyses examined the association between each outcome type and the other variables (collapsed into binary form). All of the outcome types were strongly associated with doctors’ gender (p<0.001). The remaining variables were found to be associated with each outcome type (p≤0.02), with the exception of undertakings which was not found to be associated with whether the primary medical qualification was obtained within or outside of the UK region (p=0.575); and erasure which was not found to be associated with inclusion on the GP and/or Specialist registers (p=0.226).

Mantel-Haenszel analyses were conducted and there was evidence that the true OR were different between the different categories of specialty, when examining undertakings (p=0.0041), and whether the primary medical qualification was obtained in the UK for erasure (p=0.0152), therefore these were considered as effect modifiers.

Table 3.6 presents the results from the binary logistic regression models, demonstrating that female doctors had approximately one third of the odds of having a warning imposed or being suspended or erased, when compared to
male doctors. But female doctors had half the odds of having undertakings or conditions imposed on their registration. This suggests that women are less likely to have the more severe types of sanctions (suspension and erasure) imposed against their medical registration. There was no evidence of substantial multicollinearity in the models.

Table 3.6 ORs of a female doctor having each outcome type, while controlling for number of years since PMQ, region of PMQ, and specialty (as binary variables)

<table>
<thead>
<tr>
<th>Variable</th>
<th>OR</th>
<th>95% CI</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Warning</td>
<td>0.30</td>
<td>0.25 to 0.36</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Undertakings</td>
<td>0.66</td>
<td>0.53 to 0.81</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Conditions</td>
<td>0.54</td>
<td>0.44 to 0.67</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Suspension</td>
<td>0.32</td>
<td>0.25 to 0.42</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Erasure</td>
<td>0.26</td>
<td>0.21 to 0.34</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Table 3.7 illustrates that specialty category is an effect modifier for the sanction type undertakings. Female doctors on the GP and/or Specialist registers were less likely to receive undertakings, whereas female doctors not on one of the registers were equally likely as their male counterparts to have undertakings against their medical registration.

Table 3.7 Stratum-specific ORs for receiving undertakings is the doctor is female (with specialty as a binary variable)

<table>
<thead>
<tr>
<th>Variable</th>
<th>OR</th>
<th>95% CI</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specialty category</td>
<td></td>
<td></td>
<td>0.0031</td>
</tr>
<tr>
<td>No specialty</td>
<td>0.90</td>
<td>0.67 to 1.20</td>
<td></td>
</tr>
<tr>
<td>On GP and/or Specialist register(s)</td>
<td>0.48</td>
<td>0.35 to 0.65</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.8 shows that region where primary medical qualification was obtained is an effect modifier when examining the outcome type erasure. Female doctors who had qualified in the UK had nearly one fifth of the odds of being erased from the medical register compared to the male colleagues, whereas female doctors who qualified outside of the UK had nearly one third of the odds of being erased from the medical register when compared to their male counterparts.
Table 3.8 Stratum-specific ORs for erasure from the medical register if a doctor is female (with region of PMQ as a binary variable)

<table>
<thead>
<tr>
<th>Variable</th>
<th>OR</th>
<th>95% CI</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMQ taken in UK</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No (Non-UK)</td>
<td>0.32</td>
<td>2.24 to 0.44</td>
<td></td>
</tr>
<tr>
<td>Yes (UK)</td>
<td>0.18</td>
<td>0.11 to 0.28</td>
<td></td>
</tr>
</tbody>
</table>

3.5 Discussion

3.5.1 Summary

In this large cross-sectional study, I have found strong evidence that being female is associated with a reduction in odds of having warnings or sanctions imposed (OR: 0.35, 95% CI: 0.32 to 0.38) in the unadjusted model. Controlling for years since primary medical qualification, world region where the doctor received their primary medical qualification, and specialty did slightly increase this OR (OR: 0.37, 95% CI: 0.33 to 0.41, p<0.0001), but there remained strong evidence for the association between doctor’s gender and having warnings or sanctions imposed. There was evidence that the association varied with specialty, with female GPs being the least likely to have warnings or sanctions imposed against their registration.

Female doctors were also found to be less likely to receive the most severe sanctions against their registration (suspension or erasure). Female doctors who are included on the GP and/or Specialist registers were also less likely to receive undertakings compared to their male counterparts.

To the best of my knowledge, this is the first study in the UK to examine the association between doctor’s gender and having warnings or sanctions imposed against medical registration, while adjusting for known confounders. I believe that these known confounders have only been adjusted for in one previous study, which was conducted in the USA [194].
3.5.2 Strengths and limitations

One of the major strengths of this study is that a large national database was used. The advantage of using this data set is twofold: first, since a national database was used, the findings of the study apply to all doctors registered to practise in the UK; second, having warnings or sanctions imposed is a rare outcome and using a large data set is crucial when examining rare outcomes.

A further strength of this study is that it adjusted for known confounders; all these confounders have only been adjusted for in one previous study, which was conducted in California [194]. Previous research on UK data did not adjust the measure of effect of gender for potential confounders [5].

Finally, a further advantage was the completeness of the data set. The data are collected by the GMC for inclusion on the LRMP, and not for research purposes. Doctors are required to provide the data to the GMC to be registered and as such there were no missing data. However, the fact the data were not collected for research purposes is also a limitation of the study. The study was constrained by the information collected and made available by the GMC. As such, I was only able to explore the variables available. I was not able to examine the effect of other potential confounding factors, such as ethnicity, or explore the reasons why a warning or sanction had been imposed; nor was I able to establish the date a sanction had been imposed.

It could be argued that the reason for referral to the GMC could be a source of residual confounding if systematic differences exist between the genders. The GMC may take action against a doctor’s registration for a number of reasons, which can be broadly divided into three major categories: misconduct, poor professional performance and physical or mental ill health. The data available did not provide the reasons or the category for why a sanction had been imposed, but a more detailed evaluation of the reasons for referral to the GMC may go towards explaining the gender difference observed. In their research, Alam et al [189] and Elkin et al [192],
demonstrated not only that male doctors were more likely to be subject to disciplinary action, but also that the main offence for which a doctor was being disciplined was sexual misconduct. It is possible that male doctors are more likely to commit an offence involving sexual misconduct than their female colleagues, which may go towards explaining the gender difference seen in these populations. However, other studies in this area did not find sexual misconduct to be the most common offence [191, 193, 195]. As such, exploring the offences for which a doctor may receive disciplinary action in this population may go towards explaining the gender difference observed and may help the regulatory body and medical profession to introduce targeted interventions, such as education programmes, to reduce the number of offences. I explore the difference between the genders in reason for referral to the GMC in Chapter 7.

A further source of residual confounding could be the route of referral. Doctors practising in the UK can be referred to the GMC from a wide range of sources, which can broadly divided into four main groups: the public; person acting in a public capacity (PAPC) (e.g. police, employer); doctors; and other (e.g. media) [181]. I feel it would be interesting to examine whether the referral rate for each route demonstrates any gender differences and, if so, exploring reasons for this difference. I therefore explore this hypothesis in Chapter 7.

A further limitation is that a doctor can apply for voluntary erasure from the LRMP during an investigation process. Once again, this is a potential source of residual confounding and it is possible that the gender of doctors who are subject to a complaints investigation and apply for voluntary erasure differs from those doctors who complete the investigative process and receive a warning or sanction (however, the GMC’s decision to grant a request for voluntary erasure is based on the public interest and the doctor’s health and likelihood to return to practise [200]). It is important to note that voluntary erasure is requested by doctors for multiple reasons other than being involved in an investigation process, including retiring permanently from
practising medicine or leaving the UK to work permanently abroad. To explore whether voluntary erasure requests may explain the gender difference seen, the reason why a voluntary erasure request was submitted would first have to be ascertained. This information was not available in the data set used for this research, but could be requested and explored in future studies.

It is also of interest to note that certain sanctions (erasure and some suspensions) are permanent, whereas other sanctions are time limited. As such, the permanent sanctions may be over-represented because they will never be removed from a doctor’s registration. If male doctors are more likely to receive these permanent sanctions, this could lead to male doctors being overrepresented when examining the association between gender and sanctions, and may go towards explaining the gender difference observed between doctors who had sanctions imposed against their registration.

Finally, a further limitation of the study is that nearly half of the doctors in the population were not recorded on the GP or Specialist registers and were therefore classified as not having a specialty. Doctors who are not on the GP or Specialist registers typically fall into one of two categories: either a doctor who is on a Specialty Training programme with the aim of becoming a GP or consultant in a specialty; or doctors who have chosen to work in a non-training post. It was not possible from the information made available by the GMC to examine these two categories. It would be of interest to explore if the proportions of male and female doctors differ in these two categories and to examine whether the risk of disciplinary action differs for doctors who are in a Specialty Training post compared with doctors who are working in a non-training post. It would also be of interest to explore if further information is available about the type of non-training post these doctors were working in and to examine the association with receiving warnings or sanctions.
3.5.3 Comparison with other studies

The main finding, that female doctors are less likely to be subject to disciplinary action when compared to their male colleagues, mirror the results of several studies from across the world, which have also examined the association between doctors' gender and disciplinary action [189, 191, 193-195, 192, 5]. However, the majority of these studies have been performed in the USA [191, 193-195], Canada [189], Australia and New Zealand [192], where the medical and legal systems differ from those in the UK, and therefore their findings may not be applicable to the UK population of doctors. These studies' main objective was not necessarily to explore the association between doctors' gender and disciplinary action. Some of these studies were descriptive, and those studies that did control for confounders did not control for the same confounders selected for this study, albeit for one study [194]. To the best of my knowledge, one study has been performed in the UK using national data [5]; however, when examining the association between doctors’ gender and disciplinary action, this study did not control for any potential confounders.

The findings of this study are in agreement with previous research that has shown that older doctors [194], doctors who qualified outside of the country in which they are practising [5, 163] and doctors of certain specialties [193] are more likely to be subjected to disciplinary action by a medical regulatory body. It has also been demonstrated in previous studies that female doctors are more likely to have qualified more recently than male doctors [94], are more likely to have qualified in the country in which they are practising [201] and choose different specialties to male doctors [202].

With this study I have shown the reason why male doctors receive more warnings or sanctions is not because they qualified earlier, nor because they are more likely to have qualified outside of the UK, despite both of those factors being associated with increased likelihood of warnings or sanctions.
3.5.4 Unanswered questions and future research

Through this study I have demonstrated that female doctors are less likely to receive warnings or sanctions against their medical registration compared to male doctors, and they are less likely to have the more severe sanctions imposed. However, it is not clear why women are less likely to receive warnings or sanctions when compared to men. Exploring the possible reasons for this gender difference in professional performance is required. One theory suggested by some researchers is that male and female doctors differ in communication style, and hence the interaction with patient and colleagues differs between the genders, which could affect the risk of being subject to a complaint [203, 204]. As part of this thesis I plan to conduct research to explore this further and examine whether communication styles differ between male and female doctors, and also whether the communication styles of doctors who subsequently face disciplinary action differs from those doctors who do not (see Chapter 8).

I have also demonstrated that the effect of gender on the likelihood to receive sanctions varied with specialty, with female GPs being the least likely to receive warnings or sanctions. It has been demonstrated by an observational study of primary care physicians in the USA that female primary care physicians spend more time with their patients when compared to their male colleagues, and they engaged more in conversation, displaying more positive-talk, partnership-building, question-asking and information giving [205]. These differences in communication style and time spent with patients may go towards explaining the larger gender discrepancy observed in GPs; however, future research is required to explore these possible reasons further.

My results show that doctors who had been qualified for longer were more likely to receive warnings or sanctions; it is therefore reasonable to hypothesise that the risk of receiving a warning or sanction increases with exposure. Therefore, doctors who have fewer patient encounters may be less likely to receive warnings or sanctions against their registration.
Previous studies have demonstrated that female consultants have fewer patient episodes and are more likely to work part-time (and thus have fewer patient encounters) when compared to their male colleagues [206, 207]; it is therefore reasonable to hypothesise that the gender difference observed in disciplinary action may be partly explained by the difference in work patterns between the genders. This hypothesis merits being explored further by examining whether certain work patterns (i.e. part-time compared with full-time work) are more likely to be associated with receiving warnings or sanctions.

A further possible explanation for the gender difference observed is perhaps that male and female doctors may be viewed and treated differently by the public, the profession and the regulatory body. It is possible that there is a higher threshold of tolerance for female doctors. The GMC are twice as likely to receive a complaint about a male doctor than a female doctor [181]. It is reasonable to assume that as a result male doctors are more likely to receive warnings or sanctions against their medical registration. Research examining the perception of male and female doctors would be warranted.

I have demonstrated that female doctors who are listed on the GP and/or Specialist registers are less likely to receive undertakings, when compared to their male colleagues. However, this gender difference is not observed for doctors who are neither on the GP nor Specialist registers. The association between doctors’ gender and warnings or the other types of sanctions was not found to vary with specialty registration. Examining the possible reasons behind why receiving undertakings varies according to whether a doctor is listed on the GP and/or Specialist registers would be of interest. Possibilities could be that the threshold for delivering undertakings to a doctors who are listed on the GP or Specialist registers is lower, and this lower threshold might be because undertakings need to be agreed between the doctor and the GMC and remain in place until the doctor has remediated the concerns. Undertakings may be felt to be more effective for those doctors who are fully qualified, due to the more independent nature of their work.
I have also been able to show that female doctors who qualified in the UK have one fifth of the odds of being erased from the medical register, compared to male doctors who qualified in the UK. However, female doctors who qualified outside of the UK have one third of the odds of being erased from the medical register when compared to male doctors who qualified outside of the UK. This finding suggests that doctors who qualified outside of the UK are more likely to be erased, and echo the findings of a cohort study conducted by Humphrey et al [163] who demonstrated that referral to the GMC concerning doctors who received their primary medical qualification outside of the UK was associated with the most severe sanctions against their medical registration (suspension or erasure). The difference in the odds of being erased varying by where primary medical qualification was obtained could also suggest that there may be reasons for why male UK graduates are more severely sanctioned. Perhaps they are the group who commit the more serious types of offenses, or perhaps they are treated more severely by the regulatory body.

My results also show that female doctors have one third of the odds of being erased or suspended, and at reduced odds of receiving undertakings or conditions when compared to their male colleagues. These results suggest that, with the exception of the outcome type ‘warning’, the more severe the sanction (suspension or erasure from the medical register) the less likely women will receive it. Exploring the reasons for this gender difference in the severity of sanctions imposed would be of interest. It is possible that male doctors' behaviour and actions warrant the severest types of sanctions to be imposed, whereas female doctors' behaviour and actions do not merit such severe sanctions. Examining the reasons for why doctors have sanctions imposed on their medical registration would go towards showing whether this is in fact the case.

It should also be noted that this study was observational in design and as such causality cannot be determined. It is possible that other factors, such as ethnicity, may be confounding the association between doctors’ gender
and disciplinary action. Research examining whether other potential confounders could explain the observed association is required.

The points discussed above highlight that the real interest of this research is not about the outcome of warnings or sanctions themselves, but about trying to understand the difference between male and female doctors that lead to the observed gender difference in receiving warnings or sanctions. Investigations into why and how male and female medical practices differ will in turn lead to being able to propose interventions to reduce not only the number of doctors referred to the medical regulatory body, but also the difference between the genders of doctors who are referred. Further exploration of why doctors’ gender may impact their professional performance is needed to enable the profession to develop a better understanding of the factors associated with impaired fitness to practise and, crucially, how to better support those doctors and ensure patient safety. This thesis will work towards addressing this need and build upon this research study.

3.5.5 Conclusion

In this study, I have demonstrated that female doctors practising in the UK are less likely to have warnings or sanctions imposed on their medical registration when compared to their male colleagues. These findings remained after adjusting for known confounders. The reasons behind this gender difference are not known, but possible theories for this observed difference in measure of extreme poor professional performance have been discussed. I will explore a selection of these hypotheses in greater depth throughout this thesis. In Chapter 7 I will explore whether, in a population of GPs who have been referred to the GMC with performance concerns, there are systematic differences in the route and reason for referral to the GMC. A further hypothesis is that women doctors exhibit behaviours and skills that are protective of disciplinary action, with a popular hypothesis being that the communication skills of doctors differ between the genders. I will further investigate this hypothesis in Chapter 8, where I will explore the association
between gender, communication style and disciplinary action in a population of UK medical graduates.

In Chapter 4 I will investigate whether the findings presented in this chapter are replicated in populations of doctors who practise in countries outside of the UK. I will also explore whether the observed gender difference persists when a spectrum of extreme poor performance measures are used, and whether the association with gender has varied over the last few decades, to determine whether the increasing number of women practising medicine has had any influence on this association.
Chapter 4 - Gender differences in medico-legal action against doctors: a systematic review and meta-analysis

4.1 Chapter summary

4.1.1 Background

In the large cross-sectional study, described in Chapter 3, I demonstrated that in the year 2013, female doctors registered to practise in the UK were less likely to face disciplinary action. This study provides a snapshot of the gender difference in disciplinary action amongst doctors practising in the UK in 2013. It would be of interest to explore whether the association between gender and disciplinary action amongst doctors has varied over time, given the relatively recent increase in the number of women doctors, and whether the pattern observed between gender and disciplinary action in the UK is mirrored in other countries.

4.1.2 Aims

To establish the robustness of the association between gender and poor performance in doctors, internationally and over time.

4.1.3 Design and setting

A systematic review and meta-analysis of all published original research studies examining the gender differences in medico-legal actions against doctors.

4.1.4 Methods

The electronic databases MEDLINE, Embase, and PsycINFO were searched from inception to January 2015. Backward and forward citation searching was performed. Journals that yielded the majority of the eligible articles and journals in the medical education field were electronically searched, along with the conference and poster abstracts from two of the largest international medical education conferences. Studies reporting original data, written in English or French, examining the association between gender and medico-legal action against doctors were included. Two reviewers independently

119
extracted study characteristics and outcome data from the full texts of the studies meeting the eligibility criteria. Study quality was assessed using the Newcastle-Ottawa scale. A random effects meta-analysis model was used to summarise and assess the effect of doctors’ gender on medico-legal action. Extracted outcomes included disciplinary action by a medical regulatory board, malpractice experience, referral to a medical regulatory body, complaints received by a healthcare complaints body, criminal cases, and medico-legal matter with a medical defence organisation.

4.1.5 Results

Overall, 32 reports examining the association between doctors’ gender and medico-legal action were included in the systematic review (n=4,054,551), of which 27 found that female doctors were less likely to have experienced medico-legal action. 19 reports were included in the meta-analysis (n=3,794,486, including 20,666 cases). Results showed female doctors had less than half the odds of being subject to medico-legal action than male doctors. Heterogeneity was present in all meta-analyses.

4.1.6 Conclusion

Female doctors are less likely to have had experienced medico-legal actions compared to male doctors. This finding is robust internationally, across outcomes of varying severity, and over time.
4.2 Introduction

My earlier cross-sectional study of all registered doctors in the UK in 2013 demonstrated that the gender of a doctor was an important factor associated with disciplinary action against a doctor’s medical registration [9] (see Chapter 3). Female doctors were less likely to receive disciplinary action, even after taking into account other explanatory variables such as years since qualification and specialty. That study provided a snapshot of the situation in the UK, but did not include doctors practising outside of the UK, nor did it include other measures of poor performance not resulting in disciplinary action by the GMC but nonetheless serious. Several studies examining gender differences in disciplinary action against doctors across the world have been conducted, and the results of these studies vary – some conclude female doctors are less likely to be disciplined [194, 195], but with varying effect sizes, while others have not found a significant association [58, 59]. It is not clear whether gender differences are robust across contexts and across measures of performance. Establishing the generalisability and an overall effect size internationally, over time, and on multiple measures of poor performance, will improve understanding of what factors result in poor professional performance and how to remediate it.

In the present study I completed a systematic review of the literature and meta-analysis to answer the following questions:

(1) Was the gender difference observed in UK doctors in 2013 also present in different countries, with different medical systems and cultures?
(2) Has the gender difference varied over the last four decades?
(3) Are gender differences present on measures of poor performance other than disciplinary action, such as malpractice litigation?

4.3 Methods

Guidance published by PRISMA [208] and Cochrane [209] was used to guide the study’s methodology.
4.3.1 Data sources and search strategies

Systematic searches (from inception to January 2015) of MEDLINE, Embase, and PsycINFO were conducted for studies describing the association between doctors’ gender and experience of medico-legal action. The search criteria consisted of three key concepts: (1) population of interest (medical doctors), (2) exposure of interest (female gender), and outcome of interest (measure of experience of medico-legal action). See Table 4.1. In addition, backward and forward citation searching was performed and electronic searches within the journals that yielded the majority of the eligible articles, along with journals important in the medical education field, for relevant articles. Finally, electronic searches of conference and poster abstracts published from two of the largest medical education conferences for relevant literature was conducted (See Table 4.2). Studies not published in English or French were excluded due to limited resources.
### Table 4.1 Search terms used for electronic database search

#### MEDLINE search conducted in January 2015

<table>
<thead>
<tr>
<th>Concept 1</th>
<th>Concept 2</th>
<th>Concept 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>MeSH term Physicians</td>
<td>MeSH term Gender Identity</td>
<td>MeSH term Malpractice</td>
</tr>
<tr>
<td>Doctor?</td>
<td>MeSH term Male</td>
<td>MeSH term Employee Performance Appraisal</td>
</tr>
<tr>
<td>Medical Practitioner?</td>
<td>MeSH term Female</td>
<td>Clinical competence</td>
</tr>
<tr>
<td>Medic?</td>
<td>(Male and Female MeSH terms</td>
<td>Disciplinary board</td>
</tr>
<tr>
<td>Clinician?</td>
<td>combined with AND)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MeSH term Sex</td>
<td>Disciplinary action</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Professional misconduct</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fitness to practice</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medical regulation</td>
</tr>
</tbody>
</table>

#### Embase search conducted in January 2015

<table>
<thead>
<tr>
<th>Concept 1</th>
<th>Concept 2</th>
<th>Concept 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>MeSH term Physician</td>
<td>MeSH term Gender and Sex</td>
<td>MeSH term Malpractice</td>
</tr>
<tr>
<td>Doctor?</td>
<td>MeSH term Gender Identity</td>
<td>MeSH term Job Performance</td>
</tr>
<tr>
<td>Medical Practitioner?</td>
<td>MeSH term Gender</td>
<td>MeSH term Performance</td>
</tr>
<tr>
<td>Medic?</td>
<td>MeSH term Male</td>
<td>MeSH term Professional Practice</td>
</tr>
<tr>
<td>Clinician?</td>
<td>MeSH term Female</td>
<td>MeSH term Professional Misconduct</td>
</tr>
<tr>
<td></td>
<td>(Male and Female MeSH terms</td>
<td>Clinical competence</td>
</tr>
<tr>
<td></td>
<td>combined with AND)</td>
<td>Disciplinary board</td>
</tr>
<tr>
<td></td>
<td>MeSH term Sex</td>
<td>Disciplinary action</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Professional misconduct</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fitness to practice</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medical regulation</td>
</tr>
</tbody>
</table>

#### PsycINFO search conducted January 2015

<table>
<thead>
<tr>
<th>Concept 1</th>
<th>Concept 2</th>
<th>Concept 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>MeSH term Physicians</td>
<td>MeSH term Gender Identity</td>
<td>MeSH term Professional Liability</td>
</tr>
<tr>
<td>MeSH term Family Physicians</td>
<td>MeSH term Gender Identity</td>
<td>MeSH term Professional Competence</td>
</tr>
<tr>
<td>Doctor?</td>
<td>Sex</td>
<td>Clinical competence</td>
</tr>
<tr>
<td>Medical Practitioner?</td>
<td></td>
<td>Disciplinary board</td>
</tr>
<tr>
<td>Medic?</td>
<td></td>
<td>Disciplinary action</td>
</tr>
<tr>
<td>Clinician?</td>
<td></td>
<td>Professional misconduct</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fitness to practice</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medical regulation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Professional performance</td>
</tr>
</tbody>
</table>
Table 4.2 Names of journals and conference abstracts searched electronically

<table>
<thead>
<tr>
<th>Journals that yielded the majority of eligible articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMJ Quality &amp; Safety</td>
</tr>
<tr>
<td>British Medical Journal</td>
</tr>
<tr>
<td>The American Journal of Medicine</td>
</tr>
<tr>
<td>The Journal of the American Medical Association</td>
</tr>
<tr>
<td>The Medical Journal of Australia</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Journals felt to be important in the medical education field</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic Medicine</td>
</tr>
<tr>
<td>Medical Education</td>
</tr>
<tr>
<td>Medical Teacher</td>
</tr>
<tr>
<td>The New England Journal of Medicine</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Medical education conference abstracts</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Association for Medical Education in Europe</td>
</tr>
<tr>
<td>The Association for the Study of Medical Education</td>
</tr>
</tbody>
</table>

4.3.2 Study selection

I assessed the eligibility of identified studies, without consideration of their results. Articles were considered for inclusion in the systematic review if (1) the study included data from an original and peer-reviewed study, (2) the study participants were medical doctors, (3) the authors reported an effect estimate (or provided data that enabled the calculation of an effect estimate) or the proportion of male and female participants who had experienced a medico-legal action.

I considered all studies, regardless of study design, and I used broad criteria to define the outcome term medico-legal action (see below for the definition of each outcome term used). I identified articles eligible for further review by performing an initial screen of titles and/or abstracts, followed by a full-text review. The reports were not assessed blind, and I was aware of the authors’ names, affiliations and the source of publication.

4.3.3 Data extraction

I, along with another researcher (CW) independently extracted data from all the eligible studies using a pre-determined data extraction form modified
from the Cochrane Handbook [209]. Each researcher independently recorded information on study characteristics (authors, publication year, journal, country, study design, years study conducted, sampling method, data collection method), participants’ characteristics (gender, specialty, grade, number), method by which exposure data was collected, and main outcome (type, methods by which data was collected). We also recorded information on analysis strategy and reported proportion of outcome for each gender, or odds ratio with confidence intervals. When reports contained multivariate analyses, we prioritised crude effects; however if no crude effects were reported, we included outcome measures adjusted for other variables. Any discrepancies in the data extraction process were reconciled through discussion.

4.3.4 Assessment of methodological quality

The Newcastle-Ottawa Scale (NOS) is recommended by the Cochrane collaboration to assess the quality of nonrandomised studies [209]. We independently assessed the methodological quality of each of the studies included in the systematic review using the NOS [210], any differences were resolved by consensus. The purpose of using the NOS was to systematically appraise the quality of the studies to aid comparison between the studies, and as such I decided that I would not exclude any studies from the systematic review or meta-analysis based on the findings of the appraisal of methodological quality.

4.3.5 Outcome definition and subgroup analyses

I used a variety of outcome definitions in an attempt to capture as much literature as possible and allow for the variety of terms to describe medico-legal actions used by different countries. Myself and another researcher (CW) independently selected the most relevant outcome definition for each study included in the review. Any disagreement about outcome category was resolved through consensus. Throughout this chapter, I have used the term ‘medico-legal action’ to group together and represent all of the outcomes types.
4.3.5.1 Disciplinary action by a medical regulatory body
Disciplinary action taken against the doctor by a medical regulatory board.

4.3.5.2 Malpractice experience
Either malpractice claims or malpractice cases. Malpractice claims could result in resolution, settlement, or lawsuit.

4.3.5.3 Complaints received by a medical regulatory body
Complaints or referrals received by a medical regulatory body about a doctor’s practice.

4.3.5.4 Complaints received by a healthcare complaints body
Complaints received by an organisation other than a medical regulatory body, whose function is to help investigate healthcare complaints and provide advice on how to handle the case. Patients, colleagues, healthcare organisations can refer to these bodies; self-referrals from doctors are also accepted.

4.3.5.5 Criminal case
Any sanctions imposed by the criminal justice system for criminal activities performed while practising as a doctor.

4.3.5.6 Medico-legal matter with a medical defence organisation
This umbrella term was used when a study grouped together and examined several outcome types and it was not possible to examine each outcome type separately. The included outcomes were malpractice claim, complaint to a healthcare complaints body or medical regulatory body, disciplinary hearing by a medical regulatory body, and criminal charges, among others.

I recognise that the outcome measures vary in severity (disciplinary action being taken against a doctor by a medical regulatory board is not equivalent to a complaint received by a medical regulatory body). However, I felt that a
variety of outcome measures would increase the capture of relevant literature.

I decided, a priori, to perform subgroup analyses based on study design, country where the study population were employed, type of outcome measure, grade and specialty of the doctors within the study population, the most recent year in which the data was collected (if missing, year of publication was used), and NOS rating. I chose those variables because of their potential impact on any association between doctors’ gender and medico-legal action: study design can influence types of bias introduced; medical and legal systems vary between countries, and complaints may be dealt with differently in different systems; the proportion of women practising medicine has been increasing overtime; and I have demonstrated that specialty and grade of a doctor are associated with medico-legal action [9] (see Chapter 3).

4.3.6 Statistical analysis

I performed the main analysis for all the studies combined. I then conducted subgroup analyses on variables selected a priori. To ensure there were sufficient studies in each stratum to demonstrate a meaningful result, the outcome variable was grouped into three categories: ‘Disciplinary action’, ‘Malpractice’, and ‘Other’. The variable ‘country’ was grouped into three categories that represented the continents from which the studies arose: ‘North America’, ‘Asia and Australia’, and ‘Europe’. The year variable was grouped into six 5-year bands (1985-89, 1990-94, 1995-99, 2000-04, 2005-09, and 2010-14). The variable ‘NOS rating’ was grouped into three categories: ‘8 to 9 stars’, ‘6 to 7 stars’, and ‘4 to 5 stars’.

I calculated an effect estimate for each study for the effect of female gender on medico-legal action, and performed heterogeneity tests. I then calculated a summary estimate of effect of female gender on experience of medico-legal action using the random-effects model. Meta-analyses followed to enable the provision of statistical evidence of heterogeneity.
4.4 Results

4.4.1 Systematic review

I retrieved 6,598 citations, of which 32 of the studies met the inclusion criteria for the systematic review (see Figure 4.1). A study population (including both cases and non-cases) of 4,054,551 was captured by the included studies, of which over 40,246 are cases of medico-legal actions.

All of the included studies used an observational study design (17 cohort studies [211, 189, 212-215, 192, 216-218, 193, 219, 66, 220, 221, 203, 222], 8 cross-sectional studies [223-226, 9, 5, 227, 228], and 7 case-control studies [229, 191, 194, 195, 58, 59, 74]). All the studies examined medico-legal actions as their primary outcome. Medico-legal actions included disciplinary action by a medical regulatory board (15 studies) [211, 189, 213, 229, 191-195, 66, 58, 59, 9, 5, 74], malpractice experience (9 studies) [223, 212, 224, 216, 218, 203, 227, 222, 228], referral to medical regulatory body
(2 studies) [220, 221], complaints received by healthcare complaints body (2 studies) [214, 215], medico-legal matter with a defence organisation (2 studies) [225, 226], and criminal cases (2 studies) [217, 219]. The year of publication ranged from 1988 to 2014, with the largest number of studies being published in 2011 and 2013 (4 studies in each year). Data was collected from 1971 to 2013, covering a period of 42 years. All but two of the studies were conducted in English speaking, developed countries; Australia and/or New Zealand (5 studies) [214, 192, 225, 226, 220], Canada (3 studies) [211, 189, 221], England (1 study) [215], UK (3 studies) [9, 5, 74], and USA (18 studies) [223, 212, 229, 224, 191, 216-218, 193-195, 219, 66, 58, 59, 203, 227, 222]. Of the two studies not conducted in a predominantly English speaking country, one was conducted in Denmark [213] and the other was conducted in Taiwan [228].

The 32 studies varied in how their selected population was representative of the general doctor population. Sources of data included country-wide data (13 studies) [211, 189, 223, 213, 215, 217, 219, 66, 220, 9, 5, 227, 228], data from multiple states or provinces (4 studies) [214, 224, 192, 221], data from one state (7 studies) [229, 191, 216, 193, 195, 222], data from an insurance company (4 studies) [212, 225, 226, 203], data from graduates from specific medical schools (3 studies) [58, 59, 74], and data from one medical centre (1 study) [218]. The majority of the studies examined doctors from all specialties (21 studies) [189, 214, 229, 215, 192, 217, 193-195, 226, 219, 58, 59, 220, 221, 203, 9, 5, 222, 228, 74]. When selecting their cases and non-cases, the majority of the studies selected all eligible cases (29 studies) [211, 189, 223, 212-214, 229, 224, 191, 215, 192, 216-218, 193-195, 226, 219, 66, 58, 59, 220, 221, 203, 9, 5, 222, 74] and all eligible non-cases/controls or full workforce data (18 studies) [189, 223, 212, 229, 224, 215-218, 193, 226, 219, 66, 221, 203, 9, 5, 222]. All the studies had selected appropriate non-cases/controls that were drawn from the same population from which the cases had originated. The majority of the studies used data that had been collected by other organisations, thereby reducing the risk of false ascertainment of gender or the outcome (26 studies) [211, 189, 212-214,
Study characteristics are shown in Table 4.3.
Table 4.3 Characteristics of eligible studies

<table>
<thead>
<tr>
<th>First author, year (country)</th>
<th>Study design</th>
<th>Description of study population</th>
<th>Years data collected</th>
<th>Data source</th>
<th>Outcomes assessed</th>
<th>Statistical test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alam et al. [211], 2013 (Canada)</td>
<td>Cohort</td>
<td>Anaesthetists of all grades</td>
<td>2000-2011</td>
<td>Medical regulatory authority (College of Physicians and Surgeons)</td>
<td>Disciplinary action</td>
<td>Proportion</td>
</tr>
<tr>
<td>Alam et al. [189], 2011 (Canada)</td>
<td>Cohort</td>
<td>Doctors of all specialties and grades</td>
<td>2002-2009</td>
<td>Medical regulatory authority (College of Physicians and Surgeons)</td>
<td>Disciplinary action</td>
<td>Proportion</td>
</tr>
<tr>
<td>Balch et al. [223], 2011 (USA)</td>
<td>Cross-sectional</td>
<td>Surgeons of all grades</td>
<td>2010</td>
<td>Electronic questionnaire</td>
<td>Malpractice suit in last 2 years</td>
<td>$\chi^2 \ P &lt;0.01$</td>
</tr>
<tr>
<td>Baldwin et al. [212], 1991 (USA)</td>
<td>Cohort</td>
<td>GPs and obstetricians of all grades</td>
<td>1982-1988</td>
<td>Insurance company</td>
<td>Malpractice experience</td>
<td>$\chi^2 \ P &gt;0.05$</td>
</tr>
<tr>
<td>Birkeland et al. [213], 2013 (Denmark)</td>
<td>Cohort</td>
<td>GPs; grades of doctors not stated</td>
<td>2007</td>
<td>Medical regulatory authority (Complaint handling authority)</td>
<td>Disciplinary action</td>
<td>Multivariate analysis adjusted for complaint motives, patient characteristics, GP characteristics</td>
</tr>
<tr>
<td>Bismark et al. [214], 2013 (Australia)</td>
<td>Cohort</td>
<td>All specialties of all grades</td>
<td>2000-2011</td>
<td>Health Service Commissions</td>
<td>Patient complaints</td>
<td>Proportion</td>
</tr>
<tr>
<td>Cardarelli et al. [229], 2004 (USA)</td>
<td>Case-control</td>
<td>All specialties of all grades</td>
<td>1989-1998</td>
<td>Medical regulatory authority (State Board of Medical Examiners)</td>
<td>Disciplinary action</td>
<td>Multivariate analysis adjusted for years in practice, ethnicity, international education, specialty, method of licensure</td>
</tr>
<tr>
<td>Chauhan et al. [224], 2005 (USA)</td>
<td>Cross-sectional</td>
<td>Obstetricians and gynaecologists, excluded residents</td>
<td>Not reported</td>
<td>Postal questionnaire</td>
<td>Malpractice claim</td>
<td>Multivariate analysis adjusted for age, ethnicity, years in practice, no subspecialty</td>
</tr>
<tr>
<td>First author, year (country)</td>
<td>Study design</td>
<td>Description of study population</td>
<td>Years data collected</td>
<td>Data source</td>
<td>Outcomes assessed</td>
<td>Statistical test</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------</td>
<td>---------------------------------</td>
<td>---------------------</td>
<td>-------------</td>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Clay et al. [191], 2003 (USA)</td>
<td>Case-control</td>
<td>Majority of specialties of all grades</td>
<td>1997-1999</td>
<td>Medical regulatory authority (State Medical Board)</td>
<td>Disciplinary action</td>
<td>Univariate analysis with controls matched on location</td>
</tr>
<tr>
<td>Donaldson et al. [215], 2014 (England)</td>
<td>Cohort</td>
<td>All specialties of all grades</td>
<td>2001-2012</td>
<td>National Clinical Assessment Service</td>
<td>Referral to National Clinical Assessment Service</td>
<td>Univariate analysis</td>
</tr>
<tr>
<td>Elkin et al. [192], 2011 (Australia &amp; New Zealand)</td>
<td>Cohort</td>
<td>All specialties of all grades</td>
<td>2000-2009</td>
<td>Written determinations</td>
<td>Disciplinary action</td>
<td>Rate</td>
</tr>
<tr>
<td>Ely et al. [216], 1999 (USA)</td>
<td>Cohort</td>
<td>GPs, excluding doctors who were unlicensed or recently licensed</td>
<td>1971-1994</td>
<td>Insurance company</td>
<td>Malpractice claims</td>
<td>Univariate and multivariate analysis adjusted for international education, board certification, physician's recognition award, practice location</td>
</tr>
<tr>
<td>Goldenbaum et al. [217], 2008 (USA)</td>
<td>Cohort</td>
<td>All specialties of all grades</td>
<td>1998-2006</td>
<td>Online and published sources and databases</td>
<td>Criminal and administrative cases involving controlled substances</td>
<td>$X^2 P &lt;0.001$</td>
</tr>
<tr>
<td>Hickson et al. [218], 2002 (USA)</td>
<td>Cohort</td>
<td>Majority of specialties excluding residents</td>
<td>1992-1998</td>
<td>Patient Advocates Office and Office of Insurance and Risk Management</td>
<td>Malpractice (at least one lawsuit)</td>
<td>$X^2 P &lt;0.001$</td>
</tr>
<tr>
<td>Khaliq et al. [193], 2005 (USA)</td>
<td>Cohort</td>
<td>All specialties of all grades</td>
<td>2001</td>
<td>Medical regulatory authority (State Medical Board)</td>
<td>Disciplinary action</td>
<td>Univariate and multivariate analysis adjusted for ethnicity, board certification, international education, specialty</td>
</tr>
<tr>
<td>First author, year (country)</td>
<td>Study design</td>
<td>Description of study population</td>
<td>Years data collected</td>
<td>Data source</td>
<td>Outcomes assessed</td>
<td>Statistical test</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------------</td>
<td>---------------------------------</td>
<td>----------------------</td>
<td>-------------</td>
<td>-------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Kohatsu et al. [194], 2004 (USA)</td>
<td>Case-control</td>
<td>All specialties of all grades</td>
<td>1998-2001</td>
<td>Medical regulatory authority (State Medical Board) and American Medical Association e-Physician Profiles system</td>
<td>Disciplinary action</td>
<td>Univariate and multivariate analysis adjusted for age, board certification, international education, specialty</td>
</tr>
<tr>
<td>Morrison et al. [195], 1998 (USA)</td>
<td>Case-control</td>
<td>All specialties of all grades</td>
<td>1995-1997</td>
<td>Medical regulatory authority (State Medical Board) and Directory of Physicians in the United States</td>
<td>Disciplinary action</td>
<td>Univariate analysis</td>
</tr>
<tr>
<td>Nash et al. [226], 2009 (Australia)</td>
<td>Cross-sectional</td>
<td>All specialties of all grades</td>
<td>2007</td>
<td>Postal questionnaire</td>
<td>Medico-legal matter</td>
<td>Univariate and multivariate analysis adjusted for age, marital status, specialty, international education, solo practice, hours worked per week, peer review in past 12 months, CME requirements, teaching role, AUDIT score, GHQ score</td>
</tr>
<tr>
<td>Nash et al. [225], 2009 (Australia)</td>
<td>Cross-sectional</td>
<td>GPs; grades of doctors not stated</td>
<td>2006</td>
<td>Postal questionnaire and insurance company</td>
<td>Medico-legal matter</td>
<td>$\chi^2 P &lt; 0.001$</td>
</tr>
<tr>
<td>Pande et al. [219], 2013 (USA)</td>
<td>Cohort</td>
<td>All specialties of all grades</td>
<td>2000-2011</td>
<td>Office of the Inspector General of the US Department of Health and Human Services</td>
<td>Criminal case (convicted of Medicare and Medicaid fraud)</td>
<td>Proportion</td>
</tr>
<tr>
<td>First author, year (country)</td>
<td>Study design</td>
<td>Description of study population</td>
<td>Years data collected</td>
<td>Data source</td>
<td>Outcomes assessed</td>
<td>Statistical test</td>
</tr>
<tr>
<td>------------------------------</td>
<td>--------------</td>
<td>---------------------------------</td>
<td>----------------------</td>
<td>-------------</td>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Papadakis et al. [58], 2004 (USA)</td>
<td>Case-control</td>
<td>All specialties of all grades</td>
<td>1990-2000</td>
<td>Medical regulatory authority (State Medical Board)</td>
<td>Disciplinary action</td>
<td>Univariate and multivariate analysis adjusted for undergraduate GPA, MCAT score, did not pass medical school course, professionalism severity ranking</td>
</tr>
<tr>
<td>Papadakis et al. [59], 2005 (USA)</td>
<td>Case-control</td>
<td>All specialties of all grades</td>
<td>1990-2003</td>
<td>Medical regulatory authorities (Federation of State Medical Boards)</td>
<td>Disciplinary action</td>
<td>Univariate and multivariate analysis adjusted for MCAT score, number of medical school courses not passed, unprofessional behaviour in medical school</td>
</tr>
<tr>
<td>Papadakis et al. [66]</td>
<td>Cohort</td>
<td>Internal medicine residents</td>
<td>2000-2006</td>
<td>American Board of Internal Medicine</td>
<td>Disciplinary action</td>
<td>Univariate and multivariate analysis adjusted for performance during residency, international education, no subspecialty certification</td>
</tr>
<tr>
<td>St George [220], 2003 (New Zealand)</td>
<td>Cohort</td>
<td>All specialties of all grades</td>
<td>1996-2002</td>
<td>Medical regulatory authority (Medical Council)</td>
<td>Referral to medical regulatory body</td>
<td>Proportion</td>
</tr>
<tr>
<td>Tamblyn et al. [221], 2007 (Canada)</td>
<td>Cohort</td>
<td>All specialties of all grades</td>
<td>1993-1996</td>
<td>Medical regulatory authorities</td>
<td>Referral to medical regulatory body</td>
<td>Univariate and multivariate analysis adjusted for examination score, international education, specialty, practice location</td>
</tr>
<tr>
<td>First author, year (country)</td>
<td>Study design</td>
<td>Description of study population</td>
<td>Years data collected</td>
<td>Data source</td>
<td>Outcomes assessed</td>
<td>Statistical test</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------------</td>
<td>---------------------------------</td>
<td>----------------------</td>
<td>-------------</td>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Taragin et al. [203], 1992 (USA)</td>
<td>Cohort</td>
<td>All specialties, excluding &lt;2 years of observations</td>
<td>1977-1987</td>
<td>Insurance company</td>
<td>Malpractice claims</td>
<td>Multivariate analysis adjusted for medical degree type, international education, board certification</td>
</tr>
<tr>
<td>Unwin et al. [9], 2014 (UK)</td>
<td>Cross-sectional</td>
<td>All specialties of all grades</td>
<td>2013</td>
<td>Medical regulatory authority (Medical Council)</td>
<td>Disciplinary action</td>
<td>Univariate and multivariate analysis adjusted for years since qualification, international education, specialty</td>
</tr>
<tr>
<td>Wakeford [5], 2011 (UK)</td>
<td>Cross-sectional</td>
<td>All specialties of all grades</td>
<td>2011</td>
<td>Medical regulatory authority (Medical Council)</td>
<td>Disciplinary action</td>
<td>$\chi^2 P &lt; 0.001$</td>
</tr>
<tr>
<td>Weisman et al. [227], 1988 (USA)</td>
<td>Cross-sectional</td>
<td>Obstetricians and gynaecologists excluding residents</td>
<td>1984</td>
<td>Postal questionnaire and telephone survey</td>
<td>Malpractice litigation</td>
<td>Multivariate analysis adjusted for practice type and location, years since residency, board certification, work type, patient demographics, international education</td>
</tr>
<tr>
<td>Weycker et al. [222], 2000 (USA)</td>
<td>Cohort</td>
<td>All specialties of all grades</td>
<td>1980-1989</td>
<td>Insurance company and American Medical Association Physician Masterfiles</td>
<td>Malpractice claims</td>
<td>Multivariate analysis adjusted for prior claims, educational characteristics, demographic characteristics, practice characteristics</td>
</tr>
<tr>
<td>Wu et al. [228], 2009 (Taiwan)</td>
<td>Cross-sectional</td>
<td>All specialties of all grades</td>
<td>1991 and 2005</td>
<td>Postal questionnaire</td>
<td>Malpractice claims</td>
<td>Multivariate analysis adjusted for age, specialty</td>
</tr>
<tr>
<td>First author, year (country)</td>
<td>Study design</td>
<td>Description of study population</td>
<td>Years data collected</td>
<td>Data source</td>
<td>Outcomes assessed</td>
<td>Statistical test</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------------</td>
<td>---------------------------------</td>
<td>---------------------</td>
<td>------------</td>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Yates et al. [74], 2010 (UK)</td>
<td>Case-control</td>
<td>All specialties of all grades</td>
<td>1999-2004</td>
<td>Medical regulatory authority (Medical Council)</td>
<td>Disciplinary action</td>
<td>Univariate and multivariate analysis adjusted for social class, failed exams in early/preclinical course</td>
</tr>
</tbody>
</table>

*Excluded dermatologists and physical medicine doctors*

*Excluded pathologists, radiologists, anaesthesiologists, emergency medicine doctors, and those doctors in administrative and research positions*
4.4.1.1 Disciplinary action by a medical regulatory body

Disciplinary action was measured by 15/32 studies ([211, 189, 213, 229, 191-195, 66, 58, 59, 9, 5, 74]. The majority were case-control studies (7 studies) [229, 191, 194, 195, 58, 59, 74] and were conducted in the USA (8 studies) [229, 191, 193-195, 66, 58, 59, 74]. The majority of the studies examined doctors from all specialties (11 studies) [189, 192-195, 58, 59, 9, 5, 74] and all but two of the studies examined doctors from all grades [213, 66]. All 15 of the studies obtained the outcome from data sources that had collected data prior to the studies.

Overall, 12/15 studies found that female doctors were less likely to be subject to disciplinary action [211, 189, 229, 191-195, 66, 9, 5, 74]. In 10 of those studies, the gender difference was statistically significant (p<0.05) [229, 191, 193-195, 66, 9, 5, 74], whereas the remaining two did not use inferential statistics [59, 189]. Finally, 3/15 studies found no statistically significant effect of gender [213, 58, 59].

4.4.1.2 Malpractice experience

Malpractice experience was reported by 9/32 studies [223, 212, 224, 216, 218, 203, 222, 227, 228]. A cohort study design was used in 5 of these studies [212, 216, 218, 203, 222] and a cross-sectional study design was used in the remaining studies [223, 224, 227, 228]. All except one of the studies were conducted in the USA, which was conducted in Taiwan [228]. The types of specialties included in the studies varied. The majority of the studies excluded recently qualified doctors (5 studies) [224, 216, 218, 203, 227].

6/9 studies found female doctors were significantly less likely to have malpractice experience than male doctors (p≤0.05) [223, 224, 216, 218, 203, 222]. One study examined doctors at two time points, and found female doctors were less likely to have malpractice experience in 1991 (p=0.043) but not in 2005 (P =0.168) [228]. The remaining 2/9 studies found no statistically
significant association between gender and malpractice experience [212, 227].

4.4.1.3 Referral to a medical regulatory body

Overall, 2/32 studies examined referrals to a medical regulatory body [220, 221]. These cohort studies were conducted in Canada [221] and New Zealand [220], and examined doctors from all specialties and grades.

Both studies found female doctors were less likely to be referred to a medical regulatory body, with one demonstrating a highly statistically significant association (p<0.001).

4.4.1.4 Medico-legal matter with a medical defence organisation

In total, 2/32 studies examined medico-legal matters with a medical defence organisation [225, 226]. Both of these studies were cross-sectional studies and conducted in Australia. Only one of these studies examined doctors from all specialties and grades [226].

The association between female gender and medico-legal matter was highly statistically significant (p≤0.005).

4.4.1.5 Criminal cases

Criminal cases were examined in 2/32 studies [217, 219]; both were cohort studies conducted in the USA. They examined doctors from all specialties and of all grades. Both found that female doctors were significantly less likely to experience criminal charges (p<0.05).

4.4.1.6 Complaint to a healthcare complaint body

Complaints received by a healthcare complaints body were examined by 2/32 studies [214, 215]; both were cohort studies. One study was conducted in Australia [214], and the other in England [215]. They both included doctors from all specialties and of all grades.
Both studies found that female doctors were less likely to receive complaints. One found a statistically significant effect (p<0.05), the other did not provide any inferential statistics.

4.4.1.7 Summary of findings

The majority of these studies were cohort studies (15 studies) [211, 189, 214, 215, 192, 216-218, 193, 219, 66, 220, 221, 203, 222], they were conducted in the USA (14 studies) [223, 229, 224, 191, 216-218, 193-195, 219, 66, 203, 222], and they included doctors from all specialties (19 studies) [189, 214, 229, 215, 192, 217, 193-195, 226, 219-221, 203, 9, 5, 222, 228, 74] and included doctors of all grades (21 studies) [211, 189, 223, 214, 229, 191, 215, 192, 217, 193-195, 226, 219-221, 9, 5, 222, 228, 74].

Overall, 27/32 studies found that female doctors were less likely to have had experienced at least one medico-legal action [211, 189, 223, 214, 229, 224, 191, 215, 192, 216-218, 193-195, 225, 226, 219-221, 203, 9, 5, 222, 74], although 4/27 studies did not calculate inferential statistics and did not provide sufficient data to enable the calculation of any effect size [211, 192, 220, 214]. Of the studies that provided an effect size or where it was possible to calculate an effect size from the data reported, 22/23 demonstrated that female doctors were statistically significantly less likely to have had experienced a medico-legal action (p≤0.05) [189, 223, 229, 224, 191, 215-218, 193-195, 225, 226, 219, 66, 221, 203, 9, 5, 222, 228, 74]. The remaining study examined doctors at two separate time intervals finding a significant association at the early time point only [228].

Finally, 5/32 studies found no statistically significant difference between male and female doctors [212, 213, 58, 59, 227].

4.4.2 Assessment of methodological quality

Methodological quality was assessed using the NOS [210] (see Table 4.4).
4.4.2.1 Cohort studies

Overall, the cohort studies did not show major problems of selection bias. I felt they included a truly representative cohort of working doctors, and they obtained data on both exposure and outcome from existing data records and sources. 16/17 of the studies had complete follow-up, with all subjects accounted for. 1/17 of the studies had a follow-up of 70%, which I felt was reasonable and unlikely to cause significant bias [212]. The main area of weakness for the cohort studies was failing to control for potential confounders, with 6/17 not adjusting for potential confounders [211, 189, 215, 217, 219, 220].
Table 4.4 Methodological quality assessment using the Newcastle-Ottawa scale

<table>
<thead>
<tr>
<th>Cohort studies</th>
<th>Selection</th>
<th>Comparability</th>
<th>Outcome</th>
<th>Follow-up long enough for outcomes to occur</th>
<th>Adequacy of follow up of cohorts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Representativeness of exposed cohort</td>
<td>Representativeness of non exposed cohort</td>
<td>Ascertainment of exposure</td>
<td>Demonstration outcome not present at start of study</td>
<td>Comparability of cohorts</td>
</tr>
<tr>
<td>Alam et al. [189] (2011)</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
</tr>
<tr>
<td>Alam et al. [211] (2013)</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
</tr>
<tr>
<td>Baldwin et al. [212] (1991)</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
</tr>
<tr>
<td>Bismark et al. [214] (2013)</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
</tr>
<tr>
<td>Donaldson et al. [215] (2014)</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
</tr>
<tr>
<td>Elkin et al. [192] (2011)</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
</tr>
<tr>
<td>Ely et al. [216] (1991)</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
</tr>
<tr>
<td>Goldenbaum et al. [217] (2008)</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
</tr>
<tr>
<td>Hickson et al. [218] (2002)</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
</tr>
<tr>
<td>Khaliq et al. [193] (2005)</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
</tr>
<tr>
<td>Cohort studies</td>
<td>Selection</td>
<td>Representativeness of exposed cohort</td>
<td>Selection of non exposed cohort</td>
<td>Ascertainment of exposure</td>
<td>Demonstration outcome not present at start of study</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------</td>
<td>-------------------------------------</td>
<td>-------------------------------</td>
<td>--------------------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>Pande et al. [219] (2013)</td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Papadakis et al. [66] (2008)</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td>St George [220] (2003)</td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Tamblyn et al. [221] (2007)</td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td>Taragin et al. [203] (1992)</td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Weycker et al. [222] (2000)</td>
<td>*</td>
<td>*</td>
<td></td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td>Case-control and Cross-sectional studies</td>
<td>Selection</td>
<td>Case definition</td>
<td>Representativeness of cases</td>
<td>Selection of controls</td>
<td>Definition controls</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>-----------</td>
<td>-----------------</td>
<td>-----------------------------</td>
<td>-----------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Cardarelli et al. [229] (2004)</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
</tr>
<tr>
<td>Kohatsu et al. [194] (2004)</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
</tr>
<tr>
<td>Morrison et al. [195] (1998)</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
</tr>
<tr>
<td>Papadakis et al. [59] (2004)</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
</tr>
<tr>
<td>Papadakis et al. [59] (2005)</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
</tr>
<tr>
<td>Yates et al. [74] (2010)</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
</tr>
<tr>
<td>Balch et al. [223] (2011)</td>
<td></td>
<td>★</td>
<td>★</td>
<td>★</td>
<td></td>
</tr>
<tr>
<td>Chauhan et al. [224] (2005)</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
</tr>
<tr>
<td>Nash et al. [226] (2009)</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
</tr>
<tr>
<td>Nash et al. [225] (2009)</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
</tr>
<tr>
<td>Unwin et al. [9] (2014)</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
</tr>
<tr>
<td>Case-control and Cross-sectional studies</td>
<td>Selection</td>
<td>Representativeness of cases</td>
<td>Selection of controls</td>
<td>Definition of controls</td>
<td>Comparability of cases and controls</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>-----------</td>
<td>----------------------------</td>
<td>----------------------</td>
<td>-----------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weisman et al. [227] (1988)</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wu et al. [228] (2009)</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.4.2.2 Case-control and cross-sectional studies

The case definition was made by independent validation in 10/15 studies [229, 191, 194, 195, 225, 58, 59, 9, 5, 74]. The remaining five studies obtained their case definition through self-reports [223, 224, 226-228]. 12/15 studies clearly had a representative series of cases [229, 224, 191, 194, 195, 58, 59, 9, 5, 227, 228, 74]. 3/15 studies had the potential for selection bias because they had <50% response rate to study surveys [223, 225, 226]. All the studies selected controls drawn from the same community as cases. 13/15 studies adjusted for at least two important potential confounding factors [223, 229, 224, 191, 194, 195, 226, 58, 59, 9, 227, 228, 74]. The exposure was ascertained by secure records in 9/15 studies [229, 191, 194, 195, 58, 59, 9, 5, 74]. In the remaining 6/15 studies the exposure was ascertained for cases and controls. The response rate was the same in both cases and controls in 5/15 studies [229, 195, 59, 9, 5]. In 4/15 studies the non-respondents were described [191, 225, 226, 228], and in 6/15 studies the response rate was different between the two groups, and there was no description [223, 224, 194, 58, 227, 74]. See Table 4.4.

4.4.2.3 Summary considerations on study quality

Overall, the cohort and case-control studies did not show major problems of selection bias. The main area of weakness for the cohort studies was failing to control for potential confounders. However, because I have chosen to prioritise crude results over adjusted results in the meta-analysis (I have made this decision because the factors controlled for were not consistent throughout the different studies) I felt this was of little consequence.

In the case-control studies, the area of weaknesses centred around non-response rate, with 4/7 studies having different rates of response between the controls and the cases [191, 194, 58, 74], and only one in these four studies describing the non-respondents [191].

The cross-sectional studies varied in methodological quality. The potential for selection bias was present in 3/8 studies [223, 225, 226] and 2/8 studies
did not adjust for potential confounders [225, 5], but as discussed earlier, I did not view this as a major limitation. Ascertainment of exposure and non-response rate was an area of concern in 6/8 studies [223-228].

4.4.3 Meta-analysis

Of the 32 studies that were included in the systematic review, 19 reported data that allowed the calculation of a measure of effect and were included in the meta-analysis [189, 223, 212, 229, 224, 215, 217, 193, 194, 225, 226, 219, 66, 58, 59, 203, 9, 228, 74]. The meta-analysis included 3,794,486 study participants (both cases and non-cases), of which 20,666 are cases of medico-legal action.

The studies were published from 1991 to 2014, with the majority of the studies being published in 2004, 2005, and 2009. 12/19 studies were conducted in the USA [223, 212, 229, 224, 215, 217, 193, 194, 219, 66, 58, 59, 203], 8/19 were cohort studies [189, 212, 215, 217, 193, 219, 66, 203], and 9/19 defined the primary outcome as disciplinary action by a medical regulatory body [189, 229, 193, 194, 66, 58, 59, 9, 74]. 14/19 studies included doctors from all specialties [189, 229, 215, 217, 193, 194, 226, 219, 58, 59, 203, 9, 228, 74] and 15/19 studies included doctors from all grades [189, 223, 212, 229, 215, 217, 193, 194, 226, 219, 58, 59, 9, 228, 74]. The largest study included over one million doctors [215], the smallest study examined over 250 doctors [58].

4.4.3.1 Summary effect estimates

A random-effects model found a pooled odds ratio of 0.41 (95% CI: 0.34 to 0.49). All 19 studies reported that female doctors were less likely to experience a medico-legal action than male doctors (range of odds ratios 0.61 to 0.98) [189, 223, 229, 224, 215, 217, 193, 194, 225, 226, 219, 66, 203, 9, 228, 74]; in 3/19 studies the difference was not statistically significant [212, 58, 59]. No studies showed that men were less likely to experience a medico-legal action than women (Figure 4.2). A high degree of
heterogeneity was present ($Q=233.25$, d.f.=18, $p<0.001$; $I^2=92.3\%$) which was due to differences in size rather than the direction of effect.

Figure 4.2 Results of meta-analysis of 19 reports on the association between doctors' gender and experience of medico-legal action using both fixed-effects and random-effects modelling

<table>
<thead>
<tr>
<th>Study ID</th>
<th>ES (95% CI)</th>
<th>Weight (D+L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taragin 1992</td>
<td>0.37 (0.31, 0.43)</td>
<td>6.31</td>
</tr>
<tr>
<td>Baldwin 1991</td>
<td>0.98 (0.46, 2.07)</td>
<td>3.05</td>
</tr>
<tr>
<td>Cardarelli 2004</td>
<td>0.16 (0.13, 0.21)</td>
<td>5.96</td>
</tr>
<tr>
<td>Papadakis 2004</td>
<td>0.57 (0.25, 1.30)</td>
<td>2.76</td>
</tr>
<tr>
<td>Khalil 2005</td>
<td>0.35 (0.23, 0.53)</td>
<td>4.82</td>
</tr>
<tr>
<td>Kohatsu 2004</td>
<td>0.32 (0.25, 0.40)</td>
<td>5.91</td>
</tr>
<tr>
<td>Papadakis 2005</td>
<td>0.97 (0.71, 1.33)</td>
<td>5.51</td>
</tr>
<tr>
<td>Yates 2010</td>
<td>0.18 (0.07, 0.46)</td>
<td>2.28</td>
</tr>
<tr>
<td>Chauhan 2005</td>
<td>0.56 (0.35, 0.88)</td>
<td>4.60</td>
</tr>
<tr>
<td>Wu 2009</td>
<td>0.46 (0.31, 0.68)</td>
<td>5.04</td>
</tr>
<tr>
<td>Goldenbaum 2008</td>
<td>0.32 (0.25, 0.41)</td>
<td>5.93</td>
</tr>
<tr>
<td>Nash 2009</td>
<td>0.50 (0.36, 0.72)</td>
<td>5.28</td>
</tr>
<tr>
<td>Papadakis 2008</td>
<td>0.46 (0.40, 0.58)</td>
<td>6.18</td>
</tr>
<tr>
<td>Nash (MJA 2009)</td>
<td>0.56 (0.43, 0.72)</td>
<td>5.86</td>
</tr>
<tr>
<td>Aam 2011</td>
<td>0.18 (0.14, 0.25)</td>
<td>5.64</td>
</tr>
<tr>
<td>Batch 2011</td>
<td>0.81 (0.69, 0.95)</td>
<td>6.31</td>
</tr>
<tr>
<td>Pande 2013</td>
<td>0.35 (0.26, 0.49)</td>
<td>5.44</td>
</tr>
<tr>
<td>Donaldson 2014</td>
<td>0.36 (0.36, 0.41)</td>
<td>6.59</td>
</tr>
<tr>
<td>Urwin 2014</td>
<td>0.36 (0.32, 0.38)</td>
<td>6.52</td>
</tr>
<tr>
<td>D+L Overall ($I^2$ = 92.3%, $p = 0.000$)</td>
<td>0.41 (0.34, 0.49)</td>
<td>100.00</td>
</tr>
<tr>
<td>I-V Overall</td>
<td>0.39 (0.38, 0.41)</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: Weights are from random-effects analysis

4.4.3.2 Subgroup analyses

4.4.3.2.1 Study design (cross-sectional, cohort, case-control)

There was a significant overall effect of gender in each stratum, with female doctors having decreased odds of medico-legal action (cohort OR: 0.36, 95% CI: 0.30 to 0.43; case-control studies OR: 0.35, 95% CI: 0.17 to 0.75; cross-sectional studies OR: 0.52, 95% CI: 0.36 to 0.76). There was high heterogeneity within each stratum (cohort studies $I^2$: 81.9%; case-control studies $I^2$: 95.3%; cross-sectional $I^2$: 94.1%).
4.4.3.2.2 Continent (North America, Europe, Asia and Australia)

There was a significant overall effect of gender in each stratum, with female doctors having decreased odds of medico-legal action (North America OR: 0.42, 95% CI: 0.30 to 0.57; Asia and Australia OR: 0.52, 95% CI: 0.43 to 0.62; Europe OR: 0.36, 95% CI: 0.33 to 0.40). There was high heterogeneity in the North America stratum ($I^2$: 94.5%). The degree of heterogeneity was small for both Europe ($I^2$: 55.1%) and Asia and Australia ($I^2$: 0.0%), but these strata only contained three studies each and therefore I am not able to accurately estimate heterogeneity.

4.4.3.2.3 Outcome (disciplinary action, malpractice, other)

Female doctors were significantly less likely to have experienced a medico-legal action in all three strata (disciplinary action OR: 0.34, 95% CI: 0.24 to 0.47; malpractice OR: 0.58, 95% CI: 0.37 to 0.90; other OR: 0.41, 95% CI: 0.34 to 0.49). There was considerable heterogeneity present in the strata disciplinary action ($I^2$: 93.2%) and malpractice experience ($I^2$: 92.2%).

4.4.3.2.4 Year from which data was collected

The most recent year in which the data was collected by the researchers was examined. In one study the year in which the data was collected was not reported, and therefore we used the year the study was published [224].

Female doctors were significantly less likely to have experienced a medico-legal action in all strata, apart from in 1985-89, where the result was not statistically significant (1985-89 OR: 0.56, 95% CI: 0.21 to 1.45; 1995-99 OR: 0.16, 95% CI: 0.13 to 0.21; 2000-04 OR: 0.42, 95% CI: 0.23 to 0.76; 2005-09 OR: 0.41, 95% CI: 0.30 to 0.55; 2010-14 OR: 0.44, 95% CI: 0.32 to 0.62). There was substantial heterogeneity throughout all strata.

4.4.3.2.5 Specialty and Grade

Although I had planned to conduct subgroup analyses examining grade and specialty, I decided not to proceed because the data available did not allow the formation of meaningful groups.
4.4.3.2.6 Methodological quality assessment

The overall NOS star rating for each study was reviewed. Female doctors were found to be significantly less likely to have experienced a medico-legal action, regardless of the assessed quality of the study (8-9 stars OR: 0.39, 95% CI: 0.30 to 0.52; 6-7 stars OR: 0.35, 95% CI: 0.27 to 0.44; 4-5 stars OR: 0.63, 95% CI: 0.46 to 0.86). There was heterogeneity present throughout all strata.

4.4.3.3 Bias

I used a funnel plot to assess possible bias (Figure 4.3). The distribution of the studies within the funnel plot did appear somewhat random with eight of the studies appearing outside of the funnel. This suggests heterogeneity in the studies. There was no sign of a publication bias relating to the significance of the effect found. There was a sparsity of studies at the bottom of the plot, with the majority of the studies clustering towards the top of the graph. The relative absence of studies towards the bottom of the graph could indicate an absence of smaller studies. This may reflect the nature of the data typically available to allow such studies.
4.5 Discussion

4.5.1 Summary of main results

4.5.1.1 Systematic review

The findings of the systematic review are based on 32 observational studies [211, 189, 223, 212-214, 229, 224, 191, 215, 192, 216-218, 193-195, 225, 226, 219, 66, 58, 59, 220, 221, 203, 9, 5, 222, 228, 74].

Overall the systematic review results strongly suggest that female doctors are at reduced risk of having an experience of medico-legal action when compared to male doctors, with 26/32 studies finding female doctors were less likely to experience medico-legal action [211, 189, 223, 214, 229, 224, 191, 215, 192, 216-218, 193-195, 225, 226, 219, 66, 220, 221, 203, 9, 5, 222, 74], of which 22 reported a statistically significant effect [189, 223, 229, 224, 191, 215-218, 193-195, 225, 226, 219, 66, 221, 203, 9, 5, 222, 74].
Of the remaining 6/32 studies, 5 reported that there was no evidence to suggest that there was a difference between the genders in having had experience of a medico-legal action [212, 213, 58, 59, 227]. 1/32 studies analysed the data at two separate time intervals (in 1991 and then in 2005) and found that female doctors were statistically less likely to have an experience of a medico-legal action in 1991, but that there was no difference between the genders in 2005 [228]. The results of this study suggest that the effect of gender on experiencing a medico-legal action could be affected by time. However, this study was assessed to be of moderate methodological quality (6 stars) and they had a response rate of 10.3% for the 2005 survey, which could suggest the potential for bias. The study also examines only one country.

None of the studies found that women were more likely to experience medico-legal action than men.

4.5.1.2 Meta-analysis

The meta-analysis results are based on the evidence of 19 observational studies and showed that female doctors had less than half the odds (pooled OR: 0.41, 95% CI: 0.34 to 0.49) of medico-legal action compared to male doctors. There was significant heterogeneity in the meta-analysis but this was not due to differences in the direction of effects – no studies found that women were more likely than men to experience medico-legal action.

The size of the effect of gender on experience of medico-legal action remained roughly constant in all subgroup analyses, suggesting that the effect of gender is not influenced by the study design, the country the doctor is employed in, the outcome definition, or NOS rating. The only exception was when examining the year in which the data was collected; in the 1995-99 stratum there was only one study included in this stratum and it reported that men had 6 times the odds of medico-legal action [229], however overall the effect seems stable over time.
4.5.2 Overall completeness and applicability

4.5.2.1 Literature search

The literature searching was thorough, as demonstrated by the number of reports initially identified (>6,500). I did not extensively search grey literature sources due to limited resources and the vast number of reports obtained with the search methods used. It is possible that smaller studies, or studies that did not demonstrate a gender effect, may have been overlooked. Another limitation was the exclusion of studies whose abstract and/or full-text were not available in English or French – seven studies judged to be potentially eligible based on their titles were excluded for this reason [230-236]. The exclusion of studies not available in English may partly explain why the majority of the studies included in this report are from English-speaking, high-income countries. It is possible that literature from non-English speaking countries demonstrate a different size of effect of doctors’ gender on experience of medico-legal action, and as such the results of this report may not be applicable to non-English speaking countries.

4.5.2.2 Study selection

All 32 studies included in the systematic review and all 19 studies included in the meta-analysis are observational studies. No randomised controlled studies were retrieved, which is unsurprising given the nature of the research question – it is not possible to randomly allocate doctors to be male or female.

In my systematic review, I was able to capture >40,000 cases of medico-legal action against doctors, capturing >20,500 medico-legal action cases in the meta-analysis. These large numbers of cases allowed me to draw meaningful conclusions from my results. The majority of the studies attempted to collect data that is applicable to the wider population (country-wide, state-wide etc.), a likely reflection that a doctor experiencing a medico-legal action is a relatively rare outcome, and therefore large studies are required to attempt to capture as many cases as possible. Capturing data from the wider population does mean that the results are more likely to be
generalisable. My reported studies covered eight different countries over four continents. I have demonstrated that, when stratifying the data by continent, the pooled results for each stratum was relatively stable, with female doctors having approximately half the odds of experiencing a medico-legal action. There was variability in the effect size in the North America stratum, though the direction of the effect remained consistent.

It is worth mentioning that, of the three studies included in the Asia and Australia stratum, two were from Australia and one was from Taiwan, and although the heterogeneity was small in this stratum, it did only include three studies. Within the Europe stratum, all three studies were from the UK, and therefore the stratum may not be an accurate reflection on Europe. The total number of studies and the limited range of countries from where the studies were from both in the Asia and Australia stratum and in the Europe stratum highlight the limitation of only including studies which were published in English or French.

The outcome definition used by the individual studies varied in severity. I chose to use a variety of terms to capture the outcome, with the aim of capturing as many relevant studies as possible. Because the different outcome types varied in severity, it may not be a fair to include studies together. That said, the sub-analysis examining the data by outcome type showed that the overall effect of doctors’ gender was consistent, with female doctors having approximately one third to one half of the odds of experiencing a medico-legal action in each stratum. It is also interesting to note that the two largest strata were the outcomes ‘disciplinary action’ and ‘malpractice’, both of which have severe impact on a doctor’s professional career.

The demographics of doctors in the UK and USA have been changing, with increasing numbers of women choosing to follow medicine as a career [237, 94] (see section 1.5.3.2). My results suggest that the effect of female gender on experiencing medico-legal actions has remained fairly constant over the last 15 years despite the increasing trend of women doctors (it is not possible
to comment on the years prior to 2000 due to the small number of studies in the strata). I therefore feel one can no longer argue that female doctors are less likely to face medico-legal action because there are fewer female doctors practising. If this were the case, I would expect the effect size to increase over time, to reflect the increasing number of female doctors.

Unfortunately, I was unable to explore further whether the gender difference in medico-legal action was impacted by specialty practised. Thirteen of the studies included in the meta-analysis examined whether the likelihood of medico-legal action differed between the specialties [189, 229, 215, 217, 193, 194, 226, 219, 58, 59, 203, 9, 228]; however, the specialties most and least likely to face medico-legal action varied greatly between the studies. In the studies which controlled for the effect of specialty when examining the association between gender and medico-legal action, all but one [228] demonstrated that female doctors remained less likely to have had medico-legal experience even with specialty taken into account [229, 193, 194, 226, 9].

Other variables have been shown to be both associated with doctors’ gender and experience of medico-legal action, but have not been examined by this meta-analysis. These include number of hours worked or the number of patient encounters. Studies included in this systematic review and meta-analysis have demonstrated that female doctors work less hours than male doctors [225, 226] and see less patients than their male colleagues [203]. The number of hours worked has been shown to be associated with increased likelihood of medico-legal action in three of these studies [223, 225, 226]. Exploring how the number of hours worked or number of patient encounters differ between the genders and the effect on medico-legal action may be of interest for a future review of the literature to help towards understanding the gender difference in medico-legal action.
4.5.3 Other potential biases in the review process

Only I assessed the reports for eligibility, and this was not performed blind - this could have introduced bias; however, I did use previously agreed criteria to guide my decisions, with the aim of reducing bias. Another possible source of bias is that the outcome definitions used were not wide enough, and that there may be some culturally specific terms that I am unaware of. Finally, the assessment of the methodological quality of the studies is subjective. To reduce this source of bias, two researchers independently judged the methodological quality of the studies and the NOS checklist was used to guide and support our decisions. I also chose not to exclude any studies from the systematic review or meta-analysis based on the findings of the appraisal of the methodological quality, choosing instead to present a narrative discussion of my findings.

4.5.4 Conclusions

This is the first systematic review and meta-analysis examining the association between doctors’ gender and experience of medico-legal action. It demonstrates that female doctors are less likely to have had experience of medico-legal action when compared to male doctors. This effect was demonstrated over a number of years, across a range of study designs, across different countries, and with a wide definition of outcome types, and therefore seems robust. The demonstration of a consistent effect size, present in the main analysis, as well as in the subgroup analyses, highlights that there is likely to be a fundamental reason to explain why female doctors have less than half the odds of experiencing a medico-legal action.

More detailed information is needed to understand the reasons why female doctors are less likely to experience medico-legal action. Is it an intrinsic characteristic of doctors, such as personality traits – are female doctors more likely to display certain personality characteristics or perform certain behaviours that make them less likely to be subject to medico-legal action? I will explore whether there is an association between gender, personality type and disciplinary action in Chapter 8.
There is also evidence to suggest that men and women doctors differ in practice patterns. Women have been found to be more likely to adhere to practice guidelines and evidence-based medicine [238, 239]. Complying with guidelines is likely to be protective of disciplinary action for several reasons. Firstly, the guidance is likely to be up to date and evidence based, following such guidance is likely to appropriately support their patient’s health concern. Secondly, by adhering to the advised guidelines the doctor will be able to justify the management of their patient should a concern about the patient’s management be raised. Finally, following guidance may suggest that the doctor is less likely to take risks in terms of their patient’s care. Research from the general population has shown that men are more likely to exhibit risk taking behaviour [240]. Risk taking behaviour within the population of doctors is generally not particularly high, however, doctors practising in surgery or anaesthetics were found to take greater risks [241]. Interestingly, these specialties tend not to be as popular amongst women doctors [242], particularly surgery which remains a strongly male dominated specialty [95].

As mentioned in Chapter 3, communication style between men and women doctors has also been shown to differ with women being found to display more patient centred communication and providing more psychosocial counselling [243, 244]. Patient satisfaction has been linked to effective doctor-patient communication [245]. In Chapter 8 I will explore whether communication style mediates the association between gender and disciplinary action.

Patient outcomes have also been shown to differ between men and women doctors, with women obtaining lower mortality and readmission rates for their patients [6]. Obtaining favourable clinical outcomes for patients is also likely to reduce the risk of facing future disciplinary action.

This study robustly demonstrated that female doctors are less likely to experience medico-legal action. Recognising the gender difference in a measure of extreme poor professional performance in doctors is one of the first steps towards better understanding the complex and multi-factorial...
precipitants of doctors’ professional performance. Medical schools, medical regulatory authorities, and researchers now need to work together to try and further understand the difference between the genders that could explain the difference in experience of medico-legal action, with the aim of better supporting our doctors and improving patient safety.

In the following chapter, Chapter 5, I will address a selection of the hypotheses raised to explain the observed gender difference in professional performance. I will explore the performance of populations of doctors who completed a specialty clinical assessment for the first time. This clinical assessment judges candidates’ performance on behaviours and skills, such as demonstrating empathy and effective communication, it also assesses clinical knowledge, including formulating appropriate management plans and awareness of law and ethics. Exploring whether there is a gender difference in the overall performance at this clinical assessment will provide further insight as to whether this gender difference in performance is only observed in those doctors whose fitness to practise has been found to be impaired, or whether it is also present in other measures of performance.
Chapter 5 - Passing MRCP(UK) PACES: a cross-sectional study examining the performance of doctors by gender and country

5.1 Chapter summary

5.1.1 Background

The studies conducted in Chapter 3 and Chapter 4 have established that women doctors are less likely to face medico-legal action against their medical registration. The reasons behind this gender difference are not clear. This chapter will investigate whether there is gender difference at another measure of professional performance – postgraduate medical examination pass rates.

5.1.2 Aims

This chapter will explore the association between doctors’ gender and their performance at a large international high-stakes clinical examination: MRCP(UK) PACES. It will examine how gender differences in pass rates varied by country in which the doctor received their primary medical qualification, the country in which they took the PACES examination, and by the country in which they are registered to practise.

5.1.3 Design and setting

A cross-sectional study of all candidates who attempted PACES between October 2010 and May 2013.

5.1.4 Methods

Seven thousand, six hundred and seventy-one doctors attempted PACES between October 2010 and May 2013. The gender differences in first time pass rates were analysed, controlling for ethnicity, in thee groups: (1) UK medical graduates \((N = 3,574)\); (2) non-UK medical graduates registered with the GMC, and thus likely to be working in the UK \((N = 1,067)\); and (3) non-UK medical graduates without GMC registration and so legally unable to work or train in the UK \((N = 2,179)\).
5.1.5 Results
Female doctors were statistically significantly more likely to pass at their first attempt in all three groups, with the greatest gender effect seen in non-UK medical graduates without GMC registration (OR=1.99; 95% CI=1.65 to 2.39; p<0.0001) and the smallest in the UK graduates (OR=1.18; 95% CI=1.03 to 1.35; p=0.02).

5.1.6 Conclusion
As found in a previous format of this examination, female doctors outperformed male doctors.
5.2 Introduction

A clinical examination is a measure of the candidate’s practical clinical knowledge and skills, however performance at a clinical examination is influenced by the candidate’s theoretical knowledge [111]. It has been shown in final year medical students that good performance in a clinical examination is associated with future performance as a doctor [246], it has also recently been demonstrated that the performance of a doctor at postgraduate medical examinations is associated with risk of disciplinary action by the medical regulatory body [67]. It is therefore reasonable to suggest that clinical examination outcomes can be used as a proxy measure for professional performance.

Gender differences in doctors’ academic outcomes may be confounded by other variables such as ethnicity and country of qualification. Research from the UK, USA, Netherlands, Canada, and Australia has shown that medical students and doctors from ethnic minorities and/or doctors who obtained their primary medical qualification outside of the country in which they are practising do not perform as well as their colleagues who are white or who trained in the country they are practising in, across a range of undergraduate and postgraduate medical assessments [126, 127, 247-249, 125, 250-252]. There is also evidence to suggest that male doctors who are, or have been, registered to practise medicine in the UK are more likely to have qualified in medicine outside of the UK than their female colleagues [9] (see Chapter 3).

The current study focuses on gender differences in performance at the MRCP(UK) PACES, a standardised clinical examination in internal medicine taken by around 6,000 candidates annually in the UK and 14 other countries worldwide. A previous study found female PACES candidates in 2003-4 performed statistically significantly better than male candidates after controlling for ethnicity [126]; however this population consisted only of candidates who had graduated in the UK, precluding the possibility of exploring the interaction with country of qualification and country of sitting. In
addition, that study used data from a previous format of the PACES examination: it has changed considerably since then [160].

I aimed to establish whether the gender differences in performance were present in the latest format of this clinical examination, and whether any gender differences varied by candidates’ country of primary medical qualification, whether or not they were registered with the GMC and therefore likely to be working in the UK, and whether or not they sat the examination in the UK.

5.3 Methods

5.3.1 Study design, setting and source of data

I conducted this cross-sectional study using an international database from the Federation of Royal Colleges in the UK, which organises the MRCP(UK) internal medicine specialty examinations. Permission to use the data for research purposes was obtained from MRCP(UK). The data was provided by MRCP(UK) for the purposes of this research study, after a licence agreement was signed between UCL and the MRCP(UK). The MRCP(UK) publish their data protection statement on their website, which informs candidates how their data may be stored and used [253].

5.3.2 Membership of the Royal Colleges of physicians (UK) diploma

In the UK, those doctors who wish to enter into higher specialist training in internal medicine are required to complete a three-part examination known as the MRCP(UK) Diploma, which aims to test the knowledge and skills, and behaviour of doctors in training [254]. The MRCP(UK) Diploma consists of three parts: MRCP(UK) Part 1; MRCP(UK) Part 2 Written; and MRCP(UK) Part 2 Practical Assessment of Clinical Examination Skills (PACES). Candidates are required to successfully complete all three parts of the examination before they are able to start specialist internal medicine training in the UK [254].
Doctors in many countries outside of the UK sit the examination. In some countries (e.g. India) the MRCP(UK) qualification has status similar to local graduate training programmes. In others (e.g. Hong Kong) it forms part of a conjoint qualification. In countries where doctors may have difficulty accessing any formal training programmes, the MRCP(UK) qualification is used to benchmark a doctor's knowledge and clinical skills against an internationally recognised standard. Completion of the MRCP(UK) diploma can improve a doctor's chances of getting GMC registration and a license to practise in the UK, but anecdotal evidence suggests the majority of doctors sitting internationally do not come to the UK to work. MRCP(UK) does not collect employment status from candidates in the UK or internationally. It records registration with the GMC, but not with any other regulatory bodies. GMC registration can be used as a proxy for current or previous employment in the UK.

The current study focuses on the performance of candidates at the clinical assessment component of the Diploma, PACES. PACES is a structured and standardised assessment that was first introduced in 2001, with the aim of providing a valid and reliable assessment of physical examination and communication skills [159]. PACES is run at clinical centres across the UK and 14 other countries [255]. In 2009, the format of the examination was revised considerably with the aims of ensuring that successful candidates were competent across the range of clinical skills assessed, and that they possessed the attitudes and behaviour required of a specialist trainee in internal medicine [160]. Following a transitional phase between October 2009 and July 2010, the new format of the PACES assessment was introduced for all candidates from October 2010 [160]. Today PACES consists of five stations and a total of eight patient encounters, during which seven core clinical skills relating to communication, physical examination and diagnostic reasoning are assessed (see Figure 5.1 and Figure 5.2). Each station has two examiners who independently judge the candidate’s performance (10 examiners per PACES examination). Different skills are assessed at different stations i.e. not all stations test all skills. At each
station, each examiner scores the relevant skills as 2 = satisfactory, 1 = borderline, 0 = unsatisfactory. A candidate must achieve a passing score in each skill to pass the examination overall. See [256] for more details of the examination format and scoring.
Figure 5.1 Carousel of PACES stations (source: MRCP(UK), permission obtained to reproduce figure in September 2016)
Figure 5.2 The seven core skills assessed during PACES (source: MRCP(UK), permission obtained to reproduce figure in September 2016)

<table>
<thead>
<tr>
<th>Clinical Skill</th>
<th>Skill Descriptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Physical Examination</td>
<td>Demonstrate correct, thorough, systematic (or focused in Station 5 encounters), appropriate, fluent, and professional technique of physical examination.</td>
</tr>
<tr>
<td>B Identifying Physical Signs</td>
<td>Identify physical signs correctly, and not find physical signs that are not present.</td>
</tr>
<tr>
<td>C Clinical Communication</td>
<td>Elicit a clinical history relevant to the patient’s complaints, in a systematic, thorough (or focused in Station 5 encounters), fluent and professional manner. Explain relevant clinical information in an accurate, clear, structured, comprehensive, fluent and professional manner.</td>
</tr>
<tr>
<td>D Differential Diagnosis</td>
<td>Create a sensible differential diagnosis for a patient that the candidate has personally clinically assessed.</td>
</tr>
<tr>
<td>E Clinical Judgement</td>
<td>Select or negotiate a sensible and appropriate management plan for a patient, relative or clinical situation. Select appropriate investigations or treatments for a patient that the candidate has personally clinically assessed. Apply clinical knowledge, including knowledge of law and ethics, to the case.</td>
</tr>
<tr>
<td>F Managing Patients’ Concerns</td>
<td>Seek, detect, acknowledge and address patients’ or relatives’ concerns. Listen to a patient or relative, confirm their understanding of the matter under discussion and demonstrate empathy.</td>
</tr>
<tr>
<td>G Maintaining Patient Welfare</td>
<td>Treat a patient or relative respectfully and sensitively and in a manner that ensures their comfort, safety and dignity.</td>
</tr>
</tbody>
</table>

PANCES examiners must be registered with the GMC (or regulatory equivalent in country of practise), be registered with a licence to practise, and
be in good standing. They must also be a Collegiate Member or Fellow in good standing at one of the Royal Colleges of Physicians of the UK. Collegiate Members must have achieved the certificate of completion of training (CCT), or be on the Specialist register, and be in a substantive consultant post. Physicians who are resident outside of the UK and who wish to examine must hold Fellowship of one of the UK Royal Colleges of Physicians and be in good standing. When the examination is taken outside of the UK, one of the examiners must be from the UK, and one must be from the host country. New examiners complete 3 days of training, and once trained, they must commit to at least six cycles (30 candidates) per annum averaged over 2 years. If they cannot do that, they must refresh their examiner training.

5.3.3 Population and primary outcome

The study population included all doctors attempting PACES for the first time, at any of the official MRCP(UK) examination centres in the UK or internationally between October 2010 and May 2013. The pass standard did not change over the study period. I chose to restrict the study population to those attempting PACES examination for the first time, first attempt score being a good predictor of score at subsequent attempts [257]. The exposure of interest was the candidates’ gender, as declared by the candidate to the MRCP(UK).

5.3.4 Selection of variables

The selection of co-variates was constrained to data routinely collected by MRCP(UK). The co-variates were selected prior to any statistical analysis and selection was based on published findings that suggested they might influence and confound the association between candidates’ gender and performance at examinations [127, 248, 9, 249, 125, 126]. I collapsed the variables with multiple categories to create meaningful categories in order to increase statistical power. Included co-variates were: ethnicity (white, ethnic minority, missing); country of PMQ (UK, non-UK, missing); GMC registration
(registered, not registered); country of PACES centre (UK, non-UK); year of PACES attempt; overall outcome (pass, fail).

5.3.5 Statistical methods

I first performed bivariate analyses, examining the association between the candidates’ gender and the other categorical variables; and between PACES pass rates and other categorical variables. I then completed multivariate analyses using binary logistic regression models to examine the association between candidates’ gender and overall PACES pass rates, controlling for the other variables.

5.3.6 Subgroup analyses

My initial logistic regression model included candidates’ gender and pass rates, ethnicity, primary medical qualification, GMC registration, and PACES examination centre location. Given the correlations between these variables and the risk of multicollinearity, I then decided to perform subgroups analyses to remove any possible correlation from the models. I divided the study population into three groups, representing the three broad groups of candidates who choose to complete PACES:

i) Candidates who had obtained their primary medical qualification from a university in the UK (UK medical graduates). These doctors are predominantly working in UK training posts.

ii) Candidates who obtained their primary medical qualification from an institution outside of the UK (non-UK medical graduates), and who are registered with the GMC. This population is likely to represent doctors working as doctors in the UK, some of whom will be in training posts. These doctors are likely to have had clinical experience both abroad and in the UK.

iii) Candidates who obtained their primary medical qualification from an institution outside of the UK (non-UK medical graduates), and who were not registered with the GMC. This population of doctors are currently unable to practise in the UK, and are therefore likely
to have had the majority of their clinical experience, employment and training outside the UK.

I followed guidelines set out by the STROBE statement to guide this study [199].

5.4 Results

5.4.1 Descriptive analyses

Seven thousand, six hundred and seventy-one candidates attempted PACES for the first time between October 2010 and May 2013. One candidate was excluded from all further analyses as they did not declare their gender. Of the remaining candidates 53% were men, 54% were from an ethnic minority background (11% missing ethnicity data); 52% were UK medical graduates (<1% missing primary medical qualification data); 66% were registered with the GMC to practise in the UK (0% missing GMC registration data); and 77% sat PACES in a UK-based examination centre (0% missing PACES centre data); see Table 5.1.

5.4.2 Bivariate associations between candidate gender and other variables

Table 5.1 shows the distribution of each variable by gender of the candidates. Male candidates were statistically significantly more likely to be from an ethnic minority background, more likely to be a non-UK medical graduate, less likely to be registered with the GMC, and more likely to sit PACES at a non-UK centre (all p<0.001).
Table 5.1 Distribution of variables by gender of the candidates (N=7670)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Male N=4026 (% of males)</th>
<th>Female N=3644 (% of females)</th>
<th>Statistical significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passed PACES at first attempt</td>
<td></td>
<td></td>
<td>$\chi^2 (1)=144, p&lt;0.001$</td>
</tr>
<tr>
<td>Yes</td>
<td>1681 (42)</td>
<td>2055 (56)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>2345 (58)</td>
<td>1589 (44)</td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td>$\chi^2 (1)=241, p&lt;0.001$</td>
</tr>
<tr>
<td>White</td>
<td>1101 (27)</td>
<td>1602 (44)</td>
<td></td>
</tr>
<tr>
<td>Ethnic minority</td>
<td>2466 (61)</td>
<td>1653 (45)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>459 (11)</td>
<td>389 (11)</td>
<td></td>
</tr>
<tr>
<td>World region where PMQ received</td>
<td></td>
<td></td>
<td>$\chi^2 (1)=302, p&lt;0.001$</td>
</tr>
<tr>
<td>UK</td>
<td>1692 (42)</td>
<td>2254 (62)</td>
<td></td>
</tr>
<tr>
<td>Outside of UK</td>
<td>2332 (58)</td>
<td>1390 (38)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>2 (&lt;1)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Registered with GMC</td>
<td></td>
<td></td>
<td>$\chi^2 (1)=229, p&lt;0.001$</td>
</tr>
<tr>
<td>Yes</td>
<td>2352 (58)</td>
<td>2725 (75)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1674 (42)</td>
<td>919 (25)</td>
<td></td>
</tr>
<tr>
<td>Country of examination centre</td>
<td></td>
<td></td>
<td>$\chi^2 (1)=137, p&lt;0.001$</td>
</tr>
<tr>
<td>UK</td>
<td>2862 (71)</td>
<td>3004 (77)</td>
<td></td>
</tr>
<tr>
<td>Outside of UK</td>
<td>1164 (29)</td>
<td>640 (18)</td>
<td></td>
</tr>
</tbody>
</table>

5.4.3 Bivariate associations between PACES pass rates and other variables

For the remainder of the descriptive analyses, candidates with one or more variable missing were excluded, leaving a total of 6,820 candidates. Of those, just under half (49%) passed PACES at their first attempt. Passing candidates were statistically more likely to be female rather than male [56% vs. 42%; $\chi^2(1)=144, p<0.001$]; white rather than from an ethnic minority [65% vs. 38%; $\chi^2(1)=463, p<0.001$]; be a UK rather than a non-UK medical graduate [64% vs. 32%; $\chi^2(1)=695, p<0.001$]; be registered with the GMC rather than not [55% vs. 35%; $\chi^2(1)=232, p<0.001$]; and completed the examination at a UK examination centre rather than a centre outside the UK [51% vs. 40%; $\chi^2(1)=65, p<0.001$].

In summary, ethnicity, primary medical qualification world region, GMC registration, and location of the examination centre, were associated with passing PACES at first attempt and candidate gender. I therefore
considered these four variables as confounders of the association between gender and passing PACES at first attempt.

5.4.4 Candidates with missing data

I compared the PACES performance of the 851 candidates (11% of the study population) who were missing data for at least one variable, to the PACES performance of the 6,820 candidates who no missing data. I found no evidence of a statistically significant difference between these two groups in terms of the outcome of interest (p=0.6). See Table 5.2. Given these findings I felt that candidates with missing data could be removed from the regression analyses.

<table>
<thead>
<tr>
<th>Outcome variable</th>
<th>Candidates with no missing data</th>
<th>Candidates with missing data</th>
<th>p value ($\chi^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall PACES performance</td>
<td>N=6820</td>
<td>N=851</td>
<td></td>
</tr>
<tr>
<td>Passed PACES</td>
<td></td>
<td></td>
<td>0.587</td>
</tr>
<tr>
<td>Yes</td>
<td>3329</td>
<td>407</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>3491</td>
<td>444</td>
<td></td>
</tr>
</tbody>
</table>

5.4.5 Regression analyses

5.4.5.1 Initial logistic regression model (full study population)

The initial logistic regression model included all candidates who completed PACES between October 2010 and May 2013. Candidates who had missing data for any of the co-variates included in the logistic regression model were excluded from the analyses. The co-variates included in the model were the candidates’ ethnicity, world region where primary medical qualification received, whether candidates were registered with the GMC, and the location of the MRCP examination centre. All these co-variates were found to be statistically significantly associated with both the exposure (candidates’ gender) and the outcome (passing PACES overall).
Adjusting for the other variables, female candidates were significantly more likely to pass PACES at first attempt compared with male candidates (OR=1.43, 95% CI=1.29-1.59, p<0.0001).

There was strong evidence for an interaction between candidates’ gender and candidates’ ethnicity (p=0.0012), with white female candidates having 1.15 times the odds of passing PACES at first attempt compared to white males, whereas ethnic minority female candidates having 1.66 times the odds of passing PACES at their first attempt compared to ethnic minority males.

There was evidence for multicollinearity in the initial logistic regression model, and I suspected that the co-variates were correlated. I therefore decided to divide the study population into three groups to minimise correlations (UK medical graduates; non-UK graduates registered with the GMC; non-UK graduates not registered with the GMC).

5.4.5.2 Subgroup regression analyses

5.4.5.2.1 UK medical graduates (N=3,574)

Just over half of the study population were UK medical graduates (52%), of which the majority were women (58%). The majority were of white ethnicity (67%), 0.5% were not registered with the GMC, and 1% attempted PACES at an examination centre outside of the UK. Bivariate analyses demonstrated that the latter two co-variates were not statistically significantly associated with PACES pass rate or gender, and they were therefore removed from further analyses in this group.

Female UK graduates had 1.18 times the odds of passing PACES at first attempt compared with male UK graduates, adjusting for ethnicity (OR=1.18; 95% CI=1.03 to 1.35; p=0.02); see Table 5.3. There was no evidence of an interaction between candidate gender and ethnicity (p=0.38), nor was there evidence of multicollinearity.
Table 5.3 The adjusted odds ratio (OR) for passing PACES at first attempt for female candidates compared to male candidates, and ethnic minority candidates compared to white candidates, after adjusting for all other variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Adjusted OR</th>
<th>95% CI</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>UK graduates (N=3574)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1.18</td>
<td>1.03 to 1.35</td>
<td>0.02</td>
</tr>
<tr>
<td>Ethnic minority</td>
<td>0.60</td>
<td>0.52 to 0.69</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Non-UK graduates registered with the GMC (N=1067)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1.47</td>
<td>1.11 to 1.94</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Ethnic minority</td>
<td>0.66</td>
<td>0.47 to 0.91</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>Non-UK graduates not registered with the GMC (N=2179)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1.99</td>
<td>1.65 to 2.39</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Ethnic minority</td>
<td>0.53</td>
<td>0.35 to 0.80</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Separate analyses performed for UK graduates, non-UK graduates registered with the General Medical Council (GMC), and non-UK graduates not registered with the GMC.

5.4.5.2.2 Non-UK medical graduates registered with the GMC (N=1,067)

Sixteen percent of the study population had received their primary medical qualification outside of the UK and were registered with the GMC to practise in the UK, of which men formed the majority (58%). The majority of this group declared themselves to be from an ethnic minority background (81%), and only 38 candidates had completed PACES in an examination centre outside of the UK (4%). Bivariate analyses showed that whether the candidate had completed PACES in an examination centre based in the UK was not associated with PACES pass rate or with gender, and it was removed from further analyses in this group.

Female non-UK graduates registered with the GMC had nearly one and a half times the odds of passing PACES compared to their male counterparts (OR: 1.47; 95% CI: 1.11 to 1.94; p<0.01); see Table 5.3. There was no evidence of an interaction between candidates’ gender and ethnicity (p=0.17), nor was there evidence of any multicollinearity.

5.4.5.2.3 Non-UK medical graduates not registered with the GMC (N=2,179)

Nearly one third of the study population had received their primary medical qualification from an institution outside of the UK, and were not registered with the GMC at the time of completing PACES (32%). Two thirds of this
group was male (66%), and the vast majority reported themselves to be from an ethnic minority background (95%). The majority completed PACES at an examination centre outside of the UK (65%), it is likely that the remainder travelled to the UK solely to sit the examination. Bivariate analyses demonstrated that candidates’ ethnicity was associated with PACES pass rate and gender. Centre location was found to be associated with PACES pass rate, but not gender, therefore I chose not to include it in the regression model.

Female non-UK graduates who were not registered with the GMC had nearly twice the odds of passing PACES, when compared to their male counterparts (OR=1.99; 95% CI=1.65 to 2.39; p<0.0001); see Table 5.3. There was no evidence of an interaction between candidates’ gender and ethnicity (p=0.72), nor was there evidence of any multicollinearity.

5.5 Discussion

5.5.1 Summary

In this large cross-sectional study of doctors attempting the largest high stakes postgraduate clinical examination in the world [258] I have demonstrated that female candidates are more likely to pass the examination at first attempt, even when adjusting for candidates’ ethnicity. Ethnicity has been shown to influence candidates’ performance in examinations [126, 127, 249, 125], and therefore it is important to control for its effect. The size of the effect of gender differed between the three groups examined, the largest being in non-UK medical graduates not registered with the GMC, the smallest in UK graduates registered with the GMC, and an intermediate effect size in non-UK medical graduates registered with the GMC.

5.5.2 Strengths and limitations

This large study population of over 7,500 candidates is representative of all candidates who complete PACES for the first time and increases the power of this study to detect real differences.
The data were collected for routine administrative purposes, which limited my ability to gather potentially relevant data such as doctor’s age; however this also meant that the data on many variables were complete. The variable that contributed to the majority of the missing data was ethnicity, which was self-declared; however, a comparison of candidates with and without missing data showed no evidence of a difference in terms of passing PACES.

5.5.3 Comparison with other studies

The findings that female doctors are more likely to pass PACES mirrors findings of Dewhurst and colleagues [126], who analysed data from the pre-2009 version of PACES in UK graduates only. They found that, after controlling for ethnicity, women had 1.69 the odds of passing (95% CI=1.42 to 2.02), a significantly larger order of magnitude to the finding from our logistic regression examining UK graduates only (OR=1.18; 95% CI=1.03 to 1.35).

Gender differences in the performance at UK postgraduate medical examinations in other specialties has also been shown, with women performing better than men at the General Practice specialty examination in both the written Applied Knowledge Test (AKT) and the practical Clinical Skills Assessment (CSA) [127, 259]. Women also performed better in the intercollegiate specialty board examinations for surgical training [260]. Interestingly the effect of gender is not so clear outside of the UK. In the USA, the USMLE Step 3 is the final and clinical component of the medical licensing examination. Two studies have examined the gender difference in performance at the examination in USA medical graduates, with one study finding women outperforming men [261], however the other study demonstrated no gender difference.

My finding that female international medical graduates outperform male international medical graduates is also seen in other UK and USA postgraduate examinations [259, 262-264]. To my knowledge, there are no studies looking at gender differences in the examination performance of
international medical graduates with and without registration to practise in the country in which the examination is set.

5.5.4 Unanswered questions

It is not clear why candidate demographics relate to PACES outcome. Female doctors may be better at performing the skills tested in clinical assessments [263]. It has been demonstrated that, during one clinical assessment, women ask more relevant history taking items and perform correctly more physical examination manoeuvres [264]. It could be that women are better at retraining and appropriately applying theoretical scientific and medical knowledge to clinical encounters in examinations; although scientific theoretical knowledge is formally assessed through written assessments [111]. Further, there was no evidence for a gender difference in the written components of the MRCP(UK) Diploma when examined by Dewhurst and colleagues in 2003/04 [126], although it would be of interest to analyse the gender difference on the written MRCP(UK) components during the time of period of this current study.

As mentioned in Chapters 3 and 4 a popular hypothesis is that gender differences in performance are due to differences in communication styles. A meta-analytic review of medical consultations found that female doctors are more likely to adopt a patient-centred communication style [243]. It has also been demonstrated that women doctors have greater interpersonal skills, which lead to empathic relationships [265-267]. These interpersonal skills may encourage the patient to be more forthcoming with regards to salient clinical information, enabling the doctor to correctly diagnose and manage the presenting ailment. However, this hypothesis would not completely explain the gender difference seen in performance at PACES, because there are stations where there is no sophisticated conversation with the patient. In Chapter 8 I will explore whether communication styles differ between male and female doctors practising in the UK.
A further hypothesis is that male and female doctors differ in values, and that these values lead to different motivations, which in turn influence achievement [268]. Female doctors have been shown to have higher person-related values [266, 268, 269], and one study found that performance in a clinical setting was predicted by person-related values held by the doctor [268]. It could also be that the design of the examination favours female candidates, perhaps examiners of clinical assessments may be unfairly biased towards female candidates or against male candidates. However there is no evidence to suggest a gender bias in clinical examiners in this current format of the PACES examination when assessed between 2009 and 2011 [80].

I did not examine the individual countries from which candidates had obtained their primary medical qualification, but it is possible that the female candidates were more likely to have graduated from English-speaking countries when compared to male international medical graduates. Native English speakers perform better at clinical assessments conducted in English [127], and if women international medical graduates are found to be more likely from a country where English is the dominant language, or where communication skills and cultural values are more similar to the UK when compared to men, this could go toward explaining the gender difference in performance seen in international medical graduates. It may also reflect differences in the selection and training of female doctors in countries around the world [270]. It is also possible that access to PACES and medical education in general may be biased outside of the UK. This may plausibly result in female candidates needing to be higher performers and to be more motivated than their male counterparts, to overcome obstacles that may limit their access to their examination.

Candidates’ age was not examined in this study, but it is likely that non-UK graduates were older and a previous study has demonstrated that older candidates do not perform as well as younger candidates in clinical assessments [264]. The variation between the gender difference in non-UK
graduates with and without GMC registration may reflect gendered migration patterns. For example, Lebanese medical graduates practising in the USA are significantly less likely to be female than graduates of other countries [271], and a study of Lebanese medical students by the same authors found that female students had less intention of working abroad after graduation than male students [272]. It would be of interest to explore whether the difference in PACES performance has varied year upon year, or whether the gender difference is stable. This study captures just under 3 years worth of data and therefore it is unlikely that any meaningful conclusions will able to be drawn with regards to performance over time.

It is likely that performance in large, high-stakes clinical examinations that have demonstrated good validity reflects performance in actual clinical practice. A study examining the predictive validity of two UK postgraduate medical examinations (MRCGP and MRCP(UK)) found that doctors’ performance at both examinations was predictive of fitness to practise sanctions, with the clinical assessments showing a larger predictive effect than the written components [67]. The authors reflected on which factors may be associated with performing well at a clinical examination and are protective from disciplinary action. They hypothesised that “the psychological processes involved in successfully studying and practising medicine at a high level share similar mechanisms to those underlying conduct and trust” [67]. It has been demonstrated that the personality trait conscientiousness is associated with success in a variety of occupations and with performance in medical students [107, 273]. Conscientious individuals are more likely to be careful, diligent, responsible and organised [274]. It is clear why these traits would be beneficial in individuals preparing for an examination. It would be reasonable to hypothesise that these traits would also be beneficial to doctors in their professional working lives. A doctor with these traits would be likely to ensure that they are practising to the standard expected of them and that they are maintaining and improving their clinical skills and knowledge. In Chapter 8 I will explore whether personality traits, such as conscientiousness, differ between male and female doctors.
practising in the UK and whether these personality traits differ between doctors who have had action against their medical registration and those who have not. Wakeford et al. also suggested that success at clinical examinations may be related to how individuals interact with colleagues and are able to be part of a social network of knowledge [67]. A doctor who is preparing to sit a clinical examination will benefit from interacting with other prospective candidates and previous successful candidates. These interactions will enable them to practice and improve clinical skills, and learn more about the format and expectations of the examination. Communication skills and teamwork skills are likely to support being a part of these social knowledge networks. These same skills would are also likely to be beneficial in the clinical workplace, as 8% of investigations conducted by the GMC in 2016 involved concerns about doctors’ communication skills [95].

Finally, insight may play a role. A lack of insight has been associated with doctors whose performance has raised serious concerns [275]. A lack of awareness of one’s own level of performance and the inability to identify areas of weakness could also negatively impact examination performance. It is probable that the factors underlying the gender differences in performance in clinical examinations also contribute to female doctors being less subject to medico-legal action [10].

5.5.5 Conclusions

Female doctors outperform their male counterparts in a high stakes clinical examination once adjusted for the effect of their ethnicity, and the size of this gender effect was greater in doctors who graduated outside of the UK, especially those who were not registered with the UK’s GMC. The reasons behind this gender difference in clinical examination performance is not clear, but an association between postgraduate medical examination performance and disciplinary action has been demonstrated [67], suggesting that the attributes required to pass postgraduate medical examinations might also be protective from disciplinary action being taken against a doctor’s registration. I hypothesise that the skills required to successfully complete clinical
postgraduate medical examinations might be protective from disciplinary action. I will build upon the findings of this chapter - in Chapter 8 I will investigate whether empathy, communication style and personality traits differ between the genders of a population of UK medical graduates, and whether these measures are associated with future disciplinary action. In Chapter 7 I will explore whether the gender difference in examination performance is also present at examinations used to assess fitness to practise in GPs. Firstly, however, in Chapter 6 I will explore whether the gender difference in clinical assessment performance observed in this chapter is also present in other clinical and written postgraduate medical examinations used in other specialties and in different countries.
Chapter 6 – Gender differences in postgraduate medical examination pass rates: a systematic review and meta-analysis

6.1 Chapter summary

6.1.1 Background

The large study described in Chapter 5 demonstrated that women perform better at a clinical specialty-specific UK postgraduate medical examination than men. However, the current evidence comparing the performance of male and female doctors in postgraduate medical examinations shows mixed results. It is important to establish whether a gender difference is present and whether there are any patterns in the difference in performance between the genders, to enable a greater understanding of how performance may vary between the genders.

6.1.2 Aims

This study aims to establish whether there is an association between doctors’ gender and postgraduate medical examination performance (measured by pass rates), in different populations of doctors, sitting a variety of written and clinical postgraduate medical examinations in Canada, Ireland, UK and USA.

6.1.3 Design and setting

A systematic review and meta-analysis of all published original research studies reporting on the association between gender and postgraduate medical examination performance.

6.1.4 Methods

Electronic databases MEDLINE, ERIC and CINAHL were searched, along with backward and forward citation searching. Searches were limited to January 1990 and April 2016. Studies reporting original data, written in English, and reporting on the association between gender and postgraduate medical examination performance were included. Two reviewers independently extracted data. Methodological quality of the studies was
assessed using criteria based on the NOS. A random effects meta-analysis model was used to assess the effect of gender on passing postgraduate medical examinations. The primary outcome was postgraduate medical examination pass rate.

6.1.5 Results

28 eligible studies calculating the association between doctor gender and performance in postgraduate medical examinations were included in the systematic review, capturing over 107,000 examination attempts. The majority reported that women were more likely to perform better than men. 16 findings from 14 studies were included in the meta-analysis, capturing over 68,000 examination attempts. Women had over one and a half times the odds of passing postgraduate medical examinations compared to men (OR 1.54, 95% CI 1.12 to 2.12), although there was high heterogeneity (I²=97.4%). Subgroup analyses found no significant gender effect in written examinations (OR 1.01, 95% CI 0.81 to 1.26), and a larger overall effect in clinical examinations (OR 2.63, 95% CI 1.70 to 4.06).

6.1.6 Conclusion

Overall, female doctors perform better than male doctors at postgraduate medical examinations of a clinical nature, but there is no difference in written assessments. These findings suggest that women doctors display skills that are highly regarded in clinical assessments. It could be argued that men and women doctors have slightly different skills and therefore it is plausible that the clinical assessments focus on skills in which women tend to show strength, and focus less on the skills in which the men show strength, explaining the gender difference observed in clinical assessment performance, but not in written assessment performance.
6.2 Introduction

It is well documented that female students achieve higher pass rates and scores than their male counterparts in pre-university examinations [276-280]. In higher education, female students have been shown to obtain higher grades when compared to male students [281, 282]. In medicine, studies have shown that men do not perform as well as women during their medical degree course [109, 111], or at final examinations [112, 113]. The evidence for a gender difference in performance at postgraduate medical examinations is mixed with some studies showing female doctors outperform male doctors (see Chapter 5) [126, 283] and others showing male doctors outperform female doctors, or no gender difference [284, 285]. A review of the literature in this area is required as it is important to establish whether the trend of female doctors performing better than male doctors at the undergraduate level continues through to the postgraduate level. This research will go towards establishing whether there are systematic differences in the performance of doctors at postgraduate medical examinations by gender, and will contribute to enabling the medical education community to better understand how male and female doctors may differ in terms of their academic and professional performance.

I completed a systematic review of the literature and meta-analysis to answer the following questions:

1. Is there a gender difference in doctors’ performance in postgraduate medical examinations?

2. Does the gender difference in postgraduate examination performance vary depending on the examination format?

6.3 Methods

I followed PRISMA guidance [208] on reporting.
6.3.1 Data sources and search strategies

I conducted systematic searches of three electronic databases (MEDLNE, ERIC and CINAHL) from January 1990 to April 2016, for studies describing doctors' performance in postgraduate medical examinations by their gender (see Table 6.1 for search terms used). The three electronic databases were selected to capture literature relevant to both medical and educational research. I limited the search to articles published from January 1990 because examination formats change regularly and I am keen for the findings of this review to be applicable to doctors completing postgraduate medical examinations today. In addition to electronic database searches, I performed backward and forward cited reference searching. Studies not published in English were excluded due to limited resources.
Table 6.1 Search terms used for electronic database search

MEDLINE search in April 2016

<table>
<thead>
<tr>
<th>Concept 1</th>
<th>Concept 2</th>
<th>Concept 3</th>
<th>Concept 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>MeSH term Physicians</td>
<td>MeSH term Gender</td>
<td>Post?graduate exam$</td>
<td>MeSH term Achievement</td>
</tr>
<tr>
<td>Doctor$</td>
<td>Male$</td>
<td>Medical exam$</td>
<td>Performance</td>
</tr>
<tr>
<td>Medical Practitioner$</td>
<td>Female$</td>
<td>Medical licensing exam$</td>
<td>Academic achievement</td>
</tr>
<tr>
<td>Medic$</td>
<td>Sex</td>
<td>Royal college$ exam$</td>
<td>Academic performance</td>
</tr>
<tr>
<td>Clinician$</td>
<td>Sexes</td>
<td>Specialty exam$</td>
<td></td>
</tr>
<tr>
<td>Physician$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CINAHL search April 2016

<table>
<thead>
<tr>
<th>Concept 1</th>
<th>Concept 2</th>
<th>Concept 3</th>
<th>Concept 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physician</td>
<td>Gender</td>
<td>Postgraduate exam*</td>
<td>Performance</td>
</tr>
<tr>
<td>Physician*</td>
<td>Male</td>
<td>Post-graduate exam*</td>
<td>Academic performance</td>
</tr>
<tr>
<td>Doctor*</td>
<td>Female</td>
<td>Medical exam*</td>
<td>Academic achievement</td>
</tr>
<tr>
<td>Medical practitioner*</td>
<td>Sex</td>
<td>Medical licensing exam*</td>
<td></td>
</tr>
<tr>
<td>Medic</td>
<td></td>
<td>Royal college exam*</td>
<td></td>
</tr>
<tr>
<td>Clinician*</td>
<td></td>
<td>Specialty exam*</td>
<td></td>
</tr>
<tr>
<td>Physician*</td>
<td></td>
<td>Medical college exam*</td>
<td></td>
</tr>
</tbody>
</table>

ERIC search April 2016

<table>
<thead>
<tr>
<th>Concept 1</th>
<th>Concept 2</th>
<th>Concept 3</th>
<th>Concept 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physician</td>
<td>Gender</td>
<td>Postgraduate exam*</td>
<td>Performance</td>
</tr>
<tr>
<td>Physician*</td>
<td>Male</td>
<td>Post-graduate exam</td>
<td>Academic performance</td>
</tr>
<tr>
<td>Doctor*</td>
<td>Female</td>
<td>Medical exam*</td>
<td>Academic achievement</td>
</tr>
<tr>
<td>Medical Practitioner*</td>
<td>Sex</td>
<td>Medical licensing exam*</td>
<td></td>
</tr>
<tr>
<td>Medic</td>
<td></td>
<td>Royal college exam*</td>
<td></td>
</tr>
<tr>
<td>Clinician*</td>
<td></td>
<td>Specialty exam*</td>
<td></td>
</tr>
<tr>
<td>Physician*</td>
<td></td>
<td>Medical college exam*</td>
<td></td>
</tr>
</tbody>
</table>

6.3.2 Study selection

The identified studies were assessed for inclusion using a pre-determined eligibility checklist: (1) the study included data from an original, peer-reviewed study, (2) study participants were doctors attempting a postgraduate medical examination, (3) the authors reported postgraduate medical examination performance by candidate gender.
I considered all studies, regardless of study design, and included all articles regardless of examination format. I identified articles eligible for further review by performing an initial screen of titles and/or abstracts, followed by a full-text review.

6.3.3 Data extraction

Two researchers (myself and RV) independently extracted data from all the selected eligible studies using a pre-determined data extraction form modified from the Cochrane handbook [209]. The form allowed each researcher to independently record information on study characteristics (authors, publication year, journal, study design, years study conducted, sampling strategy, recruitment method), exposure data (data collection methods), outcome data (examination type, examination format, number of attempts by candidates, data collection method), study participants (specialty, inclusion of international medical graduates, country that set the examination, representativeness, response rate, missing data), and results (overall findings, significance testing, confounders). When reports contained multivariate analyses, we prioritised crude effects; however if no crude effects were reported, we included outcome measures adjusted for other variables. Both researchers used criteria based on the NOS [210] to assess the methodological quality of each study. Any discrepancies in the data extraction process were reconciled through discussion.

6.3.4 Outcome definition and subgroup analyses

I considered all postgraduate medical examinations in this review, regardless of specialty, examination type, or examination format used. I have used the term ‘postgraduate medical examination’ to group together and represent all of the examination types and formats captured by this review.

6.3.4.1 Examination types

The different examination types captured by this review include:
6.3.4.1.1 United States Medical Licensing Examination (USMLE) Step 3

Doctors wishing to practice in the US are required to complete the United States Medical Licensing Examination (USMLE) series. USMLE Step 3 is the final examination in the series, and a doctor is required to pass it in order to obtain a license as a practising physician. Candidates must hold a medical degree to be eligible to sit the USMLE Step 3.

6.3.4.1.2 Comprehensive Osteopathic Medical Licensing Examination (COMLEX-USA) Level 3

Similar to the USMLE, the Comprehensive Osteopathic Licensing Examination of the United States (COMLEX-USA) is a series of osteopathic medical licensing examinations, of which the COMLEX-USA Level 3 is the final examination, typically taken after a doctor has started practising medicine. The COMLEX-USA is the most common route through which osteopathic physicians obtain a medical license.

6.3.4.1.3 In-Training Examination (ITE) for specialty

The In-Training Examination (ITE) for specialty is used in the USA for doctors training in certain specialties. Performance in these examinations is used to assess the doctor’s progress.

6.3.4.1.4 American Board specialty examinations

The American Board specialty examinations are used in the USA for doctors who wish to become certified in a particular specialty.

6.3.4.1.5 Educational Commission for Foreign Medical Graduates (ECFMG) Certification

The Educational Commission for Foreign Medical Graduates (ECFMG) Certification is used to evaluate international medical graduates who wish to practise in the USA. ECFMG Certification is also required for international medical graduates prior to sitting the USMLE Step 3.
6.3.4.1.6 Royal College Membership or Fellowship

The Royal College examinations are a requirement for doctors practising in the UK who wish to complete specialty training, enabling them to become hospital consultants or GPs. There is a Royal College for each specialty; each college organises the postgraduate examinations for that specialty. A candidate must hold a medical degree prior to taking the examinations.

6.3.4.2 Examination formats

The different examination formats captured by this review have been broadly divided into two categories: written knowledge tests and clinical assessments.

6.3.4.2.1 Written knowledge tests

Examinations assessing scientific theoretical medical knowledge, usually in a written format, or computer based.

6.3.4.2.2 Clinical assessments

Examinations assessing practical clinical knowledge and skills in a setting that is similar to the actual clinical setting.

6.3.4.2.3 Other examination formats

Examinations that are not in the traditional written or clinical examination format. These tend to be specific for a certain specialty. Formats included oral examinations and rapid reporting.

6.3.5 Statistical analysis

I first completed the main analysis for all studies combined, before moving on to a subgroup analysis on examination format. I chose to examine the data by examination format because I wished to investigate whether any gender difference in examination performance varies by examination format.

I calculated an effect estimate for each study for the effect of female gender on the pass rates of postgraduate medical examinations, and performed
heterogeneity tests. I calculated a summary estimate of the effect of female
gender on passing postgraduate medical examinations using a random-
effects model.

6.4 Results

6.4.1 Systematic review

I retrieved 1,445 citations, of which 28 eligible studies [286-289, 262, 284,
126, 127, 290-297, 285, 283, 260, 263, 298-303, 264, 261] met the inclusion
criteria for the review (Figure 6.1). Those 28 studies captured over 107,000
postgraduate medical examination attempts by doctors (range: 50 to 48,509;
median number of attempts per study: 1731\(^{ix}\)) over 46 years (1968 to 2014).
Study characteristics are shown in Table 6.2.

Figure 6.1 Flow chart showing reports retrieved, excluded and articles
included in the review based on the PRISMA statement

\(^{ix}\) Two of the studies did not report the total number of examination attempts [283, 263]
All 28 of the studies were observational in design, as expected given the nature of the research question. 18/28 studies used a cohort design [286-288, 284, 126, 127, 290-292, 294, 296, 297, 285, 283, 299-301, 261], and the remaining 10/28 studies were cross-sectional in design [289, 262, 293, 295, 260, 263, 298, 302, 303, 264].

All 28 studies examined the association between doctors’ gender and examination performance, though in 4/28 studies the primary aim was not to examine the relationship between doctors’ gender and postgraduate medical examination performance [127, 290, 293, 301].

The studies were published between 1991-2016, with the majority of the studies being published in the years 2007 [126, 294, 297], 2010 [285, 300, 301], 2014 [291, 283, 299], and 2015 [288, 293, 295] (three studies in each year).

The majority of studies investigated postgraduate medical examinations taken in the UK and Ireland (15 studies) [287, 289, 284, 126, 127, 291, 292, 295, 297, 283, 260, 298, 299, 302, 303], or the USA (12 studies) [286, 288, 262, 290, 293, 294, 296, 285, 300, 301, 264, 261]. The remaining study analysed a postgraduate medical examination from Canada [263].

A total of 10/28 studies analysed written postgraduate medical examinations [287, 288, 284, 290, 294-296, 298, 302, 293], 9/28 studies analysed the performance at clinical postgraduate medical examinations [286, 289, 262, 283, 263, 299, 264, 261, 127], and 4/28 studies analysed postgraduate medical examinations that included both written and clinical components [126, 297, 260, 303]. The remaining five studies either reported findings based on alternative examination formats (3/28 studies) [291, 292, 301], or did not report the examination format of the postgraduate medical examination they were analysing and I was unable to ascertain the format of the examination from other sources (2/28 studies) [285, 300].
The UK-based examinations captured by my search included: Royal College membership or fellowship specialty examinations (13/28 studies) [287, 289, 284, 126, 127, 291, 292, 295, 297, 283, 298, 302, 303], the CSA exit examination for doctors specialising in General Practice (1/28 studies) [299], and the Intercollegiate Specialty examination for doctors specialising in Surgery (1/28 studies) [260]. The American examinations captured by my search included: In-Training Examinations (ITEs) (4/28 studies) [290, 293, 294, 285], American Board Specialty examinations (3/28 studies) [285, 300, 301], Foreign Medical Graduates examinations (3/28 studies) [262, 296, 264], USMLE Step 3 (2/28 studies) [286, 261], and the COMLEX-USA Level 3 (1/28 studies) [288]. The Canadian examination captured by my search strategy was a pre-internship selection program (1/28 studies) [263].

6/28 studies examined doctors’ postgraduate examination performance from all specialties [286, 262, 296, 263, 264, 261]. Four studies examined the performance of doctors specialising in General Practice [127, 283, 299, 302], and a further four studies examined the performance of doctors specialising in Orthopaedics and Surgery [293, 285, 260, 300]. Three studies examined the postgraduate examination performance of doctors specialising in General Medicine [289, 126, 294], and Psychiatry [297, 301, 303], respectively. The performance of doctors who were specialising in Public Health was examined by two studies [287, 292]. The performance of doctors specialising in Anaesthetics [284], Emergency Medicine [290], Obstetric Gynaecology [298], Osteopathic medicine [288], Paediatrics [295], and Radiology [291] was examined by one study for each specialty.

A total of 15/28 studies examined countrywide data [262, 284, 126, 127, 291, 292, 295-297, 283, 260, 298, 302, 303, 264]. 1/28 studies examined data from multiple countries [287]. 2/28 studies analysed data from multiple states [290, 293]. The remaining studies examined data from one academic institution (5/28 studies) [286, 288, 300, 301, 261], one medical centre (2/28...
studies) [294, 285], one province (1/28 studies) [263], one health-service deanery\(^*\) (1/28 studies) [299], and one revision course (1/28 studies) [289].

26/28 studies included all doctors who were eligible to be included as the study population [286-289, 262, 284, 126, 127, 290-297, 285, 283, 260, 263, 298-300, 302, 303, 264]. 2/28 studies did not report the method by which they recruited and selected their study population [301, 261].

Data on postgraduate examination performance was collected and provided by external organisations in 20/28 studies [286-289, 284, 127, 290-292, 294, 295, 285, 283, 263, 298-303]. 1/28 studies collected postgraduate examination performance data directly from the study participants [293], and 7/28 studies did not report how the data was collected [262, 126, 296, 297, 260, 264, 261].

Half of all the studies (14/28 studies) examined candidates’ performance at an examination regardless of whether they had had an earlier attempt at the examination [287, 289, 284, 126, 127, 291-294, 296, 297, 285, 300, 303]. One quarter of the studies (7/28 studies) examined the performance of candidates completing the examination for the first time [286, 288, 262, 295, 260, 298, 264]. The remaining seven studies did not report whether they were restricting their analyses by attempt number [290, 283, 263, 299, 301, 302, 261].

Most of the studies included medical graduates, regardless of where in the world they had obtained their primary medical qualification (11/28 studies) [287, 289, 127, 291, 294, 295, 297, 299, 300, 302, 303], equal numbers of studies included either only international medical graduates [262, 296, 263, 301, 264], or only domestic medical graduates [286, 284, 126, 298, 261] (5/28 studies respectively). 7/28 studies did not report whether or not their

\(^*\) Deaneries are organisations responsible for postgraduate medical and dental training within a given region of the UK. Since 2013 deaneries in England have been replaced with Local Education and Training Boards (LETBs).
study population included international medical graduates [288, 290, 292, 293, 285, 283, 260].

Overall the 28 studies provided 31 separate findings, because 3 of the studies reported the results of two examination parts separately [126, 260, 298]. The majority of the findings reported that female doctors were statistically significantly more likely to pass postgraduate medical examinations (15/32) [287, 262, 127, 290, 291, 283, 260, 298, 299, 301, 303, 264, 261, 126]. 12/31 reported no statistically significant gender difference in examination performance [286, 288, 289, 126, 292-295, 285, 298, 300], and 2/32 reported male doctors statistically significantly outperforming female doctors [284, 260]. 4/31 did not perform significance testing [296, 297, 263, 302] xi (see Table 6.2).

I will now proceed to analyse the studies’ finding by examination format, as I feel this may provide a better insight into how doctors’ gender may influence examination performance.

---

xi Rothman et al. [263] did perform significance testing when examining the performance at individual stations, but they did not perform a significance test when examining the overall gender difference in performance. Siriwardena et al. [302] did perform significance testing when examining the performance at sex-specific questions, but they did not perform a significance test when examining the overall gender difference in performance.
<table>
<thead>
<tr>
<th>Exam format</th>
<th>First author, year (country)</th>
<th>Study design</th>
<th>Description of study population</th>
<th>Years data collected</th>
<th>Postgraduate exam</th>
<th>Sample size</th>
<th>Overall finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Written</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ayres et al. [287], 1996 (UK)</td>
<td>Cohort</td>
<td>Doctors training in Public Health, including both UK medical graduates and IMGs, and all attempts at the exam.</td>
<td>1992-1995</td>
<td>MFPHM Part 1</td>
<td>431</td>
<td>Female doctors performed better i</td>
</tr>
<tr>
<td></td>
<td>Baker et al. [288], 2015 (USA)</td>
<td>Cohort</td>
<td>Osteopathic medical practitioners, and this was their first attempt at the exam.</td>
<td>2008-2011</td>
<td>COMLEX-USA 3</td>
<td>552</td>
<td>No difference between genders i</td>
</tr>
<tr>
<td></td>
<td>Bowhay et al. [284], 2009 (UK)</td>
<td>Cohort</td>
<td>Doctors training in Anaesthetics, who were not IMGs, and all attempts at the exam.</td>
<td>1999-2008</td>
<td>FRCA Part 1</td>
<td>3303</td>
<td>Male doctors performed better i</td>
</tr>
<tr>
<td></td>
<td>Gene Hern et al. [290], 2009 (USA)</td>
<td>Cohort</td>
<td>Doctors training in Emergency Medicine.</td>
<td>2002-2008</td>
<td>ABEM ITE</td>
<td>368</td>
<td>Female doctors performed better i</td>
</tr>
<tr>
<td></td>
<td>Kim et al. [293], 2015 (USA)</td>
<td>Cross-sectional</td>
<td>Doctors training in Surgery, and all attempts at the exam.</td>
<td>2014</td>
<td>ABS ITE</td>
<td>266</td>
<td>No difference between genders i</td>
</tr>
<tr>
<td></td>
<td>McDonald et al. [294], 2007 (USA)</td>
<td>Cohort</td>
<td>Doctors training in General Medicine, including both USA medical graduates and IMGs, and all attempts at the exam.</td>
<td>2000-2003</td>
<td>IM ITE</td>
<td>195</td>
<td>No difference between genders i</td>
</tr>
<tr>
<td></td>
<td>Menzies et al. [295], 2015 (UK)</td>
<td>Cross-sectional</td>
<td>Doctors training in Paediatrics, including both UK medical graduates and IMGs, and this was their first attempt at the exam.</td>
<td>2007-2011</td>
<td>MRCPCH Part 1B</td>
<td>2056</td>
<td>No difference between genders i</td>
</tr>
<tr>
<td>Exam format</td>
<td>First author, year (country)</td>
<td>Study design</td>
<td>Description of study population</td>
<td>Years data collected</td>
<td>Postgraduate exam</td>
<td>Sample size</td>
<td>Overall finding</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------------------</td>
<td>--------------</td>
<td>---------------------------------</td>
<td>---------------------</td>
<td>-------------------</td>
<td>-------------</td>
<td>-----------------</td>
</tr>
<tr>
<td></td>
<td>Mick et al. [296], 1991 (USA)</td>
<td>Cohort</td>
<td>Doctors of all specialties, who were IMGs, and all attempts at the exam.</td>
<td>1984-1987</td>
<td>FMGEMS</td>
<td>48509</td>
<td>Male doctors performed better</td>
</tr>
</tbody>
</table>
|             | Rushd et al. [298], 2012 (UK)  | Cross-sectional | Doctors training in Obstetrics and Gynaecology, who were UK graduates, and this was their first attempt at the exam. | 1998-2008           | MRCOG Part 1 & 2 | 2156        | Part 1: No difference between genders i  
Part 2: Female doctors performed better i |
<p>|             | Siriwardena et al. [302], 2012 (UK) | Cross-sectional | Doctors training in General Practice, who were both UK medical graduates and IMGs. | Not reported         | MRCGP            | 3627        | Female doctors performed better |
| Clinical    | Andriole et al. [286], 2005 (USA) | Cohort       | Doctors of all specialties, who were not IMGs, and this was their first attempt at the exam. | 2000-2003           | USMLE 3           | 237         | No difference between genders i |
|             | Bessant et al. [289], 2006 (UK)  | Cross-sectional | Doctors training in General Medicine, including both UK medical graduates and IMGs, and all attempts at the exam. | 2002                | MRCP(UK) PACES   | 483         | No difference between genders i |
|             | Boulet et al. [262], 2005 (USA)  | Cross-sectional | Doctors of all specialties, who were IMGs, and this was their first attempt at the exam. | 2002-2003           | ECFMG            | 7746        | Female doctors performed better i |
|             | Esmail et al. [127], 2013 (UK)   | Cohort       | Doctors training in General Practice, including both UK medical graduates and IMGs, and all attempts at the exam. | 2010-2012           | MRCGP            | 5095        | Female doctors performed better i |
|             | Pope et al. [283], 2014 (UK)     | Cohort       | Doctors training in General Practice. | 2012-2013           | MRCGP CSA        | Not reported | Female doctors performed better i |</p>
<table>
<thead>
<tr>
<th>Exam format</th>
<th>First author, year (country)</th>
<th>Study design</th>
<th>Description of study population</th>
<th>Years data collected</th>
<th>Postgraduate exam</th>
<th>Sample size</th>
<th>Overall finding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rothman et al. [263], 1995 (Canada)</td>
<td>Cross-sectional</td>
<td>Doctors of all specialties, who were IMGs.</td>
<td>1987-1993</td>
<td>Ontario Pre-Internship Program</td>
<td>Not reported</td>
<td>Female doctors performed better</td>
</tr>
<tr>
<td></td>
<td>Shaw et al. [299], 2014 (UK)</td>
<td>Cohort</td>
<td>Doctors training in General Practice, including both UK medical graduates and IMGs.</td>
<td>2012</td>
<td>CSA</td>
<td>116</td>
<td>Female doctors performed better</td>
</tr>
<tr>
<td></td>
<td>van Zanten et al. [264], 2003 (USA)</td>
<td>Cross-sectional</td>
<td>Doctors of all specialties, who were IMGs, and this was their first attempt at the exam.</td>
<td>2000-2002</td>
<td>ECFMG</td>
<td>11 690</td>
<td>Female doctors performed better</td>
</tr>
<tr>
<td></td>
<td>Veloski et al. [261], 2000 (USA)</td>
<td>Cohort</td>
<td>Doctors of all specialties, who were USA medical graduates.</td>
<td>1968-1997</td>
<td>USMLE 3</td>
<td>3884</td>
<td>Female doctors performed better</td>
</tr>
</tbody>
</table>

**Written and clinical**

<p>|            | Dewhurst et al. [126], 2007 (UK) | Cohort | Doctors in training in General Medicine, who were not IMGs, and all attempts at the exam. | 2003-2004 | MRCP(UK) | 8721 | Written: No difference between genders Clinical: Female doctors performed better |
|            | Oyebode et al. [297], 2007 (UK) | Cohort | Doctors training in Psychiatry, including both UK medical graduates and IMGs, and all attempts at the exam. | 1997-2002 | MRCPsych | 10 842 | Female doctors performed better |
|            | Richens et al. [260], 2016 (UK &amp; Ireland) | Cross-sectional | Doctors training in Orthopaedics and Surgery, and this was their first attempt at the exam. | 2009-2013 | Intercollegiate Specialty Board examination | 4878 | Written: Male doctors performed better Clinical: Female doctors performed better |
|            | Tyrer et al. [303], 2002 (UK) | Cross-sectional | Doctors training in Psychiatry, including both UK medical graduates and IMGs, and all attempts at the exam. | 1999 | MRCPsych | 1891 | Female doctors performed better |</p>
<table>
<thead>
<tr>
<th>Exam format</th>
<th>First author, year (country)</th>
<th>Study design</th>
<th>Description of study population</th>
<th>Years data collected</th>
<th>Postgraduate exam</th>
<th>Sample size</th>
<th>Overall finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
<td>Hawtin et al. [291], 2014 (UK)</td>
<td>Cohort</td>
<td>Doctors training in Radiology, including both UK medical graduates and IMGs, and all attempts at the exam.</td>
<td>2006-2010</td>
<td>FRCR (UK) Part 2B</td>
<td>1571</td>
<td>Female doctors performed better (^i)</td>
</tr>
<tr>
<td></td>
<td>Holland et al. [292], 2002 (UK)</td>
<td>Cohort</td>
<td>Doctors in training in Public Health, and all attempts at the exam. Pass rates were only recorded for first attempts.</td>
<td>1996-1999</td>
<td>MFPHM Part 2</td>
<td>238</td>
<td>No difference between genders (^i)</td>
</tr>
<tr>
<td></td>
<td>Shiroma et al. [301], 2010 (USA)</td>
<td>Cohort</td>
<td>Doctors training in Psychiatry, who were IMGs.</td>
<td>1994-2004</td>
<td>ABPN</td>
<td>50</td>
<td>Female doctors performed better (^i)</td>
</tr>
<tr>
<td>Not reported</td>
<td>Pico et al. [285], 2010 (USA)</td>
<td>Cohort</td>
<td>Doctors training in Orthopaedics, and all attempts at the exam.</td>
<td>1999-2010</td>
<td>OITE &amp; ABOS</td>
<td>90</td>
<td>No difference between genders (^i)</td>
</tr>
<tr>
<td></td>
<td>Shellito et al. [300], 2010 (USA)</td>
<td>Cohort</td>
<td>Doctors training in Surgery, including both USA medical graduates and IMGs, and this was their first attempt at the exam.</td>
<td>1990-2001</td>
<td>ABS</td>
<td>62</td>
<td>No difference between genders (^i)</td>
</tr>
</tbody>
</table>

\(^i\) Significance testing performed
6.4.1.1 Results by examination format

6.4.1.1.1 Written postgraduate examinations

There were 13 results from 12/28 studies analysing performance in written examinations [287, 288, 284, 126, 290, 293-296, 260, 298, 302].

Nearly half of the results (6/13) concluded that there was not a statistically significant difference in the performance of male and female doctors in written examination performance [288, 126, 293-295, 298]. 3/13 results reported female doctors performed statistically significantly better than male doctors [287, 290, 298], whereas 2/13 findings demonstrated male doctors performed statistically significantly better than female doctors [284, 260]. Of the 2/13 findings that did not perform statistical testing, one reported women performed better [302], and the other reported men performed better [296].

6.4.1.1.2 Clinical postgraduate examinations


The majority (8/11) of the results found that the performance of female doctors at clinical examinations was statistically significantly better than male doctors [262, 126, 127, 283, 260, 299, 264, 261]. Only 2/11 of the results reported there was no evidence for a statistically significant difference in clinical examination performance between the genders [286, 289]. None of the studies found that male doctors performed better at clinical examinations than female doctors. The 1/11 studies that did not perform statistical testing found women performed better than men [263].

6.4.1.1.3 Written and clinical postgraduate examinations

Only 2 studies analysed the performance of doctors at both the written and clinical components of a medical postgraduate examination combined (it was not possible to separate the findings by examination format) [297, 303]. Both of these studies found that female doctors outperformed male doctors, however, only one of these studies performed significance testing [303].
6.4.1.1.4 Other postgraduate examination formats

3/28 studies analysed examinations that fell into this category [291, 292, 301]. 2/3 studies found that female doctors statistically significantly outperformed male doctors [291, 301]. These two studies analysed a rapid reporting examination for doctors specialising in Radiology in the UK [291], and an oral examination for international medical graduates specialising in Psychiatry in the USA [301]. The remaining study, which investigated the performance of doctors’ specialising in Public Health and is an oral examination based in the UK, found no statistically significant difference in performance between genders [292].

6.4.1.1.5 Examination format not reported

2/28 studies did not report the format of the postgraduate examination they were investigating [285, 300]. Both studies found no statistically significant difference in examination performance between the genders.

For a summary of findings for each study by examination format see Table 6.2.

6.4.1.2 Results by country

6.4.1.2.1 USA and Canada

A total of 13 studies collected and analysed data from the USA and Canada [286, 288, 262, 290, 293, 294, 296, 285, 263, 300, 301, 264, 261]. 9/13 studies were cohort in design [286, 288, 290, 294, 296, 285, 300, 301, 261], and the remaining 4/13 studies were cross-sectional studies [262, 293, 263, 264]. 5/13 studies analysed clinical examinations [286, 262-264, 261], 5/13 studies analysed written examinations [288, 290, 294, 296, 293], 1/13 studies analysed an examination with an alternative format [301], and 2/13 did not report the format of the examination they were analysing [285, 300]. 6/13 studies reported that women outperformed men in postgraduate medical examinations [262, 290, 263, 301, 264, 261], however only five of these studies performed significance testing [262, 290, 301, 264, 261]. 6/13 studies reported no gender difference in the performance at postgraduate
medical examinations [286, 288, 293, 294, 285, 300], and 1/13 studies found that men outperformed women at postgraduate medical examinations [296], however this study did not perform significance testing.

6.4.1.2.2 UK and Ireland

15 studies were conducted on postgraduate medical examinations from the UK and Ireland [287, 289, 284, 126, 127, 291, 292, 295, 297, 283, 260, 298, 299, 302, 303]. 9/15 studies were cohort in design [287, 284, 126, 127, 291, 292, 297, 299], and 6/15 studies were cross-sectional [289, 295, 260, 298, 302, 303]. 5/15 studies analysed written postgraduate medical examinations [287, 284, 295, 298, 302], 4/15 studies looked at both written and clinical examination formats [126, 297, 260, 303], 4/15 studies examined clinical examination formats [289, 283, 299, 127], and 2/15 studies examined alternative examination formats [291, 292]. 3/15 studies looked at two parts of a postgraduate medical examination [126, 260, 298], and therefore reported two separate findings (one for each part of the examination). Therefore, when reporting the overall findings of the studies I have chosen to treat each result separately and will report a total of 18 findings for the 15 studies. 11/18 results demonstrated that women outperformed men [287, 126, 127, 291, 297, 283, 260, 298, 299, 302, 303], however two of these results had not undergone significance testing [297, 302]. 5/18 results showed that there was no difference in postgraduate medical examination performance between the genders [289, 126, 292, 295, 298]. 2/18 results demonstrated that men outperformed women [284, 260], and both these results had undergone significance testing.

6.4.2 Assessment of methodological quality

Methodological quality of the cohort studies was assessed using the NOS. The NOS was slightly modified to allow the assessment of the methodological quality of the cross-sectional studies.
6.4.2.1 Cohort studies (18 studies)

Overall, the majority of the cohort studies’ participants were representative of the wider population of doctors completing postgraduate medical examinations. Only 2/18 studies were felt not to meet this criteria because it was not clearly stated how the study population was selected [296, 301]. Most of the studies obtained data about the study participants' gender via external and secure records (12/18 studies) [286, 287, 284, 127, 290, 292, 294, 285, 283, 299-301]. Only 5/18 studies examined candidates’ performance at their first attempt of an examination [286, 288, 284, 127, 294], but eight of the remaining studies did analyse their results by attempt number. Overall, I felt the cohort studies did not display major problems of selection bias. The majority of the studies adjusted their results for at least one possible confounding factor (14/18 studies) [287, 284, 126, 127, 290-292, 294, 296, 285, 283, 299, 300, 261], but the factors controlled for varied between the studies. Finally, 15/18 studies obtained examination performance data from existing and record sources [286-288, 284, 126, 127, 290-292, 294, 285, 283, 299-301], and 17/18 studies had ≥80% follow-up rate [286-288, 284, 126, 127, 290-292, 294, 296, 297, 285, 283, 299-301]. I therefore felt that the vast majority of the cohort studies did not show outcome bias. See Table 6.3.
<table>
<thead>
<tr>
<th></th>
<th>Selection</th>
<th></th>
<th></th>
<th>Demonstraten</th>
<th>Comparability</th>
<th>Outcome</th>
<th>Follow-up long</th>
<th>Adequacy of follo</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Representativeness of exposed cohort</td>
<td>Selection of non exposed cohort</td>
<td>Ascertainment of exposure</td>
<td>Demonstration outcome not present at start of study</td>
<td>Comparability of cohorts</td>
<td>Ascertainment of outcome</td>
<td>Adequacy of follow up of cohorts</td>
<td></td>
</tr>
<tr>
<td>Andriole et al. [286] (2005)</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td></td>
</tr>
<tr>
<td>Ayres et al. [287] (1996)</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td></td>
</tr>
<tr>
<td>Baker et al. [288] (2015)</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td></td>
</tr>
<tr>
<td>Bowhay et al. [284] (2009)</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td></td>
</tr>
<tr>
<td>Dewhurst et al. [126] (2007)</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td></td>
</tr>
<tr>
<td>Esmail et al. [127] (2013)</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td></td>
</tr>
<tr>
<td>Gene Hern et al. [290] (2009)</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td></td>
</tr>
<tr>
<td>Hawtin et al. [291] (2014)</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td></td>
</tr>
<tr>
<td>Holland et al. [292] (2002)</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td></td>
</tr>
<tr>
<td>McDonald et al. [294] (2007)</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td></td>
</tr>
<tr>
<td>Mick et al. [296] (1991)</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td></td>
</tr>
<tr>
<td>Oyebode et al. [297] (2007)</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td>✕</td>
<td></td>
</tr>
<tr>
<td>Selection</td>
<td>Representativeness of exposed cohort</td>
<td>Selection of non exposed cohort</td>
<td>Ascertainment of exposure</td>
<td>Demonstrator outcome not present at start of study</td>
<td>Comparability</td>
<td>Outcome</td>
<td>Follow-up long enough for outcomes to occur</td>
<td>Adequacy of follow up of cohorts</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------------------------------</td>
<td>---------------------------------</td>
<td>---------------------------</td>
<td>-------------------------------------------------</td>
<td>---------------</td>
<td>-------------</td>
<td>---------------------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Pico et al. [285] (2010)</td>
<td>*</td>
<td></td>
<td></td>
<td>**</td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Pope et al. [283] (2014)</td>
<td>*</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Shaw et al. [299] (2014)</td>
<td>*</td>
<td></td>
<td></td>
<td>**</td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Shellito et al. [300] (2010)</td>
<td>*</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Shiroma et al. [301] (2010)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Veloski et al. [261] (2000)</td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>
6.4.2.2 Cross-sectional studies (10 studies)

Overall, the majority of the cross-sectional studies selected a representative study population (9/10 studies) [262, 293, 295, 260, 263, 298, 302, 303, 264], and the study population included all eligible participants, resulting in large sample studies (7/10 studies) [262, 295, 260, 263, 298, 302, 264]. 7/10 studies compared respondents to non-respondents, or had no non-respondents [262, 295, 260, 263, 298, 303, 304]. All but one of the studies described how they obtained data on the study participants' gender [262]. Overall, I felt the cross-sectional studies were not subject to selection bias. 8/10 studies adjusted their findings for at least one potential confounding factor [262, 295, 260, 263, 298, 302, 303, 264]. I also felt that the majority of the cross-sectional studies did not display outcome bias [289, 262, 295, 260, 263, 298, 302, 303], with 8/10 studies obtaining examination performance data from external and secure records, and all but one of the studies had a response rate ≥80% [293]. See Table 6.4.
Table 6.4 Methodological assessment of cross-sectional studies using a modified version of the Newcastle-Ottawa Scale (NOS)

<table>
<thead>
<tr>
<th>Study</th>
<th>Selection</th>
<th>Comparability</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Representativeness of sample</td>
<td>Confounding factors controlled for</td>
<td>Ascertainment of outcome</td>
</tr>
<tr>
<td></td>
<td>Sample size</td>
<td>Non-respondents</td>
<td>Ascertainment of exposure</td>
</tr>
<tr>
<td>Bissant et al. [289] (2006)</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Boulet et al. [262] (2005)</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Kim et al. [293] (2015)</td>
<td>*</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>Menzies et al. [295] (2015)</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Richens et al. [260] (2016)</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Rothman et al. [263] (1995)</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Rushd et al. [298] (2012)</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Siriwardena et al. [302] (2012)</td>
<td>*</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>Tyrer et al. [303] (2002)</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Van Zanten et al. [264] (2003)</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>
6.4.2.3 Summary considerations on study quality

Overall, I felt that the majority of the cohort and cross-sectional studies were of reasonably high methodological quality, with limited evidence of selection or outcome bias.

6.4.3 Meta-analysis

The objective of this meta-analysis was to report the difference in pass rates between male and female candidates at postgraduate medical examinations. Of the 28 studies that were included in the literature review, 14 reported data that allowed the calculation of a measure of effect for pass rates and were included in the meta-analysis [287, 284, 126, 127, 291, 292, 295, 296, 298, 299, 301, 303, 289].

The corresponding authors of the studies that were not included in the meta-analysis were contacted via email (using the email address published with the study) to request further data. If the email failed to deliver or I received no response after 2 weeks, I searched the Internet (using Google) for an alternative email address for the corresponding author and I sent follow-up emails to both the original and the alternative email addresses. If the follow-up emails received no response after 2 weeks or the author replied they would look if further data was available, but did not get in contact after 6 months, I assumed that no further data was available. No response was received from 7/14 studies [288, 126, 293, 294, 283, 263, 264]. Authors from 5/14 studies responded, but either no further data was available or further data was not readily available [286, 290, 297, 285, 302]. The author from 1/14 studies provided additional data, but the data provided did not include the postgraduate medical examination of interest [261].

Of the 14/28 studies that were not included in the meta-analysis 7/14 did not provide sufficient data to be included in the meta-analysis [290, 293, 294, 297, 260, 263, 261], and 7/14 provided data on mean scores by gender, rather than pass rates [286, 288, 302, 262, 285, 283, 264].
Of the 7/14 studies that did not provide enough data, 4/7 studies only provided a statement of the direction of the gender effect [293, 297, 263, 261], and 3/7 provided a p-value to support their conclusion, but did not report absolute numbers [290, 294, 260]. 2/7 studies analysed clinical examinations [263, 261], 3/7 looked at written examinations [290, 293, 294], 2/7 studies analysed a written and clinical examination [297, 260]. Overall, 4/7 studies found women performed better than men [290, 297, 263, 261], 2/7 studies reported there was no difference in examination performance between the genders [293, 294], the remaining study found men performed better at the written assessment, but women performed better at the clinical assessment [260].

The 4/7 studies that provided mean scores by gender analysed a clinical examination [286, 262, 283, 264], 2/7 studies analysed a written examinations [288, 302] and 1/7 studies did not report the format of the examination [285]. 4/7 studies found women performed better than men [262, 283, 302, 264], and the remaining 3/7 studies reported there was no difference in examination performance between the genders [286, 288, 285].

The 14/28 studies included in the meta-analysis observed 68,411 doctors attempting a postgraduate medical examination (range: 50 to 48,509; median: 1128), of which 23,261 doctors passed the examination. 3/14 studies analysed two parts of a postgraduate medical examination separately [126, 303, 298]xii, therefore these 14 studies yielded 16 findings. For the characteristics of the studies included in the meta-analysis, see Table 6.5.

---

xii For the study by Dewhurst et al. [126] only the findings from the clinical assessment were eligible to be included in the meta-analysis.
Table 6.5 Characterisitcs of the eligible studies included in the meta-analysis

<table>
<thead>
<tr>
<th>First author, year (country)</th>
<th>Study design</th>
<th>Specialty</th>
<th>Postgraduate exam</th>
<th>Postgraduate exam type</th>
<th>Total number study participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ayres et al. [287], 1996 (UK)</td>
<td>Cohort</td>
<td>Public Health</td>
<td>MFPHM Part 1</td>
<td>Written assessment</td>
<td>431</td>
</tr>
<tr>
<td>Bessant et al. [289], 2006 (UK)</td>
<td>Cross-sectional</td>
<td>General Medicine</td>
<td>MRCP(UK) PACES</td>
<td>Clinical assessment</td>
<td>483</td>
</tr>
<tr>
<td>Bowhay et al. [284], 2009 (UK)</td>
<td>Cohort</td>
<td>Anaesthetics</td>
<td>FRCA Part 1</td>
<td>Written assessment</td>
<td>3303</td>
</tr>
<tr>
<td>Dewhurst et al. [126], 2007 (UK)</td>
<td>Cohort</td>
<td>General Medicine</td>
<td>MRCP(UK) PACES</td>
<td>Clinical assessment</td>
<td>2528</td>
</tr>
<tr>
<td>Esmail et al. [127], 2013 (UK)</td>
<td>Cohort</td>
<td>General Practice</td>
<td>MRCGP</td>
<td>Clinical assessment</td>
<td>5095</td>
</tr>
<tr>
<td>Hawtin et al. [291], 2014 (UK)</td>
<td>Cohort</td>
<td>Radiology</td>
<td>FRCR (UK) Part 2B</td>
<td>Other</td>
<td>1492</td>
</tr>
<tr>
<td>Holland et al. [292], 2002 (UK)</td>
<td>Cohort</td>
<td>Public Health</td>
<td>MFPHM Part 2</td>
<td>Other</td>
<td>238</td>
</tr>
<tr>
<td>Menzies et al. [295], 2015 (UK)</td>
<td>Cross-sectional</td>
<td>Paediatrics</td>
<td>MRCPCH Part 1B</td>
<td>Written assessment</td>
<td>2056</td>
</tr>
<tr>
<td>Mick et al. [296], 1991 (USA)</td>
<td>Cohort</td>
<td>All specialties</td>
<td>FMGEMS</td>
<td>Written assessment</td>
<td>48509</td>
</tr>
<tr>
<td>Rushd et al. (Part 1) [298], 2012 (UK)</td>
<td>Cross-sectional</td>
<td>Obstetrics &amp; Gynaecology</td>
<td>MRCOG Part 1</td>
<td>Written assessment</td>
<td>1336</td>
</tr>
<tr>
<td>Rushd et al. (Part 2) [298], 2012 (UK)</td>
<td>Cross-sectional</td>
<td>Obstetrics &amp; Gynaecology</td>
<td>MRCOG Part 2</td>
<td>Written assessment</td>
<td>821</td>
</tr>
<tr>
<td>Shaw et al. [299], 2014 (UK)</td>
<td>Cohort</td>
<td>General Practice</td>
<td>CSA</td>
<td>Clinical assessment</td>
<td>116</td>
</tr>
<tr>
<td>Shellito et al. [300], 2010 (USA)</td>
<td>Cohort</td>
<td>Surgery</td>
<td>ABS</td>
<td>Not reported</td>
<td>62</td>
</tr>
<tr>
<td>Shiroma et al. [301], 2010</td>
<td>Cohort</td>
<td>Psychiatry</td>
<td>ABPN</td>
<td>Other</td>
<td>50</td>
</tr>
<tr>
<td>First author, year (country)</td>
<td>Study design</td>
<td>Specialty</td>
<td>Postgraduate exam</td>
<td>Postgraduate exam type</td>
<td>Total number study participants</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------------</td>
<td>------------</td>
<td>-------------------</td>
<td>------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Tyrer et al. (Part 1) [303], 2002 (UK)</td>
<td>Cross-sectional</td>
<td>Psychiatry</td>
<td>MRCPsych (Part 1)</td>
<td>Both written and clinical assessments</td>
<td>1128</td>
</tr>
<tr>
<td>Tyrer et al. (Part 2) [303], 2002 (UK)</td>
<td>Cross-sectional</td>
<td>Psychiatry</td>
<td>MRCPsych (Part 2)</td>
<td>Not reported</td>
<td>763</td>
</tr>
</tbody>
</table>
6.4.3.1 Summary effect estimates

A random-effects model found a pooled odds ratio of 1.54 (95% CI 1.12 to 2.12), suggesting that female doctors obtained higher pass rates than male doctors at postgraduate medical examinations. 12/16 findings demonstrated that female doctors had higher pass rates than male doctors in postgraduate medical examinations [287, 126, 127, 291, 292, 298-301, 303, 289], however only 8/12 had findings that were statistically significant [287, 126, 127, 298, 299, 303, 289]. The range of the odds ratios was 1.42 to 4.71 for those studies that had significant results.

3/16 findings found that male doctors obtained higher pass rates than female doctors at postgraduate medical examinations [284, 296, 298], and 2/3 had findings that were statistically significant [284, 296]. The range of odds ratios for these two studies was 0.75 to 0.81.

1/16 findings found an odds ratio of 1.00, demonstrating no differences in pass rates in postgraduate medical examinations performance between the genders [295] (See Figure 6.2).

A high degree of heterogeneity was present ($Q=568.5$, d.f.=15, p<0.001; $I^2=97.4\%$).
6.4.3.2 **Subgroup analyses**

6.4.3.2.1 **Country (UK, USA)**

In the UK stratum there was a significant overall effect of gender, with female doctors having increased odds of passing postgraduate medical examinations (OR: 1.55, 95% CI: 1.14 to 2.12). There was a high degree of heterogeneity in the UK stratum ($I^2: 95.5\%$). In the USA stratum there was not a significant overall effect of gender (OR: 1.71, 95% CI: 0.49 to 6.01). However, this stratum only contained 3 findings, and therefore the result should be interpreted with caution.

6.4.3.2.2 **Examination format (clinical, written, both, other, not reported)**

There was a significant overall effect of gender in the clinical assessments, both written and clinical assessments, and examination format not reported.
strata, with female doctors having increased odds of passing postgraduate medical examinations (clinical assessments OR: 2.63, 95% CI: 1.70 to 4.06; both written and clinical assessments OR: 1.42, 95% CI: 1.11 to 1.81; not reported OR: 1.96, 95% CI: 1.46 to 2.61). It should be highlighted that all the findings reporting clinical assessments, or clinical and written assessments combined were from UK-based studies. It should also be noted that the number of findings per stratum was small, with only 3 findings in the clinical assessments stratum, 1 finding in the both written and clinical assessments stratum, and 2 findings in the not reported stratum. Both the written assessments and other assessments strata demonstrated that there was no evidence of a gender effect (written assessments OR: 1.01, 95% CI: 0.81 to 1.26; other assessments OR: 1.28, 95% CI: 0.98 to 1.67).

6.4.3.3 Bias

I used a funnel plot to assess possible bias (see Figure 6.3). The distribution of the studies did not fit a funnel shape with the majority of the studies appearing outside of the funnel. This reflects the high degree of heterogeneity in the studies. There was a clustering of studies towards the top of the funnel, with a relative sparsity towards the bottom, showing an absence of smaller studies. This may reflect the nature of the data typically available to examine this research question. There was evidence of substantial publication bias.
6.5 Discussion

6.5.1 Summary of main results

6.5.1.1 Literature review

The findings are based on 28 observational studies [286-289, 262, 284, 126, 127, 290-297, 285, 283, 260, 263, 298-303, 264, 261] and provided a total of 32 separate results. Overall women doctors perform better than men doctors at postgraduate medical examinations, with 17/32 finding women outperforming men [287, 262, 126, 127, 290, 291, 297, 283, 260, 263, 298, 299, 301-303, 264, 261], of which 15/32 report a statistically significant effect [287, 262, 126, 127, 290, 291, 283, 260, 298, 299, 301-303, 264, 261]. Only 2/32 results reported that men had statistically significantly better examination performance than women [284, 260].

The results were then analysed by country where the examination was set, to investigate whether different medical systems differ in how the genders perform at postgraduate medical examinations. When analysing the results
for USA and Canada, the majority of the studies found no significant difference between the genders and postgraduate medical examination performance [286, 288, 293, 294, 285, 300]. However, the majority of the results for UK and Ireland demonstrated that the performance of women doctors at postgraduate medical examinations was statistically significantly better than men doctors [287, 126, 127, 291, 283, 260, 298, 299, 302, 303].

When the results were examined by format I was able to demonstrate that the majority of studies that analysed performance at written examinations found no significant difference in examination performance between the genders [288, 126, 294, 295, 298, 302, 293]. However, the majority of studies investigating clinical examination performance [262, 126, 283, 260, 299, 264, 261, 127], both written and clinical examination performance together [303], and performance at alternative examination formats [291, 301] found women doctors had statistically significant better performance than men doctors. For those studies that did not report the examination format of the postgraduate medical examination being analysed the majority finding was that there was no gender difference in examination performance [285, 300].

6.5.1.2 Meta-analysis

Women doctors had over 1.5 times the odds (pooled OR: 1.54, 95% CI: 1.12 to 2.12) of passing a postgraduate medical examination compared to men. However, there was significant heterogeneity in the meta-analysis. When the results were analysed by examination format, there was no evidence for a gender difference in the pass rates at written assessments (OR: 1.01, 95% CI: 0.81 to 1.26), but women had over 2.5 times the odds of passing clinical assessments compared to men (OR: 2.63, 95% CI: 1.70 to 4.06). However, it should be noted that all the studies that analysed clinical assessment performance were from the UK. Interpretations of the results for the other examination formats must be done with care, due to the small number of findings per strata.
Results from the UK generally show a higher pass rate for women. In the meta-analysis, all the findings analysing clinical examinations and the one finding combining clinical and written assessments were from the UK, so there is a possible confound between country and assessment type. However, 3 USA findings [262, 264, 261] and one Canadian finding [263] on clinical examinations that were not included in the meta-analysis found female doctors outperformed male doctors and just one USA finding found no difference [286]. Among findings on written examinations not included in the meta-analysis, one study [260] from the UK and Ireland found male doctors outperformed female doctors and the other UK study found mixed results (no gender difference in Part 1, female doctors outperforming male doctors on Part 2) [302]. This suggests the effect is probably better explained by assessment format rather than by country.

6.5.2 Overall completeness and applicability

6.5.2.1 Literature search

I chose to limit my database searching to three electronic databases; MEDLINE, CINAHL and ERIC. It could be argued that because database searching was my main method of capturing data, I should have increased the number of databases searched in order to capture additional data. However, it has been shown that as the number of database searches increases, the additional data captured by a more exhaustive database searching will only contribute a small percentage of the overall data captured [305]. Therefore I felt that any loss from using only three electronic databases is likely to be marginal. It has also been shown that a combination of hand searching and database searching is required to identify relevant reports [306]. I thus supplemented my search strategy by choosing to perform backward and forward cited reference searching.

I chose to only include reports of original data published in peer-reviewed journals, due to limited resources and time, as such grey or unpublished literature was not included in my search. It has been reported that the exclusion of grey literature results in inflated measures of effect in meta-
analyses [307, 308], as larger studies with larger effect sizes are more likely to be submitted for publication in peer-reviewed journals. However, including grey literature brings with it the risk of introducing studies who themselves may introduce bias due to possibly lower methodological quality [309]. That said, the inclusion of grey literature may have identified reports from the individual examination bodies, and inclusion of any identified reports may have reduced the risk of publication bias.

Another limitation was the exclusion of studies not available in English. This may partly explain why all of the studies included in this report are from English-speaking countries. The exclusion of studies not published in English introduces the possible risk of bias in the findings of this review. It would be prudent to only apply the findings of this review to medical postgraduate examinations from Canada, Ireland, UK and USA.

6.5.2.2 Study selection

Over 107,000 postgraduate medical examination attempts were captured by the review. This large number allows me to draw meaningful conclusions from the results. The majority of the studies attempted to collect data that was representative of a wider population (multiple countries, country-wide, multiple states). The studies in this review capture data from four different countries, however the vast majority were from the UK and USA. I have demonstrated that when examining the results by country (USA and Canada, compared to UK and Ireland) the majority of USA and Canadian studies concluded there was no difference between men and women and their performance at postgraduate medical examinations, whereas the majority of studies from the UK and Ireland found that women had significantly better performance than men at postgraduate medical examinations. However, it must be acknowledged there may be a confounding effect of examination format.

The examination analysed by the individual studies varied in format, specialty and purpose, therefore it could be argued that it is not appropriate to
examine all the studies together. However, because of the wide variation of the examinations analysed, there were not sufficient studies per examination type to be examined and for meaningful conclusions to be drawn.

It could be argued that it may be inappropriate to conduct a meta-analysis on these data due to the variation in examination type, examination format, specialties taking the examinations, and the examinations' countries. However, I felt that despite these possible flaws, the meta-analysis is a worthwhile exercise because it gives a general feel of how the performance at postgraduate medical examinations may differ between the genders. A consistent finding from the studies' results was that female doctors perform significantly better at clinical postgraduate medical examinations than their male counterparts.

6.5.3 Conclusions

To my knowledge this is the first systematic literature review and meta-analysis examining the association between doctors' gender and performance at postgraduate medical examinations. It demonstrates that overall female doctors perform significantly better at postgraduate medical examinations, however when the studies were analysed by examination format, the effect is seen in clinical postgraduate medical examinations, but not written postgraduate medical assessments. When the studies were analysed by country, the effect is seen in UK and Ireland, but not in the USA and Canada. However, these two factors are confounded with most studies of clinical examinations done in the UK, and there is marked heterogeneity and probably publication bias.

These results strongly suggest a gender difference in the performance at clinical postgraduate medical examinations. It is likely that the performance of doctors at postgraduate medical examinations is a reflection of medical school performance, which in turn is a reflection of school performance. However, given this reasoning one would expect to observe a gender difference in all postgraduate medical examinations, and not such a marked
difference between written and clinical assessments. Clinical assessments are a measure of doctors’ practical clinical knowledge and skills, but performance at clinical assessments is also influenced by theoretical knowledge [111]. It is plausible that female doctors perform better at clinical assessments because they are better able to use and translate theoretical medical knowledge into clinical skills. Another possibility is that women may demonstrate other skills during the clinical assessment, which supports and enhances their examination performance. For example, it has been argued that female doctors differ from men in terms of communication style [243] and are more likely to have an empathic approach [310, 311]. This style of doctor-patient communication may encourage patient disclosure enabling female candidates to gain more information about the patient’s case and therefore perform better at the examination. Another possible explanation behind the observed gender difference in clinical assessment performance may be due to uncontrolled confounders. When discussing the methodological quality of the studies included in this review, I highlighted that though many of the studies did adjust their findings for confounders, the confounders adjusted for varied between studies. It is known that other demographic factors, such as ethnicity, number of years since primary medical qualification, and world region of primary medical qualification influence candidates’ performance at clinical assessments [127, 131, 111]. It is therefore plausible that other demographic factors are affecting the gender difference found in clinical assessment performance. It could also be argued that examiners of clinical assessments may be biased towards female candidates, however there is no evidence to suggest this [124, 80].

The findings of this literature review and meta-analysis provide support for my earlier findings in Chapter 5, where I demonstrated that in at a high-stakes specialty clinical examination, women performed better than men. More detailed information is needed to understand the reasons why female doctors perform better at postgraduate medical examinations, particularly clinical examinations. However, the first step is to recognise that there is an established gender difference in pass rates at postgraduate clinical medical
examinations, regardless of country, specialty and examination type. Doctors who complete these examinations and the examination boards who administer the examinations need to consider whether the gender difference in pass rates is not the result of discrimination, and whether it is preventing advancing equality of opportunity between men and women. The medical community also need to consider the possible reasons behind the existence of this gender gap.

So far this doctoral research has demonstrated that there is gender difference in medico-legal action against doctors (Chapter 3 and Chapter 4) and in postgraduate clinical examination performance (Chapter 5 and Chapter 6), with women doctors being less likely to face medico-legal action and more likely to pass postgraduate clinical examinations. Interestingly a study from the UK has found that performance at postgraduate clinical assessments is an independent predictor of disciplinary action [67], demonstrating that postgraduate clinical examination performance and the likelihood of facing medico-legal action are linked. One possible hypothesis to explain this link could be that the skills and attributes that are highly valued during a postgraduate clinical examination are protective from medico-legal action.

The following chapter of this thesis will investigate whether the gender difference in the performance at postgraduate medical examinations observed in Chapter 5 and Chapter 6 is replicated in a population of GPs whose are undergoing a performance assessment by the GMC following concerns being raised about their professional performance.
Chapter 7 – Gender differences in fitness to practise test scores – a cohort study of GPs

7.1 Chapter summary

7.1.1 Background
Every year the GMC investigates complaints made against doctors practising in the UK; those whose practise raises concerns related to issues of competence or performance may be required to complete tests of competence (ToC). For GPs, the ToC includes a written knowledge test and a practical clinical assessment called the simulated surgery.

7.1.2 Aims
To examine gender differences in the assessment scores of GPs under investigation by the GMC, compared to GPs not under investigation; and whether scores mediate any relationship between gender and sanction likelihood.

7.1.3 Design and setting
Retrospective analysis of administrative ToC GP data. ToC are written and clinical assessments taken by doctors under investigation by the GMC who have significant performance concerns, and a comparator group of volunteer doctors.

7.1.4 Methods
Analysis of variance models to compare written and clinical components ToC performance by gender and GP group (under investigation and volunteers). Path analysis to explore the relationship between gender, written and clinical ToC performance, and investigation outcome.

7.1.5 Results
On the written test, women GPs under investigation outperformed men (Cohen’s d=0.28, p=0.01); there was no gender difference in the volunteer group (Cohen’s d=0.02, p=0.93). On the clinical assessment, women
outperformed men in both GP groups (Cohen’s d=0.61, p<0.0001). In GPs under investigation a higher clinical score predicted remaining on the register without a warning or sanction; with no independent effect of gender controlling for assessment performance.

7.1.6 Conclusion

Women outperform men on clinical assessments, even among GPs with generally very poor performance. The results suggest male GPs under investigation may have particularly poor knowledge.

The finding that women GPs perform better at the clinical component of the ToC, regardless of whether they were under investigation or volunteers, mirrors the findings of Chapter 5 and Chapter 6. These findings strengthen the hypothesis that the skills that are highly valued in a postgraduate clinical assessment are protective from disciplinary action.
7.2 Introduction

Erasure or suspension from the UK medical register can result in a loss of career and income; lesser sanctions (conditions and undertakings) and warnings can also limit doctors’ careers and may hamper career progression [312, 313]. Female doctors are one third less likely to receive a warning or sanction against their medical registration compared to male doctors [9] (see Chapter 3), however the reasons underlying this difference are not clear. I will explore three hypotheses raised in the earlier chapters of this thesis.

The first hypothesis, which I raised in Chapter 3, suggested that a gender difference in the source of referral to the GMC might go toward explaining the observed gender difference in disciplinary action against doctors. The GMC receives complaints about doctors from a wide range of sources, which can be broadly divided into four main groups: the public; person acting in a public capacity (PAPC) (e.g. police, employer); doctors; and other (e.g. media). Published evidence shows that the source of the referral to the GMC is associated with the likelihood of the complaint resulting in an investigation [183].

The second hypothesis, which I also raised in Chapter 3, is that a difference in allegation types between the genders could partially explain the gender difference in disciplinary action. When the GMC receives a complaint about a doctor, the allegations are assessed and assigned to different categories. One or more allegations may be assigned to help categorise the case [183]. Evidence from previous studies suggests that certain allegations against a doctor are more likely to result in disciplinary action [189, 191-195].

The third hypothesis I will examine is whether men, on average, have greater gaps in their medical knowledge and clinical skills, and that this contributes to gender differences in disciplinary action. Evidence from the literature has found performance at both undergraduate and postgraduate medical examinations is associated with future disciplinary action [66, 74], suggesting that those doctors who face disciplinary action may have a medical
knowledge and clinical skills deficit. Research has also demonstrated that women perform better during their undergraduate training at medical school [107]. At postgraduate level it appears there is no difference between the genders at written examinations [126, 295, 298] but, as I have demonstrated in Chapter 5 and Chapter 6, women perform better at clinical assessments [126, 283, 260, 299].

7.2.1 Study aims

To gain a better understanding of the reasons for the gender difference in disciplinary action, I aimed to explore:

- Whether the referral routes to the GMC differ between the genders.
- Whether there are gender differences in the types of allegations referred to the GMC.
- The presence and magnitude of gender differences in performance on the written and clinical components of the GMC’s ToC.
- Whether ToC gender differences are comparable for doctors currently under investigation by the GMC and those doctors not currently under investigation and without restrictions on their medical registration, who had volunteered to complete the ToC.
- Whether any relationship between gender and disciplinary action is mediated by ToC performance.

7.3 Methods

7.3.1 Study design, setting and source of data

I conducted this retrospective cohort study using data from the GMC and RDME (who has held a contract with GMC for several years to develop and administer the ToC, see section 1.5.4). Permission to use the data for research purposes was obtained from both the GMC and RDME. The data was provided by the GMC for the purposes of this research study, after I signed a confidentiality agreement.
The data used to conduct this study were obtained from several sources and merged to create the research study dataset. Details of the data used to conduct this study are described below:

- All doctors under investigation by the GMC, who were required to complete ToC from 2008 to 2013. I collapsed this dataset (without loss of any information) to create one row per doctor, and then reduced it to only include those doctors who were practising as GPs at the time of referral to the GMC.
- Updated case outcome for investigated GPs.
- ToC knowledge test (KT) scores and simulated surgery (SS) scores for GPs who were investigated by the GMC from 2008 to 2013. I extracted these scores from the individual ToC reports.
- Demographic data for volunteer GPs.
- ToC KT scores for volunteer GPs, which I extracted from the individual databases for each General Practice pilot between 2008 to 2013.
- ToC SS scores for volunteer GPs.

7.3.1.1 The GMC Fitness to Practise investigation process

As described in section 2.7.1, the GMC’s role is to ensure proper standards in the practice of medicine in the UK, and holds the powers to take action against a doctor’s registration if the doctor’s fitness to practise is found to be impaired. If a doctor’s professional performance is called into question during a GMC investigation, the doctor may be required to undergo a performance assessment, consisting of two main parts: a peer review and the ToC (for details see [314]).

The GMC’s decisions about the investigated doctor’s fitness to practise, and the outcome of the investigation, are based on the doctor’s performance in both parts of the performance assessment (peer review and ToC).

At any point during an investigation it is possible for the doctor to apply to be removed from the medical register (voluntary erasure), which if granted
means the doctor does not have to complete the investigation [315]; however, should the doctor wish to return to the register and practise medicine, they would have to demonstrate they are fit to practise.

7.3.1.1.1 Tests of competence (ToC)

The ToC are designed to identify gaps in the knowledge and clinical skills of doctors under investigation for poor performance [158]. It comprises a written KT and an OSCE. In General Practice there is also the SS, which simulates a typical GP surgery, and assesses the doctor’s clinical, management and communication skills [316]. There were two versions of the SS in circulation during the timeframe of this study. Most doctors completed version 1; version 2 was primarily used to reassess doctors. I did not use the OSCE for this study because the OSCE and SS have been found to have significant overlap [316].

To ensure the ToC are fair and fit for purpose, they are regularly ‘piloted’ with doctors who have no known fitness to practise concerns and who volunteer and are paid to complete the ToC. There is no ToC pass mark: the performance of the doctor under investigation is compared with reference groups of volunteers who have completed the same questions, and this contributes to the findings of the GMC investigation [317]. For further information about the ToC see [169, 318, 158].

7.3.1.1.2 Outcome of the GMC FtP investigation

I examined whether the outcome of the GMC investigation was related to gender, KT score and SS score, and whether ToC performance mediated the relationship between gender and outcome.

I collapsed the outcome into a nominal variable: no sanction imposed, warning or sanction imposed, and no longer registered on the LRMP (see Table 7.1 for further details). I combined warnings and sanctions because warnings are recorded on the LRMP for a period of time and can affect career progression [312].
**Table 7.1 Definition of outcome types**

<table>
<thead>
<tr>
<th>Outcome type (following GMC FtP investigation)</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>No sanction imposed</td>
<td>No impairment found during the investigation and no restrictions imposed on the doctor’s medical registration.</td>
</tr>
<tr>
<td>Warning / sanction imposed</td>
<td>Warning issued, but no restrictions on the doctor’s medical registration. <strong>OR</strong> Sanction imposed resulting in a restriction on the doctor’s medical registration. Sanctions include undertakings, conditions, suspension from the LRMP, and erasure from the LRMP. (See 2.7.1 and [319, 9] for further details.)</td>
</tr>
<tr>
<td>No longer registered on the LRMP</td>
<td>Administrative erasure from the LRMP, voluntary erasure from LRMP or doctor deceased. (See [319]</td>
</tr>
</tbody>
</table>

**LRMP**: List of Registered Medical Practitioners

### 7.3.2 Population

All doctors working as GPs between 2008 and 2013 who were under investigation by the GMC and were required to complete the ToC between 2008 and 2014. I focused on General Practice because it is the largest specialty and attracts the most complaints [182].

The comparison group was GPs not currently under investigation and with no restrictions on their medical registration, who voluntarily completed either the KT and/or SS assessment as part of the ToC pilots.

### 7.3.3 Selection of variables

The selection of co-variates was constrained by the variables collected and made available by the GMC and the RDME, UCL. I selected the co-variates included in the study prior to any statistical analysis, and my selection was based on theories raised in the earlier chapters of this thesis.

The following data was available:

1. GPs being investigated by the GMC:
   Gender, the world region where they had obtained their primary medical qualification, the number of years since they had received their primary medical qualification, the referral source to the GMC, the
allegation category/ies, the number of allegations against the doctors, KT scores, SS scores, and the outcome of the investigation.

II. Volunteer GPs:
   Gender, KT scores and/or SS scores.

I proceeded by performing a variable reduction process on the selected co-variates, with the aim of reducing the number of categories into meaningful categories:

‘Time since qualification’: this continuous variable was collapsed into five categories; ‘1-10 years’ representing the bulk of time a doctor would likely be in Specialty Training. The subsequent categories were divided into 10-year blocks. This variable was renamed ‘Number of years since receipt of PMQ’.

‘Enquiry source type’: this variable was collapsed into three categories; ‘PAPC’, ‘Colleague’, and ‘Public and Other’. This variable was renamed ‘Referral source’.

‘Allegation Category’ and ‘Allegation Type’: these two variables were merged and collapsed into five categories; ‘Honesty’, ‘Professional performance’, ‘Clinical competence’, ‘Patient and colleagues’, and ‘Other’. These categories reflect the categories used by the GMC in their annual report [182]. To categorise the allegations, two researchers (myself and LM) independently allocated each allegation to one of the eight broader categories used by the GMC. The K statistic demonstrated a good level of agreement (K=0.61). Any disagreements about the allegation category were resolved through discussion. We found that only five of the possible eight categories were relevant for the study population.

‘Number of allegations’: this variable was created to capture the fact that some of the investigated doctors had more then one allegation against them. It consists of three categories; ‘1’, ‘2-3’, and ‘≥4’.
‘Last Decision’, ‘Hearing Outcome’, and ‘Case Closure Reason’: these three variables were collapsed into one variable named ‘Outcome’ and it consisted of three categories; ‘No sanction’, ‘Sanctioned’ and ‘No longer registered’.

‘KT score’: the raw scores were standardised into z-scores.

‘SS score’: percentage score.

7.3.4 Statistical methods
7.3.4.1 Categorical data
I performed bivariate analyses to look for associations between doctors’ gender and the categorical variables.

7.3.4.2 Continuous data
I z-transformed the KT scores for both GP groups (under investigation; volunteers). I performed analysis of variance to test the performance at KT and SS by doctors’ gender and GP group, checking for interaction effects.

7.3.4.3 Logistic regression
I built multinomial and binomial logistic regression models, and then a path analysis model. Path analysis was performed using multiple regression models to further explore the associations between gender, outcome type and ToC scores (KT and SS). To build a path analysis model multiple regression was used, with each variable being set as the dependent variable in turn. Paths were included in the model if they were significant at p<0.05.

I used the STROBE statement [199] to guide the study report.

7.4 Results
7.4.1 Descriptive results
7.4.1.1 Doctors under investigation
120 GPs completed a KT and SS; 24 (20%) were female and 96 (80%) were male. 3/120 (2.5%) GPs’ KT scores were missing, and a further 5/120 (4%)
GPs’ had completed version 2 of the SS; their data were excluded. See Figure 7.1. See Table 7.2 and Table 7.3 for details of these doctors’ demographics.

Figure 7.1 Flow diagram of how the population of GPs under investigation who completed KT and who completed SS

120 GPs under investigation completed a ToC

3 GPs under investigation KT missing

5 GPs under investigation completed SSv2

excluded

117 GPs under investigation used for analysis of KT (see Section 7.4.3.1)

115 GPs under investigation used for analysis of SS (see Section 7.4.3.2)

112 GPs under investigation KT and SS data used for final analysis of KT, SS and outcome (see Section 7.4.3.3)
### Table 7.2 Demographics of the 120 GPs who completed a KT and SS between 2008 and 2014 as part of a GMC investigation process, by outcome type

<table>
<thead>
<tr>
<th>Variable</th>
<th>Outcome¹</th>
<th>Total number of GPs N=120 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No sanction N=22 (%)</td>
<td>Sanction N=72 (%)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>16 (17%)</td>
<td>58 (60%)</td>
</tr>
<tr>
<td>Female</td>
<td>6 (25%)</td>
<td>14 (58%)</td>
</tr>
<tr>
<td><strong>Number of years since PMQ</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-10 years</td>
<td>0 (0%)</td>
<td>1 (1%)</td>
</tr>
<tr>
<td>11-20 years</td>
<td>6 (27%)</td>
<td>11 (15%)</td>
</tr>
<tr>
<td>21-30 years</td>
<td>5 (23%)</td>
<td>27 (38%)</td>
</tr>
<tr>
<td>31-40 years</td>
<td>11 (50%)</td>
<td>24 (33%)</td>
</tr>
<tr>
<td>≥41 years</td>
<td>0 (0%)</td>
<td>9 (13%)</td>
</tr>
<tr>
<td><strong>World region where PMQ received</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UK</td>
<td>9 (41%)</td>
<td>27 (38%)</td>
</tr>
<tr>
<td>EEA</td>
<td>2 (9%)</td>
<td>7 (10%)</td>
</tr>
<tr>
<td>International</td>
<td>11 (50%)</td>
<td>38 (53%)</td>
</tr>
<tr>
<td><strong>Referral source</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAPC</td>
<td>12 (55%)</td>
<td>55 (76%)</td>
</tr>
<tr>
<td>Colleague</td>
<td>4 (18%)</td>
<td>7 (10%)</td>
</tr>
<tr>
<td>Public and Other</td>
<td>6 (27%)</td>
<td>10 (14%)</td>
</tr>
<tr>
<td><strong>Allegation category</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acting honestly &amp; fairly</td>
<td>2 (9%)</td>
<td>4 (6%)</td>
</tr>
<tr>
<td>Professional performance</td>
<td>6 (27%)</td>
<td>30 (42%)</td>
</tr>
<tr>
<td>Clinical competence</td>
<td>3 (14%)</td>
<td>15 (21%)</td>
</tr>
<tr>
<td>Patient &amp; colleagues</td>
<td>5 (23%)</td>
<td>10 (14%)</td>
</tr>
<tr>
<td>Other</td>
<td>6 (27%)</td>
<td>13 (18%)</td>
</tr>
<tr>
<td><strong>Number of allegations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>8 (36%)</td>
<td>27 (38%)</td>
</tr>
<tr>
<td>2 to 3</td>
<td>11 (50%)</td>
<td>22 (31%)</td>
</tr>
<tr>
<td>≥4</td>
<td>3 (14%)</td>
<td>23 (32%)</td>
</tr>
</tbody>
</table>

¹See Table 7.1 for further description of outcome types.

### 7.4.1.2 Volunteer doctors

482 GPs (43% men) completed a KT between 2008 and 2014. Complete data were available for 325/349 GPs who completed a SS between 1997 and 2006. 22/349 were missing gender data, and a further 2/349 were missing complete SS data, and were excluded. See Figure 7.2.
7.4.2 Bivariate analyses

Bivariate analyses of categorical variables were performed to compare male and female GPs. There was no evidence for a statistically significant association between gender and the following variables: number of years since receipt of primary medical qualification ($p=0.232$); world region where received primary medical qualification ($p=0.982$); referral source ($p=0.493$); number of allegations against a doctor ($p=0.335$); allegation category (acting honestly & fairly $p=0.583$; professional performance $p=0.406$; clinical competence $p=0.540$; patient & colleagues $p=0.798$; other $p=0.132$); and outcome ($p=0.581$). The results are presented in Table 7.3, however to ensure that the data remained non-identifiable the variables have been collapsed into binary variables for presentation in Table 7.3. The association between gender and the variables remained not statistically significant ($p>0.05$) when the variables were collapsed into binary form. To ensure the data remains non-identifiable it is not possible to present the results of the
allegation category in Table 7.3 as the number of doctors in some of the allegation categories is less than five.

Table 7.3 The distribution of each variable by gender of the 120 GPs under investigation by the GMC, and the association of individual variables with gender

<table>
<thead>
<tr>
<th>Variable</th>
<th>Gender</th>
<th>p value (X²)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male N=96 (%)</td>
<td>Female N=24 (%)</td>
</tr>
<tr>
<td>Number of years since PMQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-30 years</td>
<td>41 (43%)</td>
<td>12 (50%)</td>
</tr>
<tr>
<td>≥31 years</td>
<td>55 (57%)</td>
<td>12 (50%)</td>
</tr>
<tr>
<td>World region where PMQ received</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UK</td>
<td>38 (40%)</td>
<td>9 (38%)</td>
</tr>
<tr>
<td>Non-UK</td>
<td>58 (60%)</td>
<td>15 (62%)</td>
</tr>
<tr>
<td>Referral source</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAPC</td>
<td>68 (71%)</td>
<td>19 (79%)</td>
</tr>
<tr>
<td>Colleague, Public or Other</td>
<td>28 (29%)</td>
<td>5 (21%)</td>
</tr>
<tr>
<td>Number of allegations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>44 (46%)</td>
<td>7 (29%)</td>
</tr>
<tr>
<td>≥2</td>
<td>52 (54%)</td>
<td>17 (71%)</td>
</tr>
<tr>
<td>Outcome¹</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No sanction</td>
<td>16 (17%)</td>
<td>6 (25%)</td>
</tr>
<tr>
<td>Sanction or no longer registered</td>
<td>80 (83%)</td>
<td>18 (75%)</td>
</tr>
</tbody>
</table>

¹See Table 7.1 for further description of outcome types.

7.4.3 The ToC results

Means and standard deviations for the KT and SS for GPs under investigation and for volunteer GPs are shown in Table 7.4 and Table 7.5, with distributions of scores shown in Figure 7.3.

7.4.3.1 Written assessment (KT)

599 GPs (117 under investigation and 482 volunteers) completed a KT. There was a significant interaction between gender and GP group on the KT: F(1, 595)=5.16, p=0.02 (see Figure 7.4). In the group of GPs under investigation women obtained higher scores than men (mean difference in z-score: -1.24, 95% CI: -3.28 to 0.81, Cohen’s d: 0.28, p=0.01), but in the group of volunteer GPs there was no evidence of a gender difference (mean difference in z-score: 0.02, 95% CI: -0.19 to 0.16, Cohen’s d: 0.02, p=0.93). See Table 7.4. When interpreting and reviewing the results presented, it is
important to be mindful of the sparsity of the data, particularly in the GP group under investigation, this is especially pertinent for the women in this group, of which there are only 23.
### Table 7.4 Mean and standard deviation of the standardised KT scores (z-scores) by GP group and gender

<table>
<thead>
<tr>
<th>GP under investigation</th>
<th></th>
<th></th>
<th>GP volunteers</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean KT z-score (SD)</td>
<td>Mean difference (95% CI)</td>
<td>p value</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>94</td>
<td>-5.16 (4.70)</td>
<td>-1.24 (-3.28 to 0.81)</td>
<td>0.01</td>
<td>206</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>-3.93 (3.20)</td>
<td></td>
<td></td>
<td>276</td>
</tr>
</tbody>
</table>
Figure 7.3 Distribution of standardised knowledge test (KT: written assessment) and Simulated Surgery (SS: clinical assessment) scores by GP group and gender.
Clinical assessment (SS)

440 GPs (115 under investigation and 325 volunteers) completed the SS. The mean score was 59.7 (SD=18.7); the range 4.2 to 94.3. See Figure 7.3. There was no interaction between gender and GP group: F(1, 436)=0.10, p=0.75. There was a main effect of gender, with women obtaining higher scores (mean difference: 11.0, 95% CI: -15.47 to -6.61, Cohen’s d: 0.61, p<0.0001); and a main effect of GP group, with volunteers obtaining higher SS scores (mean difference: 7.14, 95% CI: 3.20 to 11.07, Cohen’s d: 0.39, p=0.0004). See Table 7.5. Once again, it is important to be aware of the sparsity of data, especially when reviewing and interpreting these results of the GP group under investigation. This is particularly relevant when reviewing the results of the women in this group, of which there were only 23.


<table>
<thead>
<tr>
<th>GP under investigation</th>
<th>GP volunteers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
</tr>
<tr>
<td>Men</td>
<td>92</td>
</tr>
<tr>
<td>Women</td>
<td>23</td>
</tr>
</tbody>
</table>

7.4.3.2.1 Domains in the clinical assessment

I had planned to conduct exploratory factor analysis on the five domains of the clinical assessment to assess whether there was a difference between the genders, and a difference between the GP groups in the individual domain scores. However, after creating a correlation matrix I discovered that the five domains were strongly correlated with each other and as such decided that the exploratory factor analysis was unlikely to provide any additional useful information about how male and female GPs differ in their performance at clinical assessments. See Table 7.6.

Table 7.6 Correlation matrix of the five clinical assessment domains

<table>
<thead>
<tr>
<th>Domains</th>
<th>Data gathering</th>
<th>Management</th>
<th>Interaction</th>
<th>Anticipatory care</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data gathering</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Management</td>
<td>0.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction</td>
<td>0.83</td>
<td>0.49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anticipatory care</td>
<td>0.77</td>
<td>0.47</td>
<td>0.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explanation</td>
<td>0.83</td>
<td>0.52</td>
<td>0.90</td>
<td>0.83</td>
<td></td>
</tr>
</tbody>
</table>

7.4.3.3 KT and SS scores and the GMC investigation outcome

I then focused the study on the 112/120 investigated GPs with complete data. I decided to proceed with these analyses in light of the sparsity of the data, therefore consideration of the results in this section must be mindful of the relatively low number of GPs (GPs under investigation=112) and particularly the small number of women in this group (women GPs under investigation=22). A Fisher’s exact test showed no gender differences in
sanctions (p=0.61; of 22 females, 12 were sanctioned, 6 were not, and 4 were no longer registered; of the 90 males, 53 were sanctioned, 16 were not and 21 were no longer registered). A t-test showed no significant gender difference on the KT (p=0.26), but women significantly outperformed men on the SS (p<0.001). See Table 7.7.

Table 7.7 Univariate associations between investigation outcome, gender, KT score, and SS score

<table>
<thead>
<tr>
<th>Outcome type</th>
<th>Gender of doctor</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male (N=90)</td>
<td>Female (N=22)</td>
</tr>
<tr>
<td>No sanction</td>
<td>16</td>
<td>6</td>
</tr>
<tr>
<td>Sanction</td>
<td>53</td>
<td>12</td>
</tr>
<tr>
<td>No longer registered</td>
<td>21</td>
<td>4</td>
</tr>
</tbody>
</table>

ToC scores

<table>
<thead>
<tr>
<th></th>
<th>Male (N=90)</th>
<th>Female (N=22)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean KT z-score (SD)</td>
<td>-5.18 (4.71)</td>
<td>-3.98 (3.26)</td>
<td>0.26ii</td>
</tr>
<tr>
<td>Mean SS % score (SD)</td>
<td>52.01 (18.27)</td>
<td>66.13 (17.07)</td>
<td>&lt;0.001i</td>
</tr>
</tbody>
</table>

Statistical test used: iFisher exact test, ii t-test

Figure 7.5 shows the KT and SS scores of GPs under investigation by outcome type (no sanction, warning/sanction imposed, no longer registered). Most GPs in the ‘no sanction’ category occupy the upper right quadrant with high KT and SS scores; however, many GPs in the ‘warning/sanction’ or ‘no longer registered’ categories also have comparatively high KT and SS scores. GPs with low KT or SS scores are always in the ‘warning/sanction’ or ‘no longer registered’ categories. GPs with low SS scores have a range of KT scores, but those with low KT scores always have low SS scores.
For the path analysis, I collapsed outcome type into a binary variable (no sanction vs. sanction/no longer registered) to increase the power (univariate association with gender: p>0.05). Gender was placed at the beginning of the path model and outcome type was placed at the end. The KT score and SS score were placed in between gender and outcome type, in order of when they were performed, with the KT score being placed first followed by the SS score (during a ToC the KT is completed before the SS). Path strengths are shown as standardised beta coefficients from the multiple regression. The final path model showed being female predicted SS score; KT score predicted SS score; and SS score predicted remaining on the register without warning/sanction (see Figure 7.6). The effect of gender on final outcome was through higher SS performance, with no independent effect of gender controlling for test performance. Multinomial logistic regression (with three outcome categories) produced similar results.
Path coefficients are standardised beta coefficients from the regression analysis. Solid black arrows indicate positive effects, with the width of the arrow lines being proportional to the effect sizes. Only paths significant at $p<0.05$ are shown. A high SS score made it more likely that a GP would end up on the register without a warning/sanction following an investigation, which itself was predicted by being female, and by a high KT score.

7.5 Discussion

7.5.1 Summary

Among GPs under investigation by the GMC, women outperformed men on both the written (KT) and clinical (SS) components of the ToC. Among GPs not under investigation, women outperformed men on the SS only. Among GPs under investigation, low SS score increased the likelihood of having a warning or sanction imposed, but men were no more likely than women to have a warning or sanction imposed despite lower test scores. Warnings or sanctions are imposed based on a range of evidence of which the ToC is one component; and the outcome ‘warning or sanction imposed’ does not demonstrate the severity, which may differ by ToC score levels.

7.5.2 Strengths and limitations

A strength of this study was the inclusion of nearly all GPs investigated by the GMC for fitness to practise concerns who completed a ToC. The comparator group of doctors not under investigation enabled me to examine the gender differences at different levels of performance. A weakness is that the volunteer GPs are a self-selecting group that may not be representative of the overall GP population. It is also not possible to generalise these findings to other specialties.
A further weakness is the sparsity of data on the GPs who were under investigation by the GMC. A total of 120 GPs (of which 20% were women) undertook a KT and SS as part of a GMC investigation between 2008 and 2014. Only 112/120 of these investigated GPs had complete data (of which 22 were women), resulting in sparse data. Efforts were made to capture as much data as possible with the decision to use 6 years worth of available data (from 2008\textsuperscript{iii}, to 2014\textsuperscript{xiv}). The ToC are individually tailored to the investigated doctor’s specialty, experience and daily practice and as such the format and content of the ToC varies between specialty and doctors. This makes it very challenging to compare ToC performance across specialties, hence my decision to focus on only one specialty. My decision to focus on only one specialty does reduce the number of doctors’ ToC performance this study can explore, but I chose to focus this research study on GPs in an effort to maximise the number of investigated doctors in the study population.

I selected to focus on GPs because General Practice is not only the largest specialty, but also the specialty with the most complaints to the GMC [182]. I felt it was important to pursue the analysis of investigated GPs who completed a ToC, while being mindful of the sparsity of the data, because little is published about doctors who are investigated by the GMC and their professional performance. Access to this population can be challenging, partly because this population is unlikely to engage in research given the difficult circumstances they are currently experiencing in their professional careers. However, gaining a better understanding of the characteristics of these doctors and their professional performance is vital if the medical profession is to better understand the professional performance of doctors who perform poorly. Gaining a greater understanding of doctors who perform poorly is the first step in ensuring patient safety, supporting doctors

\textsuperscript{iii} The ToC were proposed in 1994 [158], however the three-year process to develop KT tailored for individual doctors began in 2005 [169].

\textsuperscript{xiv} 2014 was the year when work on this research study commenced.
who may struggle professionally, and educating our current and future doctors on how to improve and maintain their professional performance. Though the results of this study have arisen from sparse data, the results provides the medical education community with clues as to how those doctors who are investigated by the GMC for professional performance concerns differ from the general doctor population in terms of examination performance.

Unmeasured factors may have influenced the findings. Data from other specialties show volunteer doctors differ from doctors under investigation in terms of gender, ethnicity, world region of primary medical qualification, and seniority [317] – factors known to influence performance at clinical examinations [127, 131, 111]. Demographic data (aside from gender) was missing for many volunteer GPs so I could not adjust for these confounders.

Within the group of investigated GPs it would have been interesting to explore whether ethnicity was associated with performance or outcome, but unfortunately ethnicity data was missing for nearly one third of these GPs. It would also have been of interest to explore the relationship between number of year since primary medical qualification (or age), gender and performance to see if the recent changes in gender distribution of GPs [320, 95] and the role and format of the examinations required to gain MRCGP [321] have influenced the performance of GPs, however due to the small number of female GPs and GPs who had been qualified for 10 years or less, I was not able to explore this further. I also did not include data on the ToC OSCE, although the SS is a better predictor of outcome [316].

7.5.3 Comparison with other studies

Women generally outperform men in postgraduate medical examinations in General Practice [283, 299, 127] and other specialties [126, 260] (see Chapter 5 and Chapter 6). I have shown this holds even in a group of doctors with overall very poor performance; however there was no gender difference in the written test for GPs not under investigation. In terms of
written KT performance, a previous study has shown that on average male GPs perform more poorly than female GPs on the AKT of the Membership of the Royal College of GPs [302], which is what I found in my sample of GPs under investigation, but not among those GPs who volunteered to complete the test. It is not clear why I did not find evidence of a gender difference in KT performance in those doctors who volunteered to complete the ToC. It is important to note, however, that these GPs were a self-selecting group of volunteers. It is therefore possible that those male GPs who volunteered to complete a ToC have better medical knowledge than the average male GP population.

7.5.4 Unanswered questions and future research

These findings suggest men GPs under investigation had particularly poor knowledge compared to women GPs under investigation, even though both had been pre-selected as being likely to have poor performance. Both performed significantly worse on both test components compared to GPs not under investigation. Moreover, despite performing more poorly, male GPs were no more likely to be sanctioned than female GPs.

The reasons for gender differences found are uncertain. I propose that women and men may differ in skills and traits that are highly valued in both clinical assessments and in actual clinical practice. Women doctors in general have a more patient-centred approach and ask more psychosocial questions, which stimulates more patient disclosure [243]. Women doctors, including those under investigation, may therefore learn more information from patients, and perform better at the clinical assessment. Women have also been found, on average, to score higher on dutifulness (a facet of conscientiousness) [322], which is a predictor of performance [323, 31, 273, 324]. Female doctors may therefore demonstrate personality traits that lead them to maintaining their skills and knowledge and performing better at assessments. Women doctors have also been shown to have higher person-related values [266, 268, 269]. It has been shown that performance in a clinical setting is predicted by the person-related values held by a doctor.
It is plausible that women’s higher performance is therefore due to differences in skills and attitudes, as well as knowledge. I will explore whether personality traits, such as conscientiousness, empathy and communication style are associated with gender and disciplinary action in Chapter 8.

Male and female doctors may differ in ways I am unable to explore or measure and which influenced their performance. These possible hypotheses will be discussed below.

Lack of insight – being unaware of and not addressing deficiencies – is common amongst investigated doctors [275]. Previous research found male volunteer doctors tended to overestimate their written and clinical ToC scores, suggesting less insight [325]. If these gender differences are present in doctors under investigation, more poorly performing women may remove themselves from the medical register before taking a ToC. It would be interesting to further delve into those GPs who are no longer registered following an investigation into their fitness to practise, particularly those GPs who were voluntarily erased from the medical register, not only in terms of demographics and performance, but also the reasons behind their decision to apply for voluntary erasure.

Another factor that may have influenced doctors’ examination performance is dyslexia or another specific learning difficulty (SpLD). Doctors with a SpLD may also face extra challenges in the effective performance of their duties, especially if unrecognised or undisclosed [326]. It is uncertain whether there are gender differences in SpLDs such as dyslexia among doctors, but given that nearly 2% of medical students have dyslexia [327], it would be important to explore not only how SpLD may affect learning and performance, but also how to better identify and support those doctors with SpLD.

Organisational factors can also affect performance [31]; for example, professional isolation from peers and colleagues can limit opportunities for
feedback and development [275], and male GPs may be at more risk of professional isolation, whether in a single-handed or group practice.

This study has suggested that in the group of investigated GPs performance at KT is associated with performance at the SS. This findings is supported by previous research that has suggested that there is overlap between written and clinical assessments [246], and it seems intuitive that having a good and sound knowledge base would support performance at a clinical assessment. However, it also demonstrated that among volunteer GPs, women perform better on the clinical but not the written assessment. It may be that men in the self-selected group of volunteer GPs had better medical knowledge than the average male population.

7.5.5 Conclusion

Women GPs under investigation performed better at written and clinical assessments than their male counterparts, replicating the finding of poorer male academic performance within a highly selected population. Among GPs not under investigation, women outperformed men in the clinical but not the written assessment.

Unlike previous research, there was no evidence to suggest a gender difference in warning/sanction rates among this highly selected group of GPs who completed a ToC as part of an investigation. This suggests that the final decision on warnings or sanctions showed no gender bias beyond the differences in test performance.

I have shown that male doctors are at increased risk of facing disciplinary action against their medical registration, even after controlling for time since primary medical qualification, non-domestic primary medical qualification and specialty [9, 10] (see Chapter 3 and Chapter 4), however this study did not demonstrate evidence to suggest a difference in warning/sanction rate between the genders for GPs who completed a ToC as part of an investigation into their professional performance between 2008 and 2013.
I have also demonstrated that women perform better than men at postgraduate clinical assessments (see Chapter 5 and Chapter 6). In this study I have demonstrated that women GPs perform better than their male colleagues at the clinical assessment, regardless of whether they volunteered to complete the ToC or were being investigated by the GMC for performance concerns, mirroring the findings of both Chapter 5 and Chapter 6.

Finally, through this study I have been able to show that women GPs under investigation performed better at the written and clinical assessments than men GPs in the same situation. The phenomenon of poorer male performance in academic tests in medicine is thus replicated in this special context of doctors under investigation. It would therefore be reasonable to conclude that the final decision on warnings or sanctions appear to show no gender bias beyond the differences in test performance.

The reasons for the gender differences in clinical examination performance and in disciplinary action risk merit further exploration. Chapter 8 will explore whether certain attitudes, behaviours or skills assessed through measures of empathy, communication style and personality, are associated with gender and with risk of disciplinary action in a bid to better understand the observed difference in professional performance between the genders.
Chapter 8 – Disciplinary action against doctors by gender - associations between empathy, communication style and personality traits: a cohort study

8.1 Chapter summary

8.1.1 Background

The research presented thus far has demonstrated that female doctors are less likely to face disciplinary action against their medical registration (Chapter 3 and Chapter 4) and they outperform their male colleagues at clinical assessments (Chapter 5 and Chapter 6). Moreover, within a pre-selected population of GPs likely to have poor performance, women under investigation performed better than men at both the knowledge and clinical assessments. Despite performing more poorly, male GPs were no more likely to be sanctioned than female GPs, suggesting that the final decision on warnings or sanctions appear to show no gender bias beyond the differences in test performance (Chapter 7). Further research is required to better understand the difference in professional performance (measured by sanction risk and postgraduate medical examination pass rates) between the genders.

8.1.2 Aims

This study aims to explore whether levels of empathy, communication style and personality differ between the genders; and whether these factors mediate the association between gender and sanction risk within a population of doctors who obtained their primary medical qualification in England.

8.1.3 Design and setting

This is a prospective longitudinal study of doctors who applied to at least one of five English medical schools in 1990, representing 70% of applicants and entrants to UK medical schools that year. All study participants were accepted to study medicine, completed their medical degree, and practised as a doctor in the UK. Data from the participants were collected through five
postal questionnaires at certain points in their medical school and medical careers, and through linking data with the registration data held by the GMC.

8.1.4 Methods
A path analysis model was used to explore the relationship between doctors’ gender, empathy measure, communication style, personality traits and active or historical sanctions against their medical registration.

8.1.5 Results
Women had a reduced likelihood of active or historical sanctions against their medical registration (β coefficient = -0.01, p<0.05). There was no evidence for empathy, communication style, or personality mediating the relationship between doctors’ gender and sanction risk. Women doctors obtained higher scores on all four sub-scales of the empathy measure, approachable communication style measure, and the personality traits of neuroticism, agreeableness and conscientiousness. Women doctors obtained lower scores on the authoritative communication style measure and the personality trait openness.

8.1.6 Conclusion
In agreement with my earlier research, this study confirms that women doctors are at reduced risk of disciplinary action. Women differ from men in terms of empathy, communication style and personality traits, but there was insufficient evidence to suggest that the gender difference in these factors mediated sanction risk.
8.2 Introduction

Women are less likely to face disciplinary action against their medical registration (see Chapter 3 and Chapter 4) [9, 10] and they outperform their male colleagues at clinical postgraduate medical examinations (see Chapter 5, Chapter 6 and Chapter 7). There is published evidence demonstrating an association between performance at clinical postgraduate medical examinations and disciplinary action [67]. However, it is not clear why there is a gender difference in doctors’ professional performance.

Identifying modifiable factors that may influence a doctor’s risk of future disciplinary action will enable the medical education community to identify those doctors at greater risk and to better support, train and educate doctors to enhance the performance of the profession as a whole and to enhance patient care. With this study I have set out to explore the association between a selection of psychological measures and the future risk of disciplinary action.

8.2.1 Empathy

Empathy in doctors is defined as acknowledging and understanding the situation and emotional state of patients [328, 329]. Empathy levels have been shown to differ by gender in doctors and medical students, with women demonstrating greater empathy [330, 311, 331]. Research suggests that empathic behaviours from doctors contribute to greater patient satisfaction and compliance and have the potential to enhance doctor-patient relationships [332, 333]. However, it is also reported that engaging in the process of empathy brings with it the risk of personal distress and can contribute to burnout amongst doctors [334].

8.2.2 Communication skills

Effective communication skills are an essential part of being a doctor, as outlined in the GMC’s guidance to doctors, Good Medical Practice [19]. Research suggests that the communication that takes place during medical encounters may relate to a number of patient outcomes, including patient
satisfaction, treatment adherence, recall and understanding the consultation, and improved patient health outcomes [245, 335]. Two meta-analytic reviews of doctors’ communication style have demonstrated that men and women doctors differ in communication style, with women exhibiting more partnership building behaviours [243, 330]. This patient-centred communication style has been associated with improved adherence, greater patient satisfaction and patient enablement [336].

8.2.3 Personality

Personality is now widely considered to consist of five independent factors: ‘openness to experience’ (tendency to be curious), ‘conscientiousness’ (tendency to be organised), ‘extraversion’ (tendency to be outgoing), ‘agreeableness’ (tendency to be compassionate), and ‘neuroticism’ (tendency to be prone to psychological stress). Personality factors are regarded as largely stable traits [337], though they do change over time in response to life experiences [338]. Being aware of and taking account of personality may be beneficial in supporting doctors’ development and learning.

Of these, conscientiousness is thought most likely to be related to doctors’ performance. Conscientiousness can be defined as being persistent, planful, careful, responsible and hardworking [274]. A meta-analytic review of personality traits and occupational groups (including doctors) found that conscientiousness is associated with good job performance for all occupational groups [274]. Conscientiousness has also been found to be a predictor of high academic performance in medical students [273, 324, 269]. Interestingly, doctors with sanctions against their medical registration do not perform as well on postgraduate medical examinations, with performance at clinical assessments being a better predictor of sanctions against medical registration [67]. Poorer performance at postgraduate medical examinations has been found to be associated with patient mortality, unprofessional behaviour and sanctions [339, 66, 340, 341]. It would be reasonable to
suggest that conscientiousness may be the factor that drives both academic and clinical performance.

8.2.4 Study aims

To gain a better understanding of the gender difference in doctors’ professional performance, I aim to explore the following:

- Whether empathy, communication style or personality traits differ between the genders.
- Whether empathy, communication style or personality traits are associated with active or historical sanctions.
- Whether any relationship between gender and disciplinary action is mediated by empathy, communication style or personality traits.

8.3 Methods

8.3.1 Study design, setting and source of data

This longitudinal cohort study was conducted using survey data collected by Professor Chris McManus from UCL for a separate research project. The registration data and sanction data were provided by the GMC, after a licence agreement was signed between UCL and the GMC. Permission to use the data for research purposes was obtained from both Professor McManus and the GMC.

I merged the three datasets for the purposes of this research study.

8.3.2 The 1991 cohort dataset

A questionnaire was sent by postal service to all individuals who had a European Community postal address and had applied to any of five English medical schools taking part in the original cohort study (St Mary’s Hospital Medical School, University College and Middlesex School of Medicine\textsuperscript{v},

\textsuperscript{v} Known as UCL Medical School since 2008.

250
United Medical and Dental School of Guy’s and St Thomas’ Hospitals\textsuperscript{xvi}, University of Newcastle upon Tyne, and University of Sheffield) for admission in 1991. These individuals represented 70% of applicants and entrants to UK medical schools that year [342]. There was a 93% response rate for the first questionnaire [343].

The cohort was regularly followed up during their medical school and medical careers, and received a further four postal questionnaires: one in their final year as medical students (1995-98), which had a 56% response rate [343]; a questionnaire in their pre-registration house officer year (PRHO, equivalent to foundation training year one) (1996-99), which had a 58% response rate [344, 343]; a questionnaire when the majority would be completing their specialty training (2002/03); and the most recent questionnaire was administered in 2009. See Figure 8.1.

Each postal questionnaire was designed by Professor McManus, and contained questions about demographic data and a variety of questions measuring certain psychological factors. Each version of the questionnaire was slightly different, though some measures were repeated, allowing for an analysis of trend over time. (Copies of the questionnaires used can be found on Professor McManus’ medical education website [345].)

\textsuperscript{xvi} Known as King’s College London GKT School of Medical Education since 2015.
8.3.3 Data linkage

I linked three datasets to create the overall dataset used throughout this study:

1. Survey data from doctors who entered medical school in 1991 and were/are registered with the GMC, provided by Professor McManus, UCL (see section 8.3.2 for details about the data).
2. Registration status of the doctors as of August 2015, provided by the GMC.
3. Historical and active sanctions (erasure from the medical register, suspended from the medical register, conditions imposed against medical registration, undertakings against medical registration) and active warnings against the doctors’ medical registration from graduation to August 2015, provided by the GMC.
The data from the GMC was obtained after I provided the GMC with a list of the GMC numbers of all doctors in the study population of interest. I linked the three datasets using the doctors’ GMC number, a unique identifier, as the common variable across the three datasets. Table 8.1 describes the details of the data that was collected and methods of collection and Figure 8.2 illustrates the information flows and data linkage.

Table 8.1 Description of the 3 datasets used to create the overall study dataset

<table>
<thead>
<tr>
<th>Dataset 1</th>
<th>Description of the 3 datasets used to create the overall study dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provider</td>
<td>Survey data from doctors who entered medical school in 1991</td>
</tr>
<tr>
<td>Date of receipt</td>
<td>Professor McManus, UCL</td>
</tr>
<tr>
<td>Unique identifier variable</td>
<td>GMC number</td>
</tr>
<tr>
<td>Other variables</td>
<td>Date first survey was completed, year of birth, sex, marital status, ethnic group, children, specialty, registration status, MRCP attempts, MRCP pass, sanctions or warnings, answers to a variety of psychological measures including (but not limited to) motivation, personality, masculinity-femininity, study patterns, communication style, empathy</td>
</tr>
<tr>
<td>Dataset 2</td>
<td>GMC registration status of the cohort doctors as of August 2015</td>
</tr>
<tr>
<td>Provider</td>
<td>GMC</td>
</tr>
<tr>
<td>Date of receipt</td>
<td>August 2015</td>
</tr>
<tr>
<td>Unique identifier variable</td>
<td>GMC number</td>
</tr>
<tr>
<td>Other variables</td>
<td>GMC registration status</td>
</tr>
<tr>
<td>Dataset 3</td>
<td>Historical sanctions and active warnings of the cohort doctors from graduation to August 2015</td>
</tr>
<tr>
<td>Provider</td>
<td>GMC</td>
</tr>
<tr>
<td>Date of receipt</td>
<td>August 2015</td>
</tr>
<tr>
<td>Unique identifier variable</td>
<td>GMC number</td>
</tr>
<tr>
<td>Other variables</td>
<td>Sanctions or active warnings, date of sanction or warning</td>
</tr>
</tbody>
</table>
8.3.4 Population

All doctors who are, or have been, registered with the GMC, and were a member of the original cohort population.

8.3.5 Primary outcome and exposure

The outcome of interest was active or historical sanctions imposed against a doctor’s medical registration from time of qualification to August 2015. I chose to collapse the outcome of interest into a binary variable: doctors with active or historical sanctions against their medical registration; and doctors with no active or historical sanctions against their medical registration. The sanction types included erased after a Fitness to Practise panel hearing, suspension from the medical register, conditions imposed against medical registration, and undertakings agreed against medical registration. (See section 2.7.1 for further descriptions of sanction types.)
The exposure was the doctors’ gender, as declared by the doctors to the GMC.

8.3.6 Selection of variables

The selection of co-variates was limited to the data collected. I selected the co-variates to be included in this study prior to any statistical analysis, and the selection was based on the theories that I have raised and discussed throughout the course of this thesis. My choice of these co-variates was based on the evidence that these variables are associated with both gender and performance and therefore they were plausible candidates to mediate the relationship between gender and performance. Exploring the connection between these variables and performance and being aware of how these variables may influence performance will enable a better understanding on the different approaches that can be used to improve and maintain performance.

The variables included empathy, communication style and personality traits.

i) Empathy

Data on empathy was collected at application to medical school in the autumn of 1990. The data was collected using the 28-item Multidimensional Individual Difference Measure of Empathy that was developed by Davis in 1980 [346]. It has four sub-scales: ‘empathic concern’ (tendency to experience feelings of compassion and concern for others), ‘fantasy’ (tendency to identify with fictitious characters), ‘personal distress’ (tendency to experience discomfort and anxiety in response to distress in others), and ‘perspective-taking’ (tendency to adopt the point of view of others) [346]. Scores on these measures were treated as continuous data.

ii) Communication style
Data on communication style was collected in the cohort participants' final year of medical school (1996/97) using a version of the communicator style measure (CSM) developed by Norton in 1978 [347]. Norton defined communicator style as “the way one verbally and paraverbally interacts to signal how literal meaning should be taken, interpreted, filtered, or understood” [347]. The original CSM consists of fifty-one items, with ten factors:

- **Dominant** - Takes charge of social interactions
- **Dramatic** - Uses stylistic devices to highlight or understate content
- **Contentious** - Argumentative
- **Animated** - Provides frequent nonverbal cues
- **Impression Leaving** - Whether person is remembered because of communicative stimuli
- **Relaxed** - Not anxious or tense
- **Attentive** - Ensures the other person knows they are being listened to
- **Open** - Readily reveals personal information
- **Friendly** - Unhostile
- **Good communicator image** - At ease when interacting with others

An abbreviated version of the CSM consisting of twenty items measuring all ten variables was used in the 1996/97 questionnaire. Exploratory factor analysis was used to explore the underlying factors without setting a predefined structure, with the aim of reducing the data to a smaller set of variables (see Table 8.2 for correlation matrix of factors). Two factors were retained as both had eigenvalues greater than 1 (following Kaiser's criterion: see Figure 8.3). Varimax rotation was then performed to obtain a clearer pattern (see Table 8.3). Factor loadings with less than 0.32 were suppressed (as these variables would share too much variance), which is the value advised in the literature [348]. The first factor contained the items used to measure the variables Animated, Open, Friendly and Good communicator image. The second factor contained the items used to measure the variables Dominant, Dramatic, Contentious, Impression Leaving and Relaxed.
named the first domain “Approachable” and the second domain “Authoritative”. Together these two factors accounted for 83% of the total variance observed.
<table>
<thead>
<tr>
<th></th>
<th>F01</th>
<th>F02</th>
<th>F03</th>
<th>F04</th>
<th>F05</th>
<th>F06</th>
<th>F07</th>
<th>F08</th>
<th>F09</th>
<th>F10</th>
<th>F11</th>
<th>F12</th>
<th>F13</th>
<th>F14</th>
<th>F15</th>
<th>F16</th>
<th>F17</th>
<th>F18</th>
<th>F19</th>
<th>F20</th>
</tr>
</thead>
<tbody>
<tr>
<td>F01</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F02</td>
<td>0.25</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F03</td>
<td>0.19</td>
<td>0.41</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F04</td>
<td>0.20</td>
<td>0.25</td>
<td>0.24</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F05</td>
<td>-0.02</td>
<td>0.03</td>
<td>0.00</td>
<td>0.27</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F06</td>
<td>0.12</td>
<td>0.22</td>
<td>0.26</td>
<td>0.38</td>
<td>0.22</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F07</td>
<td>-0.01</td>
<td>0.11</td>
<td>0.18</td>
<td>0.25</td>
<td>0.12</td>
<td>0.28</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F08</td>
<td>0.07</td>
<td>-0.01</td>
<td>-0.06</td>
<td>-0.12</td>
<td>0.01</td>
<td>-0.13</td>
<td>-0.46</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F09</td>
<td>0.11</td>
<td>0.19</td>
<td>0.31</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.05</td>
<td>0.06</td>
<td>0.04</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F10</td>
<td>0.09</td>
<td>0.17</td>
<td>0.14</td>
<td>0.44</td>
<td>0.20</td>
<td>0.26</td>
<td>0.25</td>
<td>-0.13</td>
<td>0.09</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F11</td>
<td>0.14</td>
<td>0.38</td>
<td>0.21</td>
<td>0.24</td>
<td>0.12</td>
<td>0.16</td>
<td>0.02</td>
<td>0.07</td>
<td>0.14</td>
<td>0.20</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F12</td>
<td>0.11</td>
<td>0.16</td>
<td>0.10</td>
<td>0.25</td>
<td>0.15</td>
<td>0.15</td>
<td>-0.01</td>
<td>0.07</td>
<td>0.02</td>
<td>0.18</td>
<td>0.35</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F13</td>
<td>0.05</td>
<td>0.05</td>
<td>0.16</td>
<td>-0.16</td>
<td>-0.07</td>
<td>-0.03</td>
<td>0.01</td>
<td>0.03</td>
<td>0.22</td>
<td>-0.08</td>
<td>0.04</td>
<td>-0.08</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F14</td>
<td>0.08</td>
<td>0.19</td>
<td>0.19</td>
<td>0.24</td>
<td>0.16</td>
<td>0.40</td>
<td>0.21</td>
<td>-0.08</td>
<td>0.10</td>
<td>0.21</td>
<td>0.20</td>
<td>0.19</td>
<td>0.05</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F15</td>
<td>0.16</td>
<td>0.19</td>
<td>0.20</td>
<td>0.32</td>
<td>0.12</td>
<td>0.27</td>
<td>0.22</td>
<td>-0.1</td>
<td>0.16</td>
<td>0.21</td>
<td>0.19</td>
<td>0.27</td>
<td>-0.07</td>
<td>0.35</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F16</td>
<td>0.67</td>
<td>0.24</td>
<td>0.25</td>
<td>0.19</td>
<td>-0.03</td>
<td>0.15</td>
<td>0.07</td>
<td>0.03</td>
<td>0.07</td>
<td>0.07</td>
<td>0.17</td>
<td>0.13</td>
<td>0.05</td>
<td>0.15</td>
<td>0.22</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F17</td>
<td>0.08</td>
<td>0.08</td>
<td>0.03</td>
<td>0.30</td>
<td>0.35</td>
<td>0.26</td>
<td>0.11</td>
<td>-0.03</td>
<td>-0.06</td>
<td>0.25</td>
<td>0.10</td>
<td>0.15</td>
<td>-0.14</td>
<td>0.19</td>
<td>0.13</td>
<td>0.10</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F18</td>
<td>0.08</td>
<td>0.17</td>
<td>0.14</td>
<td>0.06</td>
<td>0.06</td>
<td>0.10</td>
<td>0.04</td>
<td>0.01</td>
<td>0.19</td>
<td>0.10</td>
<td>0.14</td>
<td>0.14</td>
<td>0.18</td>
<td>0.15</td>
<td>0.08</td>
<td>0.09</td>
<td>0.09</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F19</td>
<td>0.16</td>
<td>0.22</td>
<td>0.34</td>
<td>0.19</td>
<td>0.02</td>
<td>0.20</td>
<td>0.18</td>
<td>-0.14</td>
<td>0.21</td>
<td>0.17</td>
<td>0.10</td>
<td>0.03</td>
<td>0.09</td>
<td>0.15</td>
<td>0.16</td>
<td>0.21</td>
<td>0.06</td>
<td>0.20</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>F20</td>
<td>0.20</td>
<td>0.22</td>
<td>0.34</td>
<td>0.22</td>
<td>0.22</td>
<td>0.29</td>
<td>0.19</td>
<td>-0.15</td>
<td>0.17</td>
<td>0.19</td>
<td>0.12</td>
<td>0.04</td>
<td>0.06</td>
<td>0.17</td>
<td>0.20</td>
<td>0.22</td>
<td>0.11</td>
<td>0.15</td>
<td>0.52</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 8.2 Correlation matrix of the communication factors

Figure 8.3 Scree plot of the eigenvalues of factors
Table 8.3 Table showing the factor items, factor loadings and uniqueness of variance for the exploratory factor analysis on communication style

<table>
<thead>
<tr>
<th>Items</th>
<th>Factor 1 (approachable) loadings</th>
<th>Factor 2 (authoritative) loadings</th>
<th>Uniqueness of variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – open</td>
<td>0.58</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>2 – animated</td>
<td>0.48</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>3 – friendly</td>
<td>0.54</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>4 – dominant</td>
<td>0.63</td>
<td>0.57</td>
<td></td>
</tr>
<tr>
<td>5 – contentious</td>
<td>0.42</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>6 – impression leaving</td>
<td>0.55</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>7 – relaxed</td>
<td>0.45</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td>8 – relaxed</td>
<td>0.38</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>9 – friendly</td>
<td></td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>10 – dominant</td>
<td>0.52</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>11 – animated</td>
<td></td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>12 – dramatic</td>
<td></td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>13 – attentive</td>
<td></td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>14 – impression leaving</td>
<td>0.45</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>15 – dramatic</td>
<td>0.42</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>16 – open</td>
<td>0.60</td>
<td>0.63</td>
<td></td>
</tr>
<tr>
<td>17 – contentious</td>
<td>0.46</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>18 – attentive</td>
<td></td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>19 – good communicator image</td>
<td>0.46</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>20 – good communicator image</td>
<td>0.45</td>
<td>0.72</td>
<td></td>
</tr>
</tbody>
</table>

iii) Personality traits

Personality data was collected in the cohort participants’ PRHO year. An abbreviated version of the five-factor model by Costa and McCrae [349, 343] was used to collect data on personality. Scores on these measures were treated as continuous data. See section 8.2.3 for further details on the personality traits measured.

See Figure 8.4 for when data on each measure was collected.
8.3.7 Statistical methods

The distribution of the outcome and exposure variables were explored, before exploring the measures assessing empathy, communication style and personality in terms of response rate and distribution of mean scores. Bivariate analyses were conducted to explore the associations between the outcome and exposure variables, before exploring the association between the outcome and exposure variables and the co-variates of interest. A comparison of complete cases (participants with no missing data) and incomplete cases was performed to evaluate whether the complete cases were representative of the whole study population. I took the assumption that any data missing from the questionnaires were likely to be missing at random, because data on multiple different variables were collected and because the questions used in the questionnaire did not declare what they were measuring, therefore the participants were unlikely to be aware as to what was being measured. In line with this missing at random assumption, I chose to use the multiple imputation method to deal with any missing data for
the multivariate analyses. The use of multiple imputation also assumes the data have a normal distribution and therefore the distribution of the co-variates was assessed to check if this assumption would hold. Multiple imputation creates several different imputed datasets and combines the results obtained from each of them to predict the values for the missing data. The incomplete co-variates in this dataset were the four empathy sub-scales (fantasy, perspective-taking, empathic concern, and personal distress), the two communication styles (approachable and authoritative), and the five personality measures (neuroticism, extraversion, openness, agreeableness, and conscientiousness). The multiple imputations were performed for these incomplete co-variates using the multiple imputation procedure in STATA v.12/SE. In keeping with the advice from the literature, five imputations were obtained [350]. A comparison of the observed data and the imputed data was then performed prior to commencing the multivariate analysis.

Multivariate analysis was completed with a path analysis model using multiple regression. I chose to perform path analysis because it allows the specification of relationships between variables and enables a powerful pictorial way to present complex relationships. I also wanted the opportunity to gain experience of using this method of statistical analysis in my research. Path analysis is performed by using multiple regression models to further examine the associations between variables. The first path model is the saturated model containing all variables. Variables, in this instance, were placed in chronological order of when they were collected. Variables that do not reach statistical significance (p>0.05) are removed from the model until all variables in the path are significant (p<0.05). Path strengths are shown as standardised beta coefficients from the regression analysis. Further path models are created with each variable is set as the dependent variable in turn. The final path analysis model will illustrate all the relationships between outcome and the other variables, and gender and the other variables that were statistically significant.

I used the STROBE statement to guide reporting [199].
8.4 Results

8.4.1 Descriptive results

There were 2,763 cohort members who became registered with the GMC. See Table 8.4 for the details of these doctors’ demographics.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Male (%)</th>
<th>Female (%)</th>
<th>Total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=1343</td>
<td>N=1420</td>
<td>N=2763</td>
</tr>
<tr>
<td>Black, Asian &amp; minority ethnic groups</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>900 (67%)</td>
<td>1053 (74%)</td>
<td>1953 (71%)</td>
</tr>
<tr>
<td>Yes</td>
<td>337 (25%)</td>
<td>275 (19%)</td>
<td>612 (22%)</td>
</tr>
<tr>
<td>Missing</td>
<td>106 (7%)</td>
<td>92 (7%)</td>
<td>198 (7%)</td>
</tr>
<tr>
<td>Active or historical sanctions imposed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No sanction</td>
<td>1319 (98%)</td>
<td>1415 (99%)</td>
<td>2734 (99%)</td>
</tr>
<tr>
<td>Erasure following FtP panel</td>
<td>4 (&lt;0.01%)</td>
<td>0 (0%)</td>
<td>4 (&lt;0.01%)</td>
</tr>
<tr>
<td>Suspension</td>
<td>6 (&lt;0.01%)</td>
<td>2 (&lt;0.01%)</td>
<td>8 (&lt;0.01%)</td>
</tr>
<tr>
<td>Conditions</td>
<td>10 (&lt;0.01%)</td>
<td>1 (&lt;0.01%)</td>
<td>11 (&lt;0.01%)</td>
</tr>
<tr>
<td>Undertakings</td>
<td>4 (&lt;0.01%)</td>
<td>2 (&lt;0.01%)</td>
<td>6 (&lt;0.01%)</td>
</tr>
</tbody>
</table>

8.4.1.1 Response rate

The proportion of the study population who responded to the variables of interest decreased with each subsequent questionnaire: empathy (measured at application at to medical school) achieved a 88-90% response rate; communication style (measured during final year of medical school) achieved a 53-54% response rate; and personality (measured during first year post-qualification from medical school) achieved 48-49% response rate (see Figure 8.5 and Table 8.5).
Figure 8.5 Flow diagram showing when data in each measure was collected and the response rate

2763 cohort members registered with the GMC

Empathy response rate
Measured at application to medical school
- Fantasy – 2429 (88%)
- Perspective taking – 2473 (90%)
- Empathic concern – 2458 (89%)
- Personal distress – 2434 (88%)

Communication style response rate
Measured during final year at medical school
- Approachable – 1493 (54%)
- Authoritative – 1458 (53%)

Personality response rate
Measured during PRHO year
- Neuroticism – 1351 (49%)
- Extraversion – 1351 (49%)
- Openness – 1329 (48%)
- Agreeableness – 1349 (49%)
- Conscientiousness – 1350 (49%)
Table 8.5 Number and percentage of participants who responded to the questions for each co-variate, by gender and sanction status

<table>
<thead>
<tr>
<th>Total of responses</th>
<th>Gender of responders</th>
<th>Sanction status of responders</th>
</tr>
</thead>
<tbody>
<tr>
<td>N=2763 (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=1420 (%)</td>
<td>Female</td>
<td>N=2734 (%)</td>
</tr>
<tr>
<td>N=1343 (%)</td>
<td>Male</td>
<td>N=29 (%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p value (X²)</td>
</tr>
<tr>
<td>Empathy measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emp-F</td>
<td>2429 (88%)</td>
<td>2404 (88%)</td>
</tr>
<tr>
<td></td>
<td>1241 (87%)</td>
<td>2404 (88%)</td>
</tr>
<tr>
<td></td>
<td>1188 (88%)</td>
<td>25 (86%)</td>
</tr>
<tr>
<td></td>
<td>0.39</td>
<td>0.78</td>
</tr>
<tr>
<td>Emp-PT</td>
<td>2473 (90%)</td>
<td>2445 (89%)</td>
</tr>
<tr>
<td></td>
<td>1278 (90%)</td>
<td>28 (97%)</td>
</tr>
<tr>
<td></td>
<td>1195 (89%)</td>
<td>0.21</td>
</tr>
<tr>
<td>Emp-EC</td>
<td>2458 (89%)</td>
<td>2432 (89%)</td>
</tr>
<tr>
<td></td>
<td>1262 (89%)</td>
<td>26 (90%)</td>
</tr>
<tr>
<td></td>
<td>1196 (89%)</td>
<td>0.91</td>
</tr>
<tr>
<td>Emp-PD</td>
<td>2434 (88%)</td>
<td>2407 (88%)</td>
</tr>
<tr>
<td></td>
<td>1261 (89%)</td>
<td>27 (93%)</td>
</tr>
<tr>
<td></td>
<td>1173 (87%)</td>
<td>0.40</td>
</tr>
<tr>
<td>Communication style</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSM-Ap</td>
<td>1493 (54%)</td>
<td>1480 (54%)</td>
</tr>
<tr>
<td></td>
<td>824 (58%)</td>
<td>13 (45%)</td>
</tr>
<tr>
<td></td>
<td>669 (50%)</td>
<td>0.32</td>
</tr>
<tr>
<td>CSM-Au</td>
<td>1458 (53%)</td>
<td>1445 (53%)</td>
</tr>
<tr>
<td></td>
<td>803 (57%)</td>
<td>13 (45%)</td>
</tr>
<tr>
<td></td>
<td>655 (49%)</td>
<td>0.39</td>
</tr>
<tr>
<td>Personality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B5-N</td>
<td>1351 (49%)</td>
<td>1338 (49%)</td>
</tr>
<tr>
<td></td>
<td>754 (53%)</td>
<td>13 (45%)</td>
</tr>
<tr>
<td></td>
<td>597 (44%)</td>
<td>0.89</td>
</tr>
<tr>
<td>B5-E</td>
<td>1351 (49%)</td>
<td>1338 (49%)</td>
</tr>
<tr>
<td></td>
<td>753 (53%)</td>
<td>13 (45%)</td>
</tr>
<tr>
<td></td>
<td>598 (45%)</td>
<td>0.66</td>
</tr>
<tr>
<td>B5-O</td>
<td>1329 (48%)</td>
<td>1316 (48%)</td>
</tr>
<tr>
<td></td>
<td>737 (52%)</td>
<td>13 (45%)</td>
</tr>
<tr>
<td></td>
<td>592 (44%)</td>
<td>0.72</td>
</tr>
<tr>
<td>B5-A</td>
<td>1349 (49%)</td>
<td>1336 (49%)</td>
</tr>
<tr>
<td></td>
<td>753 (53%)</td>
<td>13 (45%)</td>
</tr>
<tr>
<td></td>
<td>596 (44%)</td>
<td>0.67</td>
</tr>
<tr>
<td>B5-C</td>
<td>1350 (49%)</td>
<td>1337 (49%)</td>
</tr>
<tr>
<td></td>
<td>752 (53%)</td>
<td>13 (45%)</td>
</tr>
<tr>
<td></td>
<td>598 (45%)</td>
<td>0.66</td>
</tr>
</tbody>
</table>

Emp-F: Fantasy       CSM-Ap: Approachable       B5-N: Neuroticism
Emp-PT: Perspective taking       CSM-Au: Authoritative       B5-E: Extraversion
Emp-EC: Empathic concern       B5-O: Openness
Emp-PD: Personal distress       B5-A: Agreeableness
                           B5-C: Conscientiousness

8.4.1.2 Distribution of the co-variates

I reviewed histograms of the distribution of the co-variate scores and Q-Q plots to assess the distribution of each of the co-variates, along with a comparison of the mean and mode score for each co-variate. The distribution of the scores for each of the co-variates was broadly normally distributed.

8.4.1.3 Correlation between the co-variates

The strength of the association was small for the majority of the variables (r between 0.3 to -0.3), however the strength of the association was noted to be of medium size for the following co-variates: empathy measures perspective taking and empathic concern (r=0.35); approachable and authoritative communication styles (r=0.36); authoritative communication style and
extraversion ($r=0.35$); and personality traits neuroticism and extraversion ($r=-0.36$). See Table 8.6.
Table 8.6 Correlation matrix of all variables included in the path model

<table>
<thead>
<tr>
<th></th>
<th>Sanction</th>
<th>Female</th>
<th>Emp-F</th>
<th>Emp-PT</th>
<th>Emp-EC</th>
<th>Emp-PD</th>
<th>CSM-Ap</th>
<th>CSM-Au</th>
<th>BS-N</th>
<th>BS-E</th>
<th>BS-O</th>
<th>BS-A</th>
<th>BS-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sanction</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>-0.07</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emp-F</td>
<td>0.02</td>
<td>0.19</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emp-PT</td>
<td>-0.03</td>
<td>0.14</td>
<td>0.10</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emp-EC</td>
<td>-0.03</td>
<td>0.21</td>
<td>0.25</td>
<td>0.35</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emp-PD</td>
<td>-0.04</td>
<td>-0.01</td>
<td>0.18</td>
<td>-0.21</td>
<td>&lt; -0.01</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSM-Ap</td>
<td>-0.03</td>
<td>0.21</td>
<td>0.11</td>
<td>0.15</td>
<td>0.20</td>
<td>-0.13</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSM-Au</td>
<td>0.02</td>
<td>-0.19</td>
<td>-0.05</td>
<td>-0.04</td>
<td>-0.01</td>
<td>-0.22</td>
<td>0.36</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS-N</td>
<td>-0.05</td>
<td>0.07</td>
<td>0.08</td>
<td>-0.08</td>
<td>&lt; -0.01</td>
<td>0.15</td>
<td>-0.08</td>
<td>-0.10</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS-E</td>
<td>0.03</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.04</td>
<td>0.07</td>
<td>-0.09</td>
<td>0.28</td>
<td>0.35</td>
<td>-0.36</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS-O</td>
<td>0.07</td>
<td>-0.09</td>
<td>0.20</td>
<td>0.12</td>
<td>0.05</td>
<td>0.03</td>
<td>0.05</td>
<td>0.16</td>
<td>-0.07</td>
<td>0.24</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS-A</td>
<td>-0.04</td>
<td>0.21</td>
<td>0.03</td>
<td>0.14</td>
<td>0.19</td>
<td>-0.01</td>
<td>0.20</td>
<td>-0.06</td>
<td>-0.17</td>
<td>0.17</td>
<td>0.07</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>BS-C</td>
<td>&lt; -0.01</td>
<td>0.10</td>
<td>-0.08</td>
<td>0.01</td>
<td>0.02</td>
<td>-0.17</td>
<td>0.10</td>
<td>0.19</td>
<td>-0.14</td>
<td>0.24</td>
<td>-0.05</td>
<td>0.22</td>
<td>1</td>
</tr>
</tbody>
</table>

Emp-F: Fantasy  
CSM-Ap: Approachable  
Emp-PT: Perspective taking  
CSM-Au: Authoritative  
Emp-EC: Empathic concern  
B5-N: Neuroticism  
Emp-PD: Personal distress  
B5-E: Extraversion  
B5-A: Agreeableness  
B5-O: Openness  
B5-C: Conscientiousness
8.4.2 Bivariate analyses

8.4.2.1 Active and/or historical sanctions

I performed bivariate analyses to compare male and female doctors. There was evidence for a statistically significant association between gender and active and/or historical sanctions ($\chi^2(1)=13.7$, p<0.001). However, the strength of the association between these two variables is shown to be weak (Cramer’s V = 0.07, 95% CI: 0.04 to 0.10). See Table 8.7.

Table 8.7 The association between gender and active and/or historical sanctions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Male</th>
<th>Female</th>
<th>Measure of association</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active and/or historical sanctions</td>
<td></td>
<td></td>
<td>$\chi^2(1)=13.7$, p&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>V=0.07, 95% CI: 0.04 to 0.10</td>
</tr>
<tr>
<td>No</td>
<td>1319 (48%)</td>
<td>1415 (52%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>24 (83%)</td>
<td>5 (17%)</td>
<td></td>
</tr>
</tbody>
</table>

8.4.2.2 Empathy measures

There was no statistically significant difference between gender and response rate to the empathy measures (see Table 8.5). The female participants obtained higher mean scores on all four sub-scales of the empathy measure, however the differences between the genders was only statistically significant for the sub-scales ‘fantasy’, ‘perspective-taking’, and ‘empathic concern’ (see Table 8.8).

There was no statistically significant evidence for a difference in response rate to the empathy questions between those participants with active and/or historical sanctions and those with none (see Table 8.5). Those participants with active and/or historical sanctions obtained higher mean scores on the sub-scale ‘perspective-taking’, and lower mean scores on the sub-scales ‘fantasy’, ‘empathic concern’, and ‘personal distress’. However the difference in mean scores on all four sub-scales between the two groups was not found to be statistically significant (see Table 8.9).
8.4.2.3 Communication style

Female participants were more likely to respond to the communication questions, compared to their male counterparts \( (P<0.001) \), see Table 8.5.

There was statistically significant evidence for a difference in communication style between the genders, with women demonstrating a more approachable communication style and men adopting a more authoritative communication style (see Table 8.8).

There was no evidence for a difference in response rate to the questions assessing communication style by sanction status (see Table 8.5). Those doctors with active and/or historical sanctions against their medical registration obtained slightly higher mean scores on the authoritative communication style domain and lower mean scores on the approachable communication style domain, however the difference in scores was not found to be statistically significant (see Table 8.9).

8.4.2.4 Personality traits

Female participants were significantly more likely to respond to questions assessing personality traits (see Table 8.5). Women obtained statistically significant higher mean scores for the personality trait domains of ‘neuroticism’, ‘agreeableness’, and ‘conscientiousness’. Men obtained statistically significant higher mean scores for the domain ‘openness’. There was no evidence for a difference in mean scores for the domain ‘extraversion’ between the genders (see Table 8.8).

There was no evidence for a difference in response rate to the questions measuring personality traits by sanction status (see Table 8.5). Those participants with active and/or historical sanctions obtained higher mean scores on the personality traits domains ‘extraversion’, ‘openness’, and ‘conscientiousness’ and they obtained lower mean scores in the domains ‘neuroticism’ and ‘agreeableness’. However, there was no statistically
significant evidence for a difference in mean scores between the two groups (see Table 8.9).
<table>
<thead>
<tr>
<th>Empathy measures</th>
<th>Total number of responders</th>
<th>Male score</th>
<th>Female score</th>
<th>p-value (t-test)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean (SD)</td>
<td>95% CI</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Fantasy</td>
<td>2429</td>
<td>17.93 (3.66)</td>
<td>17.72 to 18.14</td>
<td>18.87 (3.90)</td>
</tr>
<tr>
<td>Perspective taking</td>
<td>2473</td>
<td>20.44 (2.99)</td>
<td>20.27 to 20.61</td>
<td>21.40 (2.70)</td>
</tr>
<tr>
<td>Empathic concern</td>
<td>2458</td>
<td>21.70 (2.69)</td>
<td>21.55 to 21.90</td>
<td>22.80 (2.50)</td>
</tr>
<tr>
<td>Personal distress</td>
<td>2434</td>
<td>12.68 (2.81)</td>
<td>12.52 to 12.84</td>
<td>12.71 (2.67)</td>
</tr>
<tr>
<td>Communication measure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approachable</td>
<td>1493</td>
<td>19.80 (2.93)</td>
<td>19.58 to 20.02</td>
<td>20.92 (2.93)</td>
</tr>
<tr>
<td>Authoritative</td>
<td>1458</td>
<td>21.06 (3.28)</td>
<td>20.81 to 21.31</td>
<td>19.56 (3.44)</td>
</tr>
<tr>
<td>Personality</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuroticism</td>
<td>1351</td>
<td>8.57 (2.24)</td>
<td>8.39 to 8.75</td>
<td>9.09 (2.19)</td>
</tr>
<tr>
<td>Extraversion</td>
<td>1351</td>
<td>10.59 (1.85)</td>
<td>10.44 to 10.73</td>
<td>10.61 (1.87)</td>
</tr>
<tr>
<td>Openness</td>
<td>1329</td>
<td>10.56 (2.41)</td>
<td>10.36 to 10.75</td>
<td>10.07 (2.30)</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>1349</td>
<td>12.34 (1.67)</td>
<td>12.21 to 12.48</td>
<td>12.96 (1.53)</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>1350</td>
<td>11.52 (1.87)</td>
<td>11.37 to 11.67</td>
<td>11.88 (1.71)</td>
</tr>
<tr>
<td></td>
<td>Total number of responders</td>
<td>No sanctions</td>
<td>Active/historical sanctions</td>
<td>p-value (t-test)</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------------------</td>
<td>--------------</td>
<td>----------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>95% CI</td>
<td>Mean (SD)</td>
<td>95% CI</td>
</tr>
<tr>
<td><strong>Empathy measures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fantasy</td>
<td>2429</td>
<td>18.47 (3.82)</td>
<td>18.32 to 18.62</td>
<td>18.12 (3.77)</td>
</tr>
<tr>
<td>Perspective taking</td>
<td>2473</td>
<td>20.92 (2.87)</td>
<td>20.81 to 21.04</td>
<td>21.07 (3.70)</td>
</tr>
<tr>
<td>Empathic concern</td>
<td>2458</td>
<td>22.27 (2.65)</td>
<td>22.17 to 22.38</td>
<td>21.54 (3.01)</td>
</tr>
<tr>
<td>Personal distress</td>
<td>2434</td>
<td>12.70 (2.73)</td>
<td>12.59 to 12.81</td>
<td>12.22 (3.00)</td>
</tr>
<tr>
<td><strong>Communication measure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approachable</td>
<td>1493</td>
<td>20.42 (2.97)</td>
<td>20.27 to 20.57</td>
<td>19.92 (4.01)</td>
</tr>
<tr>
<td>Authoritative</td>
<td>1458</td>
<td>20.23 (3.46)</td>
<td>20.05 to 20.41</td>
<td>20.78 (1.88)</td>
</tr>
<tr>
<td><strong>Personality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuroticism</td>
<td>1351</td>
<td>8.86 (2.22)</td>
<td>8.74 to 8.98</td>
<td>8.62 (3.10)</td>
</tr>
<tr>
<td>Extraversion</td>
<td>1351</td>
<td>10.60 (1.86)</td>
<td>10.50 to 10.70</td>
<td>10.77 (1.69)</td>
</tr>
<tr>
<td>Openness</td>
<td>1329</td>
<td>10.28 (2.36)</td>
<td>10.15 to 10.41</td>
<td>11.08 (2.22)</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>1349</td>
<td>12.69 (1.62)</td>
<td>12.60 to 12.78</td>
<td>12.54 (2.18)</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>1350</td>
<td>11.72 (1.80)</td>
<td>11.62 to 11.82</td>
<td>11.85 (1.46)</td>
</tr>
</tbody>
</table>
Table 8.10 Mean scores, standard deviations and 95% confidence intervals for empathy, communication and personality measures by gender and sanction status

<table>
<thead>
<tr>
<th></th>
<th>Male no sanction</th>
<th>Male sanction</th>
<th>Female no sanction</th>
<th>Female sanction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean (SD)</td>
<td>95% CI</td>
<td>N</td>
</tr>
<tr>
<td>Empathy measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fantasy</td>
<td>1167</td>
<td>17.9 (3.7)</td>
<td>17.7 to 18.2</td>
<td>21</td>
</tr>
<tr>
<td>Perspective taking</td>
<td>1172</td>
<td>20.4 (3.0)</td>
<td>20.3 to 20.6</td>
<td>23</td>
</tr>
<tr>
<td>Empathic concern</td>
<td>1175</td>
<td>21.7 (2.7)</td>
<td>21.6 to 21.9</td>
<td>21</td>
</tr>
<tr>
<td>Personal distress</td>
<td>1150</td>
<td>12.7 (2.8)</td>
<td>12.5 to 12.9</td>
<td>23</td>
</tr>
<tr>
<td>Communication style</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approachable</td>
<td>658</td>
<td>19.8 (2.9)</td>
<td>19.6 to 20.0</td>
<td>11</td>
</tr>
<tr>
<td>Authoritative</td>
<td>644</td>
<td>21.2 (3.3)</td>
<td>20.8 to 21.3</td>
<td>11</td>
</tr>
<tr>
<td>Personality</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuroticism</td>
<td>586</td>
<td>8.6 (2.2)</td>
<td>8.4 to 8.8</td>
<td>11</td>
</tr>
<tr>
<td>Extraversion</td>
<td>587</td>
<td>10.6 (1.9)</td>
<td>10.4 to 10.7</td>
<td>11</td>
</tr>
<tr>
<td>Openness</td>
<td>581</td>
<td>10.6 (2.4)</td>
<td>10.4 to 10.8</td>
<td>11</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>585</td>
<td>12.3 (1.7)</td>
<td>12.2 to 12.5</td>
<td>11</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>587</td>
<td>11.5 (1.9)</td>
<td>11.4 to 11.7</td>
<td>11</td>
</tr>
</tbody>
</table>
8.4.3 Participants with missing data

I compared the data of the 798 participants (29% of the study population) with no missing data for the co-variates of interest to the 1,965 participants who had some missing data for those variables (see Figure 8.6). There was no evidence of a statistically significant difference between these two groups in terms of the outcome of interest (p=0.33), however there was a difference in terms of gender, with a significantly greater proportion of female participants not having any missing data (p<0.0001). See Table 8.11.

Those participants who had complete data were comparable to those with incomplete data in terms of sanction status, communication style, personality, and three of the empathy sub-scales (fantasy, empathic concern, and personal distress), but differed in terms of gender and the empathy sub-scale perspective-taking (see Table 8.12). I therefore felt that the complete cases were not fully representative of the whole study population and decided to proceed with multiple imputation to predict values for the missing data before continuing with the multivariate analysis of the data.

Figure 8.6 Flow diagram showing the number of cohort members with complete and missing data
Table 8.11 Comparison of participants with no data missing for the co-variates of interest and those participants with data missing, by gender and sanction status

<table>
<thead>
<tr>
<th>Variable</th>
<th>Participants with no missing data N=798</th>
<th>Participants with missing data N=1965</th>
<th>p-value ($X^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sanction status</td>
<td></td>
<td></td>
<td>0.33</td>
</tr>
<tr>
<td>No sanctions</td>
<td>792 (29%)</td>
<td>1942 (71%)</td>
<td></td>
</tr>
<tr>
<td>Sanctions</td>
<td>6 (21%)</td>
<td>23 (79%)</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Male</td>
<td>330 (25%)</td>
<td>1013 (75%)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>468 (33%)</td>
<td>952 (67%)</td>
<td></td>
</tr>
</tbody>
</table>

Table 8.12 Comparison of data of participants with complete data and those with incomplete data for the co-variates of interest

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number of participants with missing data</th>
<th>Mean score (SD) of participants with no missing data N=798</th>
<th>Mean score (SD) of participants with missing data</th>
<th>p-value (t-test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empathy measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fantasy</td>
<td>1631</td>
<td>18.60 (4.04)</td>
<td>18.40 (3.70)</td>
<td>0.21</td>
</tr>
<tr>
<td>Perspective-taking</td>
<td>1675</td>
<td>20.69 (2.96)</td>
<td>21.04 (2.83)</td>
<td>0.006</td>
</tr>
<tr>
<td>Empathic concern</td>
<td>1660</td>
<td>22.36 (2.59)</td>
<td>22.22 (2.69)</td>
<td>0.25</td>
</tr>
<tr>
<td>Personal distress</td>
<td>1636</td>
<td>12.79 (2.74)</td>
<td>12.65 (2.73)</td>
<td>0.22</td>
</tr>
<tr>
<td>Communication style</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approachable</td>
<td>695</td>
<td>20.51 (2.99)</td>
<td>20.31 (2.98)</td>
<td>0.20</td>
</tr>
<tr>
<td>Authoritative</td>
<td>660</td>
<td>20.24 (3.45)</td>
<td>20.22 (3.45)</td>
<td>0.93</td>
</tr>
<tr>
<td>Personality</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuroticism</td>
<td>553</td>
<td>8.90 (2.24)</td>
<td>8.90 (2.21)</td>
<td>0.39</td>
</tr>
<tr>
<td>Extraversion</td>
<td>553</td>
<td>10.61 (1.92)</td>
<td>10.57 (1.77)</td>
<td>0.70</td>
</tr>
<tr>
<td>Openness</td>
<td>531</td>
<td>10.26 (2.38)</td>
<td>10.33 (2.34)</td>
<td>0.57</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>551</td>
<td>12.73 (1.60)</td>
<td>12.63 (1.66)</td>
<td>0.25</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>552</td>
<td>11.75 (1.79)</td>
<td>11.68 (1.80)</td>
<td>0.50</td>
</tr>
</tbody>
</table>

8.4.4 Path analysis

There was a substantial amount of missing data. I decided to proceed with path analysis despite the lack of statistical evidence for a significant difference between the doctors with and without sanctions in the variables of interest. I used multiple imputation to impute values for the co-variates empathy measures, communication style and personality traits.
Path analysis was performed to further examine the associations between the doctors’ gender, sanction status and co-variate scores (empathy, communication style and personality traits). Gender was placed at the beginning of the path model and sanction status was placed at the end of the path model. The remaining co-variates were placed in between gender and sanction status, in the order of when they were collected, with the empathy measures being placed first (collected at application to medical school), followed by communication styles (collected in the final year of medical school) and finally the personality measures (collected in the first year of working after qualification from medical school). Path strengths are shown as standardised beta coefficients from the regression analysis. The final path model illustrates relationships between outcome and gender and the co-variates, and the relationships between gender and the co-variates. The relationships between the individual co-variates are not shown, partly to ensure the final model is clear and because the relationship between the co-variates is not a main area of interest in this study.

The final path analysis model showed being female made it less likely that a doctor would have active or historical sanctions (beta coefficient: -0.01, 95% CI: -0.02 to -0.01, p<0.001). No other variables in the path model were associated with sanction status. See Table 8.13 and Figure 8.7.

The final path analysis model showed evidence of a positive association between being female and the following variables: three empathy measure sub-scales including fantasy (beta coefficient: 1.08, 95% CI: 0.79 to 1.37, p<0.001), perspective-taking (beta coefficient: 0.93, 95% CI: 0.71 to 1.16, p<0.001) and empathic concern (beta coefficient: 1.09, 95% CI: 0.98 to 1.29, p<0.001), approachable communication style (beta coefficient: 0.93, 95% CI: 0.65 to 1.22, p<0.001) and conscientiousness (beta coefficient: 0.55, 95% CI: 0.33 to 0.77, p<0.001), agreeableness (beta coefficient: 0.35, 95% CI: 0.07 to 0.63, p=0.02) and neuroticism (beta coefficient: 0.49, 95% CI: 0.30 to 0.68, p<0.001). Female gender was negatively associated with sanction status (beta coefficient: -0.01, 95% CI: -0.02 to -0.01, p<0.001), openness (beta coefficient: -0.55, 95% CI: -0.97 to -0.13, p=0.02), and authoritative
communication style (beta coefficient: -1.42, 95% CI: -1.99 to -0.85, p=0.001). See Table 8.13 and Figure 8.7.

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Beta coefficients</th>
<th>95% CI</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female → Sanction</td>
<td>-0.01</td>
<td>-0.02 to -0.01</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Personality measures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female → Conscientiousness</td>
<td>0.55</td>
<td>0.33 to 0.77</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Female → Agreeableness</td>
<td>0.35</td>
<td>0.07 to 0.63</td>
<td>0.02</td>
</tr>
<tr>
<td>Female → Openness</td>
<td>-0.55</td>
<td>-0.97 to -0.13</td>
<td>0.02</td>
</tr>
<tr>
<td>Female → Extraversion</td>
<td>0.09</td>
<td>-0.16 to 0.35</td>
<td>0.40</td>
</tr>
<tr>
<td>Female → Neuroticism</td>
<td>0.49</td>
<td>0.30 to 0.68</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Communication styles</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female → Approachable</td>
<td>0.93</td>
<td>0.65 to 1.22</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Female → Authoritative</td>
<td>-1.42</td>
<td>-1.99 to -0.85</td>
<td>0.001</td>
</tr>
<tr>
<td>Empathy measures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female → Personal distress</td>
<td>0.03</td>
<td>-0.19 to 0.25</td>
<td>0.80</td>
</tr>
<tr>
<td>Female → Empathic concern</td>
<td>1.09</td>
<td>0.89 to 1.29</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Female → Perspective taking</td>
<td>0.93</td>
<td>0.71 to 1.16</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Female → Fantasy</td>
<td>1.08</td>
<td>0.79 to 1.37</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Table 8.13 Beta coefficients, 95% confidence intervals and p values of the relationships between gender, outcome, personality measures, communication styles and empathy measures used to construct path diagram (using multiple imputation)
Figure 8.7 Path diagram (using multiple imputation) showing the relationship between gender, empathy measures, communication styles, personality measures and outcome.

Path coefficients are standardised beta coefficients from the regression analysis. Solid black arrows indicate positive effects and red dashed arrows indicate negative effects. The width of the arrow lines is proportional to the effect sizes. Only paths significant at p<0.05 are shown. The relationships between the individual co-variates are not shown.

The path model was repeated without using multiple imputation for comparison with the model that used multiple imputation. The population of interest was restricted to those 798 participants who had complete data (this was to ensure that the same number of participants was included in each of the regression models in the path model). Similar to the model using multiple imputation, it showed evidence of a positive association between being female and the following variables: three empathy measure sub-scales including fantasy (beta coefficient: 1.57, 95% CI: 1.01 to 2.13, p<0.001), perspective-taking (beta coefficient: 0.84, 95% CI: 0.43 to 1.26, p<0.001) and empathic concern (beta coefficient: 1.11, 95% CI: 0.75 to 1.47, p<0.001),
approachable communication style (beta coefficient: 1.06, 95% CI: 0.65 to 
1.48, p<0.001), and conscientiousness (beta coefficient: 0.56, 95% CI: 0.31 
to 0.82, p<0.001) and agreeableness (beta coefficient: 0.40, 95% CI: 0.16 to 
0.63, p=0.001). Female gender was negatively associated with openness 
(beta coefficient: -0.57, 95% CI: -0.90 to -0.23, p=0.001), and authoritative 
communication style (beta coefficient: -1.35, 95% CI: -1.82 to -0.88, 
p<0.001). Unlike the path model using multiple imputation, no evidence for 
an association was found between gender and sanction status (beta 
coefficient: -0.01, 95% CI: -0.02 to 0.003, p=0.13), or gender and neuroticism 
(beta coefficient: 0.32, 95% CI: -0.02 to 0.66, p=0.06). This demonstrates 
that multiple imputation enabled the association between sanction status and 
the personality measure neuroticism to be shown (see Figure 8.8 and Table 
8.14).

<table>
<thead>
<tr>
<th></th>
<th>Beta coefficients</th>
<th>95% CI</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Personality measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female → Conscientiousness</td>
<td>0.56</td>
<td>0.31 to 0.82</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Female → Agreeableness</td>
<td>0.40</td>
<td>0.16 to 0.63</td>
<td>0.001</td>
</tr>
<tr>
<td>Female → Openness</td>
<td>-0.57</td>
<td>-0.90 to -0.23</td>
<td>0.001</td>
</tr>
<tr>
<td>Female → Extraversion</td>
<td>0.007</td>
<td>-0.27 to 0.28</td>
<td>0.96</td>
</tr>
<tr>
<td>Female → Neuroticism</td>
<td>0.32</td>
<td>-0.02 to 0.66</td>
<td>0.06</td>
</tr>
<tr>
<td><strong>Communication style</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female → Approachable</td>
<td>1.06</td>
<td>0.65 to 1.48</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Female → Authoritative</td>
<td>-1.35</td>
<td>-1.82 to -0.88</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Empathy measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female → Personal distress</td>
<td>-0.04</td>
<td>-0.43 to 0.35</td>
<td>0.85</td>
</tr>
<tr>
<td>Female → Empathic concern</td>
<td>1.11</td>
<td>0.75 to 1.47</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Female → Perspective taking</td>
<td>0.84</td>
<td>0.43 to 1.26</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Female → Fantasy</td>
<td>1.57</td>
<td>1.01 to 2.13</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Figure 8.8 Path diagram (not using multiple imputation) showing the relationships between gender, empathy measures, communication style, personality measures and outcome.

Path coefficients are standardised beta coefficients from the regression analysis. Solid black arrows indicate positive effects and red dashed arrows indicate negative effects. The width of the arrow lines is proportional to the effect sizes. Only paths significant at $p<0.05$ are shown. The relationships between the individual co-variates are not shown.

8.5 Discussion

8.5.1 Summary

This prospective cohort study of doctors who applied to at least one of five UK medical schools in 1990 confirms my earlier findings that women doctors are less likely to have faced disciplinary action against their medical registration (see Chapter 3 and Chapter 4) [9, 10].

Gender was shown to be associated with empathy, communication style and personality. There was no evidence for an association between disciplinary action and empathy, communication style, or personality traits. There was also no evidence to suggest that these factors may mediate the negative
association between female gender and disciplinary action likelihood in doctors.

8.5.2 Strengths and limitations

The original cohort population captured 70% of all applicants and entrants (with a European Community postal address) to UK medical schools and had a 93% response rate to the initial questionnaire. The original cohort members were therefore a representative sample of applicants to UK medical schools in 1990. It is therefore reasonable to assume that those original cohort members who successfully completed their UK medical school training and became registered with the GMC are also a representative sample of applicants to UK medical schools in 1990 who went on to qualify and practice as medical doctors in the UK.

The response rate for the questionnaires used in this study ranged from excellent (93% for the first questionnaire administered in 1990) to acceptable (56% and 58% for the questionnaires administered at the end of the final year of medical school and at the end of the pre-registration house officer (PRHO) year, respectively). These response rates, particularly to the first questionnaire, would suggest that the survey responses are representative of the study population.

A further advantage is that the co-variates of interest (empathy, communication style and personality) were measured before the outcome of interest (sanctions) occurred. The third questionnaire was administered towards the end of the cohort’s PRHO year, the first year the participants would have worked as qualified medical doctors. My earlier research has shown that the majority of sanctions occur after the doctor has been qualified for greater than ten years (see Chapter 3) [9]. It is therefore unlikely that the members of the cohort would have faced disciplinary action by the end of their first year of working. It can be safely assumed that the data on empathy, communication style and personality were collected before any of the doctors faced disciplinary action relating to their registration as a doctor.
This is important because the experience of going through a fitness to practise investigation may impact how the doctors responded to the empathy, communication style and personality questions in the questionnaires.

The ordering of the variables in the path model was based on when the data on each variable was collected, with variables collected at application to medical school being placed before variables collected in final year of medical school, which in turn was placed before variables collected in the first year post-qualification from medical school. It could be argued that the personality measures (which were collected during the participants’ first year working as a doctor post-qualification and therefore placed towards the end of the path model) could have been placed earlier in the path model as personality traits have been reported as relatively stable and that they are likely to influence empathy and communication style. However, there is evidence to suggest that personality changes do occur due to intrinsic maturation and due to social demands and experiences [338]. Due to the fact that the study population were young and the majority were likely to have been in their mid to late twenties when the personality measures were measured, and that this period of life tends to be when major life events occur, such as first job, marriage or moving in with a partner, starting a family, I felt it was reasonable to place the personality measure late in the path model.

An important limitation of the study is that very few of the participants had or have had sanctions against their medical registration: only 1% of the population. Despite this sparsity in outcome data, I chose to continue with this study. My primary reason for continuing with this study was my desire to use multiple imputation, firstly because I was keen to gain experience using this statistical method within the safety of doctoral research. Multiple imputation enabled me to retain the study participants who had or have sanctions against their medical registration, but who had missing data. Continuing with this study enabled me to explore whether the proportion of
doctors in this study who have experienced disciplinary action is in keeping with my earlier findings (see Chapter 3) [9], thereby confirming that this study’s population is reflective of the overall medical population. However, it must be highlighted that only 17 participants had active or historical sanctions, and of those only two were women. The sparsity of participants with the outcome of interest resulted in insufficient power to detect true differences in empathy, communication style and personality between those doctors without sanctions and those who have or had experienced sanctions by gender. This limitation arose because the cohort was not designed to explore gender differences in sanction risk. The population size was insufficient to explore such a rare outcome. A possible method to overcome the challenge of the sparsity of outcome data would be to follow cohorts of medical school applicants from multiple academic years.

However, the near equal proportions of men and women in the study cohort population enabled me to explore how empathy, communication style and personality differed between the genders in a population of doctors. Measures, such as personality, have been demonstrated to differ between the genders in the general population [322, 351], but little is known whether these measures differ between the genders in the highly selected population of medical doctors. This study sheds light on how men and women doctors may differ and allows for the development of possible hypotheses for the reasons why their professional performance differs in terms of sanction risk and postgraduate examination pass rates.

Though the response rate was good, due to the low power to detect true differences, any missing data would have had a large impact on findings. The decision to use multiple imputation was motivated by the desire to maintain as many study participants as possible in the regression models.

Though the study population was representative of UK trained doctors, it may not be representative of the UK medical profession as a whole, especially given the evidence that doctors who have trained outside of the UK have greater risk of facing disciplinary action (see Chapter 3) [9] and that in 2015
37% of doctors practising in the UK had obtained their primary medical qualification outside of the UK [184].

Finally, the measures used in this study were constrained by the variables, instruments of measurement and timings of data collection by of the original research project. The cohort data were not collected for this study and as such may not have been optimal for addressing this study’s aims. This study uses secondary data, which has enabled me to examine factors that are measured at the very start of the participants’ medical careers and to explore the relationship with an outcome that usually does not occur until a decade after qualification from medical school. Exploring the prospective relationship between the co-variates and the outcome would not have been possible with time constraints of doctoral research.

Throughout this study multiple statistical testing has been used. A p-value of <0.05 could be viewed as lenient given the multiple testing. However, adjusting the p-value in line with the multiple testing would have given a p-value that is likely to have been too conservative and would have increased the likelihood of a false negative association, especially in this study population of low power with a rare outcome.

8.5.3 Comparison with other studies

Despite the low number of participants with active or historical sanctions, a statistically significant association was found between female doctors being at reduced risk. This finding echoes those from my previous research (see Chapter 3 and Chapter 4) [9, 8, 10].

A positive association between the female gender and all four sub-scales of empathy measures was found. This is in keeping with previous studies of medical students [352-355].

Evidence for a difference in communication style between the genders was found, in this study, with women being more likely to report having an approachable communication style. This is in keeping with earlier research
that reported that communication styles differ between male and female doctors, with women being more empathic communicators and engaging in active partnership behaviours when communicating with patients [243, 310, 311].

Women were found to obtain higher scores for the personality traits neuroticism, agreeableness, and conscientiousness in this study. The gender difference in the personality sub-scales of neuroticism and agreeableness is replicated in studies examining the non-medical population of college students and adults [322, 351], however there was no consistent or significant evidence for a gender difference in the personality trait conscientiousness.

Interestingly, conscientiousness is a predictor of academic success in non-medical students [323] and a predictor of preclinical success in medical students [324, 273, 269, 356]. It would therefore be reasonable to hypothesise that the medical students who demonstrate high levels of conscientiousness will perform better as doctors. This hypothesis is further supported by findings from Hodgson et al. who found that there is a significant difference in the personality of medical students who demonstrate unprofessional behaviour during medical school and those who do not [75], especially given the link between unprofessional behaviour as a medical student and future disciplinary action by medical regulatory boards [59].

In this study I was unable to find evidence for an association between empathy, communication style or personality and sanction risk, however it is important to consider this study may not have had sufficient power to detect true associations between these co-variates and sanction risk. The research literature suggests that empathy, communication style and personality may influence the professional performance of medical students and therefore doctors.

Empathy is considered as the ability of a doctor to “understand a patient’s situation, perspective and feelings; communicate that understanding and
check its accuracy and to act on that in a helpful and therapeutic way" [357], and as such it is considered as the backbone of an effective doctor-patient relationship [357]. Higher levels of perceived doctor empathy are positively associated with clinical competence in medical students [354] and higher levels of patient satisfaction and compliance [332]. This in turn may be protective in a medical career and may contribute to patients being less likely to take action against their doctor [358, 332].

Earlier research has reported that the vast majority of malpractice suits against doctors are due to problems arising from interpersonal communication [359] and that those doctors with malpractice claims adopt higher communication dominance in the doctor-patient dyad [360]. These findings suggest an association between communication and malpractice and thus the communication style a doctor adopts may influence their perceived professional performance.

### 8.5.4 Unanswered questions and future research

The previous literature suggests there is a possible association between professional performance and empathy level, communication style, and personality. Unfortunately this current study has not been able to demonstrate whether these factors mediate the relationship between gender and sanction risk. It is possible that the lack of evidence is due to the low power of the study. It would be of interest to replicate the current study, but ensure there was sufficient power to detect any true associations. It is likely that multiple cohorts of medical school entrants over successive years would be required to enable this.

Previous studies suggest that empathy can be enhanced and improved through training [361-363], but that personality needs to be taken into account when designing programmes in undergraduate medical education [364]. Exploring further the different methods to deliver empathy and communication training to medical undergraduates and the impact of these methods on the communication skills of doctors may be of interest.
Interestingly, patients’ expectations of a doctor appear to be influenced by the doctor’s gender. A meta-analysis of patient satisfaction and doctors’ gender found that, when compared with male doctors, female doctors were not evaluated as highly by their patients as would be expected based on their practice style and patients’ values [365]. It has been reported that male medical students who demonstrate a patient-centred behaviours were rated higher in competence than those male students with less patient centred-behaviours. However, there was no association between patient-centred behaviours and competence ratings for female medical students [366]. This gender bias in the expectation of doctors’ behaviour suggests that any enhanced empathy and communication training is likely to only impact male medical students, and therefore male doctors. Women doctors do experience sanctions, albeit they have a reduced risk compared to men. This therefore suggests there may be other factors at play that increase a doctor’s risk of having sanctions imposed on their medical registration. It is likely that sanction risk is multi-factorial, some of which may be at the individual level, including behavioural or associated with personality, while other factors may be at an organisational level, including professional isolation [275]. Understanding the different factors that may influence professional performance and educating doctors to recognise those factors and the mechanisms to overcome or protect themselves from the negative consequences of these factors will improve the quality of care of the medical profession delivered to patients and the community as whole.

8.5.5 Conclusion

Through this study I have confirmed previous findings that women doctors are at a reduced risk of having sanctions imposed on their medical registration (see Chapter 3 and Chapter 4) [9, 10]. It has also demonstrated that there is a difference in empathy level, communication style and personality between male and female doctors. However, this study was unable to show evidence that empathy level, communication style or personality mediated the relationship between doctors’ gender and sanction risk.
Sanction risk is likely to be multifactorial and future work should focus on identifying these factors, their impact of sanction risk and whether they are modifiable or their impact can be reduced through education and training.
Chapter 9 - Discussion

“I was taught that the way of progress was neither swift nor easy.”

Marie Curie, first woman to win a Nobel Prize

“Science is not a boy’s game, it's not a girls’ game. It's everyone’s game. It’s about where we are and where we’re going.”

Nichelle Nichols, former NASA Ambassador

9.1 Chapter summary

In this final chapter I will present and discuss the findings of this doctoral research. I will discuss the strengths and the limitations of the research conducted throughout the course of this thesis and discuss my findings in context with other published research. I will close this chapter by discussing the unanswered questions my research has not addressed and my recommendations for future research.
9.2 Introduction

To answer the research question “is there a gender difference in doctors’ professional performance, and if so why?” I have applied quantitative methods, including multivariate statistical techniques to analyse large secondary datasets and meta-analyses of published research findings.

In the studies presented in Chapter 3 and Chapter 4 I investigated and quantified the gender difference in medico-legal action against doctors and established that women doctors were at reduced risk. In Chapter 5 and Chapter 6 I explored the association between gender and postgraduate medical examinations. I established that women doctors consistently outperformed men at clinical examinations, however the gender difference was not as clear in written examinations. In Chapter 7 I investigated the gender difference in the examination performance of a population of GPs who were under investigation by the GMC. I found that women doctors performed better at both the written and clinical components of the examination. I also demonstrated that clinical examination performance was a predictor of whether a warning or sanction was imposed on the doctor’s medical registration. Finally, with the study presented in Chapter 8 I explored whether empathy, communication style and personality traits mediated the relationship between gender and sanction risk. I confirmed that women doctors differed in empathy level, communication style and personality when compared to men, but these factors were not found to mediate the association between gender and sanction risk, though low power may have affected the findings of this study.

In summary, the main findings of this thesis are:

1. Women doctors are less likely to be subject to medico-legal action, across specialties, countries and over time.
2. Women doctors perform better than men at clinical examinations, across specialties.
3. Among a population of GPs being investigated for fitness to practise concerns, women doctors perform better than men at both the written and clinical examinations.
4. Gender differences exist in empathy level, communication style and personality amongst doctors.

9.3 Interpretation of findings

9.3.1 Finding 1: Women are less likely to be subject to medico-legal action

Prior to the publication of the work in this thesis, it had been established that the GMC was less likely to receive a complaint about a female doctor [180, 181]. It had also been demonstrated that complaints involving male doctors were at greater risk of being fully investigated by the GMC [367]. Previous research from Canada, USA, Australia and New Zealand had found a gender difference in the disciplinary action against doctors, with women being at reduced risk [189-195]. However, the difference in medical and legal systems in these countries compared to the UK limited the applicability of these findings to doctors practising in the UK. For doctors practising in the UK, the risk of being suspended or erased from the medical register (the severest type of sanctions) was found to be much less for women doctors [5], but this result was not adjusted for the effect of confounders, leaving the association open to speculation that the gender difference in disciplinary action amongst doctors was due differences in specialty choice, number of years of clinical practice, or country where primary medical qualification was obtained.

The study I presented in Chapter 3 showed that within the population of doctors registered to practise medicine in the UK in 2013, women were less likely to have a warning or sanction imposed against their medical registration and that this gender difference in warning or sanction risk remained after adjusting the results for the confounding effects of specialty, number of years since primary medical qualification, and area of the world where primary medical qualification was received - all potential confounders.
I also demonstrated that women doctors were at reduced risk of receiving the most severe type of sanctions (suspension and erasure). As a group, women practising as GPs were the least likely to have a warning or sanction imposed against their medical registration.

I established with the study presented in Chapter 4 that this gender difference remained across a broad spectrum of poor professional performance measures (including disciplinary action by a medical regulatory board, malpractice experience, referral to a medical regulatory body, complaints received by a healthcare complaints body, criminal cases, and medico-legal matter with a medical defence organisation), across multiple specialties, countries and over time.

In summary, I have established that women doctors were less likely to face medico-legal action and that this gender difference in a measure of extreme poor professional performance was not due to differences in specialty, years of experience, country of primary medical qualification or country in which the doctor was practising. I demonstrated that this gender difference has been constant for the last fifteen years and therefore it cannot be attributed to the shift in the gender composition of the medical workforce. These findings quash the argument that men are at greater risk of medico-legal action because there are more of them (in 2018 the UK medical register had a near equal split in men and women doctors, with 53% and 47% respectively). Nor can it be argued that men are at greater risk of medico-legal action because of the specialities they choose, the fact that they are more likely to be an international medical graduate, or that men are more likely to have greater years of experience – the gender difference in medico-legal action remained once these factors had been controlled for. Finally, these results demonstrate that the gender difference in medico-legal action is not confined to the UK, but appears to be a global phenomenon (at least in the “global West”). It would be reasonable to deduce from these findings that the professional performance of women doctors is less likely to raise serious concerns.
9.3.2 Finding 2: Women doctors perform better at clinical postgraduate examinations

Performance at professional examinations is another method of measuring professional performance in doctors. To progress in their chosen specialty and become registered on the Specialist or GP registers, doctors are required to complete postgraduate medical examinations to demonstrate that they have the knowledge and skills required to become a hospital specialist or GP.

Doctors who have had warnings or sanctions imposed against their medical registration could be viewed as a different and separate population of doctors when compared to the much larger population of doctors who have not been sanctioned. With this in mind, through Chapter 5 and Chapter 6 I sought to explore whether the gender difference observed in an extreme measure of poor professional performance (experiencing medico-legal action) is also present in a measure of an expected level of professional performance (passing postgraduate medical examinations).

In Chapter 5 I established that within a population of doctors who chose to complete the clinical component of the postgraduate medical examination required for core medical training in the UK, women candidates were more likely to pass this examination first time. This finding persisted after controlling for the effect of ethnicity, a known confounder of examination performance in doctors [127]. I also demonstrated that the effect of gender in first time pass rates was greater amongst those doctors who had obtained their primary medical qualification outside of the UK.

The study I presented in Chapter 6 echoed these findings by establishing that overall women doctors across different specialties and countries perform significantly better at clinical postgraduate medical examinations. There was no evidence to suggest a gender difference at the performance of written postgraduate medical assessments. This pattern of women performing well at clinical assessments, but no gender difference in written assessment performance was mirrored in Chapter 7 amongst GPs who volunteer to
complete the assessment used to assess doctors undergoing a fitness to practise investigation by the GMC.

Taken together, the findings of these studies reliably demonstrate that on average there is a gender difference in the performance of doctors at clinical assessments, with women outperforming men, but that there is no evidence of a gender difference in the performance of doctors at written assessments. These findings suggest that doctors differ in the skills and attributes assessed during clinical examinations. It is plausible that this difference in skills may go towards explaining the gender difference observed in disciplinary action.

9.3.3 Finding 3: Women perform better than men at both the written and clinical examination, among a population of GPs being investigated for fitness to practise concerns

The pattern of gender difference in performance at postgraduate medical examinations was established in Chapter 5 and Chapter 6, with women performing better than men at clinical assessments, but no evidence for a gender difference in performance at written assessments.

In Chapter 7 I established that this gendered pattern of performance is different amongst a population of GPs who were under investigation by the GMC for fitness to practise concerns. This study demonstrated that among this population, which has been highly selected for poor performance, women performed better than men at both the clinical and written assessments. These results replicate the finding of poorer male academic performance within this highly selected population and suggest that men under investigation for fitness to practise concerns may have particularly poor knowledge. I was also able to show that a low score in the clinical assessment was associated with an increased likelihood of having a warning or sanction imposed on the doctor’s medical registration. However, the final decision on warnings and sanctions showed no gender bias beyond the differences in test performance and therefore the gender difference in
sanctions within this population cannot be attributed differential treatment by the medical regulator by gender.

Finally, I confirmed that GPs under investigation by the GMC do not perform as well at the clinical and written assessments when compared to the GPs who volunteered to complete the assessments – an expected finding - confirming that the performance of those GPs being investigated by the GMC is below the standard of the overall GP population.

9.3.4 Finding 4: Gender differences exist in empathy level, communication style and personality amongst doctors

The findings from the earlier chapters led me to hypothesise that the gender difference in the professional performance of doctors may be due to a difference in the skills and attributes that male and female doctors bring to their clinical practice. In Chapter 8 I investigated whether empathy level, communication style and personality traits mediated the association between gender and sanction risk in a representative sample of a cohort of doctors who applied and were accepted to one of five English medical schools in 1990.

In agreement with my earlier findings, I confirmed that men were more likely to have an active or historical sanction against their medical registration. I also found a gender difference in empathy level, communication style and personality traits between male and female doctors. Women were found to have higher levels of empathy and to be more likely to adopt an approachable communication style. Women also scored more highly on the personality traits neuroticism, agreeableness and conscientiousness.

It would be reasonable to assume that the difference in empathy level, communication style and personality traits may influence practice styles between men and women, however I was unable to find any evidence that these factors were associated with sanction risk. However this finding must be interpreted with caution due to the small numbers of women doctors with
active or historical sanctions resulting in the study having low power to detect differences between the genders.

9.3.5 Overall interpretation of findings

Overall the results I have presented throughout this thesis have shown that female doctors perform better than male doctors on two measures of professional performance: being less likely to face medico-legal action; and being more likely to pass a clinical postgraduate medical examination. These gender differences in professional performance were found to be robust and remained once the results were adjusted for confounders.

I have also demonstrated that within the population of GPs who were under investigation by the GMC for fitness to practise concerns women performed better at both the written and clinical assessments than men GPs in the same situation, demonstrating that the pattern of performance between the genders differs in this special context of doctors under investigation.

Finally, I have established that men and women doctors differ in terms of empathy level, communication style and personality traits. Though I have not been able to establish any evidence that these factors are associated with sanction risk.

My findings have also shown how factors including specialty, whether the doctor qualified in medicine within the country they practice, how many years the doctor has been qualified, and ethnicity all influence the professional performance of doctors. However, when the effects of these factors are controlled for the difference in professional performance between the genders persists.

It is interesting to observe that overall the gender difference in performance at postgraduate medical examinations is only evident in the clinical components and not the written components, which tend to assess scientific and medical knowledge. This pattern in examination performance could be explained by women doctors being better at retaining and appropriately
applying theoretical scientific knowledge to clinical encounters, or perhaps a difference between the genders in their ability to communicate and demonstrate their scientific knowledge during a clinical examination. It is reasonable to hypothesise that women may exhibit certain behaviours during the clinical encounter that enable them to elicit crucial information or signs from the “patient” which result in a better overall performance.

A further and important hypothesis to consider for this difference in examination performance is the risk of examiner bias based on characteristics such as gender or ethnicity. Unlike written knowledge tests, which tend to be in a multiple-choice format and marked by computers, independent examiners assess candidate’s performance during a clinical examination. The clinical examination reviewed in Chapter 5 uses two independent examiners per station and earlier research had not found evidence for any examiner bias based on candidates’ gender [80], therefore suggesting that the gender difference in performance cannot be attributed to examiner bias in this examination. However, many clinical examinations use only one examiner per station, which may increase the possibility of examiner bias. It may be advisable for the organisers of clinical postgraduate examinations to consider how to mitigate the risk of examiner bias.

The gender difference in clinical examination performance may provide clues for the gender difference in risk of disciplinary action. It is possible that the skills required to perform well at a clinical examination are similar to the skills that are protective of disciplinary action in actual clinical practice. One hypothesis to explain the gender difference in the performance of doctors at the professional performance measures assessed could be that men and women differ in the skills that they bring to their medical practice and that modern medicine places a higher value on the skills that are demonstrated predominantly by women doctors. Though there are likely to be variations in the performance of these skills within and across both genders, it has been reported that women demonstrate skills that lead to active partnership
behaviours and empathic relationships with patients [244, 368, 310, 265-267, 311, 243]. My research supports this by finding that women doctors were more likely to score highly on empathy measures and approachable communication style (see Chapter 8). Demonstrating empathy has been associated with a higher rating of clinical competence in medical students and greater patient satisfaction and compliance [354, 332]. There is evidence for an association between communication style and malpractice risk in doctors [359, 360], with higher communication dominance and lower concern being associated with previous malpractice claims [360]. Medicine is evolving from the paternalistic model of the past to the modern expectation that doctors and patients should work in partnership, with patients sharing the decision-making responsibility in managing their healthcare needs. Patient-centred behaviour is associated with increased patient satisfaction [369, 370, 245] and therefore it would be reasonable to hypothesise that women doctors display the skills that marry with this shift in values of modern medicine thereby explaining the gender difference in the number of complaints and sanctions against doctors.

Personality differences were also shown to differ between male and female doctors in Chapter 8, with female doctors obtaining higher scores for the personality traits neuroticism, agreeableness and conscientiousness. The personality trait conscientiousness is associated with individuals being responsible, well-organised and prepared and has been associated with good job performance across a spectrum of occupations [274]. It is therefore reasonable to hypothesise that conscientious individuals are more likely to ensure that their knowledge and skills are up to date and they are less likely to undertake risky behaviour. No evidence was found for an association between personality traits and disciplinary action in Chapter 8, but conscientiousness has been found to be a predictor of academic success in medical students [356, 273, 324, 269]. Personality traits were also demonstrated to be associated with unprofessional behaviour in medical students [75].
The difference in calibre between men and women pursuing medical careers may explain the gender difference in the professional performance of doctors. It could be argued that medicine continues to attract high-achieving women, but high-achieving men are choosing alternative non-medical careers. However, the evidence to support this argument is lacking. Medicine remains one of the most competitive degree courses in the UK with stringent admission criteria including demonstration of high academic ability [46, 38]. A longitudinal study of intellectually talented individuals in the USA has demonstrated that, although there are marked gender differences in career choices, medicine is one of the few careers where there is no gender difference between these individuals, suggesting that these academically high achieving men and women are equally likely to choose medicine as a career [371].

Finally, it must be considered whether the presence of an implicit bias against men in medicine may explain the gender difference in performance. Implicit biases have been defined as “associations outside conscious awareness that lead to a negative evaluation of a person on the basis of irrelevant characteristics such as…gender” [372]. Implicit bias in medicine is most often explored from the perspective of doctors’ biases towards their patients, however it is possible that doctors who are men are regarded and treated less favourably in medicine and that there is an expectation that men’s professional performance should be higher than that of women. However, evidence suggests that it is women who are not receiving the credit afforded to their male colleagues for certain behaviours, such as patient centred skills [368]. Evidence also shows that women are underrepresented in medical leadership and academic roles in both the UK and the USA [373-375], suggesting that men are more successful in terms of promotion and career progression. Finally, the gender pay gap in medicine is in favour of men, with men earning on average over £10,000 more than their female colleagues [100]. Taken together this suggests that, within medical careers, male doctors are valued more highly in terms of career progression and financial remuneration. This pattern would be unusual in a career that was
biased against men. Nevertheless, informal concerns have been raised about the possibility of implicit bias and these concerns should be addressed.

9.4 Strengths and limitations of the research

The strengths and limitations specific to each study are described and discussed within the relevant chapters. The key strengths and limitations of the research conducted throughout this thesis are associated with the decision to use large secondary datasets. The use of secondary datasets brings with it both opportunities and potential problems. The advantages of using large secondary datasets are that the findings are representative and robust. The large numbers of participants ensured there was sufficient power, in the majority of the studies, to detect patterns. This was especially pertinent given the rarity of the outcome in some of the studies. It also enabled me to explore a population that would potentially have been difficult to access. However, a significant disadvantage was that the data was not originally collected to answer the research question of this thesis. In fact, the majority of the datasets analysed throughout this thesis were collected for administrative purposes, and not research. While the administrative datasets were fairly complete, with little missing data, these datasets were not collected for research purposes and therefore did not always contain all the variables of interest for my research question, or the data on certain variables was not collected in sufficient detail and therefore the depth and direction of the studies was limited.

The research methods I employed throughout this thesis have enabled a number of different research studies to be conducted using large quantities of data. The quantity of data, particularly the longitudinal data collected over nearly 20 years, and the number of studies performed would not have been possible within the timeframe of a PhD if the data had been collected first-hand. The use of large datasets ensured my findings were representative and robust and enabled me to confirm patterns in gender difference of professional performance and generate hypotheses to explain the observed patterns.
9.5 Unanswered question and further research

This thesis has started to understand the difference in professional performance between men and women doctors. It has confirmed that men and women doctors differ in medico-legal action risk and clinical postgraduate medical examination pass rates. It has also demonstrated that men and women doctors differed in terms of empathy level, communication style and personality traits.

This research has generated many hypotheses for the possible causes of the difference in professional performance between the genders, and these are discussed in the relevant study chapters. It is my opinion that the causes of the gender difference in professional performance in doctors are likely to be multi-faceted and as such future research could move forward in a number of directions.

I was unable to establish an association between poor professional performance and empathy, communication style and personality traits. However, it is likely that the low power of my study limited my ability to detect true associations. Exploring whether these factors are associated with the professional performance measures used throughout this thesis would be of interest. Both quantitative and qualitative studies could be used to delve deeper into this area, though any future quantitative studies would need to ensure that they have sufficient power to detect any true associations.

Patient-centred behaviours and skills are highly valued in the clinical setting with evidence suggesting an association with patient satisfaction and compliance [332]. Interestingly, several studies have demonstrated that the evaluation by patients of these patient-centred skills differs depending on the gender of the doctor [366, 365, 376, 377], suggesting that the value patients hold to certain skills demonstrated by their doctors are associated with their expectations of gender and gender roles. These findings indicate that patients hold gendered expectations of their doctors. Hall et al. [368] have questioned whether this gendered bias is also present in the medical
educators, examiners and regulators. This question merits being further explored. It would also be interesting to consider whether gendered expectations and perceptions of doctors may influence how men and women doctors behave and whether these influence which skills they bring to their daily medical practice. Further research is required to explore the ways in which men and women differ in skills and behaviours during clinical encounters, but also how the gendered expectations of patients and clinical colleagues may influence how certain skills and behaviours are acknowledged, interpreted and received.

Insight, or lack of, into one's performance may be a possible reason for the observed gender difference in professional performance. Insight in doctors has been defined as “an awareness of one’s performance in the spectrum of medical practice” [378] and it has been put forward that insight and performance may be associated [378, 275]. It could be argued that there is an equal proportion of women and men doctors whose professional performance fails to meet the expected standard, but women may be more aware of their failings and choose to take premeditative action - perhaps through retiring or leaving medicine to choose an alternative career path before their poor performance reaches a level that raises concerns, or choosing not to sit examinations until they are confident in their abilities. Delving into the previous performance of doctors who choose to remove themselves from the medical register and the reasons why these doctors have chosen voluntary erasure may shed light on this hypothesis.

It is possible that the gender difference observed in professional performance is associated with difference in practice patterns between women and men doctors. Women and men have been shown to differ in practice patterns with differences in types of employment contract, hours worked and therefore number of patient encounters [379, 224, 225]. These differences in practice patterns have been associated with the risk of medico-legal action [223-225]. It would be of interest to further examine how any gender differences in practice patterns may be associated with professional performance.

302
Both disciplinary action against medical registration and passing examinations are important measures of professional performance in practising doctors, but a crucial question to consider is whether they are measures of what make a good doctor. It is my belief that examinations, despite the great efforts made to replicate true clinical practice, will always be limited in what they can measure in terms of “real-life” performance. However, I have no doubt that when action is taken against a doctor’s medical registration, the procedures and processes that have assessed the doctor’s performance have identified when a doctor’s practise was not meeting the standard expected. Nonetheless I believe it is unlikely that all doctors whose performance falls below the expected standard are captured by this system. I therefore suggest that there may be other methods through which the medical education research community could identify and measure professional performance in doctors and I feel strongly that future research in this area should focus on different and perhaps novel ways of assessing what professional performance is and how to measure it. Examples of research in this area include work using the UKMED to examine gender differences in specialty recruitment outcomes of UK medical graduates [120] and the current work exploring the gender pay gap in medicine, which will go towards revealing any horizontal and vertical occupational segregation in medical careers. Future research employing qualitative methods would also allow for more detailed information to be collected and will facilitate a more thorough understanding with the ability to explore factors such as the influence of organisational structures and local cultures. The insights gained from qualitative research methods are likely to provide context, generate further hypotheses, and inform and develop further quantitative studies.

9.6 Conclusion

The research in this thesis describes how men and women doctors differ on two measures of professional performance: medico-legal action and clinical postgraduate medical examination pass rates. The gender difference in medico-legal action cannot be attributed to number of years since a doctor has been qualified, whether the doctor qualified outside of the UK, or
specialty. The gender difference in medico-legal action remained when the outcome definition was broadened to include outcomes of varying severity, internationally and over time. Women doctors were found to perform better at clinical postgraduate medical examinations with different objectives. Within a population of GPs being investigated for fitness to practise concerns, clinical examination performance was found to be a predictor of whether a warning or sanction was imposed on the GP’s medical registration.

Finally, there were gender differences in levels of empathy, communication style and personality, and although there was no evidence to suggest that empathy, communication style or personality were significantly correlated with sanctions this may have been due to low power.

These findings suggest that the skills women demonstrate during a clinical assessment are highly valued by the medical colleges and medical regulator. The findings also suggest that there is a gender difference amongst doctors in terms of empathy, communication skills and personality, which could suggest that men and women doctors display different behaviours when interacting with patients and clinical colleagues. It is reasonable to hypothesise that the gender difference in disciplinary action may in part be due to this gender difference in clinical skills and behaviours. Finally, this thesis has demonstrated that the increase in women studying and practising medicine has not had a negative impact on the medical profession in terms of professional performance, in fact quite the opposite with women being less likely to face disciplinary action and more likely to perform better at high-stakes clinical examinations.

Overall, though I have been able to make inroads in tackling my chosen research question, I have not been able to pin down the underlying phenomenon of what is driving the gender difference in the professional performance of doctors. I have, however, been able to generate hypotheses and future research should be performed to further understand, maintain and improve the professional performance in doctors.
References


11. Gebelhoff R. Male doctors are more likely to be sued than females, study finds. The Washington Post. 2015 18 August 2015.

12. Halkon R. Male doctors are more than TWICE as likely to be sued or struck off than female doctors. The Mirror. 2015 13 August 2015.


14. McIntosh J. Male doctors more likely to face legal action than female colleagues. Medical News Today. 2015 13 August 2015.

15. Reporter DM. Legal risk to male doctors is higher: Female peers are two-and-a-half times less likely to face being struck off or sued for negligence. Daily Mail. 2015 13 August 2015.


17. Rimmer A. Male doctors are more likely to receive sanctions than female doctors. BMJ Careers. 2014.

18. Torjesen I. Male doctors are more likely to have legal action taken against them, analysis finds. BMJ. 2015;351(h4359). doi:10.1136/bmj.h4359.


33. Griffiths R, editor. Measuring doctors’ performance - United Kingdom. 5th International Medical Workforce Conference; 2000; Sydney, Australia.


55. General Medical Council. How are students assessed at medical schools across the UK? Manchester2014.


80. McManus IC, Elder AT, Dacre J. Investigating possible ethnicity and sex bias in clinical examiners: an analysis of data from the MRCP(UK) PACES


90. Winyard G. Women will have to adapt as they become the majority, and so will the NHS. Br Med J. 2009;338(b2223).


110. Martin IG, Stark P, Jolly B. Benefiting from clinical experience: the influence of learning style and clinical experience on performance in an


115. Royal College of Physicians. MRCP(UK) and Specialty Certificate Examinations Pass Rates by Gender and Ethnicity2017.


118. Rimmer A. More female doctors than male doctors have had their revalidation deferred, GMC figures show. Br Med J. 2014;348(g2158). doi:https://doi.org/10.1136/bmj.g2158


144. Hardy J. Hospital heart unit was consumed by 'dark force' as patients were put at risk by a dysfunctional team of surgeons, leaked report reveals. Independent. 2018 4 August 2018.


148. Medical Act 1858, (1858).


150. List of Registered Medical Practitioners [database on the Internet]. Available from: https://webcache.gmc-uk.org/gmclrmp_enu/start.swe?SWECmd=GotoView&SWEBHWND=&_sn=yz8Ztm.3UhoIBD3E-ulbYiilcsoBq0bsDMzGs7pKPQnq4Aw-gKqqRFMSzqS7z0FOrivxrXHL35tfFlFJTHjdHcIlA9UqN-sAk09h7vYWgajZuBpjB48U7WSnufkiYtFLCUiPEECFu0dn3XJkoS-Ym-.naSOWF8ECK2DKuA11bgpolVV1iqhVeGcr8YqyAT7xSGdRDpxL &SWEVi ew=GMC+WEB+Doctor+Search&SRN=&SWEHo=webcache.gmc-uk.org&SWETS=1533131477&SWEApplet=GMC+WEB+Health+Provider+S earch+Applet. Accessed:


157. McManus IC, Wakeford R. PLAB and UK graduates' performance on MRCP(UK) and MRCGP examinations: data linkage study. BMJ. 2014;348:g2621. doi:10.1136/bmj.g2621.


166. Archer J, Regan de Bere S, Bryce M, Nunn S, Lynn N, Coombes L et al. Understanding the rise in Fitness to Practise complaints from members of the public: Plymouth University2014.


178. Sedgwick P. Bias in observational study designs: prospective cohort studies. BMJ. 2014;349:g7731. doi:10.1136/bmj.g7731.


234. Ozen S, Liman Z, Es H, Ozer Y, Sanli AN, Ozver I et al. Psychiatry cases alleged as malpractice that were evaluated by the 3rd specialty board of council of Turkish forensic medicine; An archival research between the years 2005-2010. Klinik Psikofarmakoloji Bulteni. 2013;23 (supplement 1)(23):S287.


257. McManus IC, Ludka K. Resitting a high-stakes postgraduate medical examination on multiple occasions: nonlinear multilevel modelling of


273. Ferguson E, James D, O'Hehir F, Sanders A, McManus IC. Pilot study of the roles of personality, references, and personal statements in relation to


313. Rimmer A. GMC warnings should be scrapped, says MDU. BMJ Careers. 2014.


342. McManus IC, Richards P, Winder BC, Sproston KA, Styles V. Medical school applicants from ethnic minority groups: identifying if and when they are disadvantaged. BMJ. 1995;310(6978):496-500.


349. Costa PT, McCrae RR. The revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual. Odessa, Florida: Psychological Assessment Resources; 1992a.


367. Humphrey C. Clarifying the factors associated with progression of cases in the GMC's Fitness to Practise process. Swindon2009.


Appendices

This section contains the documents relevant to Chapter 2:

UCL Research Ethics Committee approval documents;

UCL Data Protection Office documents.
# Chapter 2 – Study populations, context and methodology

## Ethics documentation

**UCL RESEARCH ETHICS COMMITTEE**

**APPLICATION FORM**

### SECTION A  APPLICATION DETAILS

<table>
<thead>
<tr>
<th>A1</th>
<th>Project Title: Sex differences in the professional performance of doctors practising in the UK.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Date of Submission:</strong> August 2013  <strong>Proposed Start Date:</strong> August 2013  <strong>UCL Ethics Project ID Number:</strong> 5025/001  <strong>Proposed End Date:</strong> September 2015</td>
</tr>
<tr>
<td></td>
<td><strong>If this is an application for classroom research as distinct from independent study courses, please provide the following additional details:</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Course Title:</strong>  <strong>Course Number:</strong></td>
</tr>
</tbody>
</table>

### SECTION B  PRINCIPAL RESEARCHER

<table>
<thead>
<tr>
<th>A2</th>
<th><strong>Principal Researcher</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Please note that a student – undergraduate, postgraduate or research postgraduate cannot be the Principal Researcher for Ethics purposes.</td>
</tr>
<tr>
<td></td>
<td><strong>Full Name:</strong> Professor Jane Dacre  <strong>Position Held:</strong> Director and PhD supervisor</td>
</tr>
<tr>
<td></td>
<td><strong>Email:</strong>  <strong>Telephone:</strong>  <strong>Fax:</strong></td>
</tr>
</tbody>
</table>

**Declaration To be Signed by the Principal Researcher**

- I have met with and advised the student on the ethical aspects of this project design (applicable only if the Principal Researcher is not also the Applicant).
- I understand that it is a UCL requirement for both students & staff researchers to undergo Disclosure and Barring Service (DBS) Checks when working in controlled or regulated activity with children, young people or vulnerable adults. The required DBS Check Disclosure Number(s) is: N/A
- I have obtained approval from the UCL Data Protection Officer stating that the research project is compliant with the Data Protection Act 1998. My Data Protection Registration Number is: Z6365106/2013/08/11
- I am satisfied that the research complies with current professional, departmental and university guidelines including UCL’s Risk Assessment Procedures and insurance arrangements.
- I undertake to complete and submit the ‘Continuing Review Approval Form’ on an annual basis to the UCL Research Ethics Committee.
- I will ensure that changes in approved research protocols are reported promptly and are not initiated without approval by the UCL Research Ethics Committee, except when necessary to eliminate immediate or potentially hazardous to the participant.
- I will ensure that all adverse or unforeseen problems arising from the research project are reported in a timely fashion to the UCL Research Ethics Committee.
- I will undertake to provide notification when the study is complete and if it fails to start or is abandoned.

**SIGNATURE:**

**DATE:** August 2013
A3 Applicant(s) Details
(If Applicant is not the Principal Researcher e.g. student details)

Full Name: Dr. Emily Unwin
Position Held: MPhil/PhD student

Full Name: Dr. Katherine Wood
Position Held: Lecturer and PhD supervisor

A4 Sponsor/ Other Organisations involved and Funding

a) Sponsor: ☐ UCL ☐ Other Institution
   If your project is sponsored by an institution other than UCL, please provide details: N/A

b) Other Organisations: If your study involves another organisation, please provide details. Evidence that the relevant authority has given permission should be attached or confirmed provided that this will be available upon request. General Medical Council and Royal Colleges of Physicians.

c) Funding: What are the sources of funding for this study and will the study result in financial payment or payment in kind to the department or College? If study is funded solely by UCL, this should be stated; the section should not be left blank. N/A

A5 Signature of Head of Department or Chair of the Departmental Ethics Committee
(This must not be the same signature as the Principal Researcher)

I have discussed this project with the principal researcher who is suitably qualified to carry out this research and I approve it. The project is registered with the UCL Data Protection Officer, a formal signed disk assessment form has been completed, and appropriate Insurance arrangements are in place. Links to details of UCL's policies on data protection, risk assessment, and Insurance arrangements can be found at: http://ethics.grad.ucl.ac.uk/procedures.php

UCL is required by law to ensure that researchers undergo a Disclosure and Barring Service (DBS) Check if their research project puts them in a position of trust with children under 16 or vulnerable adults.

"HEAD OF DEPARTMENT TO DELETE BELOW AS APPLICABLE"

I am satisfied that checks:  
(1) have been satisfactorily completed
(2) have been initiated
(3) are not required

If checks are not required please clarify why below.

Chair's Action Recommended: ☐ Yes ☐ No

A recommendation for Chair's action can be based only on the criteria of minimal risk as defined in the Terms of Reference of the UCL Research Ethics Committee.

PRINT NAME: David Lomas
B1 Please provide a brief summary of the project in simple prose outlining the intended value of the project, giving necessary scientific background (max 500 words).

Ensuring that today’s doctors are fit to practice has become a priority for the medical profession following high profile cases of doctors like Harold Shipman and Rodney Ledward and the Bristol Heart Scandal. Over the last twenty years performance assessments have been developed in the United Kingdom (UK) to assess the quality of care delivered by doctors with the ultimate aim of ensuring high quality patient care.

Identifying doctors who are in difficulty at an early stage is important to provide the support necessary to improve their practice [1]. Identifying reasons that doctors perform poorly and well will enable those of us working in medical education to develop educational strategies to reduce the likelihood of poor professional performance in future doctors.

Research from the USA and UK exploring the characteristics associated with doctors identified as performing poorly has shown that male doctors are more likely to be disciplined by a medical regulatory board than female doctors [2, 3, 4]. This PhD research aims to explore the reasons for sex differences in the performance of doctors by analysing national data on various measures of doctors’ academic and professional performance (source of data given in brackets):

1. General Medical Council (GMC) Fitness to Practice data between 2006-2012 (GMC).
2. List of Registered Medical Practitioners 2006-2012 (GMC).
3. Annual Review of Competence Progression data between 2010-2012 (GMC).
4. Recruitment data for Foundation and Core Training in 2012 (GMC).
5. Membership of the Royal Colleges of Physicians (MRCP) examination data 2001-2013 (Royal Colleges of Physicians, RCP)
7. Doctors’ performance in the GMC Fitness to Practice Tests of Competence (UCL).

We are also currently in the process of completing a systematic review and meta-analysis of the relationship between sex and performance in medicine.

The research aims to explore the performance of doctors practising in the UK and to determine which factors influence differences in male and female doctors’ performance outcomes.
The objectives of the research are to:

1. Examine whether male doctors practising in the UK perform more poorly than female doctors on a variety of measures of knowledge, skills and attributes and at different stages in doctors’ careers.

2. Identify factors that differ between male and female doctors and which may mediate or moderate the relationship between sex and performance outcomes and to identify whether these factors are remediable.

References:

Briefly characterise in simple prose the research protocol, type of procedure and/or research methodology (e.g. observational, survey research, experimental). Give details of any samples or measurements to be taken (max 500 words).

The research will consist of several observational studies to examine large sets of data, which have already been collected. Analyses will be conducted with the statistical software packages SPSS and STATA. Descriptive statistics will be performed to explore and describe the study population, with each variable being examined. Logistic regression will be used to examine the relationship between sex of the doctor and the outcome measure and to examine how other demographic factors impact on the differences found.

A description of the datasets that will be used for the research has been included below:

1. List of Registered Medical Practitioners May 2013:

   This is a list of the doctors who are or have been registered with the GMC. This data is collected by the GMC. This is identifiable data, which is publicly available. Permission has been sought to access and use this data for research.

   It is our understanding that research projects using this dataset meet the criteria for Exemption A.

2. GMC data including the GMC Fitness to Practice data, Annual Review of Competence Progression data, Recruitment data for Foundation and Core Training and the List of Registered Medical Practitioners:
This data includes a variety of demographics measures and performance measures of doctors who are or have been registered with the GMC. This data is collected by the GMC. Permission has been sought to access and use this data for research. The GMC will provide a pseudonymised dataset.

It is our understanding that the research projects using this dataset meet the criteria for Exemption B.

3. RCP data:

This data includes the performance of doctors who have chosen to sit the MRCP examinations along with demographic measures. This data is collected by the RCP. Permission has been sought to access and use this data for research. The RCP will provide a pseudonymised dataset.

It is our understanding that the research projects using this dataset meet the criteria for Exemption B.

4. Cohort data looking at doctors in the UK:

This data includes survey responses for a group of doctors. This data was collected by Professor McManus, UCL. Permission has been sought to access and use this data for research. Professor McManus will provide a pseudonymised dataset.

Professor McManus was informed by the UCL REC on 6th May 2008 that research using this data would be exempt from obtaining ethical approval under Exemption C.

5. Tests of Competence data and Pilot data:

This data includes the performance of doctors who have been requested by the GMC to complete a performance assessment, which includes the Tests of Competence, because their fitness to practise has been called into question. It also includes "doctors in good standing" who have volunteered to sit the Tests of Competence to test the materials. This data is collected by UCL on behalf of the GMC. The data is identifiable.

Professor Dacre applied for ethics approval to complete studies on this data and on the 23rd November 2009 the UCL Research Ethics Committee approved the study until July 2014.

*Attach any questionnaires, psychological tests, etc. (a standardised questionnaire does not need to be attached, but please provide the name and details of the questionnaire together with a published reference to its prior usage).*

---

**B3**

Where will the study take place (please provide name of institution/department)?

If the study is to be carried out overseas, what steps have been taken to ensure research and ethical permission in the study country? Is the research compliant with Data Protection legislation in the country concerned or is it compliant with the UK Data Protection Act 1998?

*Academic Centre for Medical Education, UCL Medical School, UCL.*

**B4**

Have collaborating departments whose resources will be needed been informed and agreed to participate?

*Attach any relevant correspondence.*

*Yes.*
B5 How will the results be disseminated, including communication of results with research participants?

Highlights of the research outcomes will be published in UCL research reports, on line and in newsletters. The research undertaken as part of the PhD thesis will be presented at academic and medical conferences and in clinical and medical education peer-reviewed journals. The research will be presented as part of Dr. Emily Unwin’s thesis. The researchers will liaise with the GMC and RCP to ensure they are aware of the research findings. The research participants will not be contacted.

B6 Please outline any ethical issues that might arise from the proposed study and how they will be addressed. Please note that all research projects have some ethical considerations so do not leave this section blank.

The main potential ethical issue would be regarding the Tests of Competence data. Doctors who have had their fitness to practice called into question and are requested by the GMC to complete the Tests of Competence are not informed by the GMC that their data may be used for research purposes. This dataset contains identifiable data. However, the research is interested in trends and the “bigger picture” and individual data will not be published. We feel that it is in the public’s interest to examine this data to explore poor professional performance in doctors.

SECTION C DETAILS OF PARTICIPANTS

C1 Participants to be studied

<table>
<thead>
<tr>
<th>C1a. Number of volunteers:</th>
<th>&gt;300,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper age limit:</td>
<td>N/A</td>
</tr>
<tr>
<td>Lower age limit:</td>
<td>16</td>
</tr>
</tbody>
</table>

C1b. Please justify the age range and sample size:

It is not possible to state the exact size of the study population until the datasets are received. To help give a rough estimate there were just over 329,000 doctors on the List of Registered Medical Practitioners in May 2013.

There will be no upper age limit.

The youngest study participants would be those who would have completed the first survey of Professor McManus’ Cohort in the autumn of 1990. These were students applying to medical school and therefore it is possible the youngest participant would have been 16 years old.

The vast majority of the study population will be adults, because the study population is doctors.

C2 If you are using data or information held by a third party, please explain how you will obtain this. You should confirm that the information has been obtained in accordance with the UK Data Protection Act 1998.

Data will be obtained from the GMC, RCP and Professor McManus. The data will be pseudonymised before the data is released to the researchers. The UCL Data Protection Team are aware of the format in which the data will be received.
C3 Will the research include children or vulnerable adults such as individuals with mental health problems or with learning disabilities, the elderly, prisoners or young offenders?  ☑ Yes ☐ No

How will you ensure that participants in these groups are competent to give consent to take part in this study? If you have relevant correspondence, please attach it.

It is possible that the youngest participants who completed Professor McManus’ first survey in the autumn of 1990 were 16 years old. However, this data has been collected. It would not be appropriate nor possible for the current researchers to seek consent retrospectively. The current researchers will receive a pseudonymised dataset and will therefore not be able to identify the participants.

C4 Will payment or any other incentive, such as gift service or free services, be made to any research participant?  ☑ Yes ☐ No

If yes, please specify the level of payment to be made and/or the source of the funds/gift/service to be used.

The "doctors in good standing" who sit the Tests of Competence to test the materials are given £350.00 plus travel expenses by the GMC. However, this current research does not involve any payment or other incentive. It is using data that has been collected as a result of these pilot studies, amongst others.

Please justify the payment/incentive you intend to offer.

N/A

C5 Recruitment

(i) Describe how potential participants will be identified:

N/A

(ii) Describe how potential participants will be approached:

N/A

(iii) Describe how participants will be recruited:

N/A

Attach recruitment emails/ads/urls/webpages. A data protection disclaimer should be included in the text of such literature.

C6 Will the participants participate on a fully voluntary basis?  ☑ Yes ☐ No

Will UCL students be involved as participants in the research project?  ☑ Yes ☐ No

If yes, care must be taken to ensure that they are recruited in such a way that they do not feel any obligation to a teacher or member of staff to participate.

Please state how you will bring to the attention of the participants their right to withdraw from the study without penalty?

N/A
C7 CONSENT
Please describe the process you will use when seeking and obtaining consent.

Only the “doctors in good standing” who are sitting the Tests of Competence voluntarily are asked to consent. When these doctors attend a pilot session they are asked to sign a register which states that “I agree to the data collected as part of this pilot process being used anonymously for GMC Fitness to Practise reports and research purposes. All data will be confidential and stored in accordance with the Data Protection Act 1998. I understand that I am free to withdraw from this pilot at any time.”

A copy of the participant information sheet and consent form must be attached to this application. For your convenience proformas are provided in C10 below. These should be filled in and modified as necessary.

In cases where it is not proposed to obtain the participants informed consent, please explain why below.

For the remaining datasets that will be used for this research, consent was not sought when the data was collected. However, it is felt that it is in the public’s interest and of educational value to use this data for research purposes.

C8 Will any form of deception be used that raises ethical issues? If so, please explain.

N/A

C9 Will you provide a full debriefing at the end of the data collection phase? □ Yes □ No

If No, please explain why below.

The data has been already collected from a variety of sources. There are a very large number of study participants, the majority of which are not aware that their data is being used for research purposes. The research is interested solely in trends and the “bigger picture” and no individual data will be published. It would not be feasible, nor would it be appropriate to contact each doctor whose data has been used in this research.

C10 Information Sheets And Consent Forms

A poorly written Information Sheet(s) and Consent Form(s) that lack clarity and simplicity frequently delay ethics approval of research projects. The wording and content of the Information Sheet and Consent Form must be appropriate to the age and educational level of the research participants and clarify data in simple non-technical language what the participant is agreeing to. Use the active voice e.g. “we will book” rather than “bookings will be made”. Refer to participants as “you” and yourself as “I” or “we”.

An appropriate translation of the Forms should be provided where the first language of the participants is not English. If you have different participant groups you should provide Information Sheets and Consent Forms as appropriate (e.g. one for children and one for parents/guardians) using the templates below. Where children are of a reading age, a written Information Sheet should be provided. When participants cannot read or the use of forms would be inappropriate, a description of the verbal information to be provided should be given. Please ensure that you read the forms to an age-appropriate person before you submit your application.
Information Sheet for N/A in Research Studies

You will be given a copy of this information sheet.

Title of Project: N/A

This study has been approved by the UCL Research Ethics Committee (Project ID Number): N/A

Name: N/A
Work Address: N/A
Contact Details: N/A

We would like to invite N/A to participate in this research project.

Details of Study: N/A

Please discuss the information above with others if you wish or ask us if there is anything that is not clear or if you would like more information.

It is up to you to decide whether to take part or not; choosing not to take part will not disadvantage you in any way, if you do decide to take part you are still free to withdraw at any time and without giving a reason.

All data will be collected and stored in accordance with the Data Protection Act 1998.

When you have completed your Information Sheet, please DELETE the advice section below from your application form before submitting it to the Committee.

Details of Study MUST include the following:
- Aims of the research and possible benefits.
- Who you are recruiting.
- What will happen if the participant agrees to take part (when, where, how long etc).
- Any risks (e.g., need for disclosure of information to a third party, possibility for distress).
- Possible benefits (it is good practice to offer participants a copy of the final report).
- Arrangements for ensuring anonymity and confidentiality (use optional statements below for examples). To ensure compliance with the Data Protection Act participants must be informed of what information will be held about them and who will have access to it (this relates to information that is identifiable or could potentially be linked back to an individual.)

Statements which researchers MIGHT also include as appropriate:
- A decision to withdraw at any time, or decision not to take part, will not affect the standard of care/education you receive.
- If you agree to take part you will be asked whether you are happy to be contacted about participation in future studies. Your participation in this study will not be affected should you choose not to be re-contacted.
- You may withdraw your data from the project at any time up until it is transcribed for use in the final report (insert date).
- Recorded interviews will be transcribed (written up) and the tape will then be wiped clear.
- If you decide to take part you will be given this information sheet to keep and be asked to sign a consent form.
- Submission of a completed questionnaire implies consent to participate.
- As participation is anonymous it will not be possible for us to withdraw your data once you have returned your questionnaire.
Informed Consent Form for N/A in Research Studies

Please complete this form after you have read the Information Sheet and/or listened to an explanation about the research.

Title of Project: N/A

This study has been approved by the UCL Research Ethics Committee (Project ID Number): N/A

Thank you for your interest in taking part in this research. Before you agree to take part, the person organising the research must explain the project to you.

If you have any questions arising from the Information Sheet or explanation already given to you, please ask the researcher before you to decide whether to join in. You will be given a copy of this Consent Form to keep and refer to at any time.

Participant’s Statement

I N/A

• have read the notes written above and the Information Sheet, and understand what the study involves.
• understand that if I decide at any time that I no longer wish to take part in this project, I can notify the researchers involved and withdraw immediately.
• consent to the processing of my personal information for the purposes of this research study.
• understand that such information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998.
• agree that the research project named above has been explained to me to my satisfaction and I agree to take part in this study.

Signed: ___________________________ Data: N/A

When you have completed your Informed Consent Form, please DELETE the advice section below from your application form before submitting it to the Committee.

Statements which researchers MIGHT include as appropriate:

• I understand that my participation will be taped/video recorded and I consent to use of this material as part of the project.
• I understand that I must not take part if N/A
• I agree to be contacted in the future by UCL researchers who would like to invite me to participate in follow-up studies.
• I understand that the information I have submitted will be published as a report and I will be sent a copy. Confidentiality and anonymity will be maintained and it will not be possible to identify me from any publications.
• I understand that I am being paid for my assistance in this research and that some of my personal details will be passed to UCL Finance for administration purposes.
• I agree that my non-personal research data may be used by others for future research. I am assured that the confidentiality of my personal data will be upheld through the removal of identifiers.
This is not an exhaustive list and you should consider whether you need to amend any of these statements or design different ones that are more applicable to your research.

### SECTION D DETAILS OF RISKS AND BENEFITS TO THE RESEARCHER AND THE RESEARCHED

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>D1</strong> Have UCL’s Risk Assessment Procedures been followed?</td>
<td>☐ Yes ☒ No</td>
</tr>
<tr>
<td>If No, please explain.</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>D2</strong> Does UCL’s Insurer need to be notified about your project before insurance cover can be provided?</td>
<td>☒ Yes ☐ No</td>
</tr>
<tr>
<td>The insurance for all UCL studies is provided by a commercial insurer. For the majority of studies the cover is automatic. However, for a minority of studies, in certain categories, the insurer requires prior notification of the project before cover can be provided. If Yes, please provide confirmation that the appropriate insurance cover has been agreed. Please attach your UCL insurance registration form and any related correspondence.</td>
<td></td>
</tr>
<tr>
<td><strong>D3</strong> Please state briefly any precautions being taken to protect the health and safety of researchers and others associated with the project (as distinct from the research participants).</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>D4</strong> Will these participants participate in any activities that may be potentially stressful or harmful in connection with this research?</td>
<td>☐ Yes ☒ No</td>
</tr>
<tr>
<td>If Yes, please describe the nature of the risk or stress and how you will minimise and monitor it.</td>
<td></td>
</tr>
</tbody>
</table>
### D5
Will group or individual interviews/questionnaires raise any topics or issues that might be sensitive, embarrassing or upsetting for participants?

- If Yes, please explain how you will deal with this.
- N/A

### D6
Please describe any expected benefits to the participant.

- N/A

### D7
Specify whether the following procedures are involved:
- Any invasive procedure(s)  ☐ Yes  ☒ No
- Physical contact  ☐ Yes  ☒ No
- Any procedure(s) that may cause mental distress  ☐ Yes  ☒ No

Please state briefly any precautions being taken to protect the health and safety of the research participants.

- N/A

### D8
Does the research involve the use of drugs?  ☐ Yes  ☒ No

- If Yes, please name the drug/product and its intended use in the research and then refer to Appendix I

Does the project involve the use of genetically modified materials?  ☐ Yes  ☒ No

- If Yes, has approval from the Genetic Modification Safety Committee been obtained for work?  ☐ Yes  ☒ No

- If Yes, please quote the Genetic Modification Reference Number:
**D9**

Will any non-ionising radiation be used on the research participant(s)?

- [ ] Yes
- [x] No

If Yes, please refer to Appendix II.

---

**CHECKLIST**

Please submit either 12 copies (1 original + 11 double sided photocopies) of your completed application form for full committee review or 3 copies (1 original + 2 double sided copies) for chair’s action, together with the appropriate supporting documentation from the list below to the UCL Research Ethics Committee Administrator. You should also submit your application form electronically to the Administrator at: ethics@ucl.ac.uk

### Documents to be Attached to Application Form (if applicable)

<table>
<thead>
<tr>
<th>Section B: Details of the Project</th>
<th>Ticked if attached</th>
<th>Tick if not relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Questionnaire(s) / Psychological Tests</td>
<td>[ ]</td>
<td>[x]</td>
</tr>
<tr>
<td>• Relevant correspondence relating to involvement of collaborating department/s and agreed participation in the research.</td>
<td>[x]</td>
<td>[ ]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section C: Details of Participants</th>
<th>Ticked if attached</th>
<th>Tick if not relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Parental/guardian consent form for research involving participants under 18</td>
<td>[ ]</td>
<td>[x]</td>
</tr>
<tr>
<td>• Participant’s information sheet</td>
<td>[ ]</td>
<td>[x]</td>
</tr>
<tr>
<td>• Participant’s consent form/s</td>
<td>[ ]</td>
<td>[x]</td>
</tr>
<tr>
<td>• Advertisement</td>
<td>[ ]</td>
<td>[x]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section D: Details of Risks and Benefits to the Researcher and the Researched</th>
<th>Ticked if attached</th>
<th>Tick if not relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Insurance registration form and related correspondence</td>
<td>[ ]</td>
<td>[x]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Appendix I: Research Involving the Use of Drugs</th>
<th>Ticked if attached</th>
<th>Tick if not relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Relevant correspondence relating to agreed arrangements for dispensing with the pharmacy</td>
<td>[ ]</td>
<td>[x]</td>
</tr>
<tr>
<td>• Written confirmation from the manufacturer that the drug/substance has been manufactured to GMP</td>
<td>[ ]</td>
<td>[x]</td>
</tr>
<tr>
<td>• Proposed volunteer contract</td>
<td>[ ]</td>
<td>[x]</td>
</tr>
<tr>
<td>• Full declaration of financial or direct interest</td>
<td>[ ]</td>
<td>[x]</td>
</tr>
<tr>
<td>• Copies of certificates: CTA etc...</td>
<td>[ ]</td>
<td>[x]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Appendix II: Use of Non-Ionising Radiation</th>
<th>Ticked if attached</th>
<th>Tick if not relevant</th>
</tr>
</thead>
</table>
Please note that correspondence regarding the application will normally be sent to the Principal Researcher and copied to other named individuals.
27 August 2013

Dear Professor Dacre

**Notification of Ethical Approval**

**Project ID: 5025/001: Sex differences in the professional performance of doctors practising in the UK**

I am pleased to confirm that in my capacity as Chair of the UCL Research Ethics Committee I have approved your study for the duration of the project i.e. until September 2015.

Approval is subject to the following conditions:

1. You must seek Chair’s approval for proposed amendments to the research for which this approval has been given. Ethical approval is specific to this project and must not be treated as applicable to research of a similar nature. Each research project is reviewed separately and if there are significant changes to the research protocol you should seek confirmation of continued ethical approval by completing the ‘Amendment Approval Request Form’.

The form identified above can be accessed by logging on to the ethics website homepage: http://www.grad.ucl.ac.uk/ethics/ and clicking on the button marked ‘Key Responsibilities of the Researcher Following Approval’.

2. It is your responsibility to report to the Committee any unanticipated problems or adverse events involving risks to participants or others. Both non-serious and serious adverse events must be reported.

**Reporting Non-Serious Adverse Events**

For non-serious adverse events you will need to inform Helen Dougal, Ethics Committee Administrator (ethics@ucl.ac.uk), within ten days of an adverse incident occurring and provide a full written report that should include any amendments to the participant information sheet and study protocol. The Chair or Vice-Chair of the Ethics Committee will confirm that the incident is non-serious and report to the Committee at the next meeting. The final view of the Committee will be communicated to you.

**Reporting Serious Adverse Events**

The Ethics Committee should be notified of all serious adverse events via the Ethics Committee Administrator immediately the incident occurs. Where the adverse incident is unexpected and serious, the Chair or Vice-Chair will decide whether the study should be terminated pending the opinion of an independent expert. The adverse event will be considered at the next Committee meeting and a decision will be made on the need to change the information leaflet and/or study protocol.
On completion of the research you must submit a brief report (a maximum of two sides of A4) of your findings/concluding comments to the Committee, which includes in particular issues relating to the ethical implications of the research.

With best wishes for the research.

Yours sincerely

Professor John Foreman
Chair of the UCL Research Ethics Committee

Cc:
Dr Emily Unwin & Dr Katherine Woolf, Applicants
Professor David Lomas, Dean, Faculty of Medical Sciences
### Amendment Approval Request Form

<table>
<thead>
<tr>
<th></th>
<th>Project ID Number: 5025/001</th>
<th>Name and Address of Principal Investigator:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Professor Jane Doe</td>
</tr>
</tbody>
</table>

#### Project Title:
Sex differences in the professional performance of doctors practising in the UK.

#### Type of Amendment/s (tick as appropriate)
- [ ] Research procedure/protocol (including research instruments)
- [ ] Participant group
- [ ] Sponsorship/collaborators
- [x] Extension to approval needed (extensions are given for one year)
- [ ] Information Sheet/s
- [ ] Consent form/s
- [ ] Other recruitment documents
- [ ] Principal researcher/medical supervisor*
- [ ] Other *

*Additions to the research team other than the principal researcher, student supervisor and medical supervisor do not need to be submitted as amendments but a complete list should be available upon request.

#### Justification (give the reasons why the amendment/s are needed)
We request a 2 year extension to approval.
This research project is being conducted as part of a PhD project. The PhD candidate, Dr Emily Unwin, had a period of interruption of study for 10 months (March 2014-January 2015) for maternity leave. On returning to her PhD studies Dr Emily Unwin has changed her student status from full-time to part-time. The expected completion date of her PhD studies is September 2017 (taking into account the period of interruption of study and change of student status. We request an extension of two years (up to September 2017) to reflect these changes.

#### Details of Amendments (provide full details of each amendment requested, state where the changes have been made and attach all amended and new documentation)

#### Ethical Considerations (insert details of any ethical issues raised by the proposed amendment/s)
This period of extension should have no impact on any ethical considerations.

#### Other Information (provide any other information which you believe should be taken into account during ethical review of the proposed changes)

#### Declaration (to be signed by the Principal Researcher)
- I confirm that the information in this form is accurate to the best of my knowledge and I take full responsibility for it.
- I consider that it would be reasonable for the proposed amendments to be implemented.
- For student projects I confirm that my supervisor has approved my proposed modifications.

Signature:
Date: 03/06/2015

**FOR OFFICE USE ONLY:**

Amendments to the proposed protocol have been **approved** by the Research Ethics Committee.

Signature of the REC Chair, Professor John Foreman: ...

Date: 7/7/2015.
Amendment Approval Request Form

<table>
<thead>
<tr>
<th></th>
<th>Project ID Number: 5025/001</th>
<th>Name and Address of Principal Investigator:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Professor Jane Dacre</td>
</tr>
</tbody>
</table>

2 Project Title: Sex differences in the professional performance of doctors practising in the UK.

3 Type of Amendment/s (tick as appropriate)

- Research procedure/protocol (including research instruments)
- Participant group
- Sponsorship/collaborators
- Extension to approval needed (extensions are given for one year)
- Information Sheet/s
- Consent form/s
- Other recruitment documents
- Principal researcher/medical supervisor
- Other *

*Additions to the research team other than the principal researcher, student supervisor and medical supervisor do not need to be submitted as amendments but a complete list should be available upon request.

Justification (give the reasons why the amendment/s are needed)

Amendment No. 1:
In our original ethical approval application we stated that we planned to examine the recruitment data for Foundation and Core Training in 2012. This data was to be provided to us by the General Medical Council (GMC). In November 2015 the GMC released details of their UK Medical Education Database (UKMED) and opened the application process for researchers to use this data. UKMED are data that have been collated from pre-existing databases. The aim of this database is to support research in medical education. Upon release of the details of UKMED in November 2015, we were made aware that UKMED contained details of previous academic attainment, had data on Specialty Training (also known as Core Training) from 2012-2015, and data on Fitness to Practise Declarations.
We would therefore like to request the following amendments:

1) To extend our initial research question for this project to include previous academic attainment measures, when examining the association between gender and recruitment in Specialty Training Programmes/Core Training. We already know that a gender disparity exists with regards to specialty choice. Being able to include previous academic attainment measures into our research, will enable us to examine whether the gender difference in specialty choice is mediated by previous academic achievement.

2) To extend the period in which the data was collected from one year (2012), to three years (2012-2015). This will increase the number of study participants, and thereby increase the power of our study.

Amendment No. 2:
In our original ethical approval application we stated were going to examine Annual Review of Competence Progression data. This data was to be provided to us by the GMC. We have decided that we will no longer proceed with this part of the research project.

Details of Amendments (provide full details of each amendment requested, state where the changes
have been made and attach all amended and new documentation)
Amendment No. 1:
This research project aims to understand the gender difference in specialty choice, and examine whether
previous academic attainment may influence the association between gender and specialty. The measures
of academic attainment that we would like to examine are GCSE performance, UKCAT performance (for
entry into medical school), Educational Performance Measure from medical school, and performance in
Situational Judgement Tests for entry into the Foundation Programme. This data, along with demographic
data (gender, ethnicity) and data on Specialty Training Programmes applications/offers/acceptances will be
provided by the GMC through UKMED. We also aim to explore the gender difference in doctors with a
history of Fitness to Practise Declarations, and whether academic attainment and specialty choice differs
between those doctors with a history of Fitness to Practise Declarations, and those who do not.
The data provided to us by the UKMED team will be non-identifiable.

| 6 | Ethical Considerations (insert details of any ethical issues raised by the proposed amendment/s)
| We believe that these amendments should have no impact on any ethical considerations. |

| 7 | Other Information (provide any other information which you believe should be taken into account
during ethical review of the proposed changes)
| This research project is dependent on the success of our research project application to UKMED and being
given access to the data. |

| Declaration (to be signed by the Principal Researcher) |
| - I confirm that the information in this form is accurate to the best of my knowledge and I take full
  responsibility for it. |
| - I consider that it would be reasonable for the proposed amendments to be implemented. |
| - For student projects I confirm that my supervisor has approved my proposed modifications. |

| Signature: |
| Date: 05/01/2016 |

| FOR OFFICE USE ONLY: |
| Amendments to the proposed protocol have been by the Research Ethics Committee. |
| Signature of the REC Chair, Professor John Foreman: |
| Date: 12/1/2016 |
## Amendment Approval Request Form

| 1 | Project ID Number: 5025/001 | Name and Address of Principal Investigator: | Professor Jane Dacre  
Dept. of Public Health |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Project Title: Sex differences in the professional performance of doctors practising in the UK.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Type of Amendment(s) (tick as appropriate)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Research procedure/protocol (including research instruments)</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Participant group</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sponsorship/collaborators</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Extension to approval needed (extensions are given for one year)</td>
<td>Yes</td>
<td>To 30/09/2018</td>
</tr>
<tr>
<td></td>
<td>Information Sheet(s)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Consent form(s)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other recruitment documents</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Principal researcher/medical supervisor*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Additions to the research team other than the principal researcher, student supervisor and medical supervisor do not need to be submitted as amendments but a complete list should be available upon request.

<table>
<thead>
<tr>
<th>4</th>
<th>Justification (give the reasons why the amendment(s) are needed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>We request a one-year extension to approval (up to September 2018). The project is being conducted as part of a PhD project. The PhD candidate, Dr Emily Unwin, is having a period of interruption of study for 15 months (February 2017 – February 2018) for maternity leave. The expected completion date of her PhD studies is September 2018.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5</th>
<th>Details of Amendments (provide full details of each amendment requested, state where the changes have been made and attach all amended and new documentation)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>6</th>
<th>Ethical Considerations (insert details of any ethical issues raised by the proposed amendment(s))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>This period of extension should have no impact on any ethical considerations.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>7</th>
<th>Other Information (provide any other information which you believe should be taken into account during ethical review of the proposed changes)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>8</th>
<th>Declaration (to be signed by the Principal Researcher)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I confirm that the information in this form is accurate to the best of my knowledge and I take full responsibility for it.</td>
</tr>
<tr>
<td></td>
<td>I consider that it would be reasonable for the proposed amendments to be implemented.</td>
</tr>
<tr>
<td></td>
<td>For student projects, I confirm that my supervisor has approved my proposed modifications.</td>
</tr>
</tbody>
</table>

Signature: [Blank]

Date: 07/09/2017
## Amendment Approval Request Form

<table>
<thead>
<tr>
<th></th>
<th>Project ID Number:</th>
<th>Name and Address of Principal Investigator:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5025/001</td>
<td>Professor Diana Jane Dacre</td>
</tr>
</tbody>
</table>

| 2 | Project Title: Sex differences in the professional performance of doctors practicing in the UK. |

<table>
<thead>
<tr>
<th>3</th>
<th>Type of Amendment(s) (tick as appropriate)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Research procedure/protocol (including research instruments)</td>
</tr>
<tr>
<td></td>
<td>Participant group</td>
</tr>
<tr>
<td></td>
<td>Sponsorship/collaborators</td>
</tr>
<tr>
<td></td>
<td><strong>Extension to approval needed (extensions are given for one year)</strong></td>
</tr>
<tr>
<td></td>
<td>Information Sheet/s</td>
</tr>
<tr>
<td></td>
<td>Consent form/s</td>
</tr>
<tr>
<td></td>
<td>Other recruitment documents</td>
</tr>
<tr>
<td></td>
<td>Principal researcher/medical supervisor*</td>
</tr>
<tr>
<td></td>
<td><strong>Other</strong></td>
</tr>
</tbody>
</table>

*Additions to the research team other than the principal researcher, student supervisor and medical supervisor do not need to be submitted as amendments but a complete list should be available upon request.*

<table>
<thead>
<tr>
<th>4</th>
<th>Justification (give the reasons why the amendment(s) are needed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Amendment 1:</strong> We request a 1 year extension to approval.</td>
</tr>
<tr>
<td></td>
<td>The research project is conducted as part of a PhD project. The PhD is in its final stages, but the final research study to be included in the PhD is not yet complete. A period of extension will enable the PhD candidate, Dr Emily Unwin, to complete this study. Dr Emily Unwin has been awarded a Continuing Research Status until September 2020.</td>
</tr>
<tr>
<td></td>
<td><strong>Amendment 2:</strong> We would like to provide an update of the datasets we have been able to access and use for the PhD research. The original ethics application form dated August 2013 stated that some of the datasets would be received in a pseudonymised format, but when received they contained identifiable data.</td>
</tr>
<tr>
<td></td>
<td><strong>Amendment 3:</strong> In the original ethics application form dated August 2013 we did not explicitly state that we would be merging and linking a selection of the datasets.</td>
</tr>
<tr>
<td></td>
<td><strong>Amendment 4:</strong> Supplementary data was requested from the General Medical Council (GMC) to update medical registration history of a cohort of UK doctors followed from 1991-2009. The data received from the GMC was then merged with the cohort dataset.</td>
</tr>
<tr>
<td></td>
<td><strong>Amendment 5:</strong> We originally requested ethical approval to examine the recruitment data of Foundation and Core Training. We have decided that we will no longer proceed with this part of the research project.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5</th>
<th>Details of Amendments (provide full details of each amendment requested, state where the changes have been made and attach all amended and new documentation)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Amendment 2</strong> (identifiable datasets that were recorded at pseudonymised in 2013):</td>
</tr>
<tr>
<td></td>
<td>Dataset 1: Dataset of the List of Registered Medical Practitioners on 29 May 2013. This dataset was provided by the GMC and contains identifiable data.</td>
</tr>
</tbody>
</table>
Dataset 2i
Dataset of all doctors under investigation by the GMC, who were required to complete Tests of Competence (ToC) from 2008 to 2013. Access to this dataset was given by the GMC and the RDME, UCL. The dataset contains identifiable data.

Dataset 2ii
The knowledge test scores of GPs under investigation by the GMC who had completed a Test of Competence were extracted from the individual GP reports to create this dataset. Identifiable data was used to create this dataset, but once the dataset was created, the identifiable variables were removed. Access to the GP reports was given by RDME, UCL.

Dataset 2iii
The simulated surgery scores of GPs under investigation by the GMC who had completed a Test of Competence were extracted from the individual GP reports to create this dataset. Identifiable data was used to create this dataset, but once the dataset was created, the identifiable variables were removed. Access to the GP reports was given by RDME, UCL.

Dataset 2iv
Dataset of the volunteer doctors who completed the Tests of Competence. This identifiable dataset contained the doctors’ ID and demographics. This dataset was provided by RDME, UCL.

Dataset 2v
The knowledge test scores of the volunteer GPs were extracted from the GP pilot events database held by RDME, UCL. This dataset contained identifiable data.

Dataset 2vi
The simulated surgery scores of the volunteer GPs held by RDME, UCL. This dataset is anonymised.

Dataset 2vii
The updated outcome of the investigations of GPs whose cases were not complete in Dataset 2i. This data was provided by the GMC and contains identifiable data.

Dataset 3
Dataset of the Membership of the Royal College of Physicians (MRCP) PACES data from 2010 to 2013. This dataset was provided by the MRCP and contained identifiable data. The identifiable variables of this dataset have since been removed.

Dataset 4i
Survey data from doctors who applied to medical school in 1990 and completed a questionnaire in 2009. This dataset is anonymised. (This dataset has since been deleted).

Dataset 4ii
Survey data from doctors who applied medical school in 1990 and were followed up until 2009. This dataset was anonymised. (This dataset has since been deleted).

Dataset 4iii
A copy of Dataset 4ii, but containing identifiable data. The identifiable data was since been removed.

Dataset 4iv
Registration status and active / historical sanctions against medical registration of the GMC registered doctors in Dataset 4iii. This dataset was provided by the GMC and contained identifiable data. (This dataset was deleted, once it had been merged with Dataset 4iii).

Dataset 5
Doctors who were under investigation by the GMC and were required to complete a Test of Competence (1996-2008). This dataset contained identifiable data. Access to this data was given by RDME, UCL. This dataset has since been deleted.

Amendment 3 (merging and linking of datasets):
The datasets 2i through to 2vii were merged to create a working dataset for the research project.
Doctors who were registered as General Practitioners (GPs) within the dataset of all doctors under investigation by the GMC were selected for further investigation. The Tests of Competence scores (knowledge test and simulated surgery) of these GPs were added to this dataset, along with updated case outcomes. The Test of Competence scores (knowledge test and simulated surgery) of volunteer GPs were also added to this dataset to provide a comparison.

Amendment 4 (Merging of supplementary data):
The datasets 4ii and 4iv were merged to create a working dataset for the research project.

The outcome of interest is active / historical sanctions on doctors’ medical registration. The survey data from doctors who applied to medical school in 1990 had incomplete data on the history of sanctions or warnings of the doctors in the cohort. We requested and received all historical sanctions and active warnings of the cohort doctors from the GMC (up to August 2015). This data was identifiable and was merged with the survey data. The identifiable variables of this merged dataset have since been deleted. Please note that historical and active sanctions against a doctor’s medical registration are publicly available information (via the GMC List of Registered Medical Practitioners). Active warnings against a doctor’s medical registration are also publicly available information, however there is no public record of expired warnings against a doctor’s medical registration.

<table>
<thead>
<tr>
<th>Ethical Considerations (insert details of any ethical issues raised by the proposed amendment/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>We believe that these amendments should have no impact on any ethical considerations.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other Information (provide any other information which you believe should be taken into account during ethical review of the proposed changes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>We have informed the UCL Data Protection Team about all the datasets used in this research project. The datasets used throughout this research project have been stored securely. To our knowledge, there has been no data breach.</td>
</tr>
</tbody>
</table>

It may be useful to note that Dr Emily Unwin is completing her PhD on a part-time basis.

<table>
<thead>
<tr>
<th>Declaration (to be signed by the Principal Researcher)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I confirm that the information in this form is accurate to the best of my knowledge and I take full responsibility for it.</td>
</tr>
<tr>
<td>I consider that it would be reasonable for the proposed amendments to be implemented. For student projects, I confirm that my supervisor has approved my proposed modifications.</td>
</tr>
</tbody>
</table>

Signature:  

Date:  2nd November 2018
Email from the UCL Research Ethics Committee received on 20/12/2018.

Dear Emily,

The REC Chair has approved your amendment request and the ethical approval of your study has been extended to **30/09/2020**.

**IMPORTANT: For projects collecting personal data only**
Change to legal basis for the processing of data: If you are processing (i.e. collecting, storing, using, disclosing or destroying) identifiable personal information about living individuals as part of your research then you should ensure that you comply with the requirements of the GDPR and the Common Law Duty of Confidentiality. An appropriate legal basis for the processing of your data must be identified, and you must be explicit about this and document it as part of your ethics application, and in the information you provide to your research participants. UCL’s view is that, for the vast majority of research undertaken at UCL, the appropriate legal basis will be ‘a task in the Public interest’: the processing is necessary for UCL to perform a task in the public interest - rather than ‘consent’.

However, even though the legal basis for the processing of a person’s data is most likely to be ‘a task in the public interest’ rather than ‘consent’, from an ethical perspective, obtaining a person’s informed consent for their involvement in the research is still likely to be required in order to abide by the fairness and transparency elements of principle GDPR Article 5(1)(a) or to meet confidentiality obligations.

We have recently changed the data privacy section of our template participant information sheet (PIS) to reflect this change to the legal basis for data processing – see attached. You will need to update your PIS accordingly.

With best wishes for the research,

Ed

Edward Whitfield MA, MLitt
UCL Research Ethics Administrator
### Application for inclusion of a research project Form 2

#### A. APPLICATION DETAILS

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A1</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Project Title:</strong></td>
<td>Sex differences in the professional performance of doctors practising in the UK.</td>
</tr>
<tr>
<td>Date of Submission:</td>
<td>01/08/2013</td>
</tr>
<tr>
<td>Proposed Start Date:</td>
<td>August 2013</td>
</tr>
<tr>
<td>UCL Ethics Project ID Number:</td>
<td></td>
</tr>
<tr>
<td>Proposed End Date:</td>
<td>September 2015</td>
</tr>
</tbody>
</table>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A2</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Principal Researcher</strong> (Please note that a student – undergraduate, postgraduate or research postgraduate cannot be the Principal Researcher for Ethics purposes):</td>
<td></td>
</tr>
<tr>
<td>Full Name:</td>
<td>Professor Jane Dacre</td>
</tr>
<tr>
<td>Position Held:</td>
<td>Director</td>
</tr>
<tr>
<td>Address:</td>
<td></td>
</tr>
</tbody>
</table>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A3</strong></td>
<td>Data Collector(s) Details (if Applicant is not the Principal Researcher e.g. student details):</td>
</tr>
<tr>
<td>Full Name:</td>
<td>Dr. Emily Unwin</td>
</tr>
<tr>
<td>Position Held:</td>
<td>PhD student</td>
</tr>
<tr>
<td>Address:</td>
<td></td>
</tr>
</tbody>
</table>

#### B. DETAILS OF THE PROJECT

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>B1</strong></td>
<td>Please provide a brief summary of the project</td>
</tr>
<tr>
<td>The research aims to explore the performance of doctors practising in the UK and to determine which factors influence differences in male and female doctors’ performance outcomes.</td>
<td></td>
</tr>
<tr>
<td>The objectives of the research are to:</td>
<td></td>
</tr>
<tr>
<td>1. Examine whether male doctors practising in the UK perform more poorly than female doctors on a variety of measures of knowledge, skills and attributes and at different stages in doctors’ careers.</td>
<td></td>
</tr>
<tr>
<td>2. Identify factors that differ between male and female doctors and which may mediate or moderate the relationship between sex and performance outcomes and to identify whether these factors are remediable.</td>
<td></td>
</tr>
<tr>
<td>These aims and objectives will be achieved by analysing national data on various measures of doctors’ academic and professional performance.</td>
<td></td>
</tr>
<tr>
<td>i) Doctors’ performance in the Fitness to Practice Tests of Competence (identifiable data)</td>
<td></td>
</tr>
</tbody>
</table>
C. DETAILS OF PARTICIPANTS

Data subjects
Who will the personal data be collected from?

Doctors’ performance in the Fitness to Practice Tests of Competence:
Doctors who have had their fitness to practise called into question and have been requested by the General Medical Council (the medical regulatory body) to sit a Test of Competence as part of their performance assessment.

Performance of “doctors in good standing” who pilot the Tests of Competence materials:
Doctors in good standing who have applied voluntarily to sit a Test of Competence to pilot the Tests of Competence materials.

What data will be collected
Please provide details of the type of personal data to be collected

Doctors’ performance in the Fitness to Practice Tests of Competence:
Name
GMC Number*
Specialty
Grade
Score for each question

Performance of “doctors in good standing” who pilot the Tests of Competence materials:
Name
GMC Number*
Date of birth
Sex
Ethnicity
Contact details (address, telephone number, email address)
Specialty
Grade
Score for each question

*Please note that the GMC Number will enable the merging of information (including name, sex, registration status, date of primary medical qualification, place of qualification, date if joined the Specialist or General Practice Register) from the publicly available List of Registered Medical Practitioners.
Disclosure
Who will the results of your project be disclosed to?

PhD supervisors – Professor Jane Dacre and Dr. Katherine Woolf.
Those who supplied the data - General Medical Council, Royal College of Physicians,
Professor Chris McManus.
Examiners of the PhD upgrade and the PhD.

Highlights of the research outcomes will be published in UCL research reports, on line and in
newsletters. The research undertaken as part of the PhD thesis will be presented at academic and
medical conferences and in clinical and medical peer-reviewed journals. The research will be
presented as part of a PhD thesis.

D. CONSENT

Consent
Please include the information sheet and consent forms you will be using for this project, and or
protocol

If you are not including an information sheet and consent form, please explain why:

Doctors’ performance in the Fitness to Practice Tests of Competence:
The General Medical Council do not inform doctors undergoing the Tests of Competence of the usage
of their data collected, nor how it will be processed or stored. Neither are the doctors asked to give
consent for its usage thereafter.

We feel that it is in the public interest to perform research using this data.

Performance of “doctors in good standing” who pilot the Tests of Competence materials:
Please see attached form.

E. INTERNATIONAL TRANSFER

International Transfer
The eighth principle of the Data Protection Act 1998 prohibits the transfer of personal data to countries
or territories outside the European Economic Area (which consists of the 27 EU member states,
Iceland, Liechtenstein and Norway).
At the time of writing the following countries have also been deemed adequate for the purposes of the 8th principle Argentina, Canada, Guernsey, Isle of Man, Jersey and Switzerland.

If you intend to transfer data to a country not mentioned above, please supply details of adequate safeguards below:

N/A

F. PUBLICATION

Will the results of your research be published in an academic journal or other publication? YES / NO

Please note that published results must not contain data by which an individual can be identified.

G. NOTIFICATION

G1 Notification
(Please note that notification is a prerequisite for registration)
Have you informed your department's Data Protection Coordinator about your project? YES/NO

G2 Notification
(Please note that notification is a prerequisite for registration)
Have you informed your department's computer representative about your project? YES/NO

H. ETHICS

H1 Are you applying to the UCL Research Ethics Committee? YES/NO
Date of Ethics meeting:

Finance and Business Affairs
Legal Services
6th Floor, 1-19 Torrington Place
London WC1E 7HB

January 2012
Further information

For more information and guidance on the UCL Research Committee, please visit
http://ethics.grad.ucl.ac.uk/

When all essential documents are ready to archive, contact the UCL Records Office by email at
records.office@ucl.ac.uk to arrange ongoing secure storage of your research records unless you have
made specific alternative arrangements with your department, or funder.

For information on the UCL Records Management Service, please visit
http://www.ucl.ac.uk/eld/recordsoffice/policy/records-transfer
Application for inclusion of a research project

All research projects using personal data must be registered with the UCL Data Protection Registration Service before the data is collected. This includes projects approved by the Joint Research Office (a partnership between University College London, UCL Hospitals NHS Foundation Trust and the Royal Free Hampstead NHS Trust).

UCL is required by law to comply with the data protection legislation. UCL is the Data Controller under the Act, and the Council, as the governing body, is ultimately responsible for implementation. However, students who are intending to process personal data for research purposes where they are the Data Controller, are responsible for any processing under the Act, and UCL is not responsible for any processing of personal data where it is not the Data Controller.

This form should only be completed if identifiable data is being collected and used as part of research because identifiable data a personal data and data protection law applies. The Act does not apply to data rendered anonymous (also known as de-identified), so that the data subjects are not identifiable. In this instances, registration will not be required.

As part of registering your research, we also need to be notified of changes which affect data protection compliance. You can find out more about the changes you need to tell us about by visiting the research and data protection website.

All sections must be completed before submitting this form to the data protection team.

<table>
<thead>
<tr>
<th>A. APPLICATION DETAILS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A1.</strong> Project title:</td>
</tr>
<tr>
<td>a. Proposed start date: August 2013</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. CHIEF INVESTIGATOR (C1); PRINCIPAL INVESTIGATOR (PI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>B1.</strong> (Undergraduate, postgraduate or research postgraduate cannot be a C1/P1 for Ethics purposes).</td>
</tr>
<tr>
<td>a. Full Name:</td>
</tr>
<tr>
<td>b. Position held:</td>
</tr>
<tr>
<td>c. School:</td>
</tr>
<tr>
<td>d. Faculty:</td>
</tr>
<tr>
<td>e. Department:</td>
</tr>
<tr>
<td>f. Email:</td>
</tr>
</tbody>
</table>

(please note that if the person is not a UCL employee you should provide details below of a responsible UCL employee below)

| g. Full Name: | Dr. Katherine Woolf |
| h. Position held: | Senior Lecturer |
| i. School: | |
| j. Faculty: | |
| k. Department: | |
| l. Email: | |

<table>
<thead>
<tr>
<th>C. DATA COLLECTOR (S)</th>
</tr>
</thead>
</table>

May 2018
### C1. Data Collector(s) Details (If applicant is not the PhD student leader):

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>Full Name: Dr Emily Unwin</td>
</tr>
<tr>
<td>b.</td>
<td>Position held: PhD candidate</td>
</tr>
<tr>
<td>c.</td>
<td>School:</td>
</tr>
<tr>
<td>d.</td>
<td>Faculty:</td>
</tr>
<tr>
<td>e.</td>
<td>Department:</td>
</tr>
<tr>
<td>f.</td>
<td>Email:</td>
</tr>
</tbody>
</table>
Please provide a brief summary of the project, including an explanation of the aims, design, methodology and plans for analysis that you propose to use.

Male doctors are at increased risk of disciplinary action against their medical registration compared to their female counterparts. The extent of this gender gap and the reasons behind it are unclear. The aim of this research is to better understand poor professional performance in doctors through addressing the research question: ‘are male doctors more likely to have poor professional performance, and if so, why?” The research question is explored through a positivist approach of six research studies (4 original research studies and 2 systematic reviews of the literature and meta-analyses) using quantitative epidemiological methods to analyse large secondary datasets.

The datasets that have been/will be used to answer the research question are described below:

**Datasets used for Emily Unwin’s PhD research**

**Study 1: Sex differences in sanctions received by doctors registered with the GMC on 25 May 2013.**

**Dataset 1 – The List of Registered Medical Practitioners (LRRP)**

**Provider:** General Medical Council (GMC)

**Identifiable dataset**

This dataset was registered with the UCL Data Protection Team in 2013. UCL Data Protection Registration Number ZE344103/2013/08/11. At the time of registration we erroneously reported that the data was in encrypted form.

Emily Unwin anonymised the dataset for research purposes.

**Please note:** The LRRP is publicly available. However, it is updated on a daily basis (as doctors join and leave the medical register). Sanctions (erasure, suspension, conditions and undertakings) remain on a doctor’s record indefinitely. Warnings are removed from a doctor’s record once they have expired.

**Study 2: Sex differences in the Tests of Competence scores of GPs under investigation by the GMC who are required to take a TCC, and GPs not under investigation who volunteer to take a TCC**

The datasets for Study 2 are complex, and the context in which the data are generated requires explanation:

RDME (UCL Medical School’s Research Department for Medical Education) has held a contract with the GMC (General Medical Council, the UK medical regulator) for several years to develop and administer the Tests of Competence (TCC). This contract is referred to in this document as the ‘RDME GMC Project’.

Doctors being investigated by the GMC may be required to take a TCC as part of the investigation, although not all doctors who are required to take a TCC actually end up taking one. This study focused on those doctors who were working as General Practitioners (GPs) at the time that a concern about their fitness to practice was raised, triggering an investigation by the GMC. There are different TCCs for different specialties, and the GP TCC consists of a knowledge test (KT) and a simulated surgery (SS), as well as an Objective Structured Clinical Examination (OSCE).

In addition, as part of the TCC validation process, TCCs are taken by doctors who are NOT currently under investigation by the GMC. These doctors volunteer to take the tests in exchange for a fee, and as such are called ‘volunteer doctors’. A proportion of those volunteer doctors take the GP TCCs (other volunteer doctors take TCCs designed for different medical specialties e.g. Surgery, Psychiatry, etc.). Volunteer doctors take the tests at ‘pilot events’—these events are usually held over a day at UCL or the GMC. Pilots are usually specialty-specific e.g. at a GP pilot event, several volunteer doctors will take a TCC designed for General Practitioners.
RDME therefore holds data on: the GMC numbers of all doctors required to take a ToC, the GMC numbers of all doctors who actually take a ToC (or who some are GPs), the ToC scores of all doctors under investigation who take a ToC (of who some are GPs). RDME also holds some data on the outcome of the investigation for doctors who took a ToC (e.g. was the doctor struck off or not); the GMC numbers of volunteer doctors who take a ToC (of who some are GPs); the ToC scores of volunteer doctors (of who some are GPs).

The GMC holds information on all doctors who have been on the UK Medical Register, including the case outcomes for doctors they have investigated.

The data for Study 2 came from the RDME GMC Project (RDME being Emily’s home department) and from the GMC.

**Dataset 2a** — All doctors under investigation by the GMC, who were required to complete Tests of Competence (ToCs) from 2008 to 2013.

Provider: General Medical Council (GMC)

Identifiable data

This dataset was registered with the UCL Data Protection Team in 2013, UCL Data Protection Registration Number ZG064106/2013/03/11 with the expectation that the data would be pseudonymised. However, identifiable data was provided.

**Please note:** This dataset is available to all Research Department of Medical Education (RDME) GMC researchers at UCL, as agreed by the GMC, and has been previously used in RDME research.

**Dataset 2b** — The ToC knowledge test (KT) scores of the General Practitioners (GPs) investigated by the GMC from 2008 to 2013.

Provider: ToC scores provided by RDME and extracted by Emily Unwin from the individual ToC reports.

Anonymised data (identifiable data was used to create this dataset, but once the dataset was created all identifiable variables were removed).

This dataset was created by Emily Unwin extracting the KT scores from the individual ToC reports.

**Please note:** In 2013 the UCL Data Protection team were made aware of that I would be using the ToC scores of doctors, but I did not explicitly state that I would be extracting the scores from individual reports and creating datasets myself.

The knowledge test scores of GPs being investigated by the GMC has been previously used in RDME research.

**Dataset 2c** — All doctors under investigation by the GMC (2a) plus GPs KT scores (Datasets 2a and 2b merged) identifiable data

**Dataset 2d** — ToC simulated surgery scores (SS) for the GPs investigated by the GMC from 2008 to 2013.

Provider: ToC scores provided by RDME and extracted by Emily Unwin from the individual ToC reports.

Anonymised data (identifiable data was used to create this dataset, but once the dataset was created all identifiable variables were removed).

This dataset was created by Emily Unwin extracting the SS scores from the individual ToC reports.

**Please note:** In 2013 the UCL Data Protection team were made aware of that I would be using the ToC scores of doctors, but I did not explicitly state that I would be extracting the scores from individual reports and creating datasets myself.

The simulated surgery scores of GPs being investigated by the GMC has been previously used in RDME research.

**Dataset 2e** — All doctors under investigation by the GMC plus GPs KT (2c) and SS scores (Datasets 2c and 2d)
Dataset 2f - Updated case outcome data for a selection of investigated GPs, whose cases were not completed at the time 2a was accessed.
Provider: GMC
Identifiable data

Dataset 2a did not have complete case outcomes for all the doctors at the time it was accessed by Emily Unwin. Emily Unwin extracted the GMC numbers of GPs under investigation whose case outcomes were incomplete and sent them via STFS to the GMC. The GMC matched them to the final case outcomes and sent the matched files to Emily Unwin via STFS in November 2015.

Dataset 2g - GPs investigated by the GMC plus KT and SS scores (2e with non-GPs deleted) plus updated outcomes (Datasets 2e and 2f merged)
Identifiable data
A copy of this dataset has been stored on the RDME GMC Research folder for other RDME GMC researchers to use, after permission was obtained from the GMC. Both the original 2a dataset and the individual GMC reports were available to the RDME GMC researchers prior to sharing the 2g dataset (i.e. the RDME researchers had the information available to create the same 2g dataset, but using my version would save them time).

Dataset 2h - Volunteer doctor GMC numbers and demographic data.
Provider: RDME, UCL
Identifiable data
This dataset was registered with the UCL Data Protection Team in 2013, UCL Data Protection Registration Number Z5364106/2013/08/11
Please note: Part of this data has previously been used in RDME research.

Dataset 2i - Volunteer doctor Tests of Competence knowledge test (KT) scores
Provider: RDME, UCL
Identifiable data
To create database 2i Emily Unwin extracted and imported the volunteer KT scores from the individual databases for each of the General Practice pilot events between 2008 to 2013 held by RDME. This dataset was registered with the UCL Data Protection Team in 2013, UCL Data Protection Registration Number Z5364106/2013/08/11
Please note: Part of this data has previously been used in RDME research.

Dataset 2j - GPs investigated by the GMC, including KT and SS scores plus volunteer GPs KT scores (Datasets 2g and 2i merged)
Identifiable data

Dataset 2k - Volunteer doctor Tests of Competence simulated surgery (SS) scores
Provider: Research Department of Medical Education, UCL
Identifiable data
This dataset was registered with the UCL Data Protection Team in 2013, UCL Data Protection Registration Number Z5364106/2013/08/11.

Dataset 2l - GPs investigated by the GMC, including KT and SS scores plus volunteer GPs KT and volunteer GPs SS scores (Datasets 2j and 2k merged)
Contains identifiable data

May 2016
Study 3: Sex differences in PACES scores

**Dataset 3** – Membership of the Royal College of Physicians PACES data. PACES is the clinical part of the Membership of the Royal College of Physicians (UK) Examination.
Provider: Royal College of Physicians

**Identifiable data**

**Please note:** This dataset was originally registered with the UCL Data Protection Team in 2013, UCL Data Protection Registration Number ZE364106/2013/08/11. At the time of registration it was expected that the data would be received in anonymised form. However, identifiable data enabled the creation of a binary category within the data (registered with the GMC vs. not registered with the GMC).

The provider of the data, Royal College of Physicians do provide a MRCP(UK) data protection statement on their website that has a provision for sharing data for research.

The identifiable variables from this dataset were deleted in August 2018.

Study 4: Factors predicting sex differences in GMC sanctions among a sample of doctors who entered a UK medical school in 1991

**Dataset 4a** – Survey data from doctors who entered medical school in 1991 and responded to a 2009 questionnaire
Provider: Professor Chris McManus, UCL

**Anonymised data**

**Dataset 4b** – Survey data from doctors who entered medical school in 1991
Provider: Professor Chris McManus, UCL

**Anonymised data**

This data was originally registered with the UCL Data Protection Team in 2013, UCL Data Protection Registration Number ZE364106/2013/08/11. At the time of registration it was expected that the data would be received in anonymised form. The first copy of this dataset received was anonymised, however, identifiable data was required.

**Please note:** This dataset was deleted in August 2018.

**Dataset 4c** – Survey data from doctors who entered medical school in 1991 including GMC numbers

Identifiable data
Provider: Professor Chris McManus

**Please note:** This dataset was deleted in August 2018 (though non-identifiable data from this dataset are in dataset 4f).

**Dataset 4d** – Registration status of GMC registered cohort members as of August 2015

**Dataset 4e** – Historical and active sanctions and active warnings of the GMC registered cohort members as of August 2015

Identifiable data
Provider: GMC

**Please note:** These datasets were deleted in August 2018 (though non-identifiable data from these datasets are in dataset 4f).

**Dataset 4f** – Survey data plus registration status and sanctions/warnings (Datasets 4c, 4d and 4e merged)

**Please note:** Identifiable variables were deleted from this dataset in August 2018.

**Dataset used as part of MPhil upgrade process**

**Dataset 5** – Data of doctors who were required to complete an assessment as part of an investigation into

---

May 2018
their fitness to practice (1996-2008). Provider: RDME, UCL

Identifiability data

Please note: The copy of this data set was deleted in August 2018. The master copy of this dataset was available to RDME GMC researchers.

**Datasets not used by Emily Unwin**

**Dataset 6 – Annual Review of Competence Progression data**
Provider: GMC

OpenAIRE data

Please note: This dataset was registered with the UCL Data Protection Team in 2013, UCL Data Protection Registration Number Z6364106/2013/08/11

**Dataset 7 – Recruitment data of trainee doctors**
Provider: General Medical Council

OpenAIRE data

Please note: This dataset was registered with the UCL Data Protection Team in 2013, UCL Data Protection Registration Number Z6364106/2013/08/11

**Dataset 8 – UCMS cohort data**
Provider: Dr Katherine Wood, UCL

OpenAIRE data

---

**E. PRIVACY IMPACT SCREENING QUESTIONS**

<table>
<thead>
<tr>
<th>If the answer to any of these questions is ‘yes’, then a PIA is required</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Will the project require individuals to provide information about themselves?</strong></td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td><strong>Will information about individuals be shared with organisations or people who have not previously had routine access to the information?</strong></td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td><strong>Will the project use information about individuals for a purpose it is not currently used for, or in a way it is not currently used?</strong></td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td><strong>Does the project involve you using new technology that might be perceived as being privacy intrusive? For example, the use of biometrics or facial recognition.</strong></td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td><strong>Will the project result in you making decisions or treating individuals in ways which can have a significant impact on them?</strong></td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td><strong>Is the information about individuals likely to raise privacy concerns or expectations, e.g. health records or information that people would consider to be particularly private?</strong></td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td><strong>Will the project result in you collecting and using data in ways that people might find intrusive, e.g. unexpected telephone calls?</strong></td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td><strong>Will the project require contact with individuals in ways they may find intrusive, e.g. unexpected telephone calls?</strong></td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td><strong>Will the project involve processing sensitive personal data?</strong></td>
<td>✔</td>
<td></td>
</tr>
</tbody>
</table>

**F. DETAILS OF PARTICIPANTS**

Please provide details of the potential participants for this project, including how they will be selected and recruited.

This research project does not involve the collection of new data. All data used in this research project has been obtained from secondary data sources [e.g. data that has been collected for administrative purposes or as part of an earlier non-related research project].

**Study 1**

This study population is all doctors on the GMC’s LMP on 29 May 2013. This included doctors who are or
have been registered to practice medicine in the UK since October 2005.

Study 2
All GPs between 2008 and 2013 under investigation by the GMC and required to complete the TQC between 2008 and 2014. The comparison group was GPs not currently under investigation and with no restrictions on their registration, who voluntarily completed either the KT and/or SS assessment as part of the TQC pilots.

Study 3
The study population is all doctors who were completing the clinical component (PACES) of the MRCP (UK) diplomas for the first time between October 2010 and May 2013.

Study 4
The study population is doctors who are or have been registered with the GMC, who applied to at least one of five UK medical schools for admission in 1991 and entered the research cohort through the completion of at least the first questionnaire.

G. DETAILS OF THE DATA BEING PROCESSED
Please describe the details of the personal data that is being collected, including the methods of data collection and analysis.

Study 1
Dataset 1 - The List of Registered Medical Practitioners (LRMP)
Provider: General Medical Council
Identifiable data: GMC number, Surname, Given name, Other names
Personal data: Gender, Qualification, Year of qualification, Place of qualification, Registration status, Sanctions or warnings against registration, Place of qualification

Study 2
Dataset 2a – All doctors under investigation by the GMC, who were required to complete Tests of Competence (TQC) from 2008 to 2013.
Provider: General Medical Council
Identifiable data: GMC number
Personal data: Sex, GMC registration date, nationality, ethnic group, country of medical qualification, specialty, allegation category, investigation outcome, age.

Dataset 2b - TQC knowledge test (KT) scores for the General Practitioners (GPs) who were investigated by the GMC from 2008 to 2013.
Provider: TQC scores provided by RDME and extracted by Emily Unwin from the individual TQC reports (one per doctor). Reports did not leave the RDME GMC folder on the secure S drive.
Identifiable data: anonymised
Personal data: KT score

Dataset 2c – Investigated doctors plus GPs KT scores (Datasets 2a and 2b merged)
Provider: GMC and RDME
Personal data: Sex, GMC registration date, nationality, ethnic group, country of medical qualification, specialty, allegation category, investigation outcome, age, knowledge test score, TQC date

Dataset 2d – TQC simulated surgery scores (SS) for the GPs who were investigated by the GMC from 2008 to 2013.
Provider: TQC scores provided by RDME and extracted by Emily Unwin from the individual TQC reports (one per doctor). Reports did not leave the RDME GMC folder on the secure S drive.

May 2018
Identifiable data: Anonymised
Personal data: SS score

Dataset 2e - Investigated doctors plus GPs KT and SS scores (Datasets 2c and 2d merged)
Identifiable data: GMC number
Personal data: Personal data: Sex, GMC registration data, nationality, ethnic group, country of medical qualification, specialty, allegation category, investigation outcome, age, knowledge test score, simulated surgery score, TOC date.

Dataset 2f - Updated case outcome data for investigated doctors
Identifiable data: GMC number
Personal data: Investigation outcome

Dataset 2g - GPs investigated by the GMC plus KT and SS scores (Datasets 2e and 2f merged)
Identifiable data: GMC number
Personal data: Sex, GMC registration data, nationality, ethnic group, country of medical qualification, specialty, allegation category, investigation outcome, age, knowledge test score, simulated surgery score, TOC date.

Dataset 2h - Tests of competence data (volunteer doctors, demographic data)
Identifiable data: GMC number, Surname, Forename, Date of birth
Personal data: Date of pilot, age, employment grade, Sax, University of primary medical qualification, postgraduate medical qualification, email address, mobile number, ethnicity, country of primary medical qualification.

Dataset 2i - Tests of competence data (volunteer doctors, knowledge test score)
Identifiable data: GMC number, Surname, Forename, Date of birth
Personal data: Date of pilot, age, employment grade, Sax, University of primary medical qualification, postgraduate medical qualification, email address, mobile number, ethnicity, country of primary medical qualification, KT score.

Dataset 2j - GPs investigated by the GMC, including KT and SS scores plus volunteer GPs KT scores (Datasets 2g and 2i merged)
Identifiable data: GMC number
Personal data: Sex, SS score (investigated GPs only), standardised KT score, investigated by the GMC (yes/no)

Dataset 2k - Tests of competence data (volunteer doctors, simulated surgery scores)
Identifiable data: Anonymised
Personal data: age, sex, years practicing, year qualified, list size, number of partners in practice, position, EMQ scores, SS scores

Dataset 2l - GPs investigated by the GMC, including KT and SS scores plus volunteer GPs KT and SS scores (Datasets 2j and 2k merged)
Identifiable data: GMC number
Personal data: Sex, SS domain scores, Overall SS score, KT score, investigated by the GMC (yes/no)

Study 3
Dataset 3 - Membership of the Royal College of Physicians PACES data
Provider: Royal College of Physicians
Identifiable data: RCF code number, date of birth, GMC number.

May 2018
Please note: Identifiers deleted from Emily Unwin's S drive Aug 2018.
Personal data: Attempt number, scores for each skill assessed, exam centre, exam date, pass or fail for each skill assessed and for exam overall, gender, nationality, ethnic group, living in UK or abroad, medical school, country of qualification.

Study 4
Dataset 4c – Survey data from doctors who entered medical school in 1991
Provider: Professor Chris McManus, UCL
Identifiable data: GMC number.
Please note: Dataset deleted from Emily Unwin’s S drive Aug 2018. Data with GMC number removed present in merged dataset 4e.
Personal data: date first survey was completed, year of birth, sex, marital status, ethnic group, children, specialty, registration status, MRCP attempts, MRCP pass, sanctions or warnings, answers to a variety of psychological measures including (but not limited to) motivation, personality, masculinity-femininity, study patterns, communication style, empathy.

Dataset 4d – Registration status of GMC registered cohort members as of August 2015
Provider: GMC
Identifiable data: GMC number.
Personal data: GMC registration status
Please note: Dataset deleted from Emily Unwin’s S drive Aug 2018.

Dataset 4e – Historical sanctions and active warnings of the GMC registered cohort members as of August 2015
Provider: GMC
Identifiable data: GMC number.
Personal data: Sanctions or active warnings, date of sanction or warning
Please note: Dataset deleted from Emily Unwin’s S drive Aug 2018.

Dataset 4f – Merged 4c, 4d and 4e
Personal data: date first survey was completed, year of birth, sex, marital status, ethnic group, children, specialty, registration status, MRCP attempts, MRCP pass, sanctions or warnings, answers to a variety of psychological measures including motivation, personality, masculinity-femininity, study patterns, communication style, empathy.

Dataset 5 – Data of doctors who were required to complete an assessment as part of an investigation into their fitness to practice (1995-2008).
Provider: RDOE, UCL
Identifiable data
Personal data: GMC number, Enquiry date, Outcome, Specialty, Full name, Date of birth, Reason for investigator, Age, Sex, Nationality, Primary medical qualification country, Primary medical qualification awarding body, Primary medical qualification date, Registration date, Ethnicity.

Dataset 6 – Annual Review of Competence Progression data
Provider: GMC
Anonymised data
Personal data: ARCP date, ARCP Deanery, Gender, ARCP year, Ethnicity, Age, Disability, Specialty, ARCP outcome.

Dataset 7 – Recruitment data of trainee doctors

May 2018
Provider: GMC
A non-identified data
Personal data: Gender, Age, Nationality, World region of primary medical qualification, Institution of primary medical qualification if in the UK, Ethnicity, Disability, Education level, School type, Parental education level, Deprivation quintile, Foundation School, Specialties applied to, Offer made, Accepted.

Dataset 8 - UCLMS cohort data
Provider: Dr Katherine Woolf
A non-identified data
Personal data: Sex, Country, Prior degree. Answers to a variety of questions including personality, life events and GHQ.

H. SHARING (DISCLOSURE)
Please describe how the outcomes of the research will be disseminated (for example provide an explanation as to where, and how, will the results be published, or other mechanisms you will be using to share the potential participants personal data).

All research outputs from the research projects arising from the PhD have/will not contained any identifiable information about any of members of the study population. Highlights of the research outcome will be published in UCL research reports, on line and in newsletters. The research outputs have/will include:
- Oral and poster presentations at academic and medical conferences
- Publication in clinical and medical peer-reviewed journals
- PhD thesis and oral viva

Dataset 2g - Dataset 2a plus knowledge test and simulated surgery scores (TG)
A copy of this dataset has been stored on the RDM&E GMC Research folder for other RDM&E GMC researchers to use, after permission was obtained from the 6NC. Both the original 2a dataset and the individual TG reports were available to the RDM&E GMC researchers prior to sharing the 2g dataset (i.e., the RDM&E researchers had the information available to create the same 2g dataset, but using my version would save them time). Variables available include: Date enquiry received, Case Number, GMC number, Years since qualification, Data first registered, KT score, KT standard set score, pilots’ mean score, SS domain scores, SS domain median scores, Overall SS score, SS standard set score, KT standard deviation.

I. CONSENT
Consent requirements for research projects can vary widely. Whether you are intending to use a consent form, information sheet, or verbally, it is recommended to assure compliance with the Data Protection Act and with ethical requirements.

Please include the information sheet and consent form you will be using for this project, and or protocol. If you are not including an information sheet and consent form, please explain how the consent will be recorded.

The data used in this research project came from secondary sources. Permissions to use the data were sought as required from each provider of the data.

J. DATA STORAGE
Please describe the arrangements you will make for the security of the data, including how and where it will be stored, i.e. UCL network, encrypted USB stick, encrypted laptop etc.

The data used in this research project has been stored in a folder on a secure drive (3 drive) on the UCL network. Access to this folder has been restricted to Emily’s PhD supervisors (Professor Dame Jane Dorn, Dr Katherine

May 2018
Woolf and Dr. Henry Potts), Professor Chris McEvans and Dr. Emily Unwin. When required (i.e. when no remote connection to the 3 drive was available) a copy of the data has been stored on an encrypted USB stick, but subsequently deleted when not in use. A copy of the dataset 2e, 2b, 1c, 2d, 2e, 2f, 2g, 2h, 2i, 3j, 2k, and 2l were also saved in a folder on the secure RDME GMC research drive. This folder was deleted in October 2018. A copy of the 2g dataset has been saved in a folder on the secure RDME GMC research drive.

To our knowledge there has been no breach of data protection.

*Advanced Encryption Standard (AES) 256 bit encryption which has been made a security standard within the NHS

Data Safe Haven – Identifiable Data Handling Solution

Will the personal identifiable data collected and processed as part of this research be stored in the UCL Data Safe Haven (mainly used by 3LMS divisions, institutes & departments)?

YES/NO

**If yes please ensure that you have explained how you will ensure that the data is held securely?

Further information on the Data Safe Haven service is available at: https://www.ucl.ac.uk/lit-for-3lms/research-safety/articles/data-safe-haven-reassurance

K. INTERNATIONAL TRANSFER

Will identifiable data be transferred outside the UK as part of this study? YES/NO

Data protection legislation prohibits the transfer of personal data to countries or territories outside the European Economic Area (which consists of the 27 EU member states, Iceland, Liechtenstein and Norway).

At the time of writing the following countries have also been deemed adequate – Andorra, Argentina, Canada, Faroe Islands, Guernsey, Isle of Man, Israel, Jersey, New Zealand, Switzerland and Uruguay.

The Data Protection Officer has produced guidance on the transfer of data overseas and particular to the United States. This is available from the Data Protection webpages.

If you intend to transfer data to a country not mentioned above, please supply details of adequate safeguards below:

Use of cloud computing, or the transfer of personal data to other organisations providing a specific service e.g., transcription services.

If you are intending to use, or are considering using a cloud service (defined as access to computing resources, on demand, via network), or plan on using a third-party organisation to deliver a service that will involve the transfer of personal data, you should ensure that there is an agreement in place which provides adequate levels of protection so that UCL can meets its obligations and protects the rights of the participants involved.

Please supply further details below, or seek advice by contacting the UCL Data Protection team data.protection@ucl.ac.uk

May 2018
1. Notification
(Please note that notification is a prerequisite for registration)

Have you informed your department's Data Protection Coordinator about your project? YES \[\square\] NO

2. Ethics
If you are seeking ethics approval for your research, please provide the relevant project ID number below.

Any questions regarding ethical approval should be directed to the relevant Ethics Committee or Governance Administrator.

1. UCL Ethics Project ID Number: 5020/001

2. Joint Research Office Project ID Number:

3. Other Project ID Number:

If you are not seeking ethical approval for your project, please explain why below:

3. Sponsor
Please provide details of the sponsor for this research below (if applicable). This can be an individual, company, institution, funding council, or another organization which takes responsibility for the initiation, management and/or financing of the research.

1. Proposed sponsorship arrangement

2. Details of sponsor

4. Checklist
Please submit your application form together with the appropriate supporting documentation that may be applicable from the list below.

1. Documents to be included with the application form

2. Participant information sheet(s)

3. Participant consent form(s)

4. Parent/guardian information sheet(s) and consent form(s) for research involving participants under the age of 18

5. Questionnaire

6. Advertisement of project

7. Other interview format(s)

8. Other documentation being used to invite/inform participants about the research

Approval: We may have some questions about the information you provide, but you will normally be provided with a registration number within 5 working days of submitting the form. However, the period leading up to meetings of the Ethics Committee is always very busy, and you should allow more time for your application to be processed. It is therefore very important to check in good time whether you need to register your project.

May 2018
Please note that Data Protection Registration numbers will NOT be issued when you submit an application form in person to the Data Protection Team.

Submit this form electronically and send to research.data-protection@ucl.ac.uk together with supporting documentation that you are intending to use. Please include ‘Data Protection Registration’ in the subject field.

This form will be returned to you with the appropriate registration number, which you may quote on your Ethics Application Form, or any other related forms.

<table>
<thead>
<tr>
<th>Data Protection Registration (Office use only)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>UCL Data Protection Registration Number</td>
<td>Date issued</td>
</tr>
</tbody>
</table>
UCL DATA IMPACT ASSESSMENT TEMPLATE FOR RESEARCH

Step 1 – DPIA team

<table>
<thead>
<tr>
<th>Name</th>
<th>Job Title</th>
<th>Email Address (as contact point for future privacy concerns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professor Dame Jane Dave</td>
<td>Professor of medical education</td>
<td></td>
</tr>
<tr>
<td>Dr Emily Unwalk</td>
<td>PhD candidate</td>
<td></td>
</tr>
</tbody>
</table>

Step 2 – Research summary

Project Name: Gender differences in the professional performance of doctors practicing in the UK.

Department/Entity: Research Department of Medical Education (RDMED).

Date: September 2019 to December 2020.

Step 3 – Identify the need for a DPIA

The research question is: are male doctors more likely to have poor professional performance, and if so why?

The research question is answered through a systematic review of six research studies of original research studies and a systematic review of the literature and meta-analysis using qualitative and quantitative methods to analyse large secondary dataset.

This research is important for support for doctors undergoing fitness to practice investigation and could go on to better understanding poor performance in doctors and providing better support for these doctors who are undergoing an investigation.

Please explain:
- the role of personal data in the project;
- the risks to privacy there are in your project (please list), and
- why the processing of personal data is necessary and proportional for the purposes of your project.

Role of personal data: The GMS number and a unique patient identifier (UPID) will be used to enable researchers to link or merge datasets. It was also used to identify those doctors who were registered to practice medicine in the UK and those who were not.

The outcomes of interest in this PHE project were: sanctions or warnings against a doctor’s medical registration, examination suspension rates. These outcomes were chosen because they are a measure of professional performance in doctors.

The independent variable of interest was the doctor’s gender. This independent variable was chosen to explore the difference in professional performance in doctors between the genders.

Step 4 - Please describe the information flows. If this is described in another document, please attach it to this DPIA

Information Flows: means the collection, retention, use, transfer and deletion – i.e. all types of data processing as part of the project’s lifecycle.

Please see attached document.
## Step 5 – What steps or controls are you taking to minimise risks to privacy?

**Please tick**

| a. Risks to individual privacy are minimal | j. Special category personal data is not used |
| b. Personal data is pseudonymised | k. Randomisation |
| c. Encryption of data at rest, i.e. when stored | l. Participant opt out at any stage of the research |
| d. Encryption used in transfers | m. Personal data kept in the EEA |
| e. Total number of participants is less than 50 | n. Research is not used to make decisions directly affecting individuals |
| f. Information compliance training for staff has been completed - data protection, information security, FOI | o. De-identification |
| g. Hashing or salting employed | p. Short retention limits |
| h. Adherence to privacy by design principles | q. Restricted access controls |
| i. Probabilistic risk management | r. Other (please specify) |

*Please see attached document*

## Step 6 – What steps have you taken to make sure the research is as accurate as possible and there are minimal unintended consequences? Please tick

| a. Data management plan in place | d. this study builds on a pilot study |
| b. Data management plan is peer reviewed | e. an extension to a previous similar study registered by DPO, |
| c. PI experience levels - no experience; some experience; very experienced | f. If there is, please provide the number |

## Step 7 – How have you assessed what participants will think of the research? What have you done to address concerns raised? Please tick

| a. Pilot project | b. Use of focus group | c. Information Sheet/Consent Form | d. Experience drawn from previous study |

## Step 8 – For the controls/steps specified in Step 5, who will make sure the controls are put in place? Please tick

| a. PI | b. Head of School | c. Other body (please specify) |
UCL Data Impact Assessment Template for Research

Step 4 – Please describe the information flows.

Please see the attached flow diagrams describing the information flows.

**Dataset 1 – The List of Registered Medical Practitioners (LRMP)**
Identifiable dataset

The General Medical Council (GMC) provided the data after a licence agreement was signed between Professor Dame Jane Dacre and the GMC.

The data was transferred to Dr Emily Unwin and Dr Katherine Woolf via the GMC SFTS (Secure File Transfer System) – a secure web portal, in April 2013.

Emily Unwin saved this data in a folder (Emily’s PhD research folder) on a secure drive with restricted access on the UCL network (UCLMS S drive).

Emily Unwin anonymised the dataset and saved the anonymised copy in a folder (Emily’s PhD research folder) on a secure drive with restricted access on the UCL network (UCLMS S drive). It is this anonymised copy of the dataset that Emily Unwin used to conduct her research study.

**Dataset 2a – All doctors under investigation by the GMC, who were required to complete Tests of Competence (ToC) from 2008 to 2013**
Identifiable data

This dataset is available to all Research Department of Medical Education (RDME) GMC researchers at UCL and is stored on a secure drive with restricted access on the UCL network (GMC project UCLMS S drive). In March 2015 Emily Unwin was given permission to access the dataset by the GMC after having completed a confidentiality agreement.

Emily Unwin made a copy of the dataset and saved it in a folder (Emily’s PhD research folder) on a secure drive with restricted access on the UCL network. She also kept a copy in her named folder within the GMC project UCLMS S drive folder (this folder was deleted in October 2018).

**Dataset 2b – ToC knowledge test (KT) scores for the General Practitioners (GPs) who were investigated by the GMC from 2008 to 2013**
Anonymised data

Emily Unwin was given access to the ToC reports by RDME. She extracted the knowledge test scores of the GPs included in dataset 2a from the individual ToC reports (April 2015 onwards) to create dataset 2b. Identifiable data (GMC number) was used to create this dataset, but once the dataset was complete the identifiable data was deleted.

Emily Unwin saved dataset 2b in a folder (Emily’s PhD research folder) on a secure drive with restricted access on the UCL network. She also kept a copy in her named folder within the GMC project UCLMS S drive folder (this folder was deleted in October 2018).

**Dataset 2c – Investigated doctors plus GPs KT scores (Datasets 2a and 2b merged)**
Identifiable data
Emily Unwin merged datasets 2a and 2b to create 2c. Dataset 2c was saved it in a folder (Emily’s PhD research folder) on a secure drive with restricted access on the UCL network. She also kept a copy in her named folder within the GMC project UCLMS S drive folder (this folder was deleted in October 2018).

**Dataset 2d** – ToC simulated surgery scores (SS) for the GPs who were investigated by the GMC from 2008 to 2013.

*Anonymised data*

Emily Unwin was given access to the ToC reports by RDME. She extracted the simulated surgery scores of the GPs included in dataset 2a from the individual ToC reports (April 2015 onwards) to create dataset 2d. Identifiable data (GMC number) was used to create this dataset, but once the dataset was complete the identifiable data was deleted.

Dataset 2d was saved it in a folder (Emily’s PhD research folder) on a secure drive with restricted access on the UCL network. She also kept a copy in her named folder within the GMC project UCLMS S drive folder (this folder was deleted in October 2018).

**Dataset 2e** – Investigated doctors plus GPs KT and SS scores (Datasets 2c and 2d merged)

*Identifiable data*

Emily Unwin merged datasets 2c and 2d to create 2e. Dataset 2e was saved in a folder (Emily’s PhD research folder) on a secure drive with restricted access on the UCL network. She also kept a copy in her named folder within the GMC project UCLMS S drive folder (this folder was deleted in October 2018).

**Dataset 2f** - Updated case outcome data for investigated doctors

Dataset 2a did not have complete case outcomes for all doctors.

In November 2015 Emily Unwin sent a copy of the list of GMC numbers (no other variables) of the investigated GPs whose case outcome was not complete in dataset 2a to obtain the updated case outcome. The list was sent to the GMC via the GMC SFTS. Emily Unwin then received the updated case outcomes of the doctors on the list from the GMC via the GMC SFTS (dataset 2f) in November 2015.

Dataset 2f was saved it in a folder (Emily’s PhD research folder) on a secure drive with restricted access on the UCL network. She also kept a copy in her named folder within the GMC project UCLMS S drive folder (this folder was deleted in October 2018).

**Dataset 2g** – GPs investigated by the GMC plus KT and SS scores (Datasets 2e and 2f merged)

*Identifiable data*

Emily Unwin merged datasets 2e and 2f to create 2g. Dataset 2g was saved it in a folder (Emily’s PhD research folder) on a secure drive with restricted access on the UCL network. She also kept a copy in her named folder within the GMC project UCLMS S drive folder (this folder was deleted in October 2018).
In July 2018 a copy of dataset 2g has been stored on the RDME GMC Research folder for other RDME GMC researchers to use. Permission was obtained from the GMC to do this. Both the original 2a dataset and the individual ToC reports were available to the RDME GMC researchers prior to sharing the 2g dataset (i.e. the RDME researchers had the information available to create the same 2g dataset, but using my version would save them time).

**Dataset 2h** - Tests of competence data (volunteer doctors – demographic data)

Identifiable data

The dataset of volunteer doctors (demographic data) is available to all Research Department of Medical Education (RDME) GMC researchers at UCL and is stored on a secure drive with restricted access on the UCL network. Emily Unwin accessed this dataset in August 2015 and saved a copy in a folder (Emily’s PhD research folder) on a secure drive with restricted access on the UCL network. She also kept a copy in her named folder within the GMC project UCLMS S drive folder (this folder was deleted in October 2018).

Please note: A copy of dataset 2h was also sent to Emily Unwin via UCL outlook email in October 2015, though Emily did not request and therefore did not save this data because she had already accessed the data through the secure drive and had saved a copy. Emily Unwin deleted the email along with its’ attached dataset from her UCL Outlook account in September 2018.

**Dataset 2i** – Test of competence data (volunteer doctors, knowledge test scores)

Identifiable data

Dataset 2h did not contain the knowledge test scores. Emily Unwin extracted and imported these scores from the individual databases for each of the General Practice pilots between 2006 to 2013 to create database 2i. Emily Unwin saved this database in a folder (Emily’s PhD research folder) on a secure drive with restricted access on the UCL network. She also kept a copy in her named folder within the GMC project UCLMS S drive folder (this folder was deleted in October 2018).

**Dataset 2j** – GPs investigated by the GMC, including KT and SS scores plus volunteer GPs KT scores (Datasets 2g and 2h merged)

Identifiable data

Emily Unwin merged datasets 2g and 2h to create 2j. Dataset 2j was saved it in a folder (Emily’s PhD research folder) on a secure drive with restricted access on the UCL network. She also kept a copy in her named folder within the GMC project UCLMS S drive folder (this folder was deleted in October 2018).

**Dataset 2k** - Tests of competence data (volunteer doctors, simulated surgery scores)

Anonymised data

This dataset is available to all Research Department of Medical Education (RDME) GMC researchers at UCL and is stored on a secure drive with restricted access on the UCL network. Emily Unwin accessed this dataset in August 2015 and saved a copy in a folder (Emily’s PhD research folder) on a secure drive with restricted access on the UCL network.

**Dataset 2l** – GPs investigated by the GMC, including KT and SS scores plus volunteer GPs KT and SS scores (Datasets 2i and 2j merged)

Identifiable data
In March 2016 Emily Unwin merged datasets 2j and 2k to create 2l. In March 2018 Emily Unwin re-created this dataset (this was done for purposes of checking the data had been correctly merged). Dataset 2l was saved it in a folder (Emily’s PhD research folder) on a secure drive with restricted access on the UCL network. She also kept a copy in her named folder within the GMC project UCLMS S drive folder (this folder was deleted in October 2018).

**Dataset 3** – Membership of the Royal College of Physicians PACES data
Identifiable data

The Membership of the Royal College of Physicians (MRCP) provided the data after a licence agreement was signed between Professor Dame Jane Dacre and the MRCP.

The data was transferred to Emily Unwin via the UCL Drop Box in September 2013. A passcode was required to retrieve the files.

Emily Unwin saved this data in a folder (Emily’s PhD research folder) on a secure drive with restricted access on the UCL network. In August 2018 Emily Unwin deleted the identifiable variables from this dataset.

**Dataset 4a** – Survey data from doctors who entered medical school in 1991 who completed the 2009 questionnaire
Anonymised data

Emily Unwin received a copy of these data in 2013 via UCL email.

Emily Unwin saved these data in a folder (Emily's PhD research folder) on a secure drive with restricted access on the UCL network. Emily Unwin deleted these data August 2018.

**Dataset 4b** - Survey data from doctors who entered medical school in 1991 Anonymised data

Emily Unwin received a copy of this dataset in October 2013 via UCL Drop Box.

Emily Unwin saved these data in a folder (Emily’s PhD research folder) on a secure drive with restricted access on the UCL network. Emily Unwin deleted these data August 2018.

**Dataset 4c** – Survey data from doctors who entered medical school in 1991 plus GMC numbers Identifiable data

Emily Unwin received a version of dataset 4b with GMC numbers (identifiable variable) in May 2015 via the UCL Drop Box. Emily Unwin saved these data in a folder (Emily’s PhD research folder) on a secure drive with restricted access on the UCL network. This dataset was deleted in August 2018 (though non-identifiable data from this dataset are in dataset 4f).

**Dataset 4d** – Registration status of GMC registered cohort members as of August 2015 and **Dataset 4e** – Historical sanctions and active warnings of the GMC registered cohort members as of August 2015.
Identifiable data
The General Medical Council (GMC) provided the data after a research agreement was signed between Dr Katherine Wooll and the GMC.

In August 2015 Emily Unwin sent a copy of the list of GMC numbers (no other variables) of the cohort members to the GMC to obtain current registration and historical sanctions and active warnings data. The list was sent via uCL drop Box and required a passcode to retrieve the data. The GMC did not retrieve the file and it expired after 10 days. Emily Unwin sent the list of GMC number to the GMC via the GMC SFTS in August 2015. Emily Unwin then received the current registration status and any historical sanctions and active warnings of the doctors on the list from the GMC via the GMC SFTS.

These data were saved these data in a folder (Emily's PhD research folder) on a secure drive with restricted access on the UCL network. Emily Unwin deleted these data August 2018 (though non-identifiable data from this dataset are in dataset 4f).

**Dataset 4f** – Survey data plus registration status and sanctions/warnings (Datasets 4c, 4d and 4e merged)
Identifiable variables were deleted from this dataset in August 2018.

Emily Unwin saved these data in a folder (Emily's PhD research folder) on a secure drive with restricted access on the UCL network.

**Dataset 5** – Data of doctors who were required to complete an assessment as part of an investigation into their fitness to practise (1996-2008)
Identifiable data
This dataset was available to all Research Department of Medical Education (RDME) GMC researchers at UCL and was stored on a secure drive with restricted access on the UCL network. Emily Unwin saved a copy these data in a folder (Emily's PhD research folder) on a secure drive with restricted access on the UCL network. Emily deleted this copy of the dataset in August 2018.

**Dataset 6** – Annual Review of Competence Progression Data
Anonymised data
The General Medical Council (GMC) provided the data after a research agreement was signed between Professor Dame Jane Dacre and the GMC.

Emily Unwin received these data from the GMC end of 2013 / beginning of 2014 via the GMC Connect Account.

Emily Unwin saved these data in a folder (Emily's PhD research folder) on a secure drive with restricted access on the UCL network.

**Dataset 7** – Recruitment data of trainee doctors
Anonymised data
The General Medical Council (GMC) provided the data after a research agreement was signed between Professor Dame Jane Dacre and the GMC.

Emily Unwin received these data from the GMC end of 2013 / beginning of 2014 via the GMC Connect Account.

Emily Unwin saved these data in a folder (Emily's PhD research folder) on a secure drive with restricted access on the UCL network.
Dataset 8 – UCLMS cohort data
Anonymised data

Emily Unwin received a copy of these data in February 2016 via UCL email. Emily saved these data in a folder (Emily’s PhD research folder) on a secure drive with restricted access on the UCL network.

Step 5 – What steps or controls are you taking to minimise risks to privacy?

The data used in this research project have been stored in a folder on a secure drive (S drive) on the UCL network. Access to this folder has been restricted to Emily’s PhD supervisors (Professor Dame Jane Dacre, Dr Katherine Woolf, Dr Henry Potts), Professor Chris McManus and Dr Emily Unwin. When required (i.e. when no remote connection to the S drive was available) a copy of the data has been stored on an encrypted USB stick, but subsequently deleted when not in use.
An email was received on 18/03/2019 from the UCL Data Protection Office confirming the project was now registered under reference number Z6364106/2013/08/11.