# Paediatric diagnostic test utilisation in primary care: an analysis of variation using quantitative and qualitative approaches


**Elizabeth Taruna Thomas**

Kellogg College

University of Oxford

A thesis submitted for the degree of

*Doctor of Philosophy*

Michaelmas 2023

*This thesis is dedicated to my greatest sources of love and inspiration:*

My sisters, Dr Sarah Tanusha Thomas and Anna Tashwina Thomas

My late grandparents,
Mr T T Titus, Mrs Sara Titus,
Mr Varghese K Georgekutty and Mrs Thankamma George

Finally, my parents, Dr Thomas T Titus and Dr Sara K Titus,
Whose belief in their own dreams, and their children's dreams, made this thesis possible.

# Acknowledgements

This DPhil thesis is the product of so many people who have invested in me and contributed to my development over the years.

First and foremost, I am grateful to my supervisors. Their combined skillsets, experience, and expertise in different methods provided me with a great deal of value and guidance during the DPhil.

I owe a debt of gratitude to Professor Carl Heneghan, who set the wheels in motion five years ago during my undergraduate research elective when he suggested I apply for a DPhil. Over the past few years, the questions he posed and our ensuing discussions during supervision meetings pushed me to think more critically and left me feeling inspired. His ability to see the aerial view challenged me to constantly reflect on where my research fits in and why it is important. I am also very grateful to him for providing the bridging funds that saw me through to the completion of this DPhil.

I am immensely grateful to Professor Rafael Perera for his encouragement, kindness, humour, as well as the professional and personal support he has given me during my studies. He always found time for me to patiently explain confusing statistical concepts and helped me find pragmatic solutions to my questions. He also found additional avenues to fund my research and provided thoughtful feedback and advice.

Dr Peter Gill's insights and experiences as a paediatrician and researcher were critical to my thesis. Since our very first meeting, he showed genuine interest in my intellectual and professional growth and proved to be an incredibly engaged and committed supervisor despite being based in Canada for which I am deeply grateful. Above all, I thank him for his mentorship and discussions that have extended far beyond the DPhil, shaped my future career goals and ongoing professional development.

There are several important mentors who had an impact on me before I came to Oxford. Professor Paul Glasziou introduced me to the world of research, provided me opportunities to learn by doing, and mentored me on several projects that laid the foundations for what would eventually lead to this DPhil project. He invested so much in my research and career development, and I am immeasurably grateful to him. Professor Michelle McLean mentored me during medical school and provided

invaluable support during the DPhil application process. I thank Dr Nemat Alsaba, Dr Michael Brohier, and Mrs Denise Collins who all taught me during my school and undergraduate years and later supported my application to Oxford.

I am grateful to the Clarendon Fund and Kellogg College for supporting my doctoral studies through the Clarendon scholarship. The NIHR School for Primary Care Research and Medtech and In vitro diagnostics Co-operative also funded two of the research projects within this thesis.

Dr Margaret Glogowska provided patient and encouraging supervision to me during my qualitative project for which I am so grateful. Other key collaborators I would like to thank include Dr Brian Shine, Professor Sue Moloney, and Professor Gail Hayward.

My very first friend in Oxford, Dr Georgia Richards, who I met in 2017, has been an unwavering source of support and practical advice throughout my DPhil. I am so grateful to her for her wisdom and longstanding commitment to our friendship.

A heartfelt thanks to Dr Diana Withrow, who made my daily venture into the department something to look forward to, as well as her steady stream of thoughtful advice and motivation over many cups of Common Ground coffee. I have learned so much from her in writing this DPhil.

Within the CEBM, I am immensely grateful to Ruth Davis for her support over the years for work and non-work-related matters. I am also grateful to Georgette Eaton and Amadea Turk for their friendship and motivation as we did our DPhils alongside each other. Other friends within the department I wish to thank are Kiana Collins and Annika Theodoulou, as well as my college supervisor, Dr Jamie Hartmann-Boyce.

Life in Oxford felt like a happy dream, because of my brilliant, generous, and loving friends who made the city feel like a second home. They brought the sunshine on the rainiest of days. I especially want to mention Ashleigh Barnes, Henry West, Georgia Williams, Beth Hamilton, Alli Devlin, Josechu Hernandez-Martinez, Katie Kowal, Nick Salmon, and Christina Hasenhüttl. We shared so much laughter, Michelin-quality food, adrenaline-fuelled game nights, and many memorable adventures that will forever be etched in my memory. It has been the gift of a lifetime to share this chapter of our lives together. A special thanks goes to the members of the Oxford Australian Rules

Football Club for providing me with such a welcoming and thriving community (along with a few broken bones).

Of all the roles that I get to be, it is being a sister brings me the most joy. Being away from my sisters for the past three years was not easy, especially to have missed out on the small moments as well as some big ones in their lives. My not-so-little sister, Anna, has blossomed. I am especially grateful to my sister and collaborator, Sarah, for her long-distance emotional support, as well as her intellect which is sharp and insightful. I am also grateful that she gave me a new brother(-in-law), Sajeev.

My grandparents played a formative role throughout my life and shaped my early identity and values. My sisters and I had the ultimate privilege of receiving their unconditional and steadfast love throughout our lives, until just before I left for Oxford in 2020. Milestones like this feel bittersweet as I remember them with sadness but the deepest affection and gratitude.

Finally, I would like to thank my parents. My father's bold vision, hunger for knowledge, and innate sense of duty to care for others has been a guiding influence. My mother has always been my pillar of strength and support. There are not enough words in the world to express how profoundly grateful I am to be their daughter. They made sacrifices that I still cannot fathom, to ensure we could flourish. They instilled in me a deep love for learning, nurtured my dreams, and imbued the importance of cultivating and using one's talents to be of service to others. Anything I have achieved is a testament to their love.

One of the many gifts that my father gave me was his taste in music, particularly his love for ABBA, and I can almost always find an ABBA lyric to relate to any situation. In completing this thesis, the following lyrics from their 2021 release, "I Still Have Faith in You", came to my mind:

*We stand on the summit,*
*Humbled and grateful to have survived*.

# Abstract

This thesis aimed to examine the landscape of diagnostic tests for children in primary care. My hypothesis was that substantial variation exists in how general practitioners (GPs) request tests for children, as this area entails a high level of diagnostic uncertainty. Unwarranted variation has harmful consequences; it exacerbates health inequities and strains health resources. My thesis, therefore, aimed to quantify and explore variation in paediatric test use and identify the potential drivers of variation.

First, laboratory test use in Oxfordshire was analysed from 2005 to 2019, to establish the role of paediatric diagnostic tests in general practice compared with other settings. In the subsequent study, paediatric test use was investigated in general practice using data from the Clinical Practice Research Datalink from 2007 to 2019. Tests that increased by the greatest margin and were subject to the highest between-practice variability were identified as tests that are potentially overused and warrant further investigation. Following this, semi-structured interviews were undertaken with 18 GPs and 2 GP trainees to explore their perspectives on requesting tests for children and identify some perceived drivers of variation. Finally, diagnostic guidelines for common paediatric conditions were examined, to determine whether the quality of guidelines and the quality of evidence underpinning diagnostic test recommendations could be driving variation in testing practices.

This thesis presents the first comprehensive study of paediatric test use and variation in primary care. It highlights the scarcity of evidence for diagnostic tests in children, despite their increasing use in general practice. The variability of test use that was observed between practices reflects inconsistency in guidelines and practice. In the final chapter, a research strategy to develop paediatric test indicators is proposed; these indicators seek to identify inappropriate testing and improve the quality of diagnostic services delivered to children.

# Table of contents

# Statement of Contributions

I certify that this thesis contains my own work. In cases where others have contributed to my work, their initials have been used. Their contributions are summarised below:

- Dr Subhashisa Swain (SS) conducted the Clinical Practice Research Datalink (CPRD) feasibility count in Chapter 2.
- Dr Sarah Tanusha Thomas (ST) conducted duplicate screening, extraction, and AGREE II scoring for Chapters 2, 6 and 7 as well as duplicate GRADE assessments for Chapter 7.
- Dr Julian Treadwell (JT) resolved disagreements when screening NICE guidelines relevant to primary care (Chapter 2) and voted on asthma tests that were relevant to primary care in Chapter 7.
- Dr Brian Shine (BS) extracted the laboratory testing data from Oxford University Hospitals and Oxfordshire general practices and verified the test code lists in Chapter 3. He also assisted in identifying GPs who requested the most tests in Oxfordshire to recruit them for the interview study in Chapter 5.
- Dr Patrick Fahr (PF) provided assistance with statistical analysis using R for Chapter 3.
- Prof Carl Heneghan (CH) verified the code lists for the feasibility count in Chapter 2, the test code lists and panel assignments for Chapter 4. He also resolved discrepancies related to the inclusion of guidelines and tests in Chapters 6 and 7.
- Dr Cynthia Wright Drakesmith (CWD) extracted the CPRD Aurum data for Chapter 4.
- Dr Diana Withrow (DW) provided advice on statistical analysis for Chapters 3 and 4.
- Dr Margaret Glogowska (MG) provided qualitative expertise and supervision for Chapter 5.
- Prof Gail Hayward (GH) also provided advice on the qualitative methods and analysis for Chapter 5.
- Nia Roberts (NR) conducted the searches to identify relevant international guidelines for Chapters 6 and 7.
- Professor Susan Moloney provided paediatric clinical expertise relevant to the Australian and international context for Chapters 6 and 7.
- Dr Brian Nicholson (BN), Dr Guy Bidwell (GB), Dr Stephen Bradley (SB), Dr Sharon Dixon (SD), and Dr Joseph Lee (JL) voted on the asthma tests relevant to primary care in Chapter 7.

# List of Figures

# List of Tables

# Abbreviations

| | |
|---|---|
| **AAP** | American Academy of Pediatrics |
| **AAPC** | Average Annual Percentage Change |
| **ACR** | American College of Radiology |
| **APC** | Annual Percentage Change |
| **BC** | British Columbia guideline |
| **BTS / SIGN** | British Thoracic Society / Scottish Intercollegiate Guideline Network |
| **CC** | Cincinnati Children's Hospital |
| **CMP** | Calcium, magnesium, phosphate |
| **CoV** | Coefficient of Variation |
| **CPG** | Clinical practice guideline |
| **CPRD** | Clinical Practice Research Datalink |
| **CRP** | C reactive protein |
| **CT** | Computed tomography |
| **CTS** | Canadian Thoracic Society |
| **CXR** | Chest X-ray |
| **DCGP** | Dutch College of General Practitioners |
| **EBM** | Evidence-based medicine |
| **ECG** | Electrocardiogram |
| **ED** | Emergency Department |
| **ERS** | European Respiratory Society |
| **ESPGHAN** | European Society for Paediatric Gastroenterology Hepatology and Nutrition |
| **ESPID** | European Society for Paediatric Infectious Diseases |
| **FBC** | Full blood count |
| **FeNO** | Fractional exhaled nitric oxide |
| **FEV1** | Forced expiratory volume in 1 second |
| **FMS** | Federatie Medisch Specialisten |
| **FVC** | Forced vital capacity |
| **GIN** | Guidelines international Network |
| **GINA** | Global Initiative for Asthma |
| **GLM** | Generalized linear model |
| **GP** | General practitioner |
| **HbA1c** | Glycated Haemoglobin |
| **ICGP** | Irish College of General Practitioners |
| **ICON** | International Consensus on Pediatric Asthma |
| **IDSA** | Infectious Diseases Society of America |
| **IMD** | Index of Multiple Deprivation |
| **KP** | Kaiser Permanente |
| **LFT** | Liver function test |
| **MCS** | Microscopy, culture, sensitivities |

| | |
|---|---|
| **MIC** | Medtech and In vitro diagnostics Co-operatives |
| **MRI** | Magnetic resonance imaging |
| **NAC** | National Asthma Council Australian Asthma Handbook |
| **NASPGHAN** | North American Society for Paediatric Gastroenterology Hepatology and Nutrition |
| **NHLBI** | National Heart, Lung, Blood Institute |
| **NHS** | National Health Service |
| **NICE** | National Institute for Health and Care Excellence |
| **NIHR** | National Institute for Health and Care Research |
| **NOS** | Not otherwise specified |
| **OCP** | Ova, cysts, parasites |
| **OUH** | Oxford University Hospitals |
| **PET** | Positron emission tomography |
| **PTh** | Parathyroid hormone |
| **RCGP** | Royal College of General Practitioners |
| **RCH** | Royal Children's Hospital |
| **SA** | South Australia Ministry of Health |
| **SD** | Standard deviation |
| **SPCR** | School for Primary Care Research |
| **TFT** | Thyroid function test |
| **UK** | United Kingdom |
| **US** | Ultrasound |
| **WGO** | World Gastroenterology Organization |
| **WHO** | World Health Organization |

# Chapter 1. Overview and structure of thesis

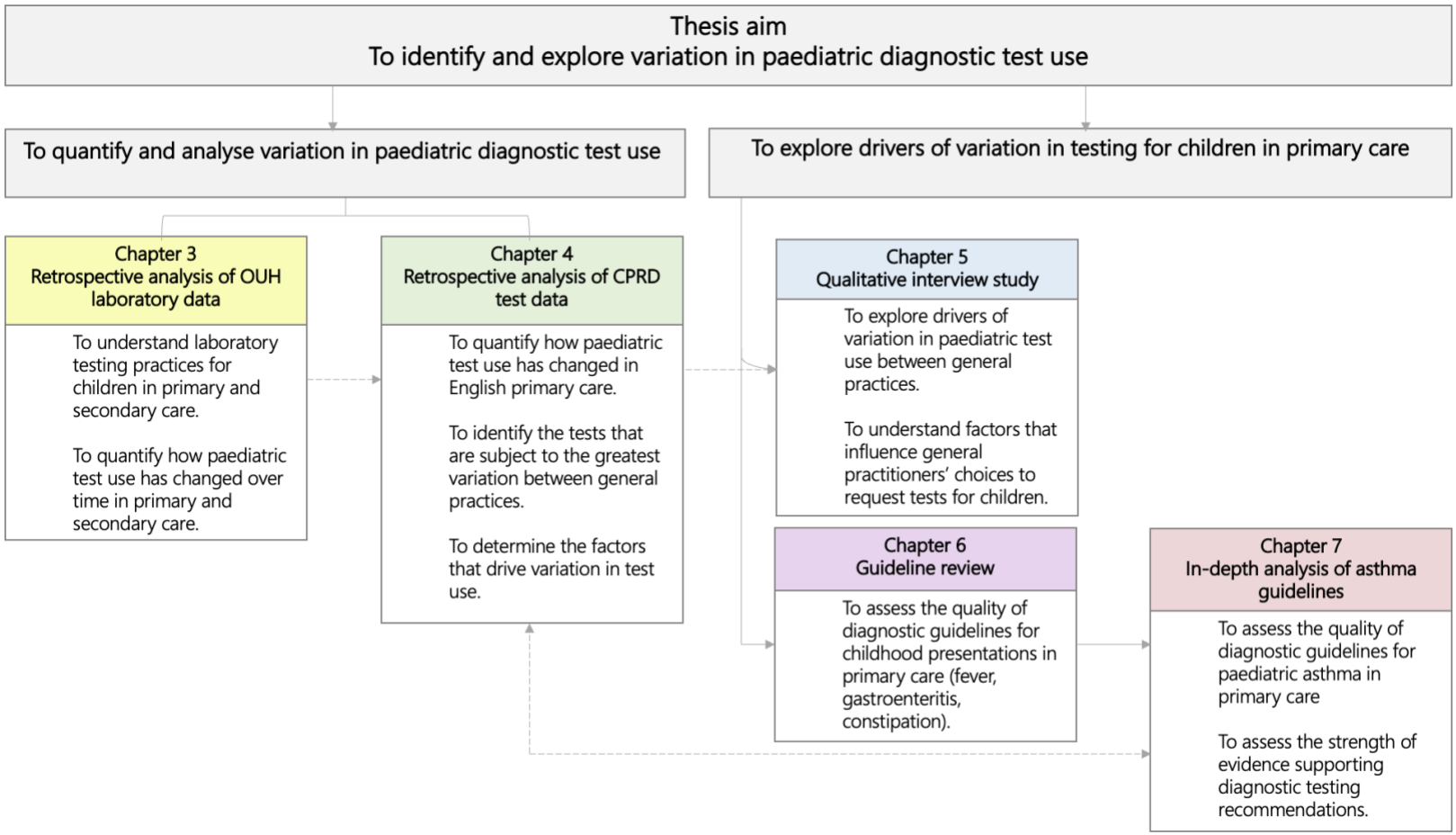
## 1.1 Aims and objectives

This thesis aimed to understand the role of diagnostic tests for children in primary care and quantify and explore variation in paediatric test use. The specific objectives were:

1. To examine laboratory testing practices for children in primary and secondary care
2. To quantify how test use has changed over time in primary and secondary care.
3. To quantify diagnostic test utilisation rates for children in English primary care.
4. To identify the tests that are subject to the greatest variation between general practices.
5. To explore drivers of variation in test use between general practices.
6. To explore factors that influence GPs' decisions to request a test for a child.
7. To assess the quality of diagnostic guidelines related to primary care childhood presentations.
8. To examine the evidence base for paediatric diagnostic testing recommendations within guidelines for children in primary care.

## 1.2 Thesis aims, objectives, and chapter summaries

**Figure 1.1 A summary of the thesis aim, objectives, and study objectives**



### 1.2.1 Chapter summaries

*Chapter 2: Introduction*

In chapter 2, I outline the rationale for this thesis. To ensure the sustainability of health systems such as the NHS, health spending needs to be more efficient and health care waste needs to be reduced. Children in the UK have worse health outcomes compared with their counterparts in the rest of Western Europe, due to wide socioeconomic inequities that result in poor health outcomes for the most disadvantaged children. I discuss the importance of measuring health care variation; to establish where differences exist and then address the sources of unwarranted variation. Most health care contacts in the UK for children take place in general practice, but the paediatric population can pose diagnostic challenges in this setting. General practitioners (GPs) have an arsenal of tools and strategies to assist them in making a diagnosis, including tests. Variation in test use in primary care has been measured and researched in adults, but to date no studies have focused on identifying and exploring variation in paediatric diagnostic test use. Reducing unwarranted variation in care will ensure tests are delivered equitably to children according to need and health spending is more efficient.

*Chapter 3: Trends in diagnostic tests ordered for children: a retrospective analysis of 1.7 million laboratory test requests in Oxfordshire, UK from 2005 to 2019.*

In Chapter 3, I sought to examine diagnostic testing practices for children in the local Oxfordshire region. I analysed temporal trends in laboratory testing for children aged 0 to 15 from 2005 to 2019 using data from the Oxfordshire University Hospital NHS Trust laboratories. I used joinpoint regression models to estimate annual percentage changes (APC) and average annual percentage change (AAPC) in test use. Temporal changes in age-adjusted rates of test use were calculated overall and stratified by healthcare setting, sex, and age. I found that overall test use demonstrated an

apparent increase in children (AAPC 1.5%, 95 confidence interval -0.8% to 3.9%), particularly after 2012. Increases were highest in females, those aged 6-15 years and outpatient and general practice settings. The most frequently requested tests were full blood count, urea and electrolytes, liver function test, C-reactive protein, and calcium magnesium phosphate. The test with the greatest increase in use was Vitamin D, which increased on average by 27% per year. Other tests that showed a significant temporal increase included parathyroid hormone, iron studies, folate, vitamin $B_{12}$ levels, glucose, HbA1c, IgA, coeliac, creatine kinase, thyroid function tests, and IgG/IgM. Test changes were not uniformly distributed across all settings and age groups, highlighting the areas where efforts could be targeted to mitigate unnecessary testing.

*Chapter 4: Temporal and practice variation in paediatric diagnostic tests in UK primary care from 2007 to 2019.*

In chapter 4, I aimed to build on the findings from chapter 3, with a focus on paediatric test use in general practice. I sought to quantify temporal changes diagnostic test utilisation rates (defined as tests per 1000 child-years) for children in English primary care, identify tests with the highest practice variation, and explore drivers of variation in testing between different general practices. I obtained general practice data from the Clinical Practice Research Datalink (CPRD) for children aged 0-15 years old who underwent tests from January 1, 2007, to December 31, 2019. I included a combination of blood tests mentioned in Chapter 3, in addition to imaging, microbiology, and other miscellaneous tests that were included in NICE guidance for children that were relevant to primary care (Chapter 6). Any tests that were specifically mentioned by participants in the qualitative interview study (Chapter 5) were also included. I analysed approximately 14 million tests in 2.5 million children across 1,464 GP practices and found that test rates increased at a rate of 3.6% per year (AAPC 95% CI 3.4 to 3.8%). Compared with imaging and other miscellaneous tests, blood tests increased by highest margin. Females aged 11-15 experienced the most consistent

increase, as well as children from more deprived areas in England. Tests with the biggest temporal increases included: faecal calprotectin, fractional exhaled nitric oxide (FeNO) testing for asthma, vitamin D, folate, vitamin $B_{12}$, and coeliac testing. Tests that were subject to the largest practice variation include fractional exhaled nitric oxide (FeNO), hearing tests, vitamin D, helicobacter testing, and monospot testing for glandular fever. I also identified tests that had an above-average rate of test use and practice variability: iron studies, vitamin $B_{12}$, coeliac test, folate, and vitamin D. The high degree of temporal increase and practice variation reflects a lack of consistency in practice and highlights the need for more evidence and informed guidance in these areas.

*Chapter 5: Exploring GP perspectives on diagnostic testing for children: a qualitative interview study.*

In this chapter, I conducted semi-structured qualitative interviews with 20 practising GPs and trainees in England to explore their views on why there are significant differences in test use for children and the factors they consider when deciding to request a test for a child. The interviews were transcribed and analysed thematically as recommended by Braun and Clarke. GPs reflected that their approach to utilising tests in children differed from adults; specifically, that their threshold to test was higher but their threshold to refer to specialists was lower. Perceptions of test utility varied between participants, including the utility of specific tests for diagnosing conditions in children (e.g., objective testing for asthma). Perceived drivers of variation in testing were grouped into 1) intrinsic (clinician) factors, relating to their individual risk tolerance and experience, and 2) extrinsic factors which encompassed disease prevalence, parental concern and differing expectations of healthcare, workforce changes leading to fragmentation in care, time constraints, local system factors, and differences in guidelines. The findings of this chapter highlight actionable issues for clinicians,

researchers, and policymakers, to address unwarranted differences in test use and improve the quality of health care delivered to children in general practice.

*Chapter 6: A quality appraisal of paediatric guidelines in general practice.*

Clinical practice guidelines (CPGs) aim to standardize care delivered when diagnosing conditions, however, the quality of paediatric CPGs relevant to diagnosis in primary care is unknown. Therefore, in this chapter, I sought to appraise the quality of diagnostic guidelines relevant to children presenting to primary care, to assess whether it could be a driver of variation in GP testing rates. I selected the top three conditions from the feasibility count undertaken in Introduction Chapter 2.8 – fever, constipation, and gastroenteritis – to be the focus of this guideline review. Two reviewers (ET, ST) assessed the quality of 16 eligible guidelines using the AGREE II tool, and we found substantial variation in the quality of their methodological reporting. Across the guidelines, major weaknesses included limited consideration of guideline applicability and inadequate reporting of competing interests.

*Chapter 7. The quality of paediatric asthma guidelines: evidence underpinning diagnostic test recommendations from a meta-epidemiological study*

In this chapter, I focussed on guidelines for asthma, which is one of the most prevalent childhood conditions for which a variety of tests have been developed to improve diagnosis. In this meta-epidemiological study, I sought to determine (1) the methodological quality and reporting of paediatric guidelines for the diagnosis of childhood asthma in primary care, and (2) the strength of evidence supporting diagnostic test recommendations. I included guidelines from the US, UK, and other high-income countries with comparable primary care systems to the UK if they contained diagnostic testing recommendations for childhood asthma in primary care. The AGREE-II tool was used to assess the quality and reporting of the guidelines. The quality of the evidence was assessed using GRADE. Eleven guidelines met the

eligibility criteria. The methodology and reporting quality varied across the AGREE II domains (median score 4.5 out of 7, range 2 to 6). The quality of evidence supporting the 50 diagnostic recommendations was generally of very low-quality. All guidelines recommended the use of spirometry and reversibility testing for children aged ≥6 years, however, the recommended spirometry thresholds for diagnosis differed across guidelines. For four of the seven included tests, there were disagreements in testing recommendations on whether to perform the test or not. The variable quality of guidelines, lack of good quality evidence and inconsistent recommendations for diagnostic tests were identified as important factors that may contribute to poor clinician adherence to guidelines and variation in testing for diagnosing childhood asthma.

*Chapter 8: Discussion and conclusions*

In Chapter 8, I summarise the findings of the overall thesis as well as its place in the context of the wider literature. I describe the strengths and limitations of the different methodological approaches that were used, along with a proposed plan for future research. The implications of this thesis for patients and parents, practice and policy are also discussed.

## 1.3 Key accomplishments

### 1.3.1 Publications

At the time of submission of this thesis, three papers from my DPhil research (Chapters 3, 6, and 7) have been published in peer-reviewed journals. A fourth paper (based on Chapter 5) has been accepted for publication in the British Journal of General Practice. I am also fortunate to have had the opportunity to lead and co-author publications that extend beyond the scope of my DPhil. As of December 2023, I have accumulated 14

peer-reviewed publications since 2020, when I commenced my DPhil. Six of these are first-author publications. The list of my publications is provided in **Chapter 1 Appendix**.

### 1.3.2 Research collaborations

My DPhil has offered the opportunity for me to collaborate with experts in evidence-based medicine, research methods, general practice, paediatrics, and pathology in Australia and throughout the UK. I also collaborated with paediatricians in the US on a review article summarising key published works in paediatric overuse in the preceding two years (see **Chapter 1 Appendix**). I also worked with a paediatrician-researcher in the US to perform a secondary analysis of a randomised controlled trial that explored the efficacy of ophthalmic antibiotics for children presenting to primary care with conjunctivitis caused by *Haemophilus influenzae* vs other bacterial causes of conjunctivitis.

### 1.3.3 Funding and Grants

I am immensely grateful to the Oxford University Clarendon Scholarship for supporting my DPhil studies. I was awarded a National Institute for Health and Care Research (NIHR) School of Primary Care Research grant (Award No 624, £23,085) to fund the studies that comprised Chapter 4 and 5 of this thesis. The Award letter is attached as in the **Chapter 1 Appendix**. The qualitative study (Chapter 5) was also supported by the NIHR MedTech and In vitro diagnostics Co-operatives (MIC).

### 1.3.4 Training and skills development

I attended courses offered by the Nuffield Department of Primary Care Health Sciences and the Medical Sciences to develop skills in research methods related to my DPhil, including evidence synthesis and critical appraisal, conducting systematic reviews, statistical analysis, qualitative analysis, and the principles of open science. I attended the University of Oxford training courses on using STATA and R statistical software, and NVivo for qualitative research. I completed the Big Data Epidemiology module (part of the MSc in Evidence-Based Health Care) to equip me with the skills required to

use the Clinical Practice Research Datalink database and analyse the data needed for chapter 4. I also completed the Teaching Evidence-Based Practice (TEBP) course in September 2023.

In addition to formal training, I completed informal training. I learned to use the AGREE guideline appraisal tool for Chapter 7 and 8, as well as the GRADE tool for Chapter 8. I consulted textbooks and academic articles, watched online training videos, and sought advice from colleagues and senior researchers who had experience of using the relevant methods.

### 1.3.5 Teaching and supervision

I have had the opportunity to take part in teaching and supervision for two groups of undergraduate medical students completing an EBM special study module. Over the course of the year they planned, designed, and carried out systematic reviews on a topic of their choosing.

Following on from the TEBP course, I also had the opportunity to facilitate workshop sessions on appraising diagnostic accuracy studies and randomised controlled trials, as part of the Practice of Evidence-Based Health Care (EBHC) module, a core module on the MSc in EBHC.

### 1.3.6 Knowledge translation (Patient and Public Involvement)

I have been grateful to actively participate in knowledge translation with other clinicians, researchers, and members of the public through my Patient and Public advisory panel. It has been one of the most rewarding aspects of the DPhil journey.

I shared some of the findings from Chapters 3, 5, 6, and 7 of this thesis at several meetings and conferences. These included: the Preventing Overdiagnosis (Calgary, June 2022), EBM Live (Oxford, July 2022), Oxford University Hospitals Paediatric

Grand Rounds (Oxford, March 2023), Preventing Overdiagnosis (Copenhagen, August 2023) and the Royal College of General Practitioners (RCGP) Optimal Testing Meeting (online, September 2023).

I recruited three members of the public, who were parents of children who had experience of undergoing medical tests to form my Patient & Public advisory panel for Chapters 4 and 5. I held three knowledge translation meetings with the advisory group. They helped to shape my planned research, provided input on the proposed interview topic guide for my qualitative study, helped recruit GPs and trainees to participate in interviews, and provided their reflections on the potential implications of my research.

## 1.4 Personal motivations for this thesis

This thesis was motivated by my reflections on my research, professional and personal experiences.

I was introduced to the concepts of unwarranted variation, overuse, and heath care harms during my undergraduate medical degree. I was supervised by Professor Paul Glasziou on a systematic review in which we estimated the prevalence of incidental breast cancer at autopsy, or in other words, the potential reservoir for breast cancer overdiagnosis. I was grateful to present this systematic review at the Preventing Overdiagnosis conference in Barcelona in 2016. At this conference, I heard from clinicians, academics, and patients about their research and personal experiences of overdiagnosis across many areas of health care. It was the first time I became aware of propensity for health care to cause harm. As a relatively “green”, idealistic medical student at the time, this challenged my preconceived notions and prompted me to start thinking deeper. For my MD project, I conducted another systematic review that demonstrated the age-related decline of normal lung function, to justify why COPD

diagnosis thresholds to be adjusted for age and sex, to prevent both under and overdiagnosis of COPD.

After I became a junior doctor, I became cognisant that my colleagues and I were contributing to the problem of too much medicine by overtesting. We routinely ordered daily bloods on patients, as was the advice given to us by well-meaning predecessors on our rotations. We were part of a culture of defensive medicine in which, we all felt we had to do more to practice safely. Part of the problem was that the harms of over testing rarely, if ever, manifested on an individual level. I experienced cognitive dissonance in realising my own role as a contributor to the problem of overtesting, when I knew that our testing practices could be ultimately harmful on a patient, family, and system level.

I later reflected on my personal experiences of overtesting when my younger sister, at the age of six years, was incidentally found to have a heart murmur when she presented to the emergency department following an episode of acute asthma. Her symptoms resolved and she was discharged with a salbutamol inhaler and a referral to the paediatric cardiologist at the children's hospital for an echocardiogram. The echocardiogram took place several weeks later at the Queensland Children's Hospital (which is one-and-a-half hours away from home), and she was found to have a benign flow murmur. It prompted me to wonder if a different clinician had reviewed my sister in the emergency department, whether they would have referred her for the same scan or deemed that the murmur had enough benign features to avoid the scan and the unnecessary associated costs.

The culmination of these experiences and reflections led me to applying for the DPhil, to explore the issue of variation in paediatric diagnostics.

# Chapter 2. Introduction

## 2.1 Health system sustainability

Health systems worldwide, particularly the National Health Service (NHS), face mounting challenges to meet the demands of a growing population and increasing medical complexity, with limited funding, resources, and a dwindling health workforce. Total health care expenditure in the NHS increased in real terms* from £115.5 billion in 2013/14 to £137.4 billion GBP in 2019/20 before the pandemic.[1] In 2024/25, NHS England spending is predicted to be £158.6 billion, a relative increase of 40% since 2013.[1] To ensure the sustainability of health systems, patients should receive effective, evidence-based, high-quality care while minimising waste and the carbon footprint of the health system. Urgent attention is needed to resolve workforce attrition and burnout. Greater investments are needed for health care services within the community and to address the socioeconomic determinants of health within the population that result in poorer health outcomes.[2,3] There should be improved access to high quality health and social care services for underserved populations. Finally, unnecessary tests and treatments that yield no clinical benefit should be dismantled.[4]

## 2.2 Inequities in child health

The NHS has made significant advances in paediatric diagnostics and treatments in the 75 years since its inception. Despite this, compared with other high-income nations in Europe, the United Kingdom's under five-mortality rate of 4.5 deaths per 1,000 live births, ranked worse than 18 other countries in 2015.[5] The significant differences in child mortality outcomes between the UK and Sweden (whose mortality rate was 1.6 per 1000 live births) were investigated in two nationally representative cohorts. The authors reported that excess child mortality in the UK was largely attributable to differences in socioeconomic status at birth.[6] Between 2010 and 2019, infant mortality

* 2022/23 prices

rates were observed to be twice as high in children in the most deprived decile in England compared with those in the least deprived decile. Similar disparities were evident in other health-related outcomes: children from the most deprived quintile were 60% more likely to attend the emergency department more than four times in a year compared with those in the most affluent quintile, children from disadvantaged backgrounds were 72% more likely to be diagnosed with chronic illness (including 2.2 times more likely to be diagnosed with asthma), and the rates of obesity and severe obesity had increased disproportionately in the most deprived quintile.[7,8]

Addressing the social determinants of child health serves a central role in reducing inequities in child health outcomes. The social determinants of child health were illustrated and adapted by Pearce and colleagues from Bronfenbrenner's Ecological Systems Theory of Child Development and Social Model of Health by Dahlgren and Whitehead. The model demonstrates the social determinants of health in expanding concentric circles centred around the child's health.[9] The innermost layer relates to the child's health behaviours, lifestyle factors, and the nature of their parent-child relationships. Next, lies the influence of the parents and carers health behaviours and lifestyle factors. Beyond this, influences include household resources (including employment and income, housing quality, family structure and relationships, qualifications), community (crime, local services, social connections, and walkability), and the living and working conditions of the parents (good quality health and social care, childcare, schools, housing, welfare, and transport). In the outermost layer, differences in the social, economic, political, cultural, and commercial structures that support a child significantly impact their allocation of resources and services and their access to health care.[9]

Measuring differences in health care, known as health care variation, is necessary to identify and compare underlying differences in the number of service providers and

resource allocation within a population. Measuring variation in health care utilisation and quality both serve as important ways to highlight the health inequities experienced by children.

## 2.3 Variation in paediatric care

The concept of variation was first explored in a 1938 landmark study by J Alison Glover upon observing a prevailing epidemic of "enlarged tonsils" in England.[10] He analysed regional data of schoolchildren and found large differences in tonsillectomy rates which continued to persist 70 years later.[11] This variation was attributed to the weak evidence supporting tonsillectomies.[12]

The work of Glover was further developed by John Wennberg who went on to publish the Dartmouth Atlas of Health Care in 1996.[13] Wennberg's Dartmouth Atlas of Children's Health care in Northern New England was released in 2013.[14] The authors reported health service utilisation and care delivered by local physicians and hospitals to 691 000 children represented in the All Payer Claims Databases of Maine, New Hampshire, and Vermont from 2007 to 2010. Seven domains of health care were measured and reported: the physician workforce, ambulatory care, effective care (see 2.4), hospitalization, common surgical procedures, diagnostic imaging, and outpatient pharmacy prescription fills.[14]

The NHS published its own Atlas of Variation in 2010 along with a series of atlases, including a special Child Health edition updated in 2016.[15] It reported variation for many indicators including; four-fold variation in rates of emergency admissions to hospital of babies within 14 days of birth, five-fold variation in the rates of children who did not receive the full course of measles-mumps-rubella (MMR) vaccination by the age of two

years, and a five-fold variation in the rates of young people with diabetes whose median glycated hemoglobin (HbA1c) was below 7.5% (glycaemic target).[15]

## 2.4 Warranted and unwarranted variation

Wennberg proposed three categories of health care delivery:

1) Effective (or high value) care which is supported by strong evidence, e.g., the MMR vaccination in reducing prevalence of measles, mumps, and rubella[16,17]
2) Preference-sensitive care which is determined by the patient or clinician
3) Supply-sensitive care which depends on the availability of resources, which includes doctors, hospital beds, equipment, radiology, or pathology[17]

Differences in health care demand according to demographic and disease prevalence will result in variation that is expected, called **warranted variation**.[16] Variation is also considered to be warranted when it results from informed patient choice. In contrast, **unwarranted variation** exists when there is clinical uncertainty amongst clinicians or patients about the most effective diagnostic or treatment strategy.[16,17] Supply factors such as limited health spending and physician availability may lead to variation that is unwarranted, however, Wennberg observed that even areas with high supply did not experience better health outcomes.[18] Identifying the sources of unwarranted variation can highlight areas that require targeted efforts to mitigate this variation.

Unwarranted variation leads to both underuse and overuse of health care. The consequences of underdiagnosis and under-treatment in children is an emotive and publicised problem. Failure to diagnose serious bacterial infections and reduced uptake of childhood immunisations has potentially catastrophic consequences, including preventable deaths. However, until recently, only a few studies explored overuse in paediatrics. It is suggested that overuse occurs in the context of asthma, food allergies,

gastroesophageal reflux, attention deficit hyperactivity disorder and viral illnesses including bronchiolitis.[19–21] Overuse of tests and treatments can result in harms, which include physical adverse effects (i.e., radiation from imaging), psychological effects, as well as the financial costs incurred.

## 2.5 Measuring variation

Variation in health care can be measured using a variety of means ranging from more crude measures to more robust.[16] These include:

- Range: the difference between the highest and lowest value, though this can be skewed by extreme values and does not consider the distribution of values.
- Interquartile range: the difference between the third and first quartile, but only accounts for these two values rather than the measure of central tendency
- Extremal quotient: ratio of the highest rate to the lowest rate which has the same limitations as the range and interquartile range.
- Standard deviation: measures the spread of the rates relative to the mean, and accounts for all observations within a given dataset however is less helpful when comparing multiple datasets as the there is no standardised unit of measurement
- Coefficient of variation: a ratio of standard deviation to the mean which can be used to compare variation of multiple datasets as there is a standardised unit, however, it does not adjust for variation within areas (or random variation).
- Systematic component of variation: a comparison of the observed value for any given indicator to the expected value, given the age and gender distribution of the population. This recognises variability both across and within areas.

Gray et al recently proposed a method of measuring variation that includes comparing values relative to the national average – defining unwarranted variation as: (a) two of

any three consecutive values greater than two standard deviations above the mean, (b) four of any five consecutive values greater than one standard deviation above the mean, and (c) eight consecutive values above the mean.[18]

## 2.6 Proposed drivers of variation

There are many proposed causes of medical variation. The 2011 Kings Fund report[16] on health care variation proposed that variation is a result of the following inter-related factors:

- Supply:
    - Clinical decisions (which are influenced by prevailing customs and clinical guidelines)
    - Government policy
    - Resource availability (beds, specialists, funding, costs, waiting lists)
    - Service configuration
    - Private provision
- Demand:
    - Clinician decisions
    - Patient decisions
    - Morbidity:
        - Determinants of illness, which include:
            - Patient attitudes to risks
            - Lifestyle of population
            - Socioeconomic status
            - Patient demographics (age, gender, ethnicity)
    - Commissioning priorities (local service gaps and areas of need)
- Data recording issues (if using routinely collected data)
- Random variation

To differentiate between variation that is warranted and meets the needs of the underlying population, and unwarranted variation, there should be adjustments for differences in the patient demographics, socio-economic status, and other potential predictors of disease. For example, urinary tract infections have established gender and age differences.[22] Other conditions are more prevalent among specific ethnic groups, such as sickle cell disease in Black children[23], nutritional rickets in children of South Asian, African, and Middle Eastern origin[24], and cystic fibrosis in White children[25]. The prevalence of chronic illness also varies according to the socioeconomic deprivation as described in Section 2.2, and the association between illness and deprivation level may be positive or negative. Rates of coeliac disease and polymyalgia rheumatica were found to be higher in the least deprived population. Graves' disease, pernicious anaemia, rheumatoid arthritis, and systemic lupus erythematosus were associated with increasing prevalence in more deprived populations. Hashimoto's thyroiditis and inflammatory bowel disease was found to have no association with socioeconomic deprivation.[26] The effects of these variables need to be explored before adjusting for differences in variation.

It is difficult to draw comparisons about variation across different health systems as underlying differences in the structures of health system influence the degree of variation and overall health outcomes. One study examined differences in tonsillectomy rates by health care systems internationally and reported significantly higher procedure rates in systems with private care provision compared to state-provided care.[27]

The extent to which differences in health systems and primary care drive clinical variation in paediatrics is unclear.[28] Barbara Starfield was a preeminent paediatrician whose body of research advocated for strengthening primary care to improve health outcomes. One of her major studies demonstrated decreasing mortality rates with a

greater supply of primary care providers, and no relationship between the number of specialist care providers and mortality.[29] Primary care structures vary by health systems internationally: in the United States, general paediatricians constitute the primary care workforce along with family physicians and general practitioners (GPs). In the UK, general paediatricians are specialists who receive referrals from GPs.[30] These structural differences would need to be accounted for when making international comparisons.

## 2.7 Children in primary care

In the UK, primary care accounts for approximately 90% of services provided by the NHS.[31] The NHS operates a 'cradle-to-grave' model, and GPs serve as the first point of care for children who are acutely ill. They also support chronic disease management and provide preventive care such as immunisations. The GP is a gatekeeper who refers children requiring specialist care. Analysis of general practice data from 1994 to 2009 showed that children below the age of 15 comprised approximately 11% of a GP's consultation workload.[31] A retrospective analysis of general practice health records showed that between 2007 and 2017, two-thirds of health care contacts by children under 15 occurred in primary care.[32] Further investigation revealed that the children from the most deprived regions (lowest index of multiple deprivation score) had fewer GP consultations than those from the less deprived group but had more emergency department attendances and admissions. Children from black or Asian ethnic groups had more GP consultations than those of white ethnic origin (with no differences in outpatient, emergency, or inpatient admissions between the groups), however, it is unclear whether this was proportionate to their health care needs.[33]

## 2.8 Common paediatric presentations in primary care

Studies have quantified primary care workload of children, however, to the best of my knowledge there are no recent studies that quantify the most common paediatric presentations in UK primary care. Other studies that reported consultation rates by presentation either grouped the presentations into broad categories that were less clinically meaningful (e.g., diseases of respiratory system) compared with specific conditions (e.g., asthma)[34], or the consultation visit data was more than fifteen years old.[35–37] Therefore, I examined clinical guidelines to provide a guide on the most frequently observed paediatric presentations in general practice.

Clinical guidelines available to GPs are likely to represent the conditions that 1) are commonly encountered in primary care and 2) entail a degree of diagnostic and management uncertainty. GPs in the UK commonly refer to the National Institute of Health and Care Excellence (NICE) clinical practice guidelines [38] to inform clinical decisions. I sought to determine the proportion of NICE guidelines that are related to children in primary care and identify the conditions that most commonly present to primary care in the UK. To do this, I searched the NICE guideline repository for guidelines related to children and young people. I designed a webscraper [39] using Python, which extracted all files related to each guideline and downloaded the pdfs to a shared encrypted drive. Two reviewers (ET, ST) conducted title screening, and included guidelines that were: 1) published or updated since 2011 2) focused on children 3) focused on conditions relevant to primary care and 4) included recommendations about diagnostic testing. Discrepancies were resolved by discussion or with a practising GP where required (JT).

A feasibility count was undertaken by SS using data from the Clinical Practice Research Datalink (CPRD) GOLD database to determine which of these conditions

were encountered most frequently in UK primary care. The incidence and prevalence of these conditions was extracted for the period from January 1 - December 31 in 2019 and are shown below. The four childhood conditions with the highest incidence of primary care presentations in 2019 were fever, gastroenteritis, constipation, and asthma. Although urinary tract infection was ranked third, there was substantial overlap in the guideline recommendations for fever in under 5s because in children and young people the primary presentation is likely to be fever, therefore, these were grouped together.

**Table 2.1 Rank order of the most common presentations in primary care with corresponding NICE guidelines for children aged less than 16 years.**

| | Incidence (1st Jan – 31st Dec 2019) | | Prevalence (31st Dec 2019) | |
|---|---|---|---|---|
| | n= 1,741,542 | | | |
| **Conditions** | **n** | **Per 1000** | **n** | **Per 1000** |
| Fever | 13343 | 7.662 | 366487 | 21.044 |
| Diarrhoea and vomiting (gastroenteritis) | 7117 | 4.087 | 333124 | 19.128 |
| Urinary tract infection | 6919 | 3.973 | 206756 | 11.872 |
| Constipation | 5720 | 3.284 | 112530 | 6.462 |
| Asthma | 5005 | 2.874 | 50358 | 2.892 |
| Headaches | 3366 | 1.933 | 69608 | 3.997 |
| Autism spectrum disorder | 2596 | 1.491 | 21824 | 1.253 |
| Drug allergy | 2541 | 1.459 | 55176 | 3.168 |
| Motor development delay (not otherwise attributed to cerebral palsy), Regression, Unsteadiness | 1859 | 1.067 | 92653 | 5.32 |
| Bronchiolitis | 1595 | 0.916 | 29832 | 1.713 |
| Bedwetting (Nocturnal enuresis) | 1430 | 0.821 | 10505 | 0.603 |
| Dizziness/vertigo | 1276 | 0.733 | 61908 | 3.555 |
| Gastro-oesophageal reflux disease | 1023 | 0.587 | 22781 | 1.308 |
| Jaundice in newborn babies | 858 | 0.493 | 9262 | 0.532 |
| Diabetes (type 1 and type 2) | 781 | 0.448 | 5137 | 0.295 |
| Blackouts/vacant spells/loss of consciousness/syncope | 605 | 0.347 | 17886 | 1.027 |
| Faltering growth (failure to thrive) | 572 | 0.328 | 1155 | 0.066 |
| Epilepsies | 484 | 0.278 | 6941 | 0.399 |
| Tremor (not otherwise attributed to cerebral palsy) | 220 | 0.126 | 2365 | 0.136 |
| Thyroid disease | 176 | 0.101 | 671 | 0.039 |
| Coeliac disease | 143 | 0.082 | 2497 | 0.143 |
| Lyme disease | 121 | 0.069 | 1078 | 0.062 |
| Tuberculosis | 110 | 0.063 | 2497 | 0.143 |
| Tinnitus | 66 | 0.038 | 176 | 0.01 |
| Non-alcoholic fatty liver disease (NAFLD) | 55 | 0.032 | 110 | 0.006 |
| Urinary incontinence in Spina bifida | 44 | 0.025 | 231 | 0.013 |
| Confusion | 33 | 0.019 | 341 | 0.02 |
| Acute kidney injury | 33 | 0.019 | 2233 | 0.128 |
| Cerebral palsy | 22 | 0.013 | 506 | 0.029 |
| Alcohol-use disorders | 22 | 0.013 | 12397 | 0.712 |
| Cystic fibrosis | 11 | 0.006 | 528 | 0.03 |
| Urinary incontinence in Spinal agenesis | 0 | 0 | 0 | 0 |
| Endometriosis | 0 | 0 | 88 | 0.005 |
| Hepatitis B (chronic) | 0 | 0 | 66 | 0.004 |
| Hepatitis C | 0 | 0 | 44 | 0.003 |
| Renal and ureteric stones | 0 | 0 | 132 | 0.008 |
| Suspected cancer | 0 | 0 | 11 | 0.001 |

## 2.9 Challenges of diagnosing children in general practice

Children present specific diagnostic challenges in the primary care setting. GPs are trained to detect and manage condition in people of all ages, but even within the paediatric population, there are physiological differences within each developmental age group. Infants, pre-schoolers, school-aged children, and adolescents have different specific communication needs and styles that require the general practitioner to tailor their approach to interpret their symptoms and signs. Children tend to present to general practice with undifferentiated illnesses which takes time to clinically manifest into recognised disease. For some common paediatric presentations, no underlying cause is identified, such as in the case of functional abdominal pain. Additionally, children have less physiological reserve and tend to decompensate much faster compared with adults. Compared with their hospital-based colleagues, GPs have less access to advanced diagnostics to help make a diagnosis. Finally, GPs serve as lynchpins in the health care system, given their gatekeeper role for patients needing to access specialist care. Therefore, it is especially important to study children in general practice.

## 2.10 Diagnostic strategies for children in primary care

The diagnostic strategies used by general practitioners were compared with those used by emergency specialists in a comparative qualitative analysis of 450 consultations of 12 GPs and 16 emergency physicians in Germany. Both groups of professionals see an unselected population and serve as gatekeepers for specialist care.[40] Compared with emergency doctors, GPs saw their role as more patient-centred with an emphasis on continuity of care, caring for a range of acute and chronic issues that are biomedical and psychosocial in nature. Emergency physicians on the other hand focused on acute, life-critical disease and patient flow, where the outcome of

decision was binary: to admit or discharge. Clinicians provided patient reassurance (information about the benign nature of symptoms and alleviating concerns and fears) in 18% of Emergency Department (ED) consultations compared with 64% of GP encounters. Probabilistic reasoning (consideration of the statistical likelihood of a particular diagnosis) was used in 30% of GP cases and 4% of ED cases, and self-labelling (where the patient or family member suggests the diagnosis) was far more common in general practice (64% vs 7% in ED). The authors observed that patients played a more active role in the GP consultation compared with ED where the physician tended to direct the consultation. Finally, the underlying assumptions with which both professional groups approached patients were starkly different – in ED, life-threatening disease was assumed and needed to be excluded, whereas in general practice, the frame of reference was 'normal', GPs looked for deviation from patient's usual health status, and serious disease was less common.[40]

A previous study by Heneghan and colleagues[41], observed and documented the different diagnostic strategies used by GPs in primary care, which they divided into three stages: 1) initiation of the diagnosis, 2) refinement of the diagnosis and 3) defining the final diagnosis. In each of these stages, different diagnostic strategies can be applied to children in primary care.

1. Initiation of the diagnosis:
    a. Spot diagnoses (characteristic rash and fever: chicken pox)
    b. Self-labelling (simple acne)
    c. Presenting complaint (abdominal pain)
    d. Pattern recognition trigger (thirst, weight loss, fatigue in an adolescent with Type 1 diabetes mellitus)
2. Refinement of the diagnosis:
    a. Restricted rule out (ruling out 'red flag' causes of headache, such as subarachnoid haemorrhage)

b. Stepwise refinement (determining whether conjunctivitis is infectious or allergic in nature)
c. Probabilistic reasoning (understanding how a positive or negative result alters the probability of disease, e.g., a positive urine dipstick for UTI)
d. Pattern recognition fit (Henoch-Schönlein purpura diagnostic triad of purpuric rash, abdominal pain, and haematuria)
e. Clinical prediction rule (CENTOR score for Group A streptococcal pharyngitis)

3. Defining the final diagnosis:
    a. Known diagnosis (having certainty of the diagnosis to make a management plan, e.g., a viral upper respiratory tract infection)
    b. Test of treatment (a proton-pump inhibitor trial for gastro-oesophageal reflux disease or trial of salbutamol treatment for suspected asthma)
    c. Test of time (a wait-and-watch approach with suspected viral gastroenteritis when a child presents with diarrhoea and vomiting and fever)
    d. No label applied: where a clear diagnosis is not present, such as the case with children who present with abdominal pain who have with no red flag features or have a recognisable pattern of symptoms.
    e. Ordering further tests

## 2.11 Diagnostic tests for children in primary care

Diagnostic tests can play a helpful role in clinical decision making for children in primary care, especially where diagnosis cannot be made based on history and examination alone. They can be used to rule in or rule out a disease, refine the diagnosis, guide management, or for monitoring disease progression. To achieve this, tests should be accurate and maximise patient benefit by ensuring timely diagnosis

and intervention. Tests can also benefit the broader population for research and disease surveillance purposes.

85% of clinical pathways in the NHS include diagnostic activity.[42] The NHS undertakes more than 1.5 billion diagnostic tests, which comprise approximately 6% of the total NHS budget.[42] Increasing demand for diagnostic services generates more workload for a healthcare workforce that is already under strain. Testing for adults in primary care increased by 9% from 2000 to 2016. Approximately one quarter of tests were thought to be unnecessary, highlighting huge potential waste in spending and resources.[43] The volume of paediatric testing and its contribution to health care waste thus far is unknown.

## 2.12 Gaps in the literature and implications for future research

Previous studies have estimated the prevalence of overuse and underuse of tests in primary care, but these focussed on adults.[44,45] There is limited knowledge about the value of diagnostic care delivered to children in primary care in the UK. If variation does exist, then further research is needed to ascertain the potential causes of variation in test use, to determine if this variation is warranted or unwarranted. Identifying unwarranted variation should prompt efforts to mitigate the source of variation. Research is also needed to understand the physical, psychological, financial, and environmental consequences of variation in test utilisation.

## 2.13 Chapter summary

- The National Health Service needs to ensure its sustainability, while balancing limited funds and resources, rising costs, and increasing population needs and demands for services.

- Despite the advances made by the NHS in paediatric diagnostics and treatments in the past 75 years, UK child mortality rates in the UK fared worse than 18 other high-income countries in Europe. This was attributed to socioeconomic inequities that resulted in worse health outcomes for children from the most deprived population compared with those in the least.
- Measuring variation in health care use and quality both serve as important ways to highlight disparities in child health outcomes.
- In the UK, most children present to a general practitioner for their health needs.
- Diagnostic tests play an important role for children in primary care and they can reduce diagnostic uncertainty if used appropriately.
- Consequently, general practice is an ideal setting in which to study variation in diagnostic services in children as it is an area where clinicians experience high levels of diagnostic uncertainty.
- Previous studies have explored variation in test use within adults, however, to date none have sought to quantify and explore variation in test use for children. This thesis aims to address these research gaps.

## 2.14 References


1. The King's Fund. The NHS budget and how it has changed. *Kings Fund*. 2023; Accessed September 10, 2023. https://www.kingsfund.org.uk/projects/nhs-in-a-nutshell/nhs-budget.
2. The King's Fund. How does the NHS compare to the health care systems of other countries? *Kings Fund.* 2023; Accessed July 20, 2023. https://www.kingsfund.org.uk/publications/nhs-compare-health-care-systems-other-countries#why-we-did-the-research.
3. The King's Fund. *The Rise and Decline of the NHS in England 2000–20*.; 2021. Accessed September 10, 2023. https://www.kingsfund.org.uk/sites/default/files/2023-04/Rise_and_Decline_of_the_NHS_April_2023.pdf.
4. Saini V, Brownlee S, Elshaug AG, Glasziou P, Heath I. Addressing overuse and underuse around the world. *The Lancet*. 2017;390(10090). doi:10.1016/S0140-6736(16)32573-9.
5. UK drops in European child mortality rankings. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/childhealth/articles/ukdropsineuropeanchildmortalityrankings/2017-10-13.
6. Zylbersztejn A, Gilbert R, Hjern A, Wijlaars L, Hardelid P. Child mortality in England compared with Sweden: a birth cohort study. *The Lancet*. 2018;391(10134). doi:10.1016/S0140-6736(18)30670-6.
7. Lee AR, Kingdon CC, Davie M, Hawcutt D, Sinha IP. Child poverty and health inequalities in the UK: A guide for paediatricians. *Arch Dis Child*. 2022;108(2). doi:10.1136/archdischild-2021-323671.
8. Spencer NJ, Blackburn CM, Read JM. Disabling chronic conditions in childhood and socioeconomic disadvantage: A systematic review and meta-analyses of observational studies. *BMJ Open*. 2015;5(9). doi:10.1136/bmjopen-2014-007062.
9. Pearce A, Dundas R, Whitehead M, Taylor-Robinson D. Pathways to inequalities in child health. *Arch Dis Child*. 2019;104(10). doi:10.1136/archdischild-2018-314808.
10. Glover JA. The Incidence of Tonsillectomy in School Children. *Proc R Soc Med*. 1938;31(10). doi:10.1177/003591573803101027.
11. Suleman M, Clark MPA, Goldacre M, Burton M. Exploring the variation in paediatric tonsillectomy rates between English regions: A 5-year NHS and independent sector data analysis. *Clinical Otolaryngology*. 2010;35(2). doi:10.1111/j.1749-4486.2010.02086.x.

12. Burton MJ, Glasziou PP, Chong LY, Venekamp RP. Tonsillectomy or adenotonsillectomy versus non-surgical treatment for chronic/recurrent acute tonsillitis. *Cochrane Database of Systematic Reviews*. 2014;2014(11). doi:10.1002/14651858.CD001802.pub3.
13. Wennberg JE, Cooper MM, Bubolz TA, et al. The Dartmouth Atlas of Health Care in the United States. 1996. The Center for the Evaluative Clinical Sciences: Dartmouth Medical School. Published 1996. Accessed October 26, 2020. https://www.dartmouthatlas.org/downloads/atlases/96Atlas.pdf.
14. Goodman D, Morden NE, Ralston SL, Chang C, Weinstein J, et al. The Dartmouth Atlas of Children's Health Care in Northern New England. The Dartmouth Institute. Published 2013. Accessed October 23, 2020. https://www.dartmouthatlas.org/downloads/atlases/NNE_Pediatric_Atlas_121113.pdf.
15. NHS RightCare. *NHS Atlas of Variation in Healthcare for Children and Young People*; 2012.
16. Appleby J, Raleigh V, Frosini F, Bevan G, Gao H, Lyscom T. Variations in Health Care. The good, the bad and the inexplicable. *Variations in health*. Published online 2011.
17. Wennberg JE. Unwarranted variations in healthcare delivery: Implications for academic medical centres. *Br Med J*. 2002;325(7370). doi:10.1136/bmj.325.7370.961.
18. Gray WK, Day J, Briggs TWR, Harrison S. Identifying unwarranted variation in clinical practice between healthcare providers in England: Analysis of administrative data over time for the Getting It Right First Time programme. *J Eval Clin Pract*. 2021;27(4). doi:10.1111/jep.13477.
19. Coon ER, Quinonez RA, Moyer VA, Schroeder AR. Overdiagnosis: How our compulsion for diagnosis may be harming children. *Pediatrics*. 2014;134(5). doi:10.1542/peds.2014-1778.
20. Coon ER, Young PC, Quinonez RA, Morgan DJ, Dhruva SS, Schroeder AR. Update on pediatric overuse. *Pediatrics*. 2017;139(2). doi:10.1542/peds.2016-2797.
21. Money NM, Schroeder AR, Quinonez RA, et al. 2021 Update on Pediatric Overuse. *Pediatrics*. 2022;149(2). doi:10.1542/peds.2021-053384.
22. Jamshidbeigi T, Adibi A, Hashemipour SMA, et al. A systematic review and meta-analysis of prevalence of urinary tract infection in childhood. *J Renal Inj Prev*. Published online 2023. doi:10.34172/jrip.2023.32160.

23. Pokhrel A, Olayemi A, Ogbonda S, Nair K, Wang JC. Racial and ethnic differences in sickle cell disease within the United States: From demographics to outcomes. *Eur J Haematol*. 2023;110(5). doi:10.1111/ejh.13936.
24. Thacher TD, Pludowski P, Shaw NJ, Mughal MZ, Munns CF, Högler W. Nutritional rickets in immigrant and refugee children. *Public Health Rev*. 2016;37(1). doi:10.1186/s40985-016-0018-3.
25. McCormick J, Green MW, Mehta G, Culross F, Mehta A. Demographics of the UK cystic fibrosis population: Implications for neonatal screening. *European Journal of Human Genetics*. 2002;10(10). doi:10.1038/sj.ejhg.5200850.
26. Conrad N, Misra S, Verbakel JY, et al. Incidence, prevalence, and co-occurrence of autoimmune disorders over time and by age, sex, and socioeconomic status: a population-based cohort study of 22 million individuals in the UK. *The Lancet*. 2023;401(10391). doi:10.1016/S0140-6736(23)00457-9.
27. Crowson MG, Ryan MA, Rocke DJ, Raynor EM, Puscas L. Variation in tonsillectomy rates by health care system type. *Int J Pediatr Otorhinolaryngol*. 2017;94. doi:10.1016/j.ijporl.2017.01.014.
28. Shi L, Macinko J, Starfield B, Wulu J, Regan J, Politzer R. The relationship between primary care, income inequality, and mortality in US States, 1980-1995. *Journal of the American Board of Family Practice*. 2003;16(5). doi:10.3122/jabfm.16.5.412.
29. Starfield B, Shi L, Grover A, Macinko J. The effects of specialist supply on populations' health: assessing the evidence. *Health Aff (Millwood)*. 2005;Suppl Web Exclusives. doi:10.1377/hlthaff.w5.97.
30. Forrest CB, Majeed A, Weiner JP, Carroll K, Bindman AB. Referral of children to specialists in the United States and the United Kingdom. *Arch Pediatr Adolesc Med*. 2003;157(3). doi:10.1001/archpedi.157.3.279.
31. Hobbs FDR, Bankhead C, Mukhtar T, et al. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007–14. *The Lancet*. 2016;387(10035). doi:10.1016/S0140-6736(16)00620-6.
32. Ruzangi J, Blair M, Cecil E, et al. Trends in healthcare use in children aged less than 15 years: A population-based cohort study in England from 2007 to 2017. *BMJ Open*. 2020;10(5). doi:10.1136/bmjopen-2019-033761.
33. Coughlan CH, Ruzangi J, Neale FK, et al. Social and ethnic group differences in healthcare use by children aged 0-14 years: A population-based cohort study in England from 2007 to 2017. *Arch Dis Child*. 2022;107(1). doi:10.1136/archdischild-2020-321045.

34. Royal College of General Practitioners. *United Kingdom - England and Wales Weekly Returns Service Annual Report 2011*.; 2011. Accessed November 27, 2023. https://ghdx.healthdata.org/record/united-kingdom-england-and-wales-weekly-returns-service-annual-report-2011.
35. Gill PJ, Wang KY, Mant D, et al. The evidence base for interventions delivered to children in primary care: An overview of Cochrane systematic reviews. *PLoS One*. 2011;6(8). doi:10.1371/journal.pone.0023051.
36. Otters HBM, van der Wouden JC, Schellevis FG, van Suijlekom-Smit LWA, Koes BW. Changing morbidity patterns in children in Dutch general practice: 1987-2001. *European Journal of General Practice*. 2005;11(1). doi:10.3109/13814780509178011
37. Charles J, Pan Y, Britt H. Trends in childhood illness and treatment in Australian general practice, 1971-2001. *Medical Journal of Australia*. 2004;180(5). doi:10.5694/j.1326-5377.2004.tb05888.x
38. National Institute for Health and Care Excellence. NICE guidance. https://www.nice.org.uk/guidance.
39. Thomas ET. NICE webscraper. https://github.com/elizabethtthomas/NICE-webscraper.
40. Bösner S, Abushi J, Feufel M, Donner-Banzhoff N. Diagnostic strategies in general practice and the emergency department: A comparative qualitative analysis. *BMJ Open*. 2019;9(5). doi:10.1136/bmjopen-2018-026222.
41. Heneghan C, Glasziou P, Thompson M, et al. Diagnostic strategies used in primary care. *BMJ*. 2009;338(apr20 1). doi:10.1136/bmj.b946.
42. Wickens C. *Why Do Diagnostics Matter? Maximising the Potential of Diagnostics Services*; 2022.
43. Watson J, Burrell A, Duncan P, et al. Exploration of reasons for primary care testing (the Why Test study): a UK-wide audit using the Primary care Academic CollaboraTive. *British Journal of General Practice*. Published online October 2, 2023. doi:10.3399/BJGP.2023.0191.
44. O'Sullivan JW, Albasri A, Nicholson BD, et al. Overtesting and undertesting in primary care: A systematic review and meta-analysis. *BMJ Open*. 2018;8(2). doi:10.1136/bmjopen-2017-018557.
45. Müskens JLJM, Kool RB, van Dulmen SA, Westert GP. Overuse of diagnostic testing in healthcare: A systematic review. *BMJ Qual Saf*. 2022;31(1). doi:10.1136/bmjqs-2020-012576.

# Chapter 3. Trends in diagnostic tests ordered for children: a retrospective analysis of 1.7 million laboratory test requests in Oxfordshire, UK from 2005 to 2019.

## 3.1 Publication statement

This chapter is based on my publication in *Archives of Disease in Childhood* (**Appendix 1.1**).

## 3.2 Introduction

Diagnostic testing plays an important role in the provision of health care. In England, laboratory and pathology services (including biochemistry, haematology, microbiology, histopathology and cytology tests) were estimated to cost 2·5 billion pounds annually, comprising 3-4% of the NHS budget according to a review of NHS Pathology Services in 2006.[1] Similar data from the United States (US) suggests that expenditure on diagnostic testing was 82·7 billion dollars in 2017 or 2% of total US healthcare spending.[2] Between 2004 and 2006 in the UK, demand for pathology services increased at an annual rate of 10%.[1]

Substantial variation has been demonstrated in diagnostic test use across primary and secondary care in the UK.[3–5] However, most of the reported diagnostic measures focused on adults instead of children, along with the broader extant literature on diagnostic tests.[5,6] There is a paucity of comprehensive data that explores laboratory testing in children.

## 3.3 Aims and objectives

The overall aim of the research reported in this chapter was to examine diagnostic testing practices for children in the local Oxfordshire region.

The specific objectives were:

1. To quantify the tests that are requested in primary care compared with secondary care for children.
2. To determine how laboratory testing for children has changed from 2005 to 2019.
3. To compare how testing rates differ by age, sex, and health care setting.
4. To determine the most frequently performed laboratory tests for children, and how these vary by age and setting.
5. To explore how specific test use has changed from 2005 to 2019.

## 3.4 Methods

### 3.4.1 Study design and data sources

I performed a retrospective observational study of laboratory test data.

### 3.4.2 Setting

I obtained laboratory testing data from Oxford University Hospitals and Oxfordshire General Practices from 1st January 2005 to 31st December 2019. This was selected as the study end date to eliminate the impacts of the COVID-19 pandemic. The laboratory is the sole referral centre for 67 general practices and four hospitals, making up over 95% of the tests carried out in the county. In this study, tests requested in the emergency department and inpatient wards were grouped by the laboratory as inpatient tests and could not be differentiated.

In Oxfordshire, children can be brought to the emergency department at the Oxford Children's Hospital by their parents or referred by their GP. From the emergency

department, children can be admitted to the inpatient ward at the Oxford Children's Hospital or the Horton General Hospital in Banbury. Paediatric specialists look after children in the inpatient wards. GPs, emergency doctors, and ward paediatric teams can also refer children to outpatient clinics at both hospitals to be reviewed by paediatric specialists.

Data from all laboratory tests conducted among children aged 0 to 15 were included. I used an age threshold of 16 for our study population as the Oxford Children's Hospital Emergency Department cares for children under the age of 16. Most adult services see children aged 16 or older, with children starting to be transitioned out of paediatric services at the age of 16. Point-of-care tests such as blood gas and glucose tests were excluded as, in practice, these can be performed at the bedside and are not consistently sent to the laboratory for analysis.

### 3.4.3 Variables and data sources

The Oxford University Hospitals (OUH) Trust database contains previously collected laboratory test data. BS extracted non-identifiable data, including the name of the test, indication for the test, patient sex and age, and whether the test was ordered in primary or secondary care. I assigned each test code (e.g., "haemoglobin") to a panel (e.g., "full blood count"), and these were verified by a clinician (BS).

### 3.4.4 Statistical analysis

I calculated the proportion of tests requested in general practice, inpatient and outpatient (hospital paediatric clinic) settings each year. I estimated crude and age-standardised test rates per 1,000 child years using the 2019 Office for National Statistics population estimates for Oxfordshire as the standard. To calculate crude rates, the numerator was the number of tests, and the denominator was mid-year

population estimates for the population aged 0 to 15 in Oxfordshire. Annual mid-year population estimates for Oxfordshire were obtained from the Office for National Statistics. To calculate age-standardised rates, the 2019 population was used as the reference standard as this was deemed to be most relevant to the current population.

Testing rates were stratified by sex and age; under 1 year (infants); 1-5 years (early childhood); 6-10 years (middle childhood), and 11-15 years (adolescence) (12). These age groups were chosen upon an initial exploratory analysis of the dataset. The crude totals of tests differed greatly for those aged 0-1 year compared to other age groups. The test numbers between those aged 1-5 years were deemed sufficiently similar to be grouped. This was also the case with children aged 6-10 and 11-15 years.

I used Joinpoint regression[8] to model temporal changes in age-adjusted rates from 2005 to 2019, similar to prior studies that have analysed trends over time.[6,9] Joinpoint regression assumes that calendar time can be sub-divided into subsets with distinct linear trends. Points where significant changes in rates occurred (joinpoints) were identified and annual percentage changes (APC) between joinpoints were estimated. I also estimated the average annual percentage change (AAPC), a summary measure of the trend from 2005 to 2019, stratified by setting, sex, and age. APCs and AAPCs were estimated for the 25 most frequently requested tests. APCs and AAPCs were modelled in Joinpoint software version 4.9.1.0[8], and all other statistical analyses were performed using R version 4.1.3.[10]

### 3.4.5 Ethics Approval

Ethics approval was not required for this study as we used retrospective de-identified data. A data protection impact assessment was conducted and approved by the Oxford University Hospitals NHS Foundation Trust.

### 3.4.6 Data availability

The study protocol was preregistered on the Open Science Framework (https://doi.org/10.17605/OSF.IO/KE6DM). All test codes, R code used for data management, analysis and creating the figures are available on GitHub (https://github.com/elizabethtthomas/paediatric-testing-oxfordshire).

## 3.5 Results

### 3.5.1 Characteristics of included participants

There were 1,749,425 tests performed on 113,607 children from 1 January 2005 to 31 December 2019, of which 46% (52,207 of 113,607) were females. 71% of tests (1,232,556 of 1,749,425) occurred in the inpatient setting, 17% in general practice, and 13% in the outpatient setting. Children had a median of five tests each (IQR 3 to 8). The median number of tests per child for each setting is shown in **Table 3.1**. 71% of tests (1,232,556 of 1,749,425) occurred in the inpatient setting, 17% in general practice, and 13% in the outpatient setting. Of the children who had at least one test, each child had a median of five blood tests (IQR 3 to 8). One-third of tests (33%, n= 580,636) were performed in the under-1 age group, of which most (96%, n=558,716) were performed in the inpatient setting (**Table 3.2**).

| Table 3.1 Characteristics of included patients and tests | | |
|---|---|---|
| | **Number of children** | **%** |
| Females | 52,207 | 46·0 |
| Males | 61,400 | 54·0 |
| Total | 113,607 | 100·0 |
| | **Number of tests** | **%** |
| Total | 1,749,425 | 100·0 |
| **Age group** | | |
| <1 year | 580,636 | 33.2 |
| 1-5 years | 439,770 | 25·1 |
| 6-10 years | 319,387 | 18·3 |
| 11-15 years | 409,632 | 23·4 |
| **Setting** | | |
| General Practice | 293,506 | 16·8 |
| Inpatient | 1,232,556 | 70·5 |
| Outpatient | 223,363 | 12·8 |
| | **Median no. of tests per child*** | **Interquartile Range** |
| Total | 5 | 3, 8 |
| General Practice | 5 | 3, 8 |
| Inpatient | 5 | 3, 7 |
| Outpatient | 4 | 2, 8 |

*In children who had at least one test

| Table 3.2 Number of tests requested by age group and setting | | | |
|---|---|---|---|
| **Age group** | **Setting** | **Number of children** | **%** |
| <1 year | Total | 580,636 | 100·0 |
| | General Practice | 4,157 | 0·7 |
| | Inpatient | 558,716 | 96·2 |
| | Outpatient | 17,763 | 3·1 |
| 1-5 years | Total | 439,770 | 100·0 |
| | General Practice | 49,123 | 11·2 |
| | Inpatient | 315,321 | 71·7 |
| | Outpatient | 75,326 | 17·1 |
| 6-10 years | Total | 319,387 | 100·0 |
| | General Practice | 85,081 | 26·6 |
| | Inpatient | 170,158 | 53·3 |
| | Outpatient | 64,148 | 20·1 |
| 11-15 years | Total | 409,632 | 100·0 |
| | General Practice | 155,145 | 37·9 |
| | Inpatient | 188,361 | 46·0 |
| | Outpatient | 66,126 | 16·1 |

### 3.5.2 Temporal change in test use

The age-adjusted rate of total test use increased from 878 tests per 1,000 child years in 2005 to 1,107 tests per 1,000 child years in 2019, though this change was not statistically significant (**Figure 3.1a** AAPC 1·5% [95% CI -0·8 to 3·9%, p=0·2], **Chapter 3 Appendix Table 3.2**). Test rates initially decreased by 2·2% per year between 2005 and 2012 (95% CI -0·6 to -3·8%, p=0·01, **Figure 3.1**). From 2012 to 2015, the APC was 9·0% per year (95% CI -3·0 to 22·5%) and then changed to 2·8% per year (95% CI -0·5 to 6·3%) between 2015 and 2019.

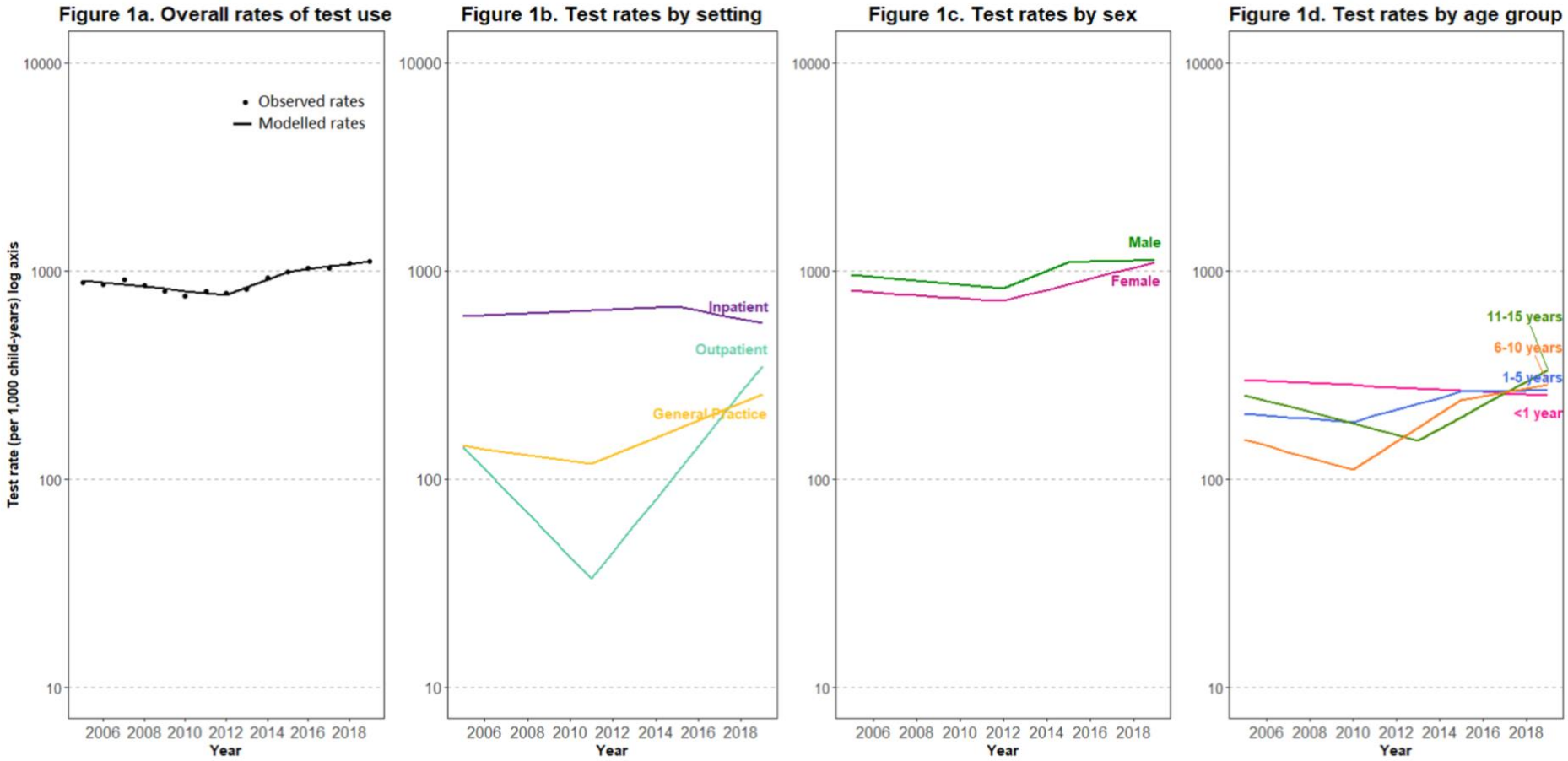
**Figure 3.1b** shows a temporal change in test use by setting. Testing rates remained steady in the inpatient setting (AAPC -0.6%, 95% CI –2.1 to 0.9%, p=0.4). In general practice, test use was stable until 2011 (AAPC -3.3%, 95% CI -8·3 to 1·9%, p=0·2), and then increased by 10·1% per year (95% CI 6·5 to 14·0%, p<0·001). Testing in

outpatients decreased by 21.6% per year until 2011 (95% CI -28·3 to -14·2%, p<0.001) and then sharply increased by 34.1% per year (95% CI 26·6 to 42·2%, p<0·001).

**Figure 3.1c** illustrates test use by sex. Testing rates for males and females followed similar trends until 2015 when test use in males stabilized (AAPC 1.0%, 95% CI -2·6 to 4·7%, p=0·5), whereas testing in females continued to rise by 6·4% per year from 2012 (95% CI 4·7 to 8·0%, p<0·001).

The rates of test use by age group are presented in **Figure 3.1d**. Test use declined overall in children under 1 year (AAPC -1.2%, 95% CI -2.2 to -0.2%, p=0.02). Testing in all other age groups increased after 2010; this was particularly striking for children aged 6-10 years (AAPC 4.4%, 95% CI 2.6 to 6.3%, p <0.001) and children aged 11-15 years from 2013 (APC 13.8%, 95% CI 6·9 to 21·1%, p<0·001).

**Figure 3.1 Test use among children in Oxfordshire from 2005 to 2019**




**AAPC: Average annual percentage change; APC: annual percentage change**

Figure 1a: AAPC = 1.5% (95% CI -0.8 to 3.9%, p=0.2)
APC 2005 – 2012 = -2.2% (95% CI -3.8 to -0.6%, p=0.01)
APC 2012 – 2015 = 9.0% (95% CI -3.0 to 22.5%, p=0.1)
APC 2015 – 2019 = 2.8% (95% CI -0.5 to 6.3%, p=0.1)

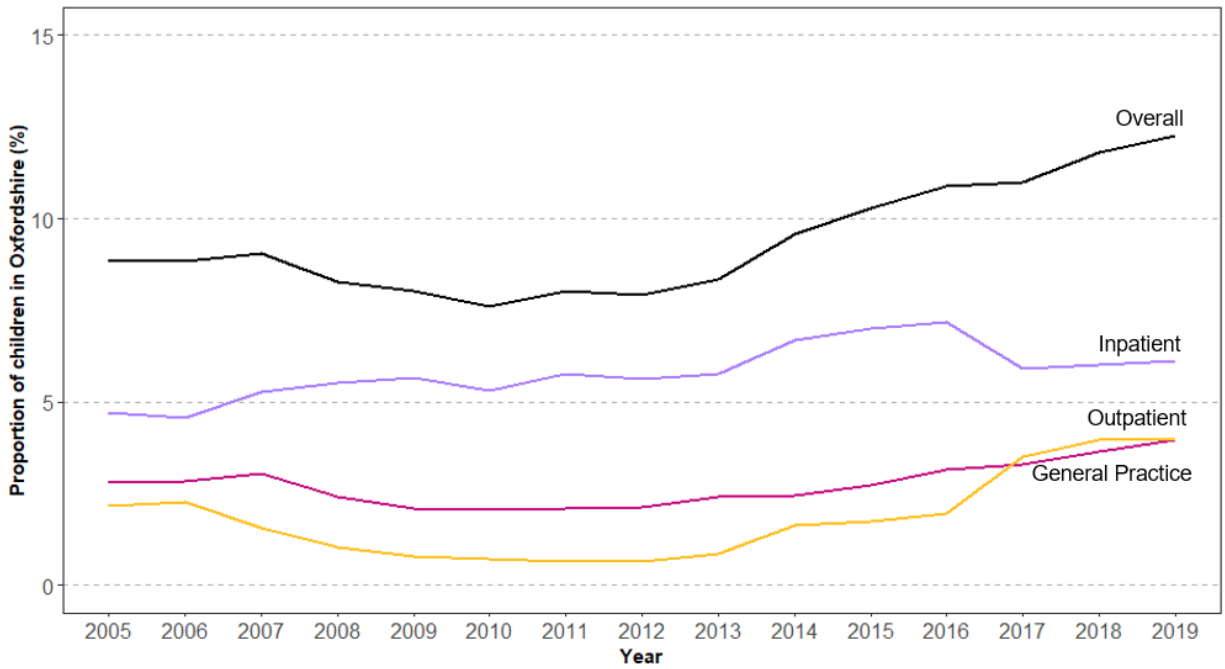
1b: AAPC by setting. General Practice = 4.2% (95% CI 1.5 to 6.9%, p= 0.002); Inpatient = -0.6% (95% CI -2.1 to 0.9%, p=0.4); Outpatient = 6.6% (95% CI 1.9 to 11.5%, p=0.005

1c: AAPC by sex. Female = 2.3% (95% CI 1.3 to 3.3%, p <0.001); Male = 1.2% (95% CI -1.3 to 3.8%, p=0.3)

1d: AAPC by age group. <1 year = -1.2% (95% CI -2.2 to -0.2%, p=0.02); 1-5 years= 1.9% (95% CI -0.8 to 4.6%, p=0.2); 6-10 years = 4.4% (95% CI 2.6 to 6.3%, p <0.001); 11-15 years = 2.0% (95% CI -1.1 to 5.2%, p =0.2)

The proportion of children in Oxfordshire receiving at least one test in any setting increased by 39% (from 8·8% to 12·3%, **Figure 3.2**, **Chapter 3 Appendix Table 3.3**). Increases were highest in the outpatient setting, where the proportion of children receiving at least one test increased by 84% (from 2·2% to 4·0%).

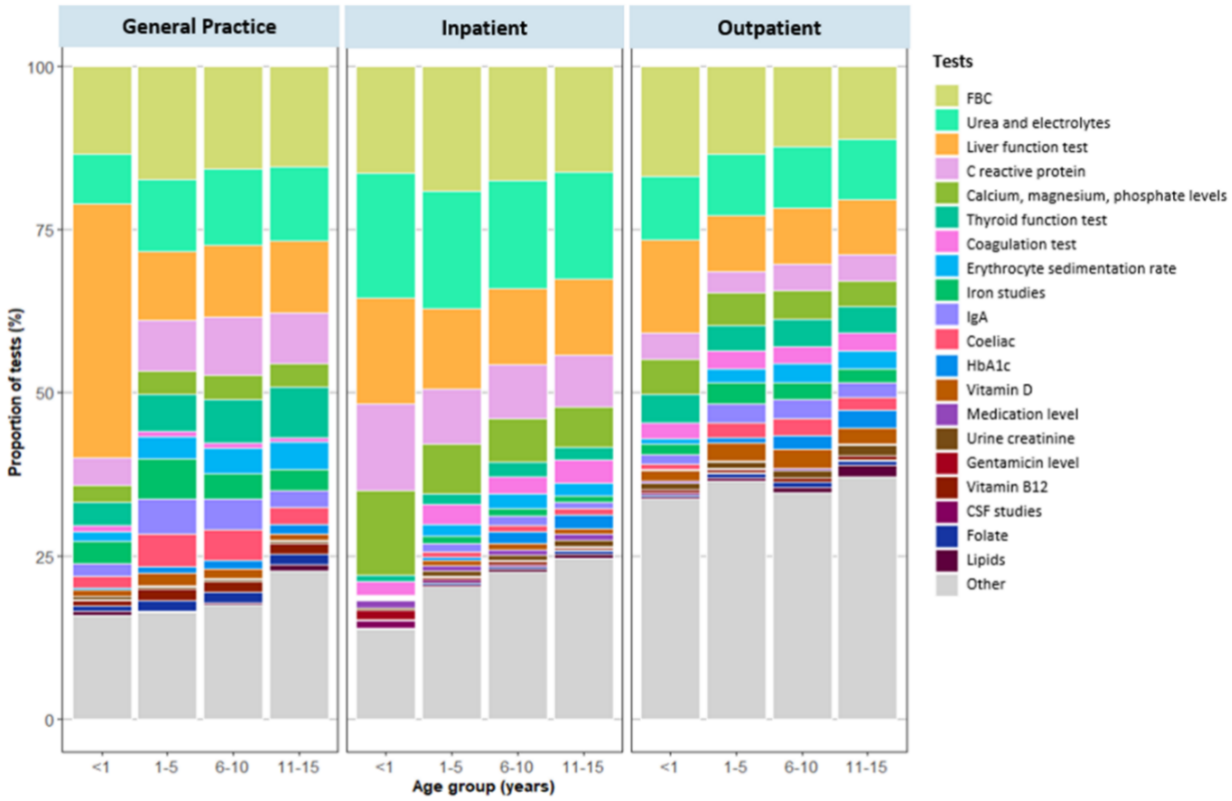
**Figure 3.2 The proportion of children in Oxfordshire who had at least 1 test from 2005 to 2019 overall, and in each healthcare setting**




### 3.5.3 Test ranking

The most frequently ordered tests are shown by setting and age group in **Figure 3.3**. The top five tests were: full blood count (FBC), urea and electrolytes, liver function tests (LFT), C-reactive protein (CRP) and calcium, magnesium, phosphate (CMP) levels. The top five tests were reasonably consistent across the age groups and settings **(Figure 3.3, Table 3.3**).

**Figure 3.3 The most frequently requested tests for children in Oxfordshire from 2005 to 2019, by setting and age**




Abbreviations: CSF studies – cerebrospinal fluid studies; FBC – full blood count

**Table 3.3 The 25 most frequently requested tests from 2005 to 2019, stratified by setting.**

| | Total | General Practice | | Inpatient | | Outpatient | |
|---|---|---|---|---|---|---|---|
| **Test** | **n** | **n** | **%** | **n** | **%** | **n** | **%** |
| Full blood count | 321,421 | 53,288 | 16·6 | 234,446 | 72·9 | 33,687 | 10·5 |
| Urea and electrolytes | 305,959 | 38,420 | 12·6 | 242,947 | 79·4 | 24,592 | 8·0 |
| Liver function tests | 250,612 | 38,183 | 15·2 | 188,658 | 75·3 | 23,771 | 9·5 |
| C reactive protein | 180,448 | 27,342 | 15·2 | 143,092 | 79·3 | 10,014 | 5·5 |
| Calcium magnesium phosphate | 154,806 | 11,842 | 7·6 | 131,150 | 84·7 | 11,814 | 7·6 |
| Thyroid function tests | 54,322 | 23,950 | 44·1 | 19,627 | 36·1 | 10,745 | 19·8 |
| Coagulation tests | 45,291 | 2,451 | 5·4 | 35,769 | 79·0 | 7,071 | 15·6 |
| Erythrocyte sedimentation rate | 35,200 | 13,271 | 37·7 | 15,515 | 44·1 | 6,414 | 18·2 |
| Iron studies | 29,576 | 13,283 | 44·9 | 9,541 | 32·3 | 6,752 | 22·8 |
| IgA | 28,838 | 12,335 | 42·8 | 9,656 | 33·5 | 6,847 | 23·7 |
| Coeliac | 24,804 | 11,890 | 47·9 | 7,309 | 29·5 | 5,605 | 22·6 |
| HbA1c | 18,769 | 4,620 | 24·6 | 9,543 | 50·8 | 4,606 | 24·5 |
| Vitamin D | 18,435 | 4,331 | 23·5 | 7,266 | 39·4 | 6,838 | 37·1 |
| Medication level | 14,401 | 301 | 2·1 | 13,516 | 93·9 | 584 | 4·1 |
| Urine creatinine | 12,998 | 1,040 | 8·0 | 8,742 | 67·3 | 3,216 | 24·7 |
| Gentamicin level | 10,629 | 19 | 0·2 | 10,560 | 99·4 | 50 | 0·5 |
| Vitamin $B_{12}$ | 10,478 | 5,570 | 53·2 | 3,193 | 30·5 | 1,715 | 16·4 |
| CSF studies | 9,835 | 68 | 0·7 | 9,469 | 96·3 | 298 | 3·0 |
| Folate | 9,825 | 5,281 | 53·8 | 2,849 | 29·0 | 1,695 | 17·3 |
| Lipids | 9,676 | 2,388 | 24·7 | 4,682 | 48·4 | 2,606 | 26·9 |
| Amylase | 8,108 | 640 | 7·9 | 6,614 | 81·6 | 854 | 10·5 |
| IgG/IgM | 7,992 | 1,135 | 14·2 | 4,455 | 55·7 | 2,402 | 30·1 |
| Creatine Kinase | 7,236 | 557 | 7·7 | 4,480 | 61·9 | 2,199 | 30·4 |
| Parathyroid hormone | 6,739 | 282 | 4·2 | 4,122 | 61·2 | 2,335 | 34·6 |
| Monospot test | 6,214 | 4,670 | 75·2 | 1,346 | 21·7 | 198 | 3·2 |

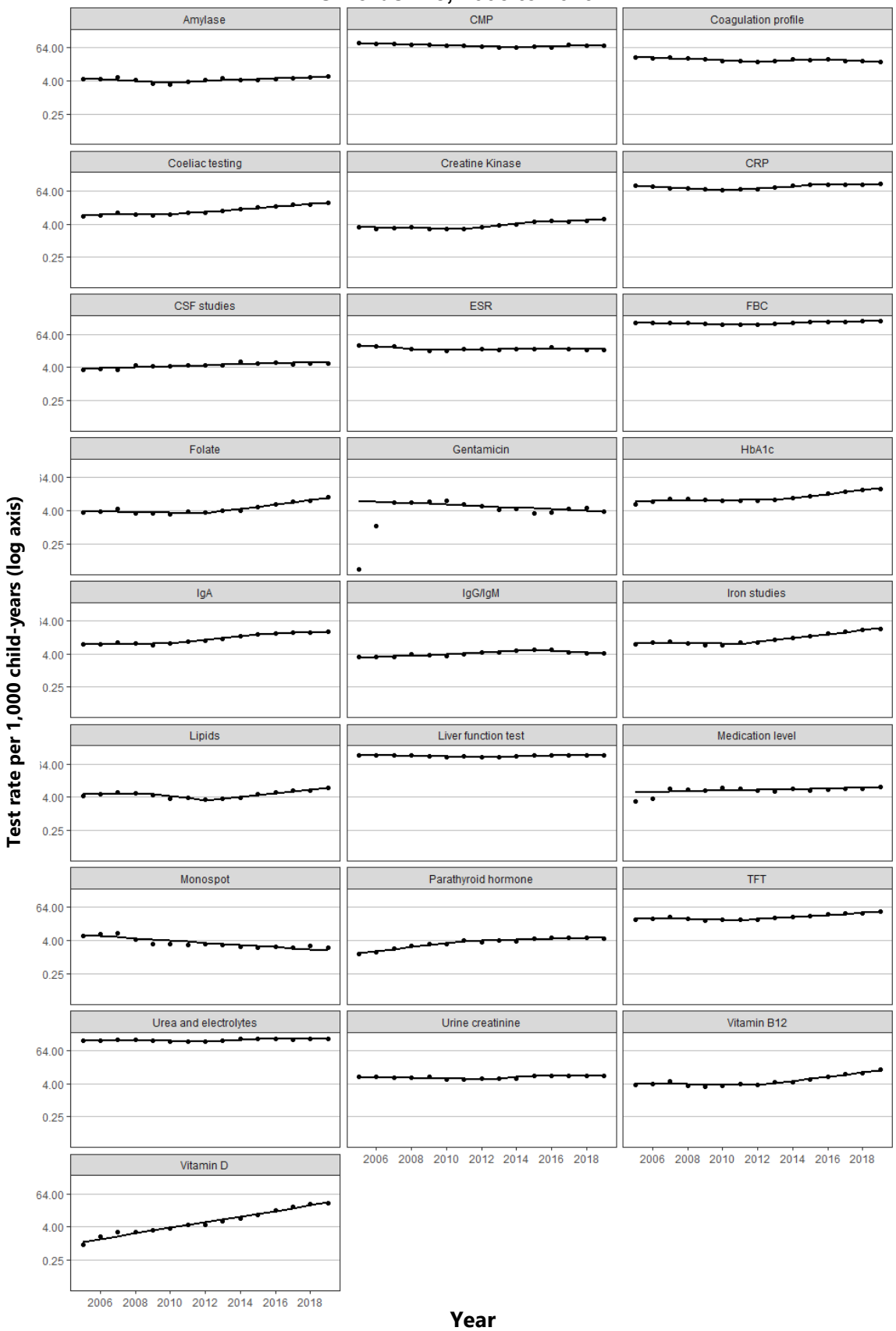
Abbreviations: CSF studies – cerebrospinal fluid studies

### 3.5.4 Trends in specific test use

The temporal changes in the 25 most frequently requested tests are shown in **Figure 3.4**. Test use followed specific patterns:

- Continuous increase: coeliac testing, creatine kinase, CSF studies, folate, HbA1c, IgA, iron studies, medication level, parathyroid hormone, thyroid function test, vitamin $B_{12}$, vitamin D.
- Continuous decrease: coagulation profile, gentamicin levels, monospot test for glandular fever.
- Mixed: amylase, calcium, magnesium, phosphate levels (CMP), C-reactive protein, creatine kinase (CK), IgG/IgM, full blood count (FBC), glucose, lipids, liver function tests (LFT), urea and electrolytes, thyroid function tests (TFT), urine creatinine.

**Figure 3.4 Temporal patterns in test use for the top 25 tests by children in Oxfordshire, 2005 to 2019**




Abbreviations: CMP – Calcium, magnesium, phosphate levels; CRP – C reactive protein; CSF studies – cerebrospinal fluid studies; ESR – erythrocyte sedimentation rate; FBC – full blood count; TFT – Thyroid function test

The average annual percentage change for each test is presented in **Figure 3.5,** and **Chapter 3 Appendix Table 3.4.** Vitamin D testing had the largest average annual change, increasing by 26·5% per year (95% CI 23·7 to 29·3%, p<0·001) from 0.9 tests per 1,000 child-years to 28.9 tests per 1,000 child-years, followed by parathyroid hormone testing, which increased by 9·8% per year (95% CI 6·8 to 12·9%, p<0·001) from 1.3 tests per 1,000 child-years to 4.9 tests per 1,000 child-years, and iron studies, rising by 9·3% per year (95% CI 7·3 to 11·4%, p<0·001) from 9.2 tests per 1,000 child-years to 33.3 tests per 1,000 child-years. Other tests that demonstrated a significant increase were (in decreasing order of AAPC): Vitamin $B_{12}$, folate, HbA1c, IgA levels, coeliac test, creatine kinase, CSF studies, thyroid function tests, and IgG/IgM. For those tests with decreasing rates, testing for glandular fever (monospot) fell by the largest margin of 8·8% per year (95% CI -11·4 to -6·1%, p<0·001), followed by gentamicin testing, which decreased by 6·1% per year (95% CI -10·9 to -1·1%, p=0·02) and coagulation tests which declined by 2·9% per year (95% CI -6·8 to 1·2%, p=0·2).

### 3.5.5 Changes in test use by age and setting

When stratified by age group (**Figure 3.6**) and setting (**Figure 3.7**), testing increased consistently for vitamin D. For other tests, trends were not uniformly distributed across ages and settings. For example, parathyroid hormone levels, iron studies, folate, and vitamin $B_{12}$ testing increased in the 1–15-year-olds, more so in general practice compared with other settings. HbA1c testing increased in general practice and inpatient settings but decreased in the outpatient setting. CRP testing also significantly increased in general practice from 2011, with an annual percentage increase of 9% per year after this point (95% CI 5·1% to 13·0%, p<0·001).

**Figure 3.5 Average annual percentage change in test use for 25 specific tests from 2005 to 2019**

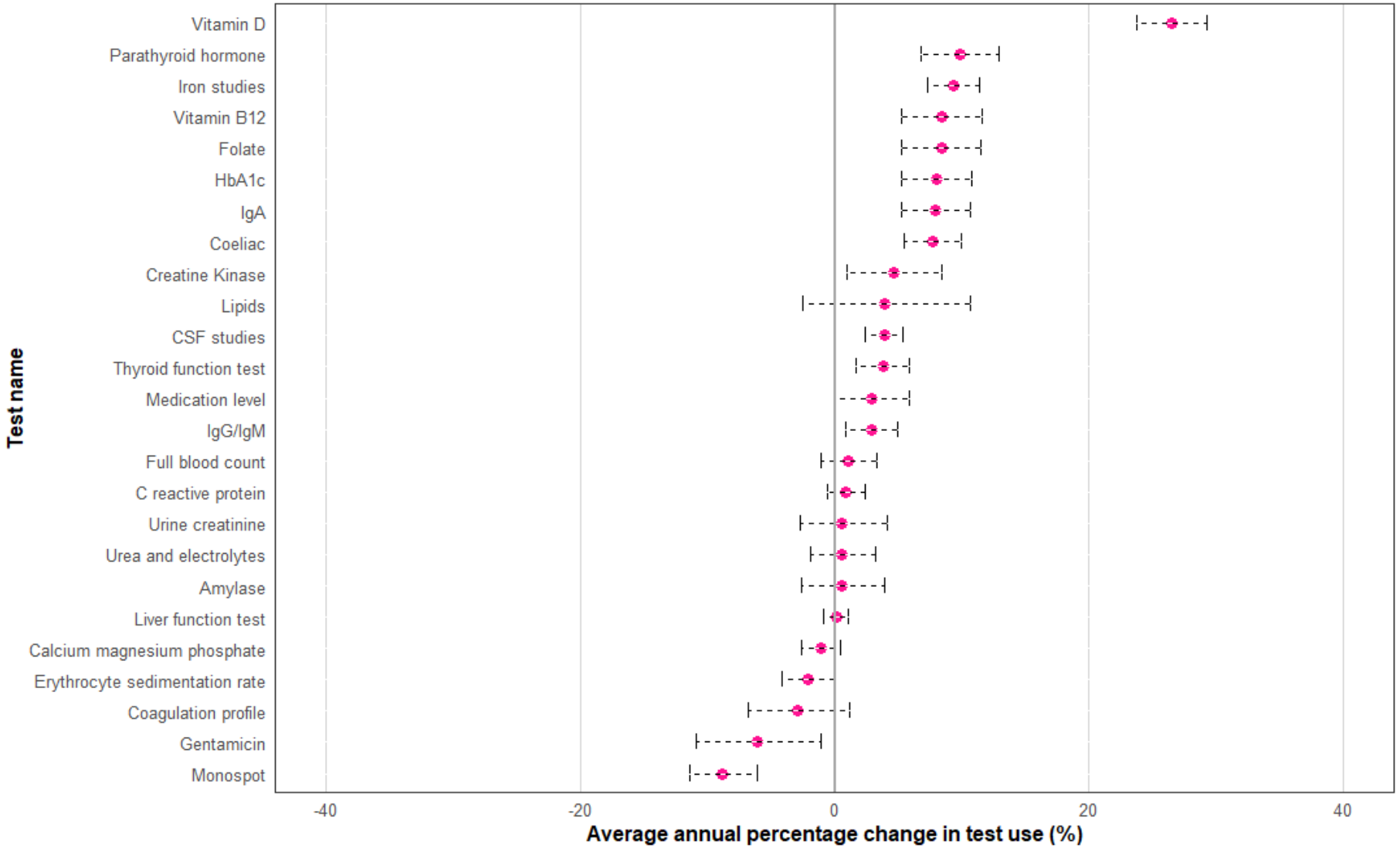
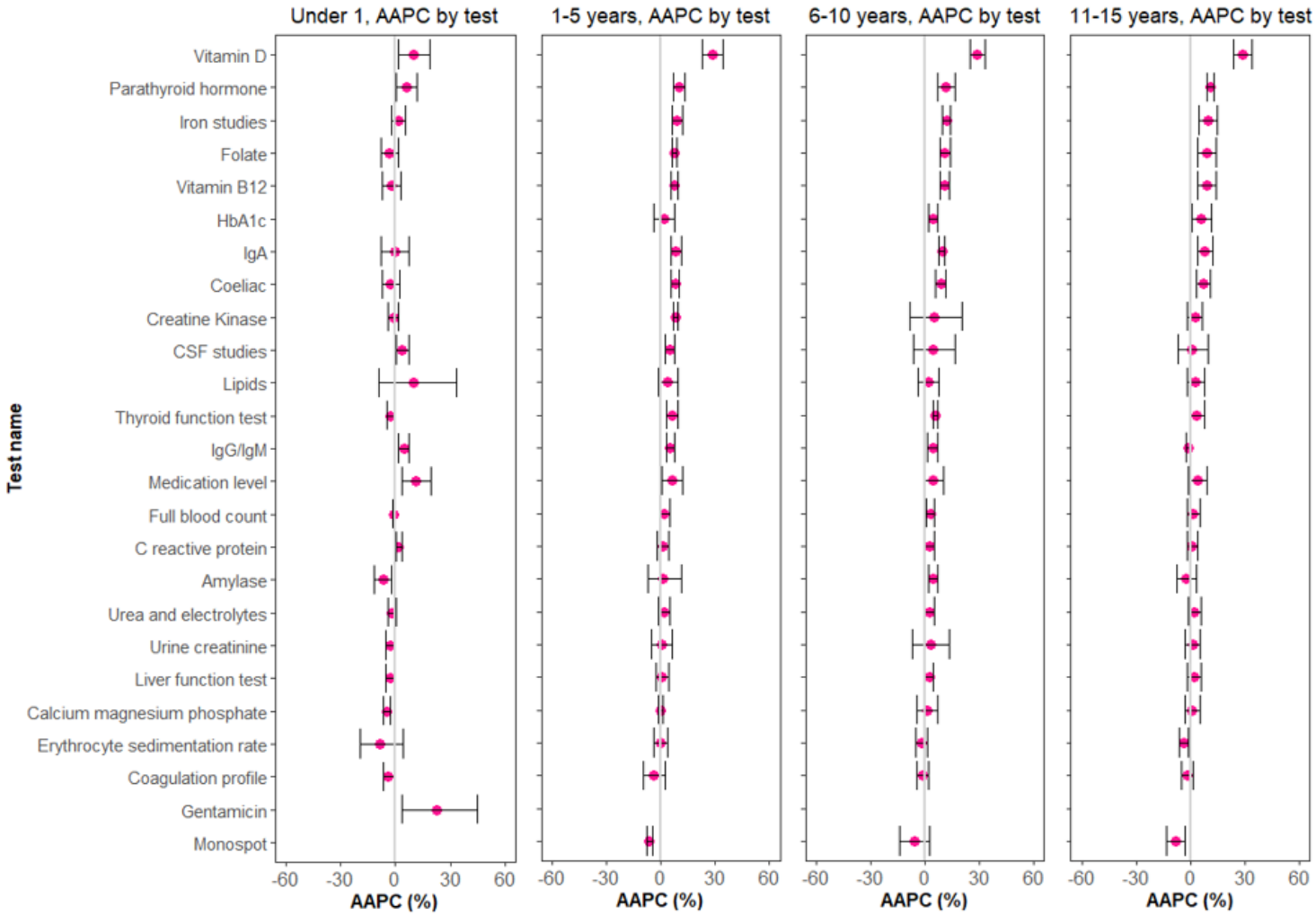
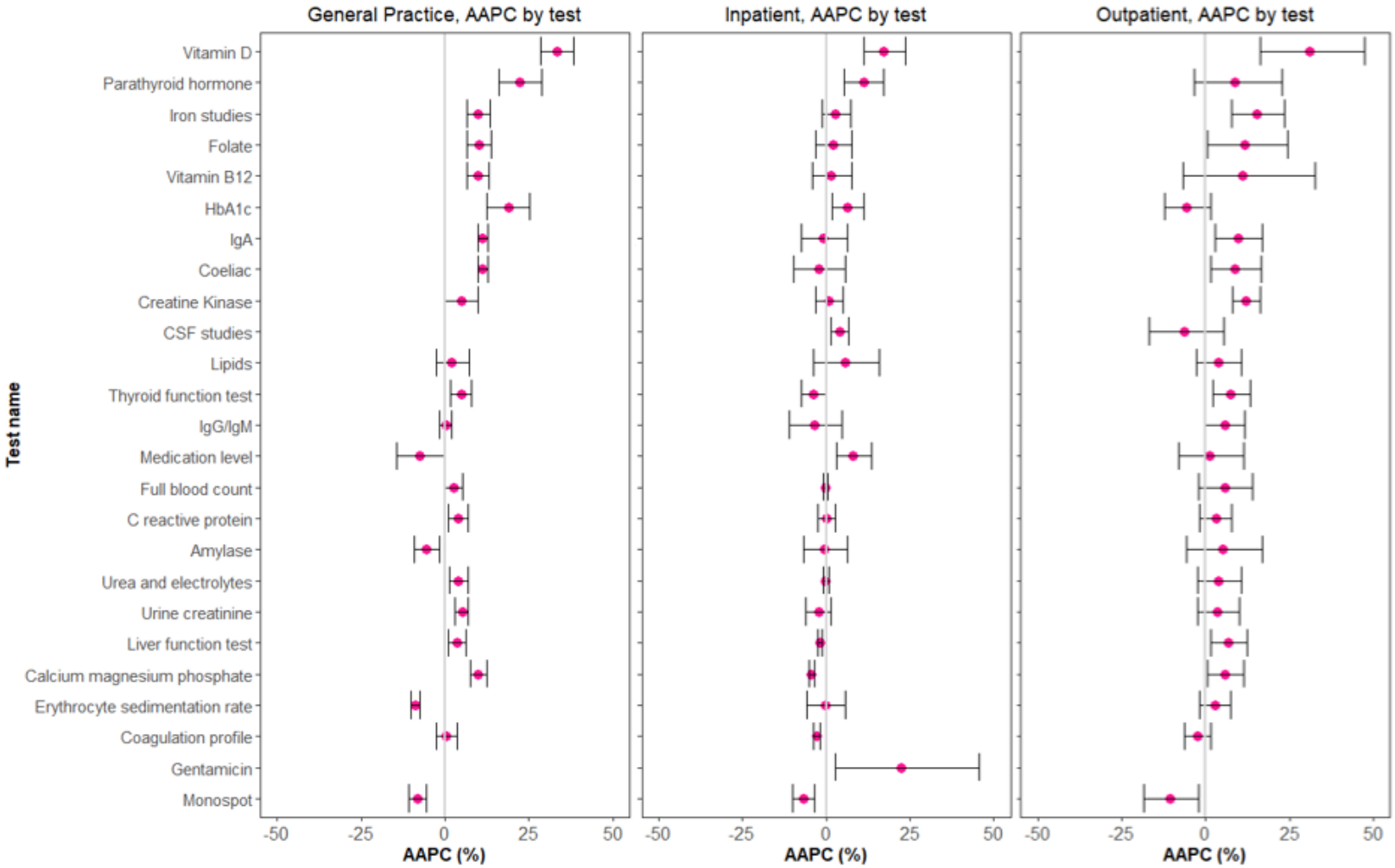


**Figure 3.6 Average annual percentage change in specific test use by age in Oxfordshire, 2005 to 2019**




Tests were excluded from the joinpoint analysis if there were fewer than 5 tests performed in a given year, hence there are missing AAPC data points for some tests (HbA1c, monospot, gentamicin).

**Figure 3.7 Average annual percentage change in specific test use by setting in Oxfordshire, 2005 to 2019**




Tests were excluded from the joinpoint analysis if there were fewer than 5 tests performed in a given year, hence there are missing AAPC data points for some tests (CSF studies, gentamicin).

## 3.6 Discussion

### 3.6.1 Statement of principal findings

In this descriptive study, I identified trends in test use for children from 2005 to 2019 in Oxfordshire by sex, age, and setting. The results demonstrated that after an initial decline, testing rates for children appeared to increase from 2012 to 2019. Testing increases were more pronounced in females than males, and in children aged between 6 and 15 compared with the other age groups. The largest relative increases occurred in the outpatient setting followed by general practice. The drivers of the changes in overall testing rates can likely be attributed to testing increases in these settings and subgroups.

Of the most common tests, testing for vitamin D, parathyroid hormone, iron studies, folate, and vitamin $B_{12}$ increased by the greatest proportion annually. Relative increases in these tests were most pronounced in general practice. These are consistent with temporal changes in test use by adults in UK primary care from 2000 to 2015[6], where O'Sullivan and colleagues reported increases in testing for vitamin D (which increased by 54% per year). Iron, ferritin, vitamin $B_{12}$, and folate testing in adults increased by 16-19% per year. In recent years, there has been greater awareness of some conditions such as vitamin D deficiency and iron deficiency, making it more likely that doctors test for these conditions. Increased disease prevalence may also explain the rise in testing.[11] Iron studies, folate, vitamin $B_{12}$, thyroid function tests and creatine kinase form part of the workup for fatigue. A 2007 prospective study of British adolescents reported the point prevalence of fatigue was 38%, but recent estimates are lacking.[12] Most of these tests (vitamin D, iron studies, folate, vitamin $B_{12}$, coeliac test, serum IgA) are also included in investigations for malnutrition and faltering growth, suggesting increasing clinician concern and/or incidence of these conditions.

Increasing testing rates in females compared with males may be explained by sex differences in health symptoms. A systematic review of international studies analysing sex-differences in children and adolescents found an emerging or increasing “excess” morbidity in females during the transition to adolescence for self-reported general health, and specific symptoms including headache, abdominal pain, tiredness, back pain and dizziness.[13] A follow-up prospective cohort study also found a female “excess” in common presentations: high temperature, rash, earache, stomach-ache, head-ache, constipation and eczema as well as health-related days off school.[14] The study authors discussed that excess reported morbidity in females may be explained by biological differences with the onset of menstruation, or sex differences in illness-related attitudes and expectations of children and their parents.[14] Any of these presentations could potentially lead to more tests. Higher testing rates in adolescent females also correspond to patterns of outpatient appointment attendances in England in 2019-20. While females aged 0 to 4 (45%) and 5 to 9 (46%) attended less frequently than males, this pattern reversed in 10-14-year-olds females (53%) and 15-19-year-olds females (61%) who attended more frequently than their males of the same age.[15]

Overall testing rates in Oxfordshire decreased between 2005 and 2012 and then appeared to increase (APC 2012-2015 9.0%, 95% CI -3.0 to 22.5%). Decreases in the early period mirrored decreasing inpatient and outpatient attendances at Oxford University Hospitals.[16,17] The inflection point in testing rates aligns with a major service change in 2011, with eleven new general paediatric consultants appointed in Oxfordshire. Whilst all staff were appropriately qualified, newer members of staff may be inclined to test more[18,19], which may be one explanation for change in trends in testing rates. The findings of this chapter were presented at the Oxford University Hospitals Department of Paediatrics Grand Rounds in March 2023. In the discussion that ensued with the paediatric consultants who attended, it was perceived that increased parental expectations and anxiety levels have contributed to increased test rates. In primary care,

there has been a decline in experienced family doctors who are more likely to reassure parents without testing and referring children, and a concurrent rise in less experienced GPs who may test and refer more.[20] Parents of children referred by these GPs may present to outpatient appointments expecting a diagnostic workup. The median number of tests per child did not differ between general practice and inpatient settings. This could reflect differences in the number of tests undertaken at each health care encounter in these two settings, where children may have had fewer tests per health care encounter but visited their GP several times over the study period, compared with inpatient settings, where children are likely to have had more tests performed in one encounter.

The most pronounced increases in testing occurred in the outpatient setting, which expanded during the study period. According to NHS digital data, from 2009 to 2020, the number of children seen in OUH outpatient clinics increased by 63% (with a sharp increase in 2013).[16] Since the Oxford Children's Hospital opened in 2007, the tertiary specialty workload has increased, with an increased number of referrals from all around the region outside of Oxfordshire (including Berkshire, Buckinghamshire, Wiltshire, Milton Keynes). We could not distinguish between tests that were conducted among Oxfordshire residents and referrals for tests from other areas, or which tests were for specialty or general paediatric patients. As a result, the appropriateness of the denominator may have changed over time, influencing the observed trends.

Some of the observed increases in testing rates warrant further investigation as to their appropriateness. NICE guidance only recommends vitamin D testing in children if they have musculoskeletal symptoms, abnormal serum bone profile or X-Ray findings, suspected bone disease such as osteomalacia or known bone disease such as osteoporosis.[21] A retrospective analysis of vitamin D testing in the Northumbria Healthcare NHS Trust from 2002 to 2017 found that over 75% of the tests performed on those aged below 30 had an inappropriate clinical indication.[22]

The clinical implications of these reported trends should be considered on a test-by-test basis. My findings of vitamin D testing trends mirrored those of a recently published study from Australia, which reported that vitamin D test requests for children increased 30-fold in general practice from 2003 to 2018 with the odds of detecting low levels remaining stable.[23] The incidence of nutritional rickets in children (the sequelae of vitamin D deficiency) was similarly reported to be low, with an annual incidence of 0.48 per 100,00 children under 16 in the UK in a surveillance study from 2015 to 2017[24], reinforcing the need to mitigate this low-value practice. Instead, for asymptomatic children, routine supplementation could be considered rather than testing. Similarly, comparing testing rates to rates of abnormal tests for ferritin, vitamin $B_{12}$, and folate would help to assess whether this represents low-value testing. I did not conduct such analyses because information on past laboratory analytic methods was unavailable. Efforts to reduce unnecessary tests should also target specific settings based on the observed trends, for instance, haematinics in general practice.

Unnecessary testing in children has important cost implications. Assuming the cost of a vitamin D test was £10 (as per the unit costings from laboratory manager at the OUH Trust Laboratory), expenditure on vitamin D tests for children in Oxfordshire in 2019 was £37,900. In UK primary care alone, an estimated £1.7 million was spent on vitamin D tests on children aged 0-17 in 2014.[25] Given the subsequent increases in vitamin D testing, it is likely that current national expenditure on vitamin D tests in children is far higher than this figure across primary and secondary care. The financial consequences of unnecessary tests also extend beyond the cost of tests alone; tests generate increased workload for laboratory staff as well as the clinicians who must review and action the result. An abnormal result often leads to further investigation and monitoring, treatment and/or referrals, which also costs the health system.

### 3.6.2 Strengths and limitations

This study is the first to describe long-term trends in test use in a population-based study of children. Test variation over time indicates potential changes in disease prevalence but also suggests potentially inappropriate testing.

The generalisability of this study's findings beyond Oxfordshire is unclear. Despite having 17 out of 83 areas classified among the most deprived 20% nationally, Oxfordshire, on average, is a less socially deprived county with high educational attainment, and increased test use over time may reflect greater access to laboratory tests in this area. However, other regions of the UK may show different trends and/or lower testing rates. If this is the case, it may support a hypothesis of over-testing in Oxfordshire.

I excluded glucose and blood gas tests because they could be unreliably coded in the laboratory data (in some cases these are performed as bedside point-of-care care tests rather than being sent to the laboratory for analysis); however, these were the 8th and 15th most common tests, based on the potentially under-reported data. Estimations of temporal trends in test use and the proportion of children receiving at least one test, therefore, may be generalizable to laboratory tests but not point-of-care testing.

I did not have access to individual patient-level data or patient records containing complete information on the clinical indications for each test. This would have provided insight into the appropriateness of tests. For example, it would be advantageous to compare whether children who have complex conditions and frequently visit the hospital (e.g., for cystic fibrosis pulmonary optimization) are more likely to have blood tests than others. If patient-level data were available, sensitivity analyses excluding children with high testing rates could have provided a more accurate assessment of testing trends in the general population.

### 3.6.3 Implications for research and clinical practice

I limited testing data until the end of 2019 to eliminate the impacts of the COVID-19 pandemic. However, further research should be conducted to explore the pandemic's impacts. A systematic review examining changes in global healthcare use during the pandemic, demonstrated a 31% decline in diagnostic imaging and tests, based on 12 included studies.[28] The pandemic serves as a natural experiment, allowing an examination into the impacts of decreased healthcare utilization. Reduced testing almost certainly resulted in missed diagnoses and deleterious outcomes in some cases, but the authors of this review found that reductions often tended to be greater for milder or less severe illness, suggesting that in some cases, forgoing tests resulted in reduced harms from unnecessary testing. The review did not find any primary studies conducted in children, and future research focusing on how the use of childhood diagnostics changed during the pandemic may determine which tests, if any, could be considered unnecessary.

Laboratory tests do not account for all the tests that children receive. Further analyses of testing variation should include other tests, including urine testing, microbiology and infection, imaging, and spirometry. The findings of this chapter can be compared with other settings to examine if the changes in test use are consistent across England and in other places with similar paediatric healthcare systems. Larger datasets including individual patient-level data and demographics can be used to determine if testing rates are linked to deprivation levels and ethnicity.

The analyses I have presented allow clinicians to become aware of their test-ordering practices. Increased test use exacerbates the burden on physicians with increased patient workload and time pressures. It also puts considerable strain on health expenditure. While testing is crucial in certain situations, every clinician should consider if

a test is likely to yield more benefit than harm to the child, their family, and the overall health system.

### 3.6.4 How this chapter fits within the wider thesis

My thesis aims to quantify and analyse variation in test use for children in primary care. In this chapter I sought to appreciate the landscape of laboratory testing in children; to understand what proportion of tests are performed in primary care compared to secondary care; determine which tests were subject to the greatest temporal increases; and identify the specific settings and groups in which the greatest temporal changes occurred. The findings of this chapter informed the tests I chose to focus on in Chapter 4, where I analysed English general practice data to determine temporal changes and between-practice variation in paediatric testing, to thereby detect tests that may be subject to overuse.

## 3.7 Conclusions

Laboratory test use by children in Oxfordshire demonstrated an apparent increase since 2012 after an initial decline. Vitamin D tests increased by the highest margin, which reflects greater clinician awareness but potential overuse. Further research should compare the observed trends with other regions to determine if there is also geographic and practice variation in test utilisation.

## 3.8 Chapter Summary

- I used laboratory testing data from the Oxford University Hospitals (OUH) Trust database from 2005 to 2019 to investigate how test use has changed over time in general practice and hospital settings.

- After an initial decline, laboratory test use by children in Oxfordshire demonstrated an apparent increase since 2012 (APC 1.6%, 95 confidence interval -0.8% to 4.1%).
- Increases were highest in females, in those aged 6-15 years and in general practice and outpatient settings.
- The most frequently requested tests were full blood count, urea and electrolytes, liver function test, C-reactive protein, and calcium magnesium phosphate levels.
- Vitamin D tests increased by the greatest margin (27% per year on average), followed by parathyroid hormone, iron studies, vitamin $B_{12}$, and folate levels.
- Measuring temporal increase in test use is an indirect method of identifying overuse, prompting the need for further research and improved clinical guidance in these areas.
- The findings of this chapter also highlight the specific settings where efforts should be targeted to mitigate unnecessary testing.

## 3.9 References


1 UK Department of Health. Report of the Review of NHS Pathology Services in England chaired by Lord Carter of Coles. London, 2006.

2 Ducatman BS, Ducatman AM, Crawford JM, Laposata M, Sanfilippo F. The Value Proposition for Pathologists: A Population Health Approach. Acad Pathol 2020; **7**. DOI:10.1177/2374289519898857.

3 Public Health England. The 2nd Atlas of Variation in NHS Diagnostic Services in England. NHS RightCare 2017.

4 Wennberg J, Cooper M. Dartmouth Atlas of Health. 1998.

5 Busby J, Schroeder K, Woltersdorf W, et al. Temporal growth and geographic variation in the use of laboratory tests by NHS general practices: Using routine data to identify research priorities. Br J Gen Prac 2013; **63**. DOI:10.3399/bjgp13X665224.

6 O'Sullivan JW, Stevens S, Hobbs FDR, et al. Temporal trends in use of tests in UK primary care, 2000-15: Retrospective analysis of 250 million tests. The BMJ 2018; **363**. DOI:10.1136/bmj.k4666.

7 Hiscock H, Neely RJ, Warren H, Soon J, Georgiou A. Reducing unnecessary imaging and pathology tests: A systematic review. Pediatrics. 2018; **141**. DOI:10.1542/peds.2017-2862.

8 National Cancer Institute, Division of Cancer Control and Population Sciences. Joinpoint Regression Program. USA: Surveillance Research Program. 2020.

9 Clift AK, Coupland CAC, Hippisley-Cox J. Prostate-specific antigen testing and opportunistic prostate cancer screening: A cohort study in England, 1998–2017. British Journal of General Practice. 2021; **71**. DOI:10.3399/bjgp20X713957.

10 R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

11 Brüggmann D, Alafi A, Jaque J, et al. World-wide research architecture of Vitamin D research: Density-equalizing mapping studies and socio-economic analysis. Nutr J 2018; **17**. DOI:10.1186/s12937-018-0313-6.

12 Rimes KA, Goodman R, Hotopf M, Wessely S, Meltzer H, Chalder T. Incidence, prognosis, and risk factors for fatigue and chronic fatigue syndrome in adolescents: A prospective community study. Pediatrics 2007; **119**. DOI:10.1542/peds.2006-2231.

13 MacLean A, Sweeting H, Egan M, Der G, Adamson J, Hunt K. How robust is the evidence of an emerging or increasing female excess in physical morbidity between

childhood and adolescence? Results of a systematic literature review and meta-analyses. Soc Sci Med. 2013; **78**. DOI:10.1016/j.socscimed.2012.11.039.

14 Sweeting H, Whitley E, Teyhan A, Hunt K. Sex differences in child and adolescent physical morbidity: Cohort study. BMJ Paediatr Open 2017; **1**. DOI:10.1136/bmjpo-2017-000191.

15 NHS Digital. Hospital Outpatient Activity 2019-20. Hospital Outpatient Activity. https://digital.nhs.uk/data-and-information/publications/statistical/hospital-outpatient-activity/2019-20. 2020; published online Oct 8 (accessed July 11, 2023).

16 NHS Digital. Hospital Outpatient Activity. https://digital.nhs.uk/data-and-information/publications/statistical/hospital-outpatient-activity. 2022; published online Sept 29. (accessed July 11, 2023).

17 NHS Digital. Hospital Admitted Patient Care Activity. Hospital Admitted Patient Care Activity. 2021. https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity (accessed July 11, 2023).

18 McGillivray DL, Roberts Bräuer R, Kramer MS. Diagnostic Test Ordering in the Evaluation of Febrile Children: Physician and Environmental Factors. American Journal of Diseases of Children 1993; **147**. DOI:10.1001/archpedi.1993.02160320072022.

19 Griffith CH, Wilson JF, Desai NS, Rich EC. Does pediatric housestaff experience influence tests ordered for infants in the neonatal intensive care unit? Crit Care Med 1997; **25**. DOI:10.1097/00003246-199704000-00024.

20 NHS Digital. General Practice Workforce. 2022. https://digital.nhs.uk/data-and-information/publications/statistical/general-and-personal-medical-services (accessed Nov 4, 2022).

21 National Institute of Health and Care Excellence. Vitamin D deficiency in children. https://cks.nice.org.uk/topics/vitamin-d-deficiency-in-children/. 2022; published online Jan. (accessed Oct 4, 2022).

22 Woodford HJ, Barrett S, Pattman S. Vitamin D: Too much testing and treating? Clinical Medicine, Journal of the Royal College of Physicians of London 2018; **18**. DOI:10.7861/clinmedicine.18-3-196.

23 Zurynski Y, Munns CF, Sezgin G, Imai C, Georgiou A. Vitamin D testing in children and adolescents in Victoria, Australia: are testing practices in line with global recommendations? Arch Dis Child 2023; archdischild-2022-325000.

24 Julies P, Lynn RM, Lynn RM, et al. Nutritional rickets under 16 years: UK surveillance results. Arch Dis Child 2020; **105**. doi:10.1136/archdischild-2019-317934.

25 Basatemur E, Hunter R, Horsfall L, Sutcliffe A, Rait G. Costs of vitamin D testing and prescribing among children in primary care. Eur J Pediatr 2017; **176**. DOI:10.1007/s00431-017-2986-9.

26 Oxfordshire County Council. Oxfordshire Insight: Deprivation. https://insight.oxfordshire.gov.uk/cms/deprivation-dashboard. 2019.

27 Office for National Statistics. National statistics: English indices of deprivation 2019. Ministry of Housing, Communities & Local Government. https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019. 2019; published online Sept 26.

28 Moynihan R, Sanders S, Michaleff ZA, et al. Impact of COVID-19 pandemic on utilisation of healthcare services: A systematic review. BMJ Open 2021; **11**. DOI:10.1136/bmjopen-2020-045343.

# Chapter 4. Temporal and practice variation in paediatric diagnostic tests in UK primary care from 2007 to 2019

## 4.1 Introduction

Many clinical decisions involve the use of diagnostic tests[1], but their use in children, specifically in UK general practice, has not been well characterised. Measuring variation in testing can identify tests that are potentially overused or underused, which both have harmful consequences. Overuse can result in testing cascades, harm from unnecessary treatment, and physical and emotional trauma for the patient and their family. It generates additional clinical workload for clinicians and lead to increases in unnecessary referrals, healthcare contacts, and health spending in a system that is already overburdened. Underuse of tests can result in missed or delayed diagnosis and treatment with potentially serious physical, emotional, and financial consequences for patients, families, and clinicians. It is therefore important to identify variation in test use to detect potential underuse and overuse of tests, as it has implications for the rates of disease diagnosis, referral rates, hospital activity, treatment, and cost to the health system.

In Chapter 3, I examined temporal trends in blood test use for children in Oxfordshire and identified the tests with the greatest increase in use. The current chapter examines data from English general practices to analyse temporal trends and practice variation in all types of paediatric tests and investigate potential demographic factors that drive variation in testing.

## 4.2 Aims and objectives

The overall aim of this chapter was to quantify and analyse temporal change and practice variation in paediatric diagnostic tests in English primary care.

The specific objectives were:

1. To quantify how paediatric diagnostic test use has changed over time in English primary care.
2. To quantify how paediatric test use varies across general practices in England.
3. To identify tests that are subject to the greatest variation between practices.
4. To determine how demographic and socioeconomic factors drive variation in test use.

## 4.3 Methods

### 4.3.1 Study design and setting

I performed a retrospective population-based observational study, analysing routinely collected data from the electronic health records of children aged 0 – 15 years presenting to general practices in England.

### 4.3.2 Data sources

The Clinical Practice Research Datalink Aurum contains routinely collected data from GP practices in England that use EMIS Web® electronic patient record system software.[2] The data extraction build from Feb 2022 encompasses approximately 19.9% of the UK population and 16.6% of UK general practices.[3]

I also requested and obtained linked data for practice-level index of multiple deprivation (IMD) deciles.

### 4.3.3 Participants

I included all children aged 0 to 15 years in CPRD-registered general practices during the study period from 1 January 2007 to 31 December 2019. During this period, children were followed up from birth or registration date until they reached the age of 16 years, died, reached the end of the study period or until they transferred out of the practice. If a patient

transferred to another CPRD-registered practice, then they were registered as a new patient in the data, but this would only have occurred in a very small fraction of the overall study population. Patients were included if their records were considered acceptable for research purposes (a CPRD data quality indicator) and were registered at practices defined as up-to-standard in terms of data reporting by CPRD during the study period.

### 4.3.4 Variables

The following variables were extracted: practice ID, patient ID, age, gender, practice registration date, registration end date, name of test, date of test, and practice-level index of multiple deprivation (1-10).

The exposures of interest were: 1) test year and 2) practice ID.

Covariates included age, gender, and socioeconomic status. Socioeconomic status was defined by the Index of Multiple Deprivation 2019 deciles, where 1 represents practices in the least deprived areas and 10 represents practices in the most deprived areas.[2,4]

The primary outcomes of this study were:

1) Temporal changes in test rates from 2007 to 2019, stratified by test type, gender, age group and deprivation level, measured by the average annual percent change (AAPC) and
2) Practice variability in test use, measured by the coefficient of variation (CoV) in 2019.

In exploratory analyses, I examined:

1) Temporal changes in specific tests by gender and age group between 2007 and 2019
2) The influence of gender, age, and deprivation level on practice testing rates in 2019.
3) The relationship between practice variation and rates of testing for specific tests

4) The relationship between practice variation and temporal changes for specific tests
5) Changes in testing rates of practices over time

### 4.3.5 Included tests

For overall metrics of test use, all tests were included. When looking at specific tests, I chose a subset of tests that comprised the following:

- The 25 most frequently recorded tests between 2007 and 2019. These tests made up more than three quarters (76%) of the total tests (or ~82% of tests after removing unspecified tests), suggesting that these would be the tests most likely to have an impact on overall trends.
- Tests specifically reported by GPs or trainee GPs as tests they request for children, or tests they perceived to be subject to temporal or practice variation during the interviews conducted in Chapter 5.
- Tests that were highlighted in Chapter 3 as being subject to significant temporal increases in Oxfordshire (Chapter 3)

This resulted in 35 included tests, comprising 79% of the total tests that were conducted (or 85%, after excluding unspecified tests). The list of included tests and cumulative frequency of all tests by rank is provided in the **Chapter 4 Appendix (Table 4.1 and Figure 4.1).**

### 4.3.6 Statistical methods

I calculated the crude rates of test use by dividing the total number of diagnostic tests by the overall number of children-years of observation contributed during each year of the study period and multiplying by 1,000. I calculated age-adjusted annual rates by standardising the population to the number of children aged 0 to 15 years of age who

were registered to CPRD practices in 2019. 2019 was selected for standardisation as this was deemed to be most relevant to the current population.

*Temporal variation*

Following a similar approach to Chapter 3, I used joinpoint regression to model temporal changes in age-adjusted test rates from 2007 to 2019. Points where significant changes in rates occurred (called joinpoints) were identified, and annual percentage changes (APC) between joinpoints were estimated. The joinpoint regression model also provided an estimate of the average annual percentage change (AAPC), a summary measure of the trend from 2007 to 2019. The overall age-adjusted rates, APC, and AAPCs were stratified by test type, gender, age group (less than 1, 1-5, 6-10, 11-15 years), and index of multiple deprivation (IMD) level (where 1 represents the least deprived group). The IMD deciles provided in the original dataset were collapsed into quintiles for the purposes of interpretability and clarity in the graphs. For all other analyses using IMD level (including practice variation), deciles were used.

*Practice variation*

To determine the crude rates of test use by practice, the total number of tests by each practice (numerator) was divided by the child-years contributed by the practice in 2019 (denominator), where a child who was registered for a year contributed 1 person-year of observation (or a fraction if registered for less than a full year). The unadjusted coefficient of variation (CoV) was calculated by dividing the standard deviation of the unadjusted test rates by the mean. I calculated the adjusted rates of test use by using a generalised linear model with Poisson errors to estimate the number of tests ordered from each practice adjusted for practice-level differences in gender (proportion of females), median age of the study population, and deprivation index (IMD decile). These adjusted rates of test use were then used to calculate the adjusted CoV (see **Box 4.1**)

**Box 4.1. Calculations for adjusted rates of test use, coefficient of variation and confidence intervals**

The Poisson log-linear regression model for rate data was calculated as follows:

$$\log\frac{\mu}{t} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

Where $\mu$ is the number of tests requested by each practice in 2019,
$t$ is the person-years contributed by each practice in 2019.
$\beta_0$ is the intercept.
$x_1$ is the practice index of multiple deprivation decile.
$x_2$ is the proportion of female patients aged 0 to 15 in the practice in 2019
$x_3$ is the median age of children aged 0 to 15 in the practice in 2019

$$\frac{\mu}{t} = e^{\beta_0} e^{\beta_1 x_1} e^{\beta_2 x_2} e^{\beta_3 x_3}$$
$$\mu = t e^{\beta_0} e^{\beta_1 x_1} e^{\beta_2 x_2} e^{\beta_3 x_3}$$

The adjusted mean ($x$) and standard deviation (SD) were computed from the GLM model fitted rate.
The adjusted CoV of test use for each practice was calculated by dividing the adjusted mean ($x$) by the adjusted standard deviation (SD).
The following calculations were used for the confidence intervals:

$$Standard\ error_{CoV} = \frac{CoV_{adjusted}}{\sqrt{(2n)}}$$

$\alpha = 0.05, df = n - 1$ - where $n$ is the number of observations

$$95\%CI\ (CoV) = CoV_{adjusted} \pm t\,\alpha_{=0.05;\, df=(n-1)}\, SE_{CoV}$$

Tests with the most variation between practices were identified and summarised graphically in a dot chart (ranked from tests with the greatest variation to those with the least).

*Additional analyses*

Relationships between the age-adjusted 2019 test rate and adjusted CoV as well as AAPC vs adjusted CoV were explored after visualising the association on scatterplots using linear regression.

Finally, to explore the changes in practice testing rates over time, I calculated the three-year average test rate for the beginning of the study period (2007 – 2009) and the end of the study period (2017 – 2019). I stratified the resulting rates into quintiles from 1 (lowest testing quintile) to 5 (highest testing quintile) for both time ranges (2007 - 2009, 2017 - 2019). These were presented as a Sankey chart to visualise the changes in practice testing rates over time.

APCs and AAPCs were modelled in Joinpoint Regression software version 5.0.2.[5] Data cleaning, management and all other analyses were performed using R version 4.3.1.[6]

### 4.3.7 Sparse or missing data

In some cases when performing analyses for the joinpoint regression, there were zero tests performed in some years at the beginning of the study, as the tests were introduced in primary care during the study period (faecal calprotectin and fractional exhaled nitric oxide [FeNO] testing). The Joinpoint Regression Program is not designed to handle missing data and does not compute AAPCs in groups where there are zero cases, therefore, I imputed the number of tests as 1 in those years for the purposes of bypassing the error and enabling the program to proceed with the analysis.[7] Imputation was performed for the overall test-specific calculations, however, for the subgroup analyses (AAPC by gender, age, and index of multiple deprivation quintile), tests were excluded if there were too few tests to perform a meaningful analysis, or the test was not technically possible to perform, leaving the outcome blank in the graph (e.g., spirometry in <1 year old).

There were cases in which there were unspecified diagnostic test codes. Examples include “Diagnostic procedure” which was coded as “Diagnostic test NOS”, or “Bacterial culture and sensitivity” which was assigned to the test code “Microbiology NOS”. I attempted to make sense of the unspecified tests by examining the context in which they

occurred, and determining which other tests the child received that day. If the unspecified test co-existed (occurred on the same day for the same child) with other matched specified test codes, then I grouped these together. These were then de-duplicated, so that tests were not double counted. For example, if there was a code for Microbiology NOS and Urine MCS on the same day for a given child, I re-assigned the Microbiology NOS as Urine MCS and then performed de-duplication. Unfortunately, not all the unspecified test codes could be explained by other tests that were concurrently requested on the same day and so these codes remained unspecified. These tests were included in the overall calculations of test use, but they were excluded from the test-specific analyses.

#### 4.3.8 Data access, cleaning, and management methods

The patient cohort was extracted by the CPRD team (CWD) based on the age eligibility criteria and the acceptability of the data. I performed further data cleaning to exclude those outside the age eligibility criteria. The age for each patient was calculated by subtracting the date of observation from the year of birth (assuming January 1$^{st}$ if the month was not provided).

To extract the test codes, all EMIS observations were mined from the relevant EMIS code categories: 3 (Diagnostics), 11 (Haematology), 20 (Cytology/Histology), 33 (Immunology), 34 (Biochemistry), 41 (Radiology), 46 (Investigation requests) and 47 (Microbiology) along with the total number of occurrences in EMIS across all patients and all years. I excluded codes with less than 10,000 occurrences as these were unlikely to represent important tests (<0.2% of the defined cohort). I went through each code with a practising clinician (CH) to determine whether it represented a test ("Serum folate level") or not a test ("Alcohol use disorders identification test score"). Of the 12,392 codes, there were 4,123 test codes deemed relevant for inclusion.

Test codes were assigned to test panels as previously described in **Chapter 3**, where I independently assigned each test code to a panel, and these were verified by a clinician (CH).

### 4.3.9 Reporting standards

This chapter is written according to the RECORD statement checklist for use in observational studies using routinely collected health data.[8]

### 4.3.10 Ethics approval

The protocol was approved via the CPRD's Research Data Governance process (study reference 22_001998) and is available on the Open Science Framework (osf.io/pwgtf).

### 4.3.11 Patient and public involvement

Parent representatives on my PPI panel were involved in the planning and design of this study, including consideration of which tests to include for the test-specific analyses. They were also involved in discussions on future directions of this research.

### 4.3.12 Deviations from the protocol

In the original protocol, I registered the intention to include test data from 2006 to 2021 to capture a 15-year period. Later, the study end date was changed to 31 December 2019 to exclude pandemic effects. After receiving the extracted data for tests performed in 2006, I observed an abnormally low proportion of children aged under 1 (approximately half of the 2007 population). Reviewing the distribution of the months of birth of these children revealed that the data had excluded a substantial portion of children who had been born in 2005. Consequently, to get a more accurate understanding of temporal trends in each age group, the study start date was changed to 1 January 2007.

### 4.3.13 Data availability

All test codes, R code used for data management, analysis and creating the figures is available on GitHub (https://github.com/elizabethtthomas/cprd-paediatric-tests).

## 4.4 Results

### 4.4.1 Characteristics of included participants and tests

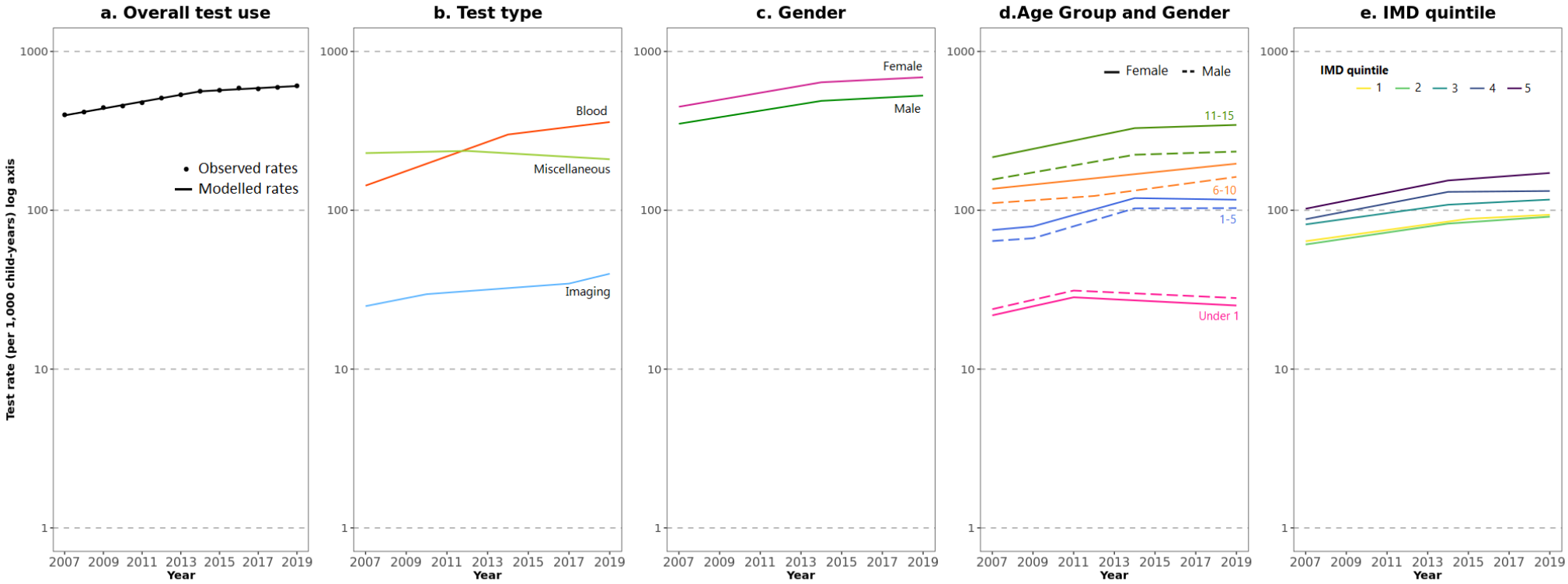
There were 14,299,598 tests performed on 2,542,101 children from 1 January 2007 to 31 December 2019, of which 50.4% (1,282,072 of 2,542,101) were females, see **Table 4.1**. As children were followed up over time, they were present in more than one age category, generating the discrepancy in total number of children across age categories. 54.5% of the total tests (7,794,755 of 14,299,598 tests) were performed in females. The highest proportion of tests (39.2%) was conducted for children aged 11-15, and blood tests were the most frequently performed type of diagnostic tests. Across the study population, the median number of tests per child per year was 2 (IQR 1,3). Once stratified by age group, the median number of tests per year was 1 (IQR 1 to 2), 1 (IQR 1 to 3), 1 (IQR 1 to 3) and 2 (IQR 1 to 5) for children aged <1, 1-5, 6-10 and 11-15 years respectively. The deprivation level of the contributing practices ranged from 1 to 10 (median 6, IQR 4 to 9).

**Table 4.1 Characteristics of included participants and tests**

| | Number of people | % |
|---|---|---|
| **Total** | 2,542,101 | 100 |
| **Gender** | | |
| Male | 1,259,930 | 49.6 |
| Female | 1,282,072 | 50.4 |
| Other | 99 | 0.0 |
| **IMD** | | |
| 1 (least deprived) | 192,958 | 7.6 |
| 2 | 215,874 | 8.5 |
| 3 | 196,193 | 7.7 |
| 4 | 193,098 | 7.6 |
| 5 | 243,847 | 9.6 |
| 6 | 264,146 | 10.4 |
| 7 | 255,668 | 10.1 |
| 8 | 312,028 | 12.3 |
| 9 | 330,837 | 13.0 |
| 10 (most deprived) | 337,452 | 13.3 |
| **Total** | 3,441,017 | 100.0 |
| **Age group** | | |
| <1 | 632,664 | 18.4 |
| 1-5 | 904,987 | 26.3 |
| 6-10 | 952,642 | 27.7 |
| 11-15 | 950,724 | 27.6 |
| | **Number of tests** | **%** |
| **Total** | 14,299,598 | 100.0 |
| **Gender** | | |
| Male | 6,504,010 | 45.5 |
| Female | 7,794,755 | 54.5 |
| Other | 833 | 0.0 |
| **Age group** | | |
| Under 1 | 1,460,767 | 10.2 |
| 1-5 | 3,186,435 | 22.3 |
| 6-10 | 4,046,845 | 28.3 |
| 11-15 | 5,605,551 | 39.2 |
| **Type of test** | | |
| Blood | 7,157,882 | 50.1 |
| Imaging | 885,709 | 6.2 |
| Miscellaneous | 6,256,007 | 43.7 |
| | **Median tests per person per year** | **IQR** |
| **Overall** | 2 | 1,3 |
| **Gender** | | |
| Male | 1 | 1,3 |
| Female | 2 | 1,3 |
| **Age group** | | |
| Under 1 | 1 | 1,2 |
| 1-5 | 1 | 1,3 |
| 6-10 | 1 | 1,3 |
| 11-15 | 2 | 1,5 |

Miscellaneous tests include laboratory analysis of non-serum samples (e.g., urine, stool) and physiological measurements.

## Figure 4.1 Temporal trends in paediatric test use in general practice from 2007 to 2019




| Average Annual Percentage Change | | | (% per year [95% CI]) | | (% per year [95% CI]) | | (% per year [95% CI]) | | (% per year [95% CI]) | | (% per year [95% CI]) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **A. Overall.** | | | 3.6 (3.4 to 3.8) | | | | | | | | |
| **B. Test type.** | | **Blood** | 8.0 (7.7 to 8.4) | **Imaging** | 4.0 (3.7 to 4.3) | **Misc.** | -0.7 (-1.1 to -0.3) | | | | |
| **C. Gender.** | | **Female** | 3.6 (3.4 to 3.9) | **Male** | 3.5 (3.2 to 3.7) | | | | | | |
| **D. Age group and gender.** | **Female** | **<1** | 1.2 (0.6 to 1.8) | **1-5** | 3.7 (3.0 to 4.4) | **6-10** | 3.1 (2.3 to 3.9) | **11-15** | 4.0 (3.7 to 4.3) | | |
| | **Male** | **<1** | 1.3 (0.7 to 2.1) | **1-5** | 4.1 (3.4 to 4.6) | **6-10** | 3.2 (2.8 to 3.7) | **11-15** | 3.4 (3.0 to 3.9) | | |
| **E. IMD quintile.** | | **1** | 3.2 (3.0 to 3.5) | **2** | 3.4 (3.1 to 3.8) | **3** | 3.0 (2.8 to 3.4) | **4** | 3.5 (3.1 to 3.9) | **5** | 4.4 (4.1 to 4.8) |

### 4.4.2 Temporal trends in overall test use

The age-adjusted rate of total test use increased from 399 tests per 1,000 child years in 2007 to 608 tests per 1,000 child years in 2019, an overall increase of 52% and an average annual percentage increase of 3.6% per year (AAPC 95% CI 3.4 to 3.8%) which was statistically significant, see **Figure 4.1a**, **Table 4.2**. Test rates initially increased by 5.1% per year (APC 95% CI 4.7 to 5.6%) between 2007 to 2014, then increased by 1.6% per year (APC 95% CI 0.9 to 2.1%) between 2014 and 2019.

**Figure 4.1b** shows temporal changes in test use stratified by test type. The greatest increase was observed for blood tests which increased by 8.0% per year (AAPC 95% 7.7 to 8.4%), followed by imaging tests which increased on average by 4.0% per year (AAPC 95% CI 3.7 to 4.3). Miscellaneous tests (which include urine, stool, and physiological tests) decreased by 0.7% per year (AAPC 95% CI -1.1 to -0.3%).

Rates of test use by gender and age group are shown in **Figure 4.1c** and **4.1d**. Test rates were consistently higher in females compared with males. When stratified by age group, the rates of change were similar for both genders and age groups, except for 11-15-year-old females, where testing increased by 4.0% per year (AAPC 95% CI 3.7 to 4.3%), 0.6% per year higher than 11-15-year-old males in whom testing increased by 3.4% per year (AAPC 95% CI 3.0 to 3.9%).

**Figure 4.1e** demonstrates test use by IMD deprivation quintile, in which a higher index represents greater levels of deprivation. Test rates were highest in the most deprived cohort and increased the most, with an AAPC of 4.4% (95% CI 4.1 to 4.8%), compared with those from the lower quintiles of deprivation.

**Table 4.2 Standardised rates of test use overall; stratified by test type, gender, age, and deprivation quintile; annual percentage change and average annual percentage change from 2007 to 2019**

| | | | Rate* (per 1,000 child years) | | | 95% Confidence Interval | | |
|---|---|---|---|---|---|---|---|---|
| | Start | End | Start | End | APC (%) | Lower Limit | Upper Limit | AAPC (%) and 95% CI |
| Overall | 2007 | 2014 | 399.2 | 561.9 | 5.1 | 4.7 | 5.6 | 3.6 (3.4, 3.8) |
| | 2014 | 2019 | 561.9 | 607.7 | 1.6 | 0.9 | 2.1 | |
| Blood Tests | 2007 | 2014 | 144,6 | 296.7 | 11.1 | 10.3 | 12 | 8.0 (7.7, 8.4) |
| | 2014 | 2019 | 296.7 | 356.9 | 3.7 | 2.8 | 4.6 | |
| Imaging | 2007 | 2010 | 25.3 | 29.8 | 5.9 | 4.3 | 8.8 | 4.0 (3.7, 4.3) |
| | 2010 | 2017 | 29.8 | 34.4 | 2.2 | 1.4 | 2.6 | |
| | 2017 | 2019 | 34.4 | 39.9 | 7.4 | 5.2 | 9.1 | |
| Miscellaneous | 2007 | 2012 | 229.3 | 236.2 | 0.6 | -0.2 | 2.8 | -0.7 (-1.1, -0.3) |
| | 2012 | 2019 | 236.2 | 210.8 | -1.7 | -2.7 | -1.2 | |
| Female | 2007 | 2014 | 448.6 | 638.7 | 5.2 | 4.7 | 5.7 | 3.6 (3.4, 3.9) |
| | 2014 | 2019 | 638.7 | 690.6 | 1.5 | 0.8 | 2.1 | |
| Male | 2007 | 2014 | 356.3 | 491.2 | 4.9 | 4.4 | 5.4 | 3.5 (3.2, 3.7) |
| | 2014 | 2019 | 491.2 | 528.4 | 1.5 | 0.8 | 2.2 | |
| Female | | | | | | | | |
| <1 | 2007 | 2011 | 21.5 | 27.3 | 6.8 | 4.4 | 10.3 | 1.2 (0.6, 1.8) |
| | 2011 | 2019 | 27.3 | 25.5 | -1.5 | -2.5 | -0.6 | |
| 1-5 | 2007 | 2009 | 75.6 | 78.8 | 2.7 | -1.0 | 7.5 | 3.7 (3.0, 4.4) |
| | 2009 | 2014 | 78.8 | 117.9 | 8.6 | 6.5 | 10.0 | |
| | 2014 | 2019 | 117.9 | 118.6 | -0.5 | -2.9 | 0.8 | |
| 6-10 | 2007 | 2019 | 136.3 | 196.4 | 3.1 | 2.3 | 3.9 | 3.1 (2.3, 3.9) |
| 11-15 | 2007 | 2014 | 215.2 | 328.6 | 6.2 | 5.6 | 6.8 | 4.0 (3.7, 4.3) |
| | 2014 | 2019 | 328.6 | 350.1 | 0.9 | -0.1 | 1.9 | |
| Male | | | | | | | | |
| <1 | 2007 | 2011 | 23.5 | 30.0 | 7.0 | 4.3 | 11.5 | 1.3 (0.7, 2.1) |
| | 2011 | 2019 | 30.0 | 28.7 | -1.4 | -2.6 | -0.4 | |

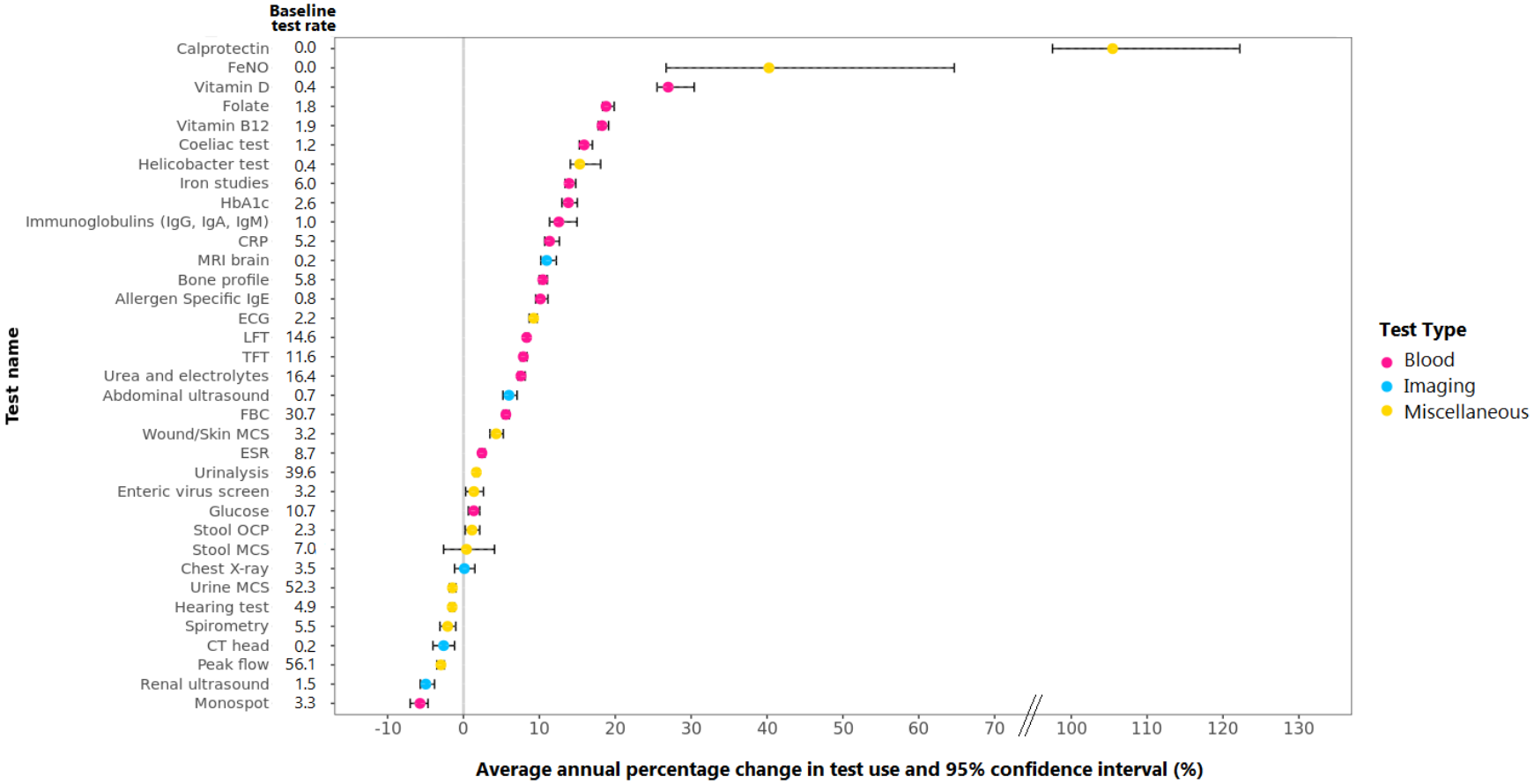
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1-5 | 2007 | 2009 | 64.8 | 66.9 | 2.0 | -1.4 | 5.9 | 4.1 (3.4, 4.6) |
| | 2009 | 2014 | 66.9 | 101.8 | 9.1 | 7.4 | 10.3 | |
| | 2014 | 2019 | 101.8 | 105.2 | 0.1 | -2.1 | 1.3 | |
| 6-10 | 2007 | 2012 | 108.6 | 121.0 | 2.1 | -0.8 | 3.3 | 3.2 (2.8, 3.7) |
| | 2012 | 2019 | 121.0 | 159.1 | 4.0 | 3.3 | 6.7 | |
| 11-15 | 2007 | 2014 | 159.5 | 226.7 | 5.3 | 4.5 | 6.4 | 3.4 (3.0, 3.9) |
| | 2014 | 2019 | 226.7 | 235.5 | 0.9 | -0.9 | 2.1 | |
| IMD 1 (least deprived | 2007 | 2015 | 63.5 | 87.9 | 4.1 | 3.7 | 4.7 | 3.2 (3.0, 3.5) |
| quintile) | 2015 | 2019 | 87.9 | 94.6 | 1.5 | 0.1 | 2.4 | |
| IMD 2 | 2007 | 2014 | 61.4 | 83.5 | 4.4 | 3.9 | 5.6 | 3.4 (3.1, 3.8) |
| | 2014 | 2019 | 83.5 | 92.4 | 2.0 | 0.5 | 2.8 | |
| IMD 3 | 2007 | 2014 | 82.4 | 109.0 | 4.2 | 3.7 | 5.0 | 3.0 (2.8, 3.4) |
| | 2014 | 2019 | 109.0 | 116.7 | 1.5 | 0.4 | 2.2 | |
| IMD 4 | 2007 | 2014 | 88.5 | 130.1 | 5.8 | 5.1 | 6.7 | 3.5 (3.1, 3.9) |
| | 2014 | 2019 | 130.1 | 134.4 | 0.3 | -0.9 | 1.3 | |
| IMD 5 (most deprived | 2007 | 2014 | 103.4 | 154.4 | 6.0 | 5.4 | 7.0 | 4.4 (4.1, 4.8) |
| quintile) | 2014 | 2019 | 154.4 | 169.6 | 2.2 | 0.9 | 3.1 | |

*Observed rates; Abbreviations: APC – annual percentage change; AAPC – average annual percentage change;

### 4.4.3 Temporal trends in specific tests

The average annual percentage change (AAPC) for each test is presented in **Figure 4.2.** Faecal calprotectin testing was subject to the greatest average annual change, increasing from 0 tests/1,000 child-years in 2007 year to 1.8 tests/1,000 child-years by 2019, equivalent to 105.5% per year (AAPC 95% CI 97.5 to 122.2%). This was followed by fractional exhaled nitric oxide (FeNO) tests which increased by 40.3% per year (AAPC 95% CI 26.7 to 64.7%) from 0 tests/1,000 child-years in 2007 to 0.2 tests/1,000 child-years in 2019, then vitamin D tests which increased by 27.0% per year (AAPC 95% CI 25.5 to 30.4%) from 0.5 tests/1,000 child-years in 2007 to 8.5 tests/1,000 child-years in 2019. The following tests increased by greater than 10% per year, in descending order: folate, vitamin $B_{12}$, coeliac testing, helicobacter testing, iron studies, HbA1c, immunoglobulins, C reactive protein, MRI brain, bone profile, and allergen-specific IgE (see **Table 4.3**). Tests that demonstrated a consistent decrease included urine microscopy/culture/sensitivities, hearing test, spirometry, CT head, peak flow measurements, renal ultrasound, and monospot testing for glandular fever. The changes were analysed by gender and age group which showed similar changes (see **Chapter 4 Appendix Figure 4.3**). There were also no large differences in temporal changes by IMD quintile (see **Chapter 4 Appendix Figure 4.4**)

**Figure 4.2 Temporal change in specific tests for children aged 0 to 15 years in English primary care from 2007 to 2019**




Abbreviations: CRP – C reactive protein; CT – Computed tomography; ECG – Electrocardiogram; ESR – Erythrocyte sedimentation rate; FBC – Full blood count; FeNO – Fractional exhaled nitric oxide; LFT – Liver function test; MCS – Microscopy, culture, sensitivities; MRI – Magnetic resonance imaging; NOS – Not otherwise specified; OCP – Ova/cysts/parasites; TFT – Thyroid function test; US – Ultrasound

**Table 4.3 Average annual percentage change and test rates of 35 specific tests for children aged 0 to 15 years in general practice from 2007 to 2019**

| Test name | Test type | AAPC | 95% CI lower limit | 95% CI upper limit | 2007 test rate* (tests/1,000 child years) | 2019 test rate* (tests/1,000 child years) |
|---|---|---|---|---|---|---|
| Calprotectin | Miscellaneous | 105.5 | 97.5 | 122.2 | 0[a] | 1.8 |
| FeNO | Miscellaneous | 40.3 | 26.7 | 64.7 | 0[a] | 0.2 |
| Vitamin D | Blood | 27.0 | 25.5 | 30.4 | 0.4 | 8.5 |
| Folate | Blood | 18.8 | 18.4 | 19.9 | 1.8 | 14.5 |
| Vitamin $B_{12}$ | Blood | 18.3 | 17.7 | 19.1 | 1.9 | 15 |
| Coeliac test | Blood | 15.9 | 15.3 | 17.0 | 1.2 | 7.5 |
| Helicobacter test | Miscellaneous | 15.3 | 14.1 | 18.1 | 0.4 | 2.1 |
| Iron studies | Blood | 13.9 | 13.4 | 14.8 | 6 | 29.5 |
| HbA1c | Blood | 13.8 | 13.0 | 15.0 | 2.6 | 12.3 |
| Immunoglobulins (IgG, IgA, IgM) | Blood | 12.6 | 11.3 | 15.0 | 1 | 4 |
| C reactive protein | Blood | 11.3 | 10.7 | 12.6 | 5.2 | 18.8 |
| MRI brain | Imaging | 11.0 | 10.2 | 12.2 | 0.2 | 0.7 |
| Bone profile | Blood | 10.5 | 10.0 | 11.0 | 5.8 | 19.7 |
| Allergen Specific IgE | Blood | 10.1 | 9.5 | 11.1 | 0.8 | 2.37 |
| Electrocardiogram | Miscellaneous | 9.2 | 8.7 | 9.7 | 2.2 | 6.4 |
| Liver function test | Blood | 8.3 | 8.1 | 8.6 | 14.6 | 38.1 |
| Thyroid function test | Blood | 7.9 | 7.5 | 8.4 | 11.6 | 28.5 |
| Urea and electrolytes | Blood | 7.6 | 7.2 | 8.1 | 16.4 | 38.9 |
| Abdominal ultrasound | Imaging | 6.0 | 5.2 | 7.1 | 0.7 | 1.56 |
| Full blood count | Blood | 5.6 | 5.2 | 6.0 | 30.7 | 58.5 |
| Wound/Skin MCS | Miscellaneous | 4.3 | 3.5 | 5.2 | 3.2 | 5.1 |
| Erythrocyte sedimentation rate | Blood | 2.4 | 2.1 | 2.8 | 8.7 | 11.6 |
| Urinalysis | Miscellaneous | 1.7 | 1.4 | 2.0 | 39.6 | 48 |
| Enteric virus screen | Miscellaneous | 1.4 | 0.3 | 2.6 | 3.2 | 3.8 |
| Glucose | Blood | 1.4 | 0.6 | 2.1 | 10.7 | 12.8 |
| Stool OCP | Miscellaneous | 1.1 | 0.2 | 2.1 | 2.3 | 2.4 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Stool MCS | Miscellaneous | 0.4 | -2.6 | 4.1 | 7 | 6.9 |
| Chest X-ray | Imaging | 0.1 | -1.2 | 1.5 | 3.5 | 3.8 |
| Urine MCS | Miscellaneous | -1.5 | -1.9 | -1.0 | 52.3 | 44.9 |
| Hearing test | Miscellaneous | -1.5 | -1.8 | -1.2 | 4.9 | 4.1 |
| Spirometry | Miscellaneous | -2.1 | -3.1 | -1.0 | 5.5 | 4.6 |
| CT head | Imaging | -2.6 | -4.0 | -1.2 | 0.2 | 0.2 |
| Peak flow | Miscellaneous | -3.0 | -3.5 | -2.5 | 56.1 | 39.4 |
| Renal ultrasound | Imaging | -5.0 | -5.7 | -3.8 | 1.5 | 0.8 |
| Monospot | Blood | -5.8 | -7.0 | -4.7 | 3.3 | 1.6 |

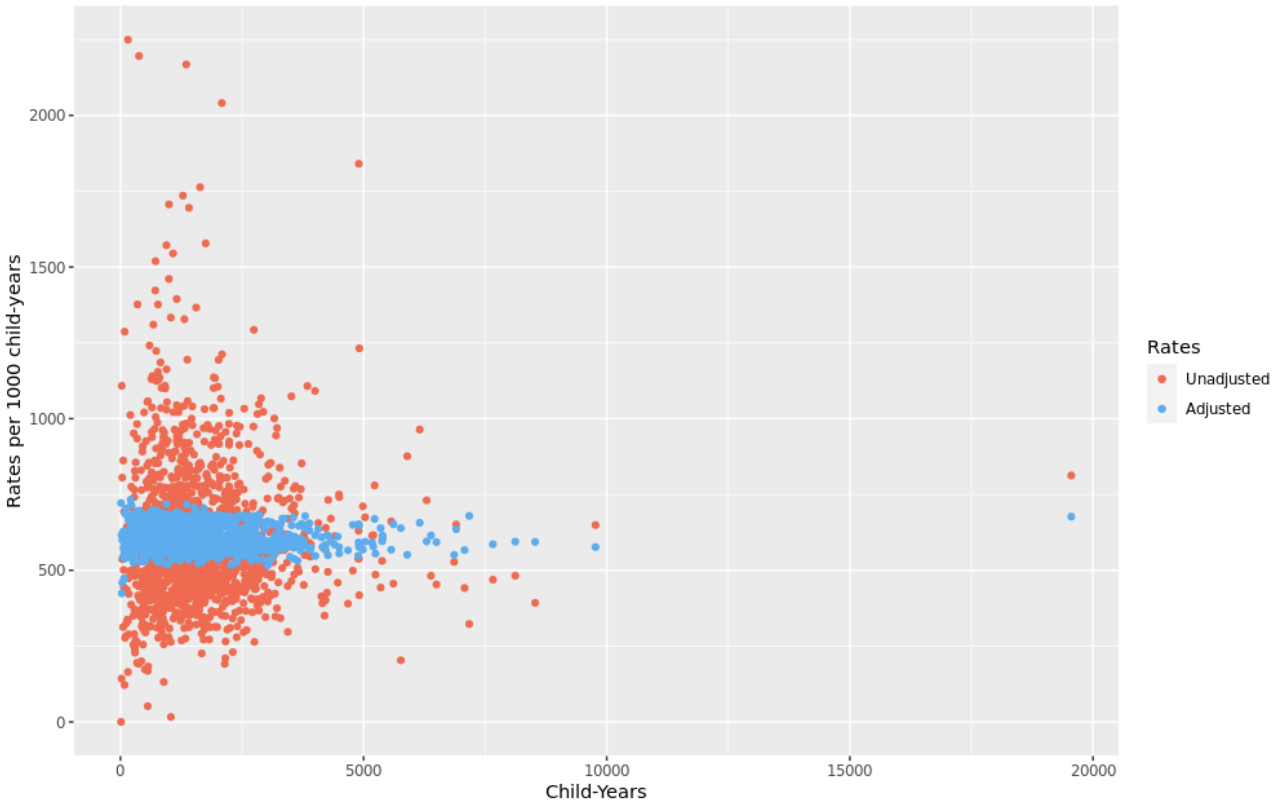
*Observed rates; Abbreviations: AAPC – average annual percentage change; CT – computed tomography; FeNO – fractional exhaled nitric oxide test for asthma; MCS – microscopy, culture, and sensitivities; MRI – magnetic resonance imaging; OCP – ova, cysts, and parasites

[a] There were zero tests in 2007, the first recorded calprotectin test was in 2009 and the first recorded FeNO test was in 2010

### 4.4.4 Practice variation in test use

In 2019, 1,464 practices contributed 2,406,042 child-years of observation (ranging from 11 to 19,553 child-years per practice). The mean rate of test use by practice (adjusted for median age, gender, and deprivation level) was 609 tests per 1,000 child-years (standard deviation [SD] 41). **Figure 4.3** shows the adjusted and unadjusted rates of test use for each practice.

**Figure 4.3 Crude and adjusted practice-specific test request rates for children aged 0 to 15 in 2019; Adjusted for gender, age, and deprivation**




### 4.4.5 Rank order of practice variability of specific tests

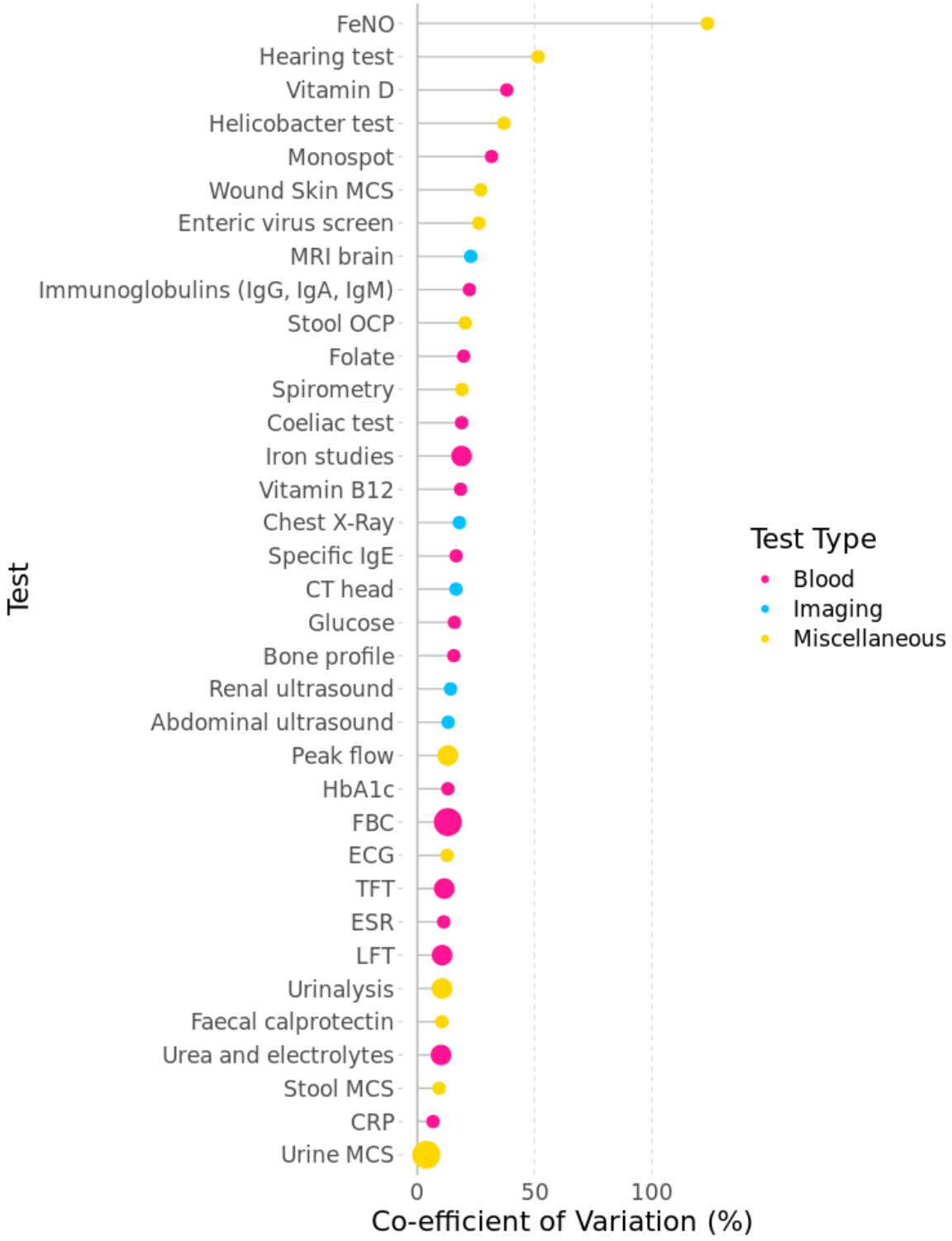
**Figure 4.4** shows the rank order of the tests from highest to lowest practice variability (CoV). FeNO was subject to the greatest practice variability that was substantially higher than the other tests, with an adjusted CoV of 123.7% (95% CI 123.6 to 123.9%). This was followed by hearing tests (CoV 51.6%, 95% CI 51.4% to 51.7%) and Vitamin D tests (CoV

38.1%, 95% CI 38.0 to 38.3%). Tests with higher rates of use (represented by larger bubbles in **Figure 4.4**) were, on average, subject to lower practice variability.

### 4.4.6 Practice variation of tests compared with test rate

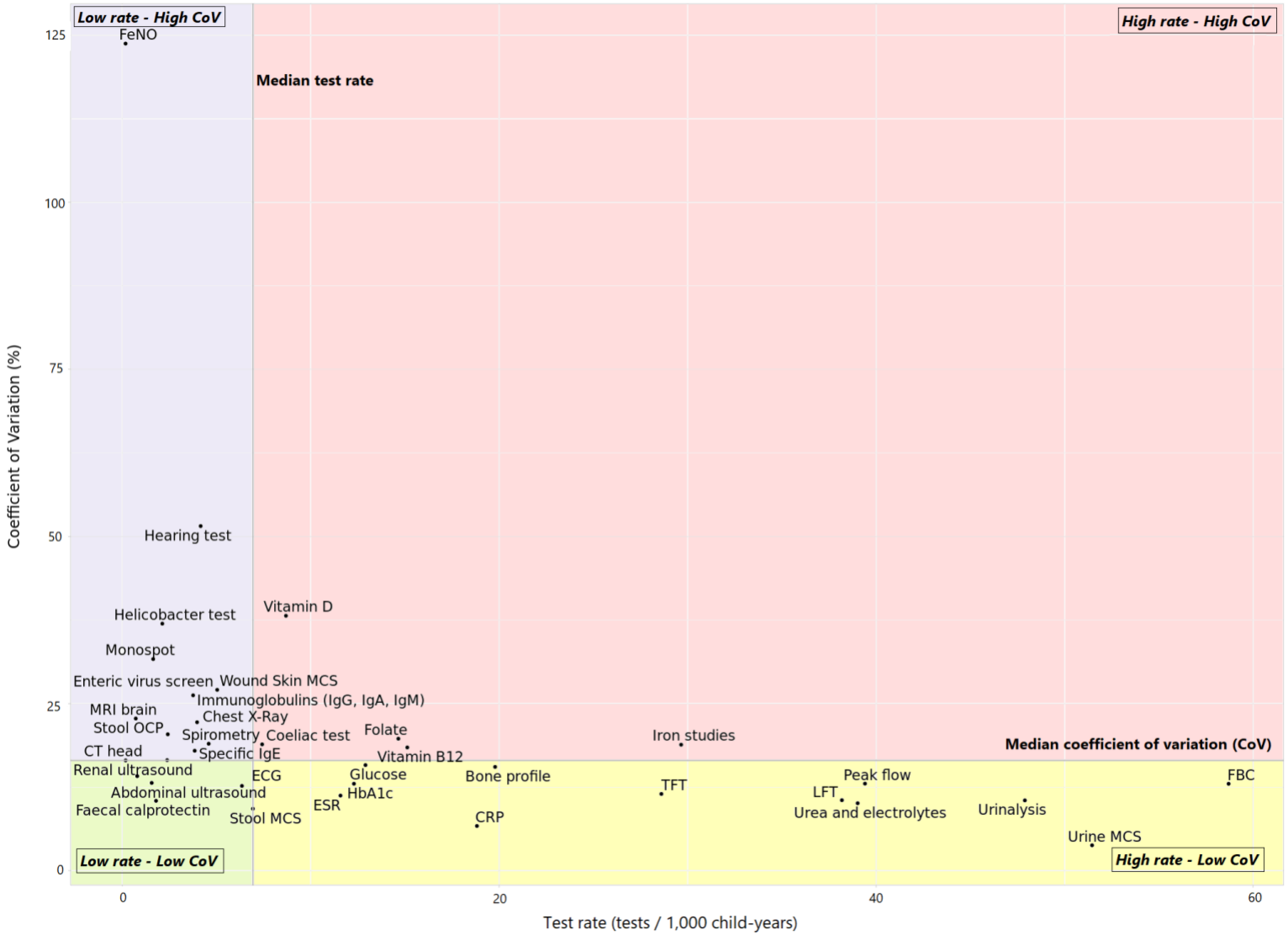
**Figure 4.5** plots the adjusted coefficient of variability of each of the 35 tests against their test rate. The median coefficient of variation of test use was 16.5% (IQR 12.1 to 21.3%) and the rate of test use was 6.9 tests/1000 child-years (IQR 2.4 to 19.3%). Most tests were either classified as low-test rate-high variability (37%, 13 out of 35) or high test rate-low variability (37%, 13 out of 35), see **Chapter 4 Appendix Table 4.2**. The following five tests were classified as high test rate – high variability: Iron studies, Coeliac test, Vitamin $B_{12}$, Folate, and Vitamin D. A strong linear association between the two variables was not observed ($r^2$= 0.12, or 0.21 when excluding FeNO as an outlier due its varying availability in primary care [**Chapter 4 Appendix Figure 4.5**]).

**Figure 4.4 Rank order of between-practice variability of tests in 2019; adjusted for gender, age, and deprivation**



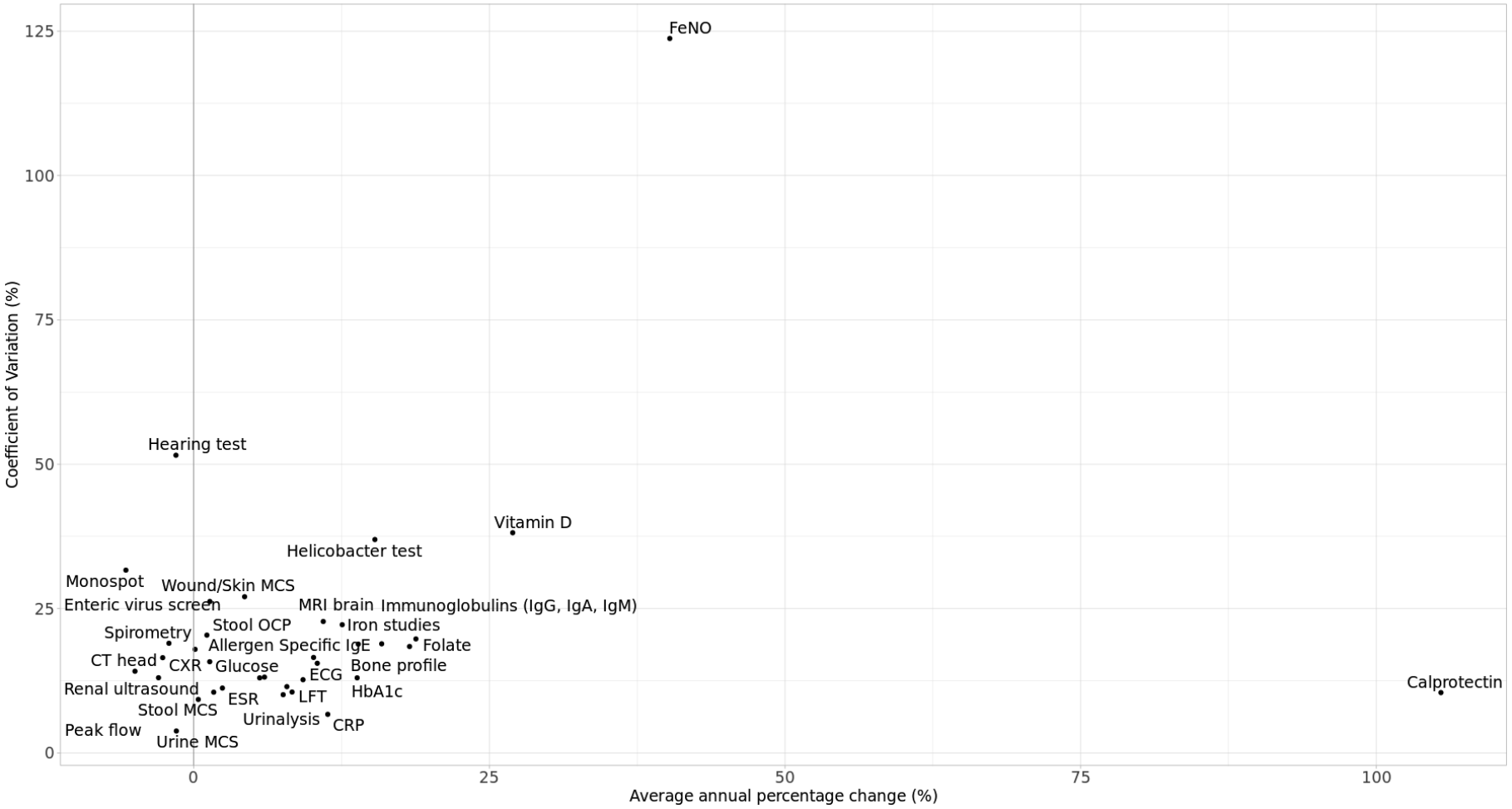

Tests with higher rates of use represented by larger bubbles
Abbreviations: CRP – C reactive protein; CXR – Chest X-ray; ECG – Electrocardiogram; ESR – Erythrocyte sedimentation rate; FBC – Full blood count; FeNO – Fractional exhaled nitric oxide; LFT – Liver function test; MCS – Microscopy, culture, sensitivities; MRI – Magnetic resonance imaging; NOS – Not otherwise specified; OCP – Ova/cysts/parasites; TFT – Thyroid function test

**Figure 4.5 Test rate and degree of practice variability for specific tests in 2019**




Abbreviations: CRP – C reactive protein; CXR – Chest X-ray; ECG – Electrocardiogram; ESR – Erythrocyte sedimentation rate; FBC – Full blood count; FeNO – Fractional exhaled nitric oxide; LFT – Liver function test; MCS – Microscopy, culture, sensitivities; MRI – Magnetic resonance imaging; NOS – Not otherwise specified; OCP – Ova/cysts/parasites; TFT – Thyroid function test

**Figure 4.6 Temporal and practice variability of tests used by children aged 0 to 15**




Abbreviations: CRP – C reactive protein; CXR – Chest X-ray; ECG – Electrocardiogram; ESR – Erythrocyte sedimentation rate; FBC – Full blood count; FeNO – Fractional exhaled nitric oxide; LFT – Liver function test; MCS – Microscopy, culture, sensitivities; MRI – Magnetic resonance imaging; NOS – Not otherwise specified; OCP – Ova/cysts/parasites; TFT – Thyroid function test

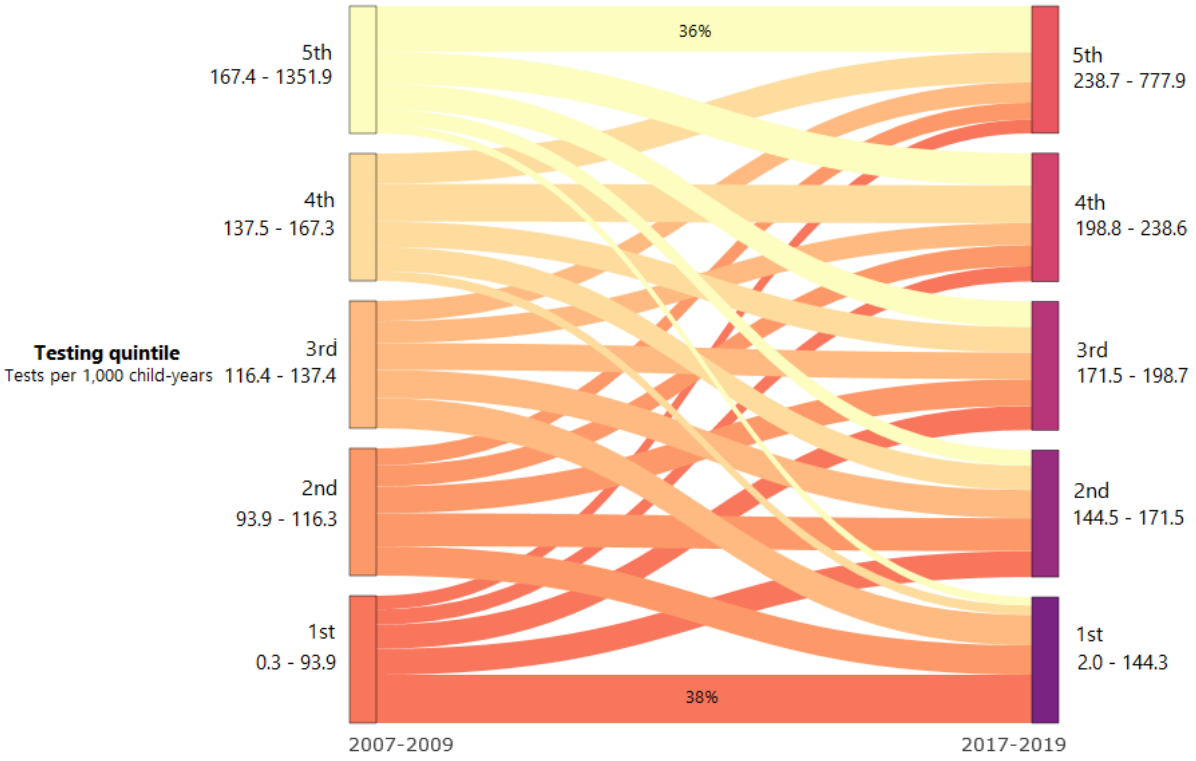
### 4.4.7 Temporal variability and practice variation

The temporal change (AAPC) and practice variation in 2019 (CoV) of all tests were plotted in **Figure 4.6**. There appeared to be an association between the two measures of variation, where a higher degree of temporal change was positively correlated with a higher degree of practice variation, except for calprotectin which had a relatively low CoV despite having the greatest temporal increase. Both FeNO testing and calprotectin were introduced during the study period but while calprotectin had relatively low CoV in 2019, there was a very high degree of practice variability in FeNO testing. No linear association was observed between AAPC and CoV ($r^2$= 0.04) when including all the tests. A strong association between the two variables may have indicated that temporal increases in overall test use were driven by certain practices, however, this was not observed. After excluding calprotectin as an outlier, the $r^2$ co-efficient increased to 0.34 which suggests a moderate association, however, when excluding both FeNO and calprotectin as potential outliers, the $r^2$ co-efficient was 0.02 (see **Chapter 4 Appendix Figure 4.6** for the plotted linear regression models).

### 4.4.8 Practice testing rates over time

The change in testing rates of practices from the beginning of the study period to the end of the study period is represented in **Figure 4.7**. 36% of practices that were categorised as being in the highest quintile of testing remained in the highest quintile. Similarly, 38% of practices in the lowest quintile of testing at the beginning remained “low testers”. There was a substantial degree of flux in testing rates for other practices, though only a minority moved between the highest and lowest quintiles (7-11%).

**Figure 4.7 Temporal trends in testing rates of practices by quintile**




## 4.5 Discussion

### 4.5.1 Statement of principal findings

In this descriptive study, I investigated temporal trends and practice variation in paediatric test use between 2007 and 2019. The findings from this chapter which analysed 14 million tests in 2.5 million children across nearly 1500 GP practices showed that testing rates increased at a rate of 4% per year. Blood tests increased by the highest margin, and females aged 11-15 experienced the most consistent increase, as well as children from more deprived areas in England. Tests with the biggest temporal increases included: faecal calprotectin, FeNO, vitamin D, folate, vitamin $B_{12}$, coeliac, and helicobacter tests. Tests that were subject to the largest practice variation include fractional exhaled nitric oxide (FeNO), hearing tests, vitamin D, helicobacter testing, and monospot testing for

glandular fever. I also identified tests that had an above-average rate of test use and variability: iron studies, vitamin $B_{12}$, coeliac test, folate, and vitamin D.

### 4.5.2 Strengths and limitations in relation to previous work

These findings address the generalisability limitations of the previous chapter set in Oxfordshire by extending the study to general practices across England. In addition to blood tests, this chapter also includes other types of diagnostic tests including imaging, urine, stool, and other physiological measurements. I also analysed between-practice variation to explore another dimension of variation and improve understanding of paediatric test variation. The previous chapter analysed test use by age and gender; this chapter additionally explores the impact of socioeconomic status by including the deprivation level of the practice to which patients were registered.

The findings of this chapter concur with the findings of increased blood test use in Oxfordshire general practices; specifically for vitamin D, folate, vitamin $B_{12}$, iron studies, coeliac testing, HbA1c, bone profile, CRP, thyroid function tests, urea and electrolytes, and liver function tests (**Figure 3.7**). Notably, the increases were more pronounced in this study compared with the previous chapter. In addition, I identified other blood tests that demonstrated a temporal increase in general practice: allergen-specific IgE, full blood count, erythrocyte sedimentation rate and glucose (the latter had been excluded in the previous chapter). It is unclear why allergen-specific IgE was increasing in the broader population and not in Oxfordshire. This may be the result of local policies discouraging allergy testing in general practice or the referral of children to allergy services for testing.

The increasing use of both calprotectin and FeNO tests reflect the implementation of these tests in primary care during the study period. **Figure 4.6** shows that by 2019 there was little practice variability in calprotectin tests but a large degree of variation for FeNO testing, therefore, these two tests serve as interesting case studies. Several factors may explain this discrepancy in practice variability in 2019, including the ease with which the different test samples can be obtained and analysed. Additionally, it could be due to the lack of equipment, time, access, and funding for FeNO testing in primary care despite evidence of their feasibility and acceptability, and NICE guidance recommending FeNO testing for the diagnostic workup of childhood asthma.[9,10]

The increase observed in vitamin D tests over this 13-year study period corresponds to the 30-fold increase observed in Australian general practice over 16 years from 2003 to 2018.[11] Similar studies that can corroborate this chapter's findings for other tests, such as the eight-fold increase in folate and vitamin $B_{12}$ tests, six-fold increase in coeliac tests, and five-fold increase in iron studies have not been identified. To our knowledge, this is the first nationwide study to report on increased paediatric testing in general practice.

A major strength of this study is the use of CPRD data which has been shown to be representative of the UK population due to its breadth and coverage, is quality-assured, and has been extensively validated for use in observational research.[12] Most previous work on paediatric diagnostic testing focuses on variation in testing for a specific test or clinical condition or is limited to specific centres or geographic areas.[11,13–16] For example, in a single-centre study of paediatricians in Netherlands, there was high intra- and inter-observer variability and 67% guideline adherence in the diagnostic workup of children with recurrent abdominal pain.[13] Similar hospital-based studies that focused on specific conditions such as bronchiolitis[17], pneumonia[18], and head injury[19] demonstrated high

variation in diagnostic practices. An international study of five European emergency departments (ED) found high variability in diagnosis and management for consecutive paediatric presentations; the ED in Vienna performed lab tests in 36% of presentations against 9% in London.[15] Additionally, there were substantial differences in the rates of imaging, admission, IV medication, and fluid administration.[15] Patterns of management were reported to be hospital-specific across a range of clinical presentations, suggesting that underlying differences in paediatric healthcare delivery were due to a combination of factors including differences in primary care systems, health culture, availability of resources, and patient expectations.

The methods used in this paper are robust and provide interpretable measures of changes over time (AAPC) and variation between practices (CoV). These indirect measures of overuse were used in prior studies that utilised CPRD data to measure diagnostic test variation in adults. [20,21] Joinpoint regression is used in many population-based studies analysing trends over time with changing population parameters.[20,22] It does not, however, handle missing data, so years in which there were zero tests (including calprotectin and FeNO) had to be addressed individually by imputation, as specified in the methods. The CoV provides a relative measure of variability that allows for comparison across practices and tests. It does not become artificially inflated with a larger sample size, as standard deviations do, but since the CoV is a ratio, it can become inflated as the test rate approaches zero. Other measures of variation, such as the systematic component of variation (SCV) and the chi-squared statistic $\chi^2$ have been shown to be more robust measures of variation in small area variation studies compared to the CoV[23], however, they compare an observed value to the expected value, and there are no known indicators to represent the expected values of testing for the specified tests in our analysis.

CPRD data relies on the quality of the electronic data input by clinicians using electronic health record software. The quality of coding has evolved over time with increased automation. As a consequence, miscoding and misclassification may have occurred, particularly earlier in the study period, leading to missed tests. Double counting could also have occurred if both the request and the completed test were coded separately, or if tests that were performed in specialist outpatient settings were transcribed into the electronic health record by the treating GP from clinic letters and results. It is unclear whether these coding issues affected our study dataset but if this were the case then they would impact the accuracy of the reported testing rates and trends in this chapter. For example, changes in how hearing tests were coded during the study period could offer a possible explanation for their observed temporal decline and high practice variability in 2019.

The code list of included tests (and their associated panels) was developed and cross-checked with 1) existing NHS trust laboratory test lists, 2) the test codes from Chapter 3 (which was corroborated by a consultant chemical pathologist) and 3) checked with another clinician-researcher. However, some of the coding choices of which terms to include/exclude, and which test panel to assign were subjective. For example, some tests were assigned to a grouped test panel, such as “Immunoglobulins (IgA, IgG, IgM)” or “Medication levels” which included multiple test components. These grouped codes may have obscured more subtle trends, in contrast with single condition-specific test codes, like the monospot test for glandular fever. Condition-specific test codes can reveal more about the diagnostic strategies for a particular diagnosis, e.g., the decline in monospot testing for glandular fever may reflect a change in clinician preferences for different diagnostic strategies, such as clinical diagnosis or Epstein-Barr Virus serology for antibodies.
Some tests were assigned unspecified codes, for example, the Medcode term “Microscopy, culture and sensitivities” was coded as “Microbiology NOS (not otherwise

specified)" as it was not possible to determine what type of test was requested. This specific test code comprised 4.6% of the total tests requested (as shown in **Chapter 4 Appendix Figure 4.2**). Other examples of unspecified codes included "Urine test NOS", "Ultrasound NOS" and "Diagnostic test NOS". Altogether, these unspecified codes comprised 8.2% of the total test data. I attempted to identify these unspecified tests by examining the tests that the unspecified codes co-existed with (for the same patient on the same day) and if there were test codes that should have been counted as 1 test rather than multiple, these were recoded. An example of this was that Microbiology NOS was recoded as a Stool MCS if they occurred on the same day for the same patient. These test codes were then de-duplicated. Despite taking these measures, it is possible that there were errors in this process of test assignment, potentially influencing the accuracy of the observed trends.

Testing rates were found to be largely influenced by gender, age, and deprivation level, however, there are other factors that can contribute to variation and were not explored within these data. These factors include varying levels of disease prevalence, patient preferences, local policies and pathways, availability and access of tests, uptake and adoption of newer diagnostic technologies, and clinician preferences in the setting of uncertain evidence.[24]

### 4.5.3 Implications for clinical practice and research

Increased requests for haematinic tests (vitamin $B_{12}$, folate, iron) suggest increasing clinician concern for nutritional deficiencies and anaemia. Though there are no longitudinal studies investigating trends in childhood and adolescent anaemia in the UK, the Global Burden of Disease 2019 study reported that the prevalence of anaemia in high-income

countries increased from 10% in 2000 to 12% in 2019 in children aged 6 – 59 months, increased from 11% in 2000 to 13% in 2019 in non-pregnant women aged 15-49 years.[25] Haematinic tests, in addition to testing for coeliac disease, helicobacter pylori, vitamin D, CRP, and HbA1c may also be requested for more non-specific symptoms or 'unexplained complaints', including fatigue, musculoskeletal issues or abdominal symptoms. Similar increases have been observed in adults.[20] In a cross-sectional study of Dutch general practitioners, GPs were 26 times more likely to request at least one blood test in patients with fatigue, and 2.5 times more likely to order a blood test in patients with musculoskeletal complaints, compared to patients with abdominal complaints. Factors associated with an increased likelihood of blood test requests in patients with unexpected complaints were fatigue, duration of complaint for over four weeks, the absence of a psychosocial contributor, and the absence of a working hypothesis of a syndrome.[26]

Increased coeliac testing corresponds with higher rates of coeliac disease.[4,27] A population-based cohort study using CPRD data from 2000 to 2019 also identified that there was an increasing incidence of Type 1 diabetes mellitus and other autoimmune diseases. In most cases, the incidence of autoimmune diseases was higher in more deprived socioeconomic groups (with the exception of coeliac disease for which there was a higher incidence in the least deprived socioeconomic group).[4] Our study findings of increased testing rates correlate with higher rates of diagnosis of autoimmune disease (excluding coeliac disease) in more deprived socioeconomic groups, although our study is limited to children. It is unclear whether there is a causal link, i.e., if increased testing is leading to higher detection of disease, or increasing prevalence is resulting in more testing, or both are occurring concurrently; this needs to be studied prospectively.

Previous research has shown that GPs who are less comfortable with uncertainty tend to request more tests and make more hospital referrals which contribute to increased healthcare costs.[28] Factors that have been found to be associated with greater intolerance of diagnostic uncertainty in GPs and trainees include having less experience, as well as clinicians who are less resilient and more prone to burnout. [29,30] Diagnostic uncertainties may also arise from sub-optimal paediatrics exposure and experience in the GP training program, highlighting areas for future policy intervention.[31]

Further research is needed to understand the implications of this study's findings and examine the appropriateness of increased testing by comparing the clinical indication of the test with evidence-based guidance (such as NICE). In the previous chapter, I examined the indications for Vitamin D testing and identified that observed increases were likely to represent overuse. Future research should examine the specific testing indications for vitamin $B_{12}$, folate, iron studies, coeliac disease, helicobacter pylori and HbA1c testing which increased significantly in the 14-year study period and determine whether testing was guideline concordant or discordant. Similarly, analysing test results to determine if increasing test rates correspond with increasing abnormal results and possible disease can highlight potential overuse. For example, the study by Zurynski et al described earlier, which reported a 30-fold increase in vitamin D testing across Australian general practices, found that the odds of detecting low levels remained low, further demonstrating that it was likely requested inappropriately. [11]

This study did not examine whether increased testing has led to an increase in referral rates. A follow-up analysis of the data from this study cohort will be undertaken to address this research question. Understanding the drivers and appropriateness of referrals is complex; they are likely to be influenced by child's needs, parental or caregiver requests,

and the GP's level of comfort in managing the patient in primary care. If an increase in referrals to hospitals is observed alongside increases in testing rates, then testing may be considered a driver of increasing hospital activity.

I specifically looked at test utilisation in the period preceding the pandemic, however, testing rates likely declined substantially during the pandemic with decreased paediatric consultations in primary care. In the case that test rates have recovered since the pandemic, then if test rates were to continue at the rate of growth since 2014 (APC 1.6%), then by 2024 test rates in general practice would be 656 tests/1,000 children. This has considerable cost implications. A Vitamin D test costs approximately £10 according to the Oxford University Hospitals NHS Trust Laboratories. Applying the rates of vitamin D requests to the rest of the UK population in 2019 (according to the mid-2019 ONS estimates – 12,697,836 people aged 0 to 15), vitamin D tests requested for children in general practice cost £2.2 million in 2019. If rates continued to increase at the AAPC of 29.6% per year, by 2024, the rate of vitamin D testing may be estimated at 42.3 tests/1,000 child-years, costing £5.4 million across UK primary care. Folate tests costs £3.40/unit, equating to an estimated cost of £1.3 million in 2019, and if the rate of testing continues to increase at 19.3%/year, it would cost an estimated £2.5 million in 2024. CRP tests cost £1.40 per unit, costing approximately £670,000 in 2019 and forecasted to cost £1.1 million by 2024 if the AAPC remains 11.7%/year. While these are estimates (and assume testing rates recovered post-pandemic), they provide an indication of the potential scale of the financial consequences if testing rates continued to increase. Future research is therefore needed to evaluate the economic implications of increased testing in children.

### 4.5.4 How this chapter fits within the wider thesis

In this chapter, I aimed to assess the generalisability of the findings from Chapter 3 by examining paediatric test use across general practices in England. I examined trends in paediatric diagnostic test use by test type, gender, sex, and socioeconomic level of deprivation. I identified tests that increased substantially in their use since 2007. I also compared test use between different practices and identified tests that were subject to greatest practice variability. I examined the influence of demographics on test variation; however, there are other possible drivers of variation that I will explore in the following chapter.

## 4.6 Conclusions

The analyses presented in this chapter provide a broader picture of testing practices based on individual-level data across general practices in England. Increased testing rates can generate more clinical activity including more specialist referrals, and the potential cost implications are substantial. Future research should compare tests against clinical guideline standards and examine the test results to judge whether test increases are warranted or unwarranted.

## 4.7 Chapter Summary

- I used CPRD data to analyse temporal trends and practice variability in paediatric test use across general practices in England from 2007 to 2019 and examined the influence of demographic factors and socioeconomic status on test rates.
- I presented tests in rank order of tests by their magnitude of temporal increase and practice variability and identified (1) tests subject to the greatest increase over

time, (2) the highest variation in use between general practices as well as (3) those that have a higher-than-average rate of use and variability.

- The greatest temporal increases were observed in faecal calprotectin and fractional exhaled nitric oxide (FeNO) testing, which were both introduced in primary care during the study period, followed by vitamin D tests.
- Tests with the highest practice variability included FeNO, hearing tests, and vitamin D levels.
- Tests identified as high test rate-high variability were iron studies, coeliac test, vitamin $B_{12}$, folate, and vitamin D.
- This results from this chapter represent the first UK-wide descriptive study of paediatric test use in general practice. Future research should evaluate whether temporal increases and variation is warranted on a test-by-test basis by comparing testing to evidence-based indicators or examining test results.

## 4.8 References


1. UK Department of Health. Report of the Review of NHS Pathology Services in England Chaired by Lord Carter of Coles; 2006.
2. Wolf A, Dedman D, Campbell J, et al. Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. Int J Epidemiol. 2019;48(6). doi:10.1093/ije/dyz034.
3. Clinical Practice Research Datalink. (2022). CPRD Aurum February 2022 (Version 2022.02.001). Clinical Practice Research Datalink. https://doi.org/10.48329/gcgx-f815.
4. Conrad N, Misra S, Verbakel JY, et al. Incidence, prevalence, and co-occurrence of autoimmune disorders over time and by age, sex, and socioeconomic status: a population-based cohort study of 22 million individuals in the UK. The Lancet. 2023;401(10391). doi:10.1016/S0140-6736(23)00457-9.
5. National Cancer Institute, Division of Cancer Control and Population Sciences. Joinpoint Regression Program. USA: Surveillance Research Program. Published online 2020.
6. R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
7. Xu J, Lin Y, Yang M, Zhang L. Statistics and pitfalls of trend analysis in cancer research: A review focused on statistical packages. J Cancer. 2020;11(10). doi:10.7150/jca.43521.
8. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. PLoS Med. 2015;12(10). doi:10.1371/journal.pmed.1001885.
9. Lo D, Beardsmore C, Roland D, et al. Spirometry and FeNO testing for asthma in children in UK primary care: A prospective observational cohort study of feasibility and acceptability. British Journal of General Practice. 2020;70(700). doi:10.3399/BJGP20X713033.
10. NICE. Overview | Asthma: diagnosis, monitoring and chronic asthma management | Guidance | NICE. National Institute for Health and Care Excellence. 2017;(November).
11. Zurynski Y, Munns CF, Sezgin G, Imai C, Georgiou A. Vitamin D testing in children and adolescents in Victoria, Australia: are testing practices in line with global

recommendations? Arch Dis Child. Published online May 17, 2023:archdischild-2022-325000. doi:10.1136/archdischild-2022-325000.

12. Herrett E, Gallagher AM, Bhaskaran K, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015;44(3). doi:10.1093/ije/dyv098.
13. van Kalleveen MW, Noordhuis EJ, Lasham C, Plötz FB. Large variation in clinical practice amongst pediatricians in treating children with recurrent abdominal pain. Pediatr Gastroenterol Hepatol Nutr. 2019;22(3). doi:10.5223/pghn.2019.22.3.225.
14. Sheridan DC, Meckler GD, Spiro DM, Koch TK, Hansen ML. Diagnostic testing and treatment of pediatric headache in the emergency department. Journal of Pediatrics. 2013;163(6). doi:10.1016/j.jpeds.2013.07.006.
15. Ropers F, Bossuyt P, MacOnochie I, et al. Practice variation across five European paediatric emergency departments: a prospective observational study. BMJ Open. 2022;12(3). doi:10.1136/bmjopen-2021-053382.
16. Zanetto L, van de Maat J, Nieboer D, et al. Diagnostic variation for febrile children in European emergency departments. Eur J Pediatr. 2022;181(6). doi:10.1007/s00431-022-04417-8.
17. Lirette MP, Kuppermann N, Finkelstein Y, et al. International variation in evidence-based emergency department management of bronchiolitis: a retrospective cohort study. BMJ Open. 2022;12(12). doi:10.1136/bmjopen-2021-059784.
18. Florin TA, French B, Zorc JJ, Alpern ER, Shah SS. Variation in emergency department diagnostic testing and disposition outcomes in pneumonia. Pediatrics. 2013;132(2). doi:10.1542/peds.2013-0179.
19. Niele N, Willemars L, van Houten M, Plötz FB. National survey on managing minor childhood traumatic head injuries in the Netherlands shows low guideline adherence and large interhospital variations. Acta Paediatrica, International Journal of Paediatrics. 2018;107(1). doi:10.1111/apa.14076.
20. O'Sullivan JW, Stevens S, Hobbs FDR, et al. Temporal trends in use of tests in UK primary care, 2000-15: Retrospective analysis of 250 million tests. The BMJ. 2018;363. doi:10.1136/bmj.k4666.
21. O'Sullivan JW, Stevens S, Oke J, et al. Practice variation in the use of tests in UK primary care: A retrospective analysis of 16 million tests performed over 3.3 million patient years in 2015/16. BMC Med. 2018;16(1). doi:10.1186/s12916-018-1217-1.

22. Clift AK, Coupland CAC, Hippisley-Cox J. Prostate-specific antigen testing and opportunistic prostate cancer screening: A cohort study in England, 1998–2017. British Journal of General Practice. 2021;71(703). doi:10.3399/bjgp20X713957.
23. Ibãez B, Librero J, Bernal-Delgado E, et al. Is there much variation in variation? Revisiting statistics of small area variation in health services research. BMC Health Serv Res. 2009;9. doi:10.1186/1472-6963-9-60.
24. Public Health England. The 2nd Atlas of Variation in NHS Diagnostic Services in England. NHS RightCare. 2017;(January).
25. Stevens GA, Paciorek CJ, Flores-Urrutia MC, et al. National, regional, and global estimates of anaemia by severity in women and children for 2000–19: a pooled analysis of population-representative data. Lancet Glob Health. 2022;10(5). doi:10.1016/S2214-109X(22)00084-5.
26. Koch H, van Bokhoven MA, ter Riet G, et al. What makes general practitioners order blood tests for patients with unexplained complaints? A cross-sectional study. European Journal of General Practice. 2009;15(1). doi:10.1080/13814780902855762.
27. Zingone F, West J, Crooks CJ, et al. Socioeconomic variation in the incidence of childhood coeliac disease in the UK. Arch Dis Child. 2015;100(5). doi:10.1136/archdischild-2014-307105.
28. Alam R, Cheraghi-Sohi S, Panagioti M, Esmail A, Campbell S, Panagopoulou E. Managing diagnostic uncertainty in primary care: A systematic critical review. BMC Fam Pract. 2017;18(1). doi:10.1186/s12875-017-0650-0.
29. Nevalainen M, Kuikka L, Pitkälä K. Medical errors and uncertainty in primary healthcare: A comparative study of coping strategies among young and experienced GPs. Scand J Prim Health Care. 2014;32(2). doi:10.3109/02813432.2014.929820.
30. Cooke GPE, Doust JA, Steele MC. A survey of resilience, burnout, and tolerance of uncertainty in Australian general practice registrars. BMC Med Educ. 2013;13. doi:10.1186/1472-6920-13-2.
31. Modi N, Simon C. Child health care: Adequate training for all UK GPs is long overdue. British Journal of General Practice. 2016;66(646). doi:10.3399/bjgp16X684853.

# Chapter 5. General practitioners' perspectives on diagnostic tests for children: a qualitative interview study

## 5.1 Publication statement

This chapter is based on my article which has been accepted for publication in the *British Journal of General Practice* (**Appendix 1.1**).

## 5.2 Introduction

85% of healthcare contacts for children aged below 15 occur in general practice[1], and this age group comprises one-tenth of a general practitioner's consultation workload in the UK.[2] General practitioners encounter different paediatric diagnostic challenges compared with other clinicians as children present with more undifferentiated symptoms, and common presentations like abdominal pain, headache, and fatigue may have no identifiable underlying cause. GPs must balance the risks of over-investigation and unnecessary referrals, with missing or delaying a diagnosis.

It is difficult to achieve the right balance of care, and the threshold to test, treat or refer varies among clinicians.[3] Consequently, there is substantial variation in the health care delivered to children.[4] In the preceding chapter, I found that overall paediatric test use in UK general practice increased by 52% between 2007 and 2019. There was also a substantial growth in requests for specific tests including folate, vitamin $B_{12}$, coeliac test, iron studies, and especially vitamin D, which was subject to a 32-fold increase. Variability in test use between general practices was also observed for tests such as fractional exhaled nitric oxide (FeNO) for asthma, hearing tests, vitamin D levels, and helicobacter pylori tests.

In the previous chapter, I explored how demographic differences such as gender, age, and socioeconomic status contribute to paediatric test variation in general practice. Other reported measures that generally drive variation in health care include disease prevalence, resource availability, and clinician and patient preferences.[5,6] Whilst some degree of variation in test use is expected and in keeping with population demographics and disease prevalence, unwarranted variation highlights tests that may be either low value or delivered inequitably.[7]

Seeking general practitioners' (GPs') views on paediatric testing and variation is important for several reasons:

1) They encounter more diagnostic uncertainty than hospital-based clinicians in caring for children as their population is more unselected and the probability of encountering disease is lower.
2) There are widespread differences in how testing processes can occur in general practice due to the structure of practices being run autonomously within specific deaneries
3) GPs can comment and offer possible explanations on the previously described testing trends as it pertains to their testing behaviours
4) GPs may identify specific tests that they believe are subject to more differences in use and offer possible explanations as to why there is greater variation with these tests.

While it is important to establish where variation in testing exists, it is especially important to understand the factors that drive variation, as they highlight areas where clinicians, researchers, and policymakers may target efforts to improve testing.

## 5.3 Aims and objectives

The overall aim of the research reported in this chapter was to explore GP perspectives on using diagnostic tests for children in primary care and understand the factors that underpin their diagnostic decisions.

The specific objectives were:

1. To understand factors that influence general practitioners' choices to request tests for children.
2. To explore potential reasons for variation in test use between general practices.

## 5.4 Methods

### 5.4.1 Study design

I conducted a qualitative study using semi-structured telephone or online interviews with general practitioners in England. I decided to use a qualitative approach to address the research aim as there is a paucity of literature addressing this issue, and it is an appropriate method to explore people's experiences and perceptions of phenomena. The study received ethics approval from the University of Oxford Medical Sciences Interdivisional Research Committee (R85324/RE001).

### 5.4.2 Sampling and recruitment

I recruited GPs and GP trainees through GP groups on social media, GP email lists, and through informal contacts within the authors' academic primary care department. To address potential volunteer bias (i.e., GPs who test more thoughtfully and are interested in test variation may be more likely to participate), I contacted the senior chemical pathologist at the Oxford University Hospital clinical laboratories (BS) to identify GPs who requested the most tests in Oxfordshire and email them with our advertising material. From these

initial contacts, I conducted “snowball” sampling to identify additional participants. Of those who agreed to participate by filling in the expression-of-interest form, a purposive sample was selected to obtain a maximum variation sample of GPs from diverse backgrounds, including gender, location, practice partner status (i.e., partner, salaried GP, locum, trainee) and years since GP qualification for non-trainees. Information sheets and consent forms were emailed in advance of the interview. I obtained verbal consent at the start of the interview; I then signed the consent form on the participant’s behalf and then scanned and sent the form to each participant after the interview.

### 5.4.3 Data collection

I developed a topic guide based on available literature, expertise within the research team (MG, GH, PG, CH), and input from the PPI advisory group. Initial questions explored the GP’s paediatric consultation workload, testing frequency, and most frequently performed tests in children. The topic guide also included questions relating to the considerations that a GP makes when deciding whether to request a diagnostic test and their perspectives on what drives variation in test use between clinicians. A pilot interview was completed with a GP who had extensive experience of conducting qualitative interviews with GP participants. This helped me to refine the interview guide and hone my interview techniques. The topic guide (see **Chapter 5 Appendix 5.1**) was modified iteratively as the interviews progressed, based on new issues that emerged from the interviews.

I conducted the semi-structured interviews between April and June 2023. Data collection ended when, in consultation with the research team, I determined that adequate information power for the conceptualised themes had been achieved. [8]

### 5.4.4 Data analysis

The audio recordings were transcribed verbatim by Bristol Transcription and Translation Services, a university-authorised transcription service. I checked the accuracy of each transcript and anonymised potentially identifiable information. I analysed the data using reflexive thematic analysis as recommended by Braun and Clarke.[9,10] This method provided an effective means of answering the study's objectives by capturing individuals' experiences within the organisational context of general practice. It also allowed for the recognition of patterns across the data set, to formulate a set of practical and actionable outcomes to apply to clinical practice.[10] The analysis was supported by NVivo software to organise, manage, and analyse the data (version 1.6).[11] I read each transcript multiple times to ensure familiarity with the data and coded all the transcripts. I also maintained a diary during the study design, interview, and analysis process to reflect on my subjectivity and reflexivity as a researcher. I shared some of these reflections and engaged in discussions with my supervisory team about the influences of my previous clinical experiences working in paediatrics and research experience in overdiagnosis.

Initial inductive coding was developed into a coding framework, which was iteratively developed throughout the analysis. The relationships between codes were explored and developed into categories and themes. These were organised into a thematic mind map as recommended by Braun and Clarke[9] to explore connections between concepts, themes, and sub-themes. The themes were repeatedly checked for commonality and coherence amongst the data to assess whether the individual themes accurately represented the entire body of data and related to each other appropriately [9] I shared and discussed the findings with the research team, which included three PPI panel members.

## 5.5 Results

### 5.5.1 Characteristics of included participants

Twenty GPs participated in the interviews between April and June 2023 (interview duration 27 to 82 minutes). Their characteristics are summarised in **Table 5.1**.

The number of GPs who responded to the study advertisement and participated in the study are provided in **Table 5.2**. Three participating GPs were identified as 'high testers' in Oxfordshire, based on county testing numbers in 2019 (preceding the COVID-19 pandemic). There was a roughly equal proportion of female (55%) and male GPs (45%). There were six GP partners, seven salaried GPs, seven locum GPs and two GP trainees. The span of GP experience ranged from 0 to 32 years.

The results of the interviews are presented under the following main themes:

- How decisions to test differ in children compared with adults
- Perceived utility of tests
    - General considerations
    - Test-specific considerations
- Perceived drivers of variation
    - Intrinsic factors
    - Extrinsic factors

**Table 5.1 Self-reported characteristics of participating GPs and GP trainees and their respective practice populations (n=20)**

| Characteristic | n (%) |
|---|---|
| Gender | |
| Female | 11 (55) |
| Male | 9 (45) |
| GP status[Ω] | |
| Partner | 6 |
| Salaried | 7 |
| Locum | 7 |
| Trainee | 2 |
| Time since qualification (non-trainees) | |
| <5 years | 6 (33) |
| 5-14 years | 6 (33) |
| >15 years | 6 (33) |
| Geographic area* | |
| Urban | 16 |
| Semi-rural | 6 |
| Rural | 1 |
| Deprivation level* | |
| Low | 3 |
| Low-moderate | 2 |
| Moderate | 5 |
| Moderate-high | 11 |
| High | 2 |
| % of patients from ethnic minority background | |
| <10% | 5 |
| 10-19% | 4 |
| 20-29% | 3 |
| 30% or more | 8 |

* Totals do not add up as locums practised across different areas

[Ω] Two GPs worked in both salaried and locum roles

Abbreviations: GP: general practitioner

**Table 5.2 Number of GPs invited and interviewed**

| | Invited | Expressed Interest | Participated |
|---|---|---|---|
| Informal contacts | 6 | 5 | 5 |
| Email lists | 77 | 7 | 6 |
| Social media | - | 25 | 6 |
| High testers* | 5 | 1 | 1* |
| Snowball recruitment (through participants) | - | 2 | 2 |

*High testers relate to GPs who, by crude numbers, requested the most tests in Oxfordshire in 2019. Two GPs identified as being a high tester volunteered for the study through other channels

### 5.5.2 How decisions to test differ in children compared with adults

The sample contained GPs and trainees who described that their experiences requesting and performing tests for children differed from adults. In a few cases, doctors themselves performed blood tests; in other cases, the practices had a phlebotomy service that set their own age cut-offs for doing blood tests on children. Below these thresholds, children would be referred to the local hospital to receive blood tests. The participating GPs and trainees had different thresholds for requesting tests in children compared with adults for various reasons. Some perceived that the burden of testing on children is higher compared with adults.

> *"I mean, kids I would say I tend to have a bit of a higher threshold before doing bloods, for various reasons, especially young kids, because of the traumatic aspect of having their bloods taken. Obviously young kids can't understand why someone's pricking them with a needle, so there's definitely that to take into account. I think… with an adult or someone older, you'd be a little bit more willing to [test]. Often they want blood tests and request that, so it definitely makes that process a little bit easier."* (L4, Locum GP, practising <5 years)

Another GP mentioned the challenge of managing unexpected results.

> *"The risk that if you get an abnormal result you're then going to have to do more testing - is there in my mind… for a child, it's a more kind of scary and potentially unpleasant thing to have a test, whatever that is, whether it's a blood test or an x ray, you know, it means going to an unfamiliar environment and having something done to you"* (L1, Locum GP, 5-14 years practising)

Another participant believed that children tend to present with conditions for which alternative diagnostic strategies can be employed.

> *"I think we probably do less tests – or feel tests are less useful in children. They're a self-selecting group… the vast majority of children don't come to us with conditions that need tests. The vast majority are acute illnesses that are a clinical diagnosis with either a 'Direct to treatment,' or a 'Watch and wait' strategy with them."* (T1, GP trainee)

Several participants reflected that in children, the question of whether to test or not prompts them to also consider whether they should refer to paediatrics; hence, their threshold to refer was lower than in adults.

> *"I think that with kids it's different because we do have luckily a paediatric service locally who will see them fairly rapidly if we feel they need to. And I think that that in many ways defers the question of testing to someone else… A better test of whether this child needs testing is whether I think I'm prepared to send them to my secondary care colleagues and that decision is made much more easily, because I know my secondary care colleagues would see them quicker."* (P2, GP partner, 5-14 years)

To avoid two sets of blood draws, referral was also seen as a preferred route to testing, in case GPs overlooked an important test.

> *'If we're having to do bloods on children – so for the younger children – we're of the feeling that, probably, they need a paediatric referral – and the worry is that, if you're doing a blood test, you really want to do the right blood test at the right time, so that you don't want to have to bleed the child twice."*
> (P6, GP partner, 5-14 years)

### 5.5.3 Perceived utility of tests

#### *5.5.3.1 General considerations*

Generally, GPs considered tests useful when needing to rule out a specific condition (i.e., coeliac disease), rule in a suspected condition (i.e., diabetes), or reassure parents. Tests were also seen as supportive in guiding management, including informing treatment decisions where conditions could be managed in primary care, or supporting specialist referrals.

> *"The only time I do a test in primary care is if I think it will keep them out of hospital. I don't tend to do tests in primary care if I'm going to refer anyway, because there's no point…so I only do tests in primary care if it's something that I think I can manage in primary care, or it's going to direct whether I need to refer or not refer."*
> (S2, Salaried GP, 5-14 years)

On the other hand, GPs raised several concerns about testing in children. In addition to the physical and emotional burden of testing for children and their caregivers, GPs described the cost of testing for families including transport, parking, and taking time off work. Additionally, they considered the broader cost implications for the health system, and the potential environmental cost of using these resources. From the clinician's perspective, doing tests in children generates additional workload and can be more technically challenging, requiring more time, equipment and staff compared with testing in adults.

#### *5.5.3.2 Test-specific considerations*

GPs had different perceptions of which tests are considered more useful for children than others. Though not strictly a diagnostic test, growth and developmental assessments were deemed appropriate as an objective measure for determining how sick a child was, and

two GPs specifically emphasised that they need to be utilised more and documented appropriately.

> "*I think in paediatrics - I mean the tests that they [paediatricians] would ask for are things like growth charts. I think that's entirely appropriate…And another test which is a developmental assessment. And so I think that those are actually more important tests when it comes to chronic disease in children than frankly a [C reactive protein]. If a child's growing properly and developing in the direction you'd expect them to, then that is probably a better test and I hope my specialist colleagues appreciated that test more than most biochemical lab tests"* (P2, GP partner, 5-14 years)

Urine and stool testing were perceived to be valuable as they are cheap and non-invasive. Faecal calprotectin was also considered a helpful test in the appropriately selected population if the GP wanted to rule out a differential diagnosis of inflammatory bowel disease (IBD).

> *"I would do a faecal calprotectin because it's cheap, it's not invasive and it will help me rule out things like IBD."* (S5, Salaried GP, <5 years)

There were mixed views relating to the use of objective tests for asthma, including fractional exhaled nitric oxide (FeNO) testing and spirometry, for various reasons. These included differing opinions on the clinical utility of the test, discrepancies in testing recommendations between local and national guidelines, and variation in access to the tests in general practice (especially following the COVID-19 pandemic when these tests were not performed). One of the participants felt that objective testing needed to be utilised more in children to prevent misdiagnosis of asthma and guide treatment decisions.

> *"I think we should be using more FeNO testing because I think we label a huge number of people asthmatic that shouldn't be."* (P2, GP partner, 5-14 years)

Other GPs believed that objective tests for asthma are generally inaccurate. One GP spoke of the pitfalls of spirometry testing.

> *"I think some of it [the reason for not using spirometry] is an awareness of the test as being actually not that helpful and clouding the waters as well as, I mean... I've never tried to do spirometry in a young person, but I don't know how good they would be at doing it"* (L1, GP locum, 5-14 years)

The role of policy and incentives for using diagnostic tests in childhood asthma was explored, with one GP stating that the quality measures should not necessarily be focused on performing tests but rather on improving the quality of clinical assessments to ameliorate asthma outcomes.

> *"There's a big push from 'QOF' [Quality Outcomes Framework] – that we should be doing either spirometry or FeNO – or a peak flow reversibility – as an objective measure. I think their requirement is spirometry or FeNO – so that puts us in a very difficult position – but also, it's not necessarily clinically indicated. Our priority is improving the quality of assessments – because that's likely to be more helpful in terms of managing children's asthma than access to tests, which are of some benefit, but that's not the game-changer, in our view. In our view, it's about the quality of assessments that are going on. So, again, we're looking at hubs – you know, there's lots of money around for that. I'm concerned that's just going to generate activity rather than improve the quality of care"* (S4, Salaried GP, 15-24 years)

Aside from asthma tests, other tests that were considered less useful included vitamin D, allergy tests and C reactive protein (CRP) tests.

GPs generally felt that testing for vitamin D has increased in recent years because of increased public health awareness. One GP explained the challenges related to increased requests for vitamin D tests, including uncertainty about the appropriate thresholds in children and whether supplementation would be a better strategy than testing.

> *"The Vitamin D I find more difficult because I think that stems really from… public campaigns where people have said for example most people in the UK are probably deficient of Vitamin D… but instead of using … Vitamin D supplements for example they would rather want a test because obviously they have symptoms (for example, growth symptoms and tiredness symptoms) they ascribe to having low Vitamin D levels. And the test itself you know. I'm not so sure we've figured out what the right thresholds are for Vitamin D, so it's all a good debate."* (S1, Salaried GP, 5-14 years)

Another GP described their use of allergy tests as having decreased during their clinical practice after recognising that they are not helpful, difficult to interpret and costly.

> *"I never do RAST tests – looking for allergy – and I know that some people use them a lot...They're not easy tests to interpret at the best of times – we're not convinced it's a cost-effective way of assessing allergy, and so there'll be a range of views on that, but that's one of the things which I might have done in the past, and now I'm not doing at all."* (S4, Salaried GP, 15-24 years)

Some GPs mentioned C reactive protein (CRP), a non-specific inflammatory marker, can be helpful in some specific contexts, e.g., when referring a child with rheumatological symptoms or signs. Outside these contexts, they perceived that CRP was too vague, with an abnormal result often triggering a cascade of further tests.

> *'Very rarely would I test CRP. I think the only times I would use it is if I thought that the child had some kind of inflammatory condition, like an inflammatory arthritis or inflammatory bowel disease, where the CRP would really help with the diagnosis, because otherwise it's an extremely vague marker. It often leads to a cascade of testing because the CRP will come back slightly high for no particular reason, so I don't personally think it's a very good test to use. It's just kind of a blanket test as part of a panel of other things.* (S3, Salaried GP, <5 years)

Some tests, like vitamin $B_{12}$ and autoimmune tests were perceived to be inappropriate in the primary care setting and only valuable for secondary care, where specialists have the expertise to interpret the result in the context of the child's signs and symptoms and manage abnormal results.

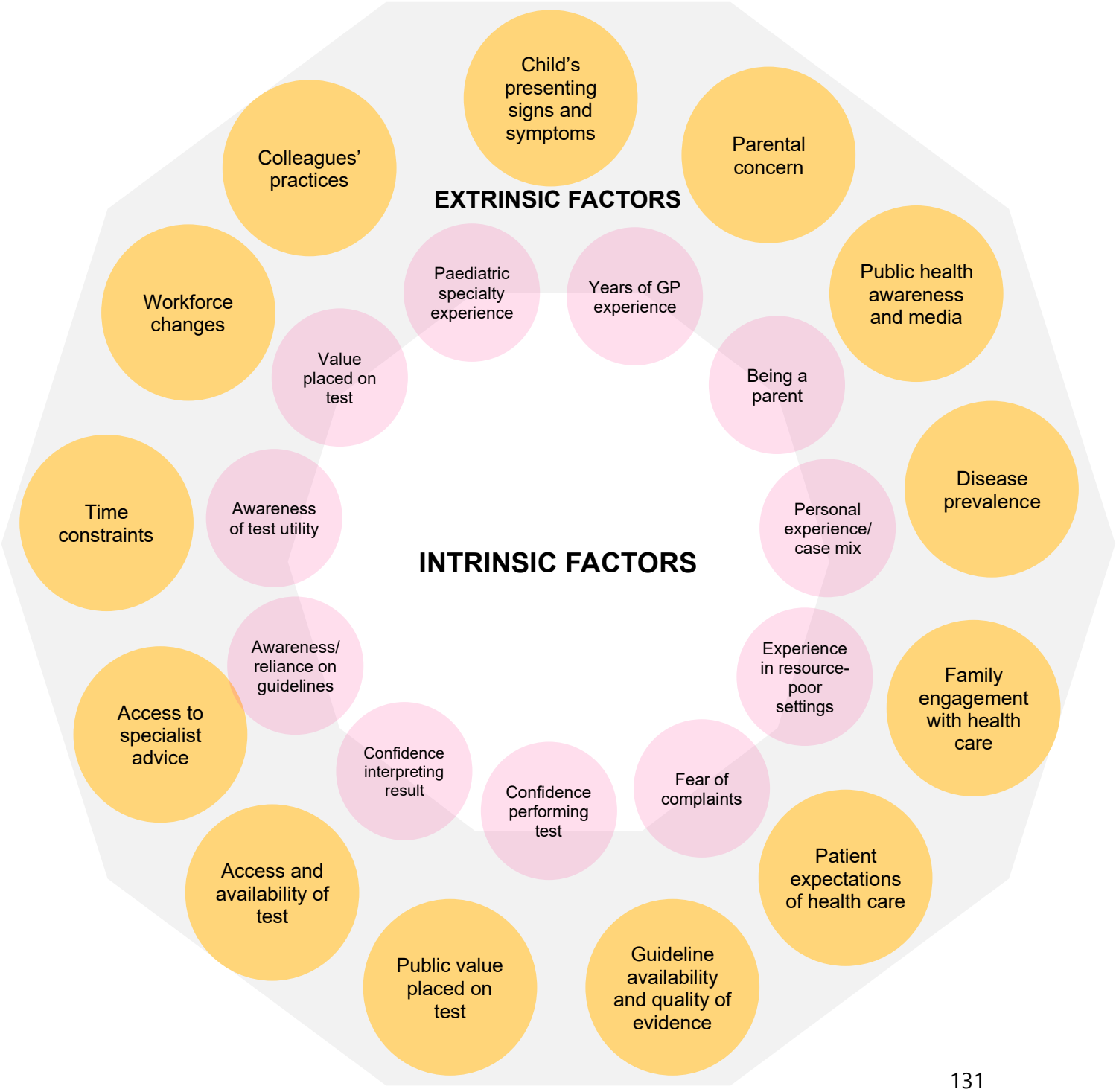
GPs spoke about the challenges of navigating conversations with parents who request what GPs perceive to be inappropriate tests, such as tests for nutritional deficiencies (measuring levels of trace elements like zinc and magnesium), allergy tests, or immunoglobulin tests, alongside tests that they would consider to be appropriate when investigating a child's presenting symptoms. There was tension between wanting to comply with parents' requests and just 'ticking' another box on the request form despite knowing the test was unlikely to be of clinical benefit.

> *"I think the harder scenario is when we're being asked to do tests alongside ones we think are reasonable to add in other tests which we don't think are reasonable. We don't really, necessarily, understand the result and…. I suppose nutritional deficiencies would be part of that, and I guess certain allergy tests where, actually, the allergy test isn't that helpful – or looking at different immunoglobulins – not really that helpful."* (P6, GP partner, 5-14 years)

### 5.5.4 Perceived drivers of variation

The perceived drivers of variation are summarised in **Figure 5.1** and include intrinsic factors specific to the clinician and extrinsic factors.

**Figure 5.1 Factors that drive variation in testing**



#### *5.5.4.1 Intrinsic factors (Clinician factors)*

GPs reported that variation in practice and testing can be partly attributed to the individual GP's confidence in managing children and their parents. Factors contributing to confidence levels included personal paediatric specialty experience, duration of clinical experience as a GP, and having personal experience of being a parent. They reflected that individual clinicians' past experiences may colour their future behaviours; for example, a missed diagnosis in one patient may lead to searching for that diagnosis in subsequent patients. GPs who had previously worked in resource-poor settings noted that they were judicious with their testing decisions because of their experience of having limited access to tests.

Some GPs mentioned the technical challenges associated with doing blood tests in children and that clinician confidence to perform phlebotomy varied (if there was no in-house phlebotomist). GPs also described the challenges of interpreting specific tests like paediatric ECGs, X-rays, and antibody screens and that they would only request tests for which they were confident interpreting and managing the results, as the following clinician discussed regarding ECG.

> *"I mean I was taught that an ECG in a child is not easy to interpret and you shouldn't really be doing them. I would do it if I emailed a cardiologist and they asked me to do it and then they were willing to look at it. But I feel that - I mean, as GP's, I think we're all fairly open about the fact that we don't do ECGs enough to feel confident about doing them and interpreting them. So we're very dependent on other people to interpret them and I don't do them in kids."* (L1, Locum GP, 5-14 years)

Several GPs spoke about the fact that variation in testing practices may be related to individual clinicians' awareness of the utility and accuracy of tests, including the rates of false positives, and that an 'abnormal' result simply refers to a statistical deviation from the population mean.

> *"It depends on the doctor's perception of how valuable the blood tests, the information that the blood tests give, what value they provide. I mean I think a lot of people are not aware of the low positive predicted value of things that come with certain scenarios"*
> (SL1, Salaried and locum GP, 5-14 years)
>
> *"A lot of GPs don't understand that an 'abnormal result' – simply means that it's outside of 2.5 standard deviations from the mean, and… that one in 20 results in a normal population will be classed as 'abnormal' on the computer screen…"*
> (L2, Locum GP, 15-24 years)

Another intrinsic factor that varied between clinicians was their awareness of and reliance on guidelines. While GPs generally knew of local guidelines and/or national guidelines such as the NICE Clinical Knowledge Summaries (the most frequently cited guidelines) and found them helpful, some GPs felt that guidelines could be limited in terms of providing holistic patient care. One GP explained why they occasionally felt they had to deviate from clinical guidelines to provide patient-centred care.

> *"I think being a GP you have to juggle quite a lot to ensure that you are providing the care that your patient needs and sometimes that means going off-piste, off the script. So the guidelines are incredibly helpful to understand where you lie and how you cannot get sued, essentially, or not get struck off, but sometimes you will have to…GPs own a particularly different role compared to the rest of the hospital doctors and other sort of NHS employed staff, we are independent contractors that are subcontracted to the NHS… So, while we work for the NHS and we are part of the NHS, we are not directly employed by the NHS, and because of that we hold a privileged position of… being able to be the advocate for a patient beyond what the NHS sees as being normal."*
> (SL1, Salaried and locum GP, 5-14 years)

Additionally, GPs and trainees commented that though they practised autonomously, testing practices become normative, and they had adopted testing practices from their supervisors and fellow colleagues. One trainee reported:

> *"We do meet every day... for coffee, and that's the sort of forum to bring up challenging cases. I do have formal supervision with my supervisor but, often, I find it more useful – in those informal things – and, even if I've not got specific question, somebody else might say, 'Oh, I saw a child with this' or, 'I've not done this – what do people think – would anybody have done anything differently?' That's where… I'd [started to] do the complement [testing] – for parents worried about immunodeficiency. So, discussion like that is useful."* (T1, GP trainee)

#### *5.5.4.2 Extrinsic factors (non-clinician factors)*

Aside from the severity of the child's presenting condition, the most frequently cited factor that influenced the decision to test was the parent's degree of concern and the likelihood of the result reassuring the parent.

> *I think people are less likely to be reassured based on clinical findings – they sort of give you…, 'Well, that's all very well, but I'd still like the test.' So, just generally – it's 'Fatigue,'… or 'I just don't feel well,' where… because the differential diagnosis nest is so broad, you end up doing a lot of tests. I find myself doing quite a lot of $B_{12}$ and folate, and Vitamin D, and checking thyroid function quite a lot, and they very rarely have an abnormality – or, if I do find an abnormality – say, in folate, or $B_{12}$ – 'Do I really think that abnormality is attributable to their symptoms?' – probably not – but then, of course, you found anyone with this, you're obliged to treat it. People don't tend to come back and say, 'I've had my $B_{12}$, and I don't feel any different,' so whether there is a placebo-type element of, 'I've had a test, there's been an abnormality, I've had treatment and now I'm better'.* (T1, GP trainee)

GPs reported that differences in the population that they serve can influence testing behaviours. Some ethnic minority groups have a higher risk of having specific conditions, prompting the GP to consider testing for these conditions in certain patients, such as vitamin D tests for children of Bangladeshi or African origin. One GP mentioned that they were more likely to accommodate parental requests with their Somali patients, to forge positive therapeutic relationships and trust between families and health care providers.

> *"We don't have a good relationship with our Somalian population, and…in the past (I'm talking about 40 or 50 years ago), they were underserved and underrepresented in their medical pathways. So, they were misdiagnosed for years… and because of that there is still a deep mistrust between that community and the GPs in our area. So, a lot of work has gone into bridging that gap, so any time we see any of the Somali [patients] we try our best to accommodate for their needs and their asks, I suppose, their requests…but we don't do things outside of reasonable expectations or reasonable suggestions. So, if they're asking for some wildly inappropriate tests or assessments then we will say no, but it's more a case of it's important for them to experience normal healthcare." (SL2, Salaried and locum GP, <5 years)*

Similarly, patient groups who are more disadvantaged do not present to their GP frequently, and GPs felt they had to test opportunistically in these cases.

> *"Sometimes we see families or children who you worry about, and they might be difficult people to get back in touch with. So, say for example, a gross generalisation, but say if they're non-English speaking and if I say to them come back in a month, and I'm not too sure whether they really will, then I'll have a lower threshold just to kind of get the tests done, so I know that they're done. Use that opportunity, they're there in front of you…not too sure how compliant they'll be with my plan, so it's better to catch all then. That might have some influence as well occasionally, because as I say, we sometimes see patients who are complex socially, and so you have to also always think about safeguarding and stuff like that as well."*
> (S5, Salaried GP, <5 years)

Conversely, GPs perceived that patients from some ethnic backgrounds hold different expectations of health care based on their experience of health systems in their country of origin, and view GPs as a barrier to receiving tests and specialist care.

> *"I find that - so [in one of their locum practices] lots of patients are from East Timor and Sudan and there's just a different cultural, a set of assumptions coming to the consultation from [them] and they're much more likely to… say… you're the doctor, I'll do what you say, that is …how they are behaving in the conversation. So making shared decisions is actually much more complicated because they just want you to tell them what to do. And it's often through an interpreter as well, but consequently someone who, if you don't want to do a task, it's much easier to say to that person, we're not going to do it and these are the reasons why and they'll be incredibly accepting and kind about that and understanding. But the other side of that population is the usually white, usually Eastern European …and people from mainland Europe generally who are used to going direct to specialists and GPs are just in the way. They're like … "I want to see a gynaecologist because I want a chlamydia test." And so that's a different conversation again … you're trying to talk them down."* (L1, Locum GP, 5-14 years)

Public beliefs that tests are high value and technically accurate were also thought to be a driver of increased parental requests for tests.

> *"I think as a society, both medics and patients overvalue tests. We know that most … 90% of diagnoses are made by history alone type thing but we, if you look at tests and how they shift the probability of disease, it is not as much as clinicians think of it. It's definitely not so much as patient thinks and I do think that part of my feeling when it comes to tests is often the pressure comes from patient's parents and they don't understand tests. They think that the test is the answer which is often not the case, and they mis-value the history. And in fact, you can almost feel it when you're trying to take a history from the patient to try to examine them, there's almost this impatience where they're like, well, let's just cut straight to the tests."* (P2, GP partner, 5-14 years)

Other external factors related to funding, accessibility, availability of tests, and access to specialist advice.

> *"We don't do spirometry in our area anymore. It has to go to Respiratory. That's for adults and children. We used to do spirometry before COVID, but that funding has been taken away from us now, so we don't do it in-house anymore."* (S2, Salaried GP, 5-14 years)
>
> *"If [accessing specialist care] was a very long delay, then I perhaps would have to examine my risk thresholds when it came to testing a bit more closely."* (P2, GP partner, 5-14 years)

GPs also believed that unnecessary testing might result from time constraints hampered GPs' from performing a focussed history and physical examination to narrow down the differential diagnosis, causing them instead to rely on tests. A lack of time also limited comprehensive discussions relating to the benefits and harms of testing and its uncertainties.

> *"If a patient comes in or a parent asking for a blood test, the easy way to finish that consultation is to agree to the blood test and they move on, they get out of the room, and you're onto your next patient, whereas if you actually want to discuss it and try and help, you know, because you've got to reach a shared understanding, if they walk out the room unhappy with you then that's not great because they're just going to go, well, it causes problems for you and it causes problems for the next doctor that they're going to go and see and just request it. So, if you want to help them understand and reach a shared understanding of the plan, then that takes time and effort and energy."* (SL1, Salaried and locum GP, 5-14 years)

Limited time and working in a pressured environment may also lead to testing that is inappropriate when the boundaries of adult care become blurred with paediatric care. This might occur in the case of vitamin $B_{12}$ testing, for example, which GPs believe is usually unnecessary for children in the primary care setting.

> *"I think sometimes also you get mixed up between adults and kids, and sometimes you will do what you do for adults for kids, except we manage adults in primary care, but we may not manage children in the same way. The threshold's slightly different and the significance of the results is slightly different… so I think sometimes, those tests, like I said, if I test for anaemia, I probably would do $B_{12}$ and everything, but there are times when I'm like, if the $B_{12}$ is low, I don't know how to replace $B_{12}$ in children"* (S2, Salaried GP, 5-14 years)

Workforce changes were highlighted as potential contributors to unnecessary testing. One GP mentioned that erosions in the continuity of care with fewer doctors in substantive roles and more locum doctors may lead to more testing, where clinicians try to be as "thorough" as possible and "paint a complete picture" by utilising tests.

> *"There's lots more less than full-time working and lots more fragmentation, therefore, of patient-doctor relationships and experience…If you're the only person – you've never met these people before… you've got to take everything they say at face value. You're probably never going to see them again. That definitely changes your threshold for testing... So then feeding into that is not just less than full time working. It's - there's fewer people in substantive roles and I'm part of the problem as a locum so I don't know if it's, you know my instinct is (and it's certainly the comment of lots of my colleagues) … that locums request more tests and I think there are more locum workers now."* (L2, Locum GP, 5-14 years)

Finally, GPs explained that guidelines influenced their decisions to test but that guideline recommendations varied by local area, in some instances contradicting NICE guidance. Guidelines did not always consider the local context and accessibility of tests.

> *"For asthma, there's always a bit of a tension with this locally, because the guidelines say that we should have spirometry results for children to be able to diagnose asthma, but the local children's hospital clinical viewpoint is that we should be just diagnosing and treating it clinically, so there's a bit of a tension between the two things of what we should be doing in primary care and what the hospital will accept in terms of a referral. So yes, usually we would refer for spirometry to diagnose asthma if the child is old enough to be able to do it, but that doesn't always necessarily happen, and then you end up having to rely on other features of asthma and then just treating the child without the result anyway."* (S3, Salaried GP, <5 years)

Guidelines were criticised for sometimes misrepresenting different recommendations with mixed quality of evidence uniformly and being less applicable to primary care.

> *"I mean the evidence behind guidelines is quite varied… It's not necessarily going to be that applicable to the child that you're seeing but it is useful as a kind of touch point and again, if you are making a referral to somebody, and you know that they will want X, Y or Z before they see the child, then it's helpful to know that to smooth the process."* (L3, Locum GP, <5 years)

While some GPs were unaware whether there were local guidelines for children in primary care, other GPs expressed they were overwhelmed by the volume of guidelines they were expected to stay abreast of. In some areas, local paediatricians and/or GPs had developed guidelines related to caring for children in the community that was accessible to local GPs on their practice intranet. Variation in GP awareness of guidelines and guideline accessibility may therefore be factors that contribute to testing variation.

## 5.6 Discussion

### 5.6.1 Summary of findings

Testing in children requires specific considerations and may carry substantial consequences. GPs' diagnostic decision-making for children differed from adults; their threshold to request a test was higher for children, and the threshold to refer to specialty care was lower than adults. There was ambivalence relating to the appropriate level of testing for children. Some tests, like faecal calprotectin and urine tests were considered useful, and other tests, like CRP, vitamin D and spirometry, were seen to be of less value in children. Multiple factors were suggested to account for variation between clinicians and practices in the rates of test use in children, including disease prevalence, workforce changes, individual clinicians' risk tolerance, differing patient expectations of health care, local system factors and guidelines. Parental requests were emphasised by most GPs as one the most frequent reasons to test in children and a potential driver of variation.

### 5.6.2 Comparison with existing literature

No studies have explored the use of diagnostic tests for children in primary care. This is likely due to the relative infrequency of testing compared to adults, where over-testing has been identified as a major problem and source of waste in the NHS.[12] A previous qualitative interview study by Watson et al explored and compared GPs' and adult patients' expectations, experiences, and understanding of tests within primary care.[13] In both my study and Watson's, GPs believed that time and workload pressures presented barriers to a shared understanding of testing and uncertainty.[13] Their study found that in most cases, patients rarely requested tests. Still, they viewed them as a positive step in moving forward with the consultation and affirmation that their GP took their concerns seriously. They reported that decisions to request tests were led mainly by doctors, with no

examples of shared decision-making or information sharing. In contrast, many of the GPs interviewed within our study felt that parental request was one of the most common reasons for them to order a test, where parents sometimes entered consultations with the expectation that the outcome would be their child receiving a test. A systematic review examined “parental concern” or gut feeling as a diagnostic accuracy tool for sepsis in children and found that parental concern performed well as a diagnostic marker for serious illness with a high positive likelihood ratio of 16.4.[14] Another recent prospective cohort study found that while parental concern is not associated with sepsis in isolation, it was associated with PICU admission and bacterial infection. Therefore, parental concern could be a useful diagnostic adjunct when combined with other markers, but it has not yet been explored in general practice.[15]

Gill and colleagues conducted an interview study to examine GPs’ views on quality markers for children in primary care.[16] Participating GPs, who all practised in the Thames Valley region, shared the concerns of the GPs in our study that a lack of standardised training influenced decision-making and quality of care. GPs within their study also felt that the principles of decision-making would need to be modified to meet the needs of vulnerable children at risk, a finding that was validated by GPs in my interview sample.

Another qualitative study in UK primary care explored GPs’ and nurse practitioners’ approaches to inflammatory marker testing in primary care.[17] The authors identified issues that were similarly present within my study, including ambivalence relating to the appropriate use of inflammatory markers and the tension between not wanting to miss a diagnosis but also being wary of unexpected results that are borderline or inconclusive and lead to testing cascades. Unfortunately, there is a lack of evidence quantifying how often these testing cascades occur. Clinicians doubted the ’right’ level of testing, reporting a lack

of guidance and education which is consistent with our findings. Participants in my sample also specifically highlighted a lack of public and clinician awareness about test utility. This finding is supported by two systematic reviews by Hoffmann and colleagues, which reported that patients[18] and clinicians[19] tend to overestimate the benefits of tests/screening and underestimate the harms.

It was interesting to note differences in the perceived utility of tests by GPs within my study. Generally, participants agreed that faecal calprotectin as a helpful test in primary care. This is supported by a diagnostic accuracy study of faecal calprotectin in UK general practice, which found that the test had 100% sensitivity and 91% specificity, distinguishing between IBD and functional gut disorder in children aged between 4 and 18, resulting in fewer unnecessary referrals and diagnostic tests in secondary care.[20] This is interesting to compare with the previous chapter which found low practice variability (10%) in rates of faecal calprotectin requests and suggests that the evidence for its use has facilitated its uptake in practice. Participants, however, expressed mixed views on using FeNO tests to diagnose childhood asthma. Two diagnostic accuracy meta-analyses examining the diagnostic accuracy of FeNO testing in diagnosing asthma found that it performed moderately well in terms of diagnostic accuracy, with estimates of sensitivity of 0.80 and specificity estimates ranging from 0.64 to 0.81 using a FeNO threshold of less than 20 parts per million. [21,22] GPs also reported varying access to the test within their practices and conflicting guidance on whether to use objective asthma testing or rely on clinical features alone. In Chapter 4, I reported that FeNO testing was subject to the highest practice variability (124%). Variation in FeNO use may therefore be attributed in part to inconsistency of the available evidence, clinical guidance, availability, and access to the test.

### 5.6.3 Strengths and limitations

This study is the first to explore GPs' perspectives on how, when, and why they choose to employ (or not employ) diagnostic tests in children. Participants worked in various geographic settings throughout England in practices that served diverse population groups. I used rigorous and systematic methods to collect and analyse the data. To ensure the trustworthiness of the findings, a senior qualitative researcher (MG), checked the transcripts, codes, and coding framework. The analysis was supported by practising GPs (CH, GH) within the research team who supported the credibility of the identified themes and sub-themes.

Recruitment for this study was advertised through GP groups including the academic primary care department and the Overdiagnosis working group of the Royal College of GPs, which was identified as an important source of bias in my study sample as these groups are likely to have an interest in reducing unnecessary testing. To address this, I expanded the study recruitment to general GP groups on social media, where a substantial proportion (30%) of interview participants were subsequently recruited. I also included GPs identified as high testers (a further 15% of the participant sample) to address this volunteer bias and add to the range of perspectives captured within my interview sample.

This study interviewed GPs to gather their experiences and perspectives on healthcare encounters with children and their families. A limitation of the study was that it was restricted to GP perspectives. Though most test requests originate from GPs in primary care, they do not account for all primary care healthcare personnel who request tests, including advanced nurse practitioners, specialist and practice nurses, pharmacists, and paramedics. [23]

### 5.6.4 Implications for clinical practice, policy, and research

The use of tests varies by clinician and practice, and trends have changed substantially over time. Such variation can lead to poor health outcomes and worsen health inequities.[5] This study explores some of the reasons for this variation. The GPs in our interview sample described many factors that underpin testing variation. The specific drivers differ according to the child, parent, test, and clinician. This study highlights areas for clinicians and policymakers to target to improve testing and reduce unwarranted variation.

These include:

- Emphasising clinical history taking and examination for a more accurate pre-test probability (the likelihood of a child having the pre-specified condition before receiving the diagnostic test), and how it differs in the context of ethnic variation, such as vitamin D levels.
- Developing specific strategies or tools for clinicians to communicate uncertainty in diagnosis and test utility to a child and their parent/guardian.
- Enhancing GP awareness and education surrounding the utility of specific tests and different reference ranges for children.
- Increasing paediatric exposure in general practice training to improve clinician confidence in managing childhood presentations.
- Implementing electronic clinician decision support alerts to prompt clinicians if requesting tests that require specific considerations or may not be indicated (i.e., Vitamin $B_{12}$ tests in children).
- Having a unified repository for clinical guidelines related to children and young people which incorporates local and national guidance.

- Refining local guidance to account for the access and availability of tests and the realities of working in primary care.
- Strengthening the quality of the evidence base for diagnostics in children.

Future research should explore the perspectives of parents and caregivers of children who have undergone tests in primary care to gain a more holistic understanding of the experience of diagnostic tests in this setting. The generalisability of some of the observations and experiences of GPs from this study could be studied using a data-driven approach, for example, assessing whether less experienced clinicians test more frequently than more experienced clinicians.

### 5.6.5 How this chapter fits within the wider thesis

In this chapter, I sought to understand paediatric testing practices from the perspectives of GPs and make sense of the quantitative data presented in Chapter 3 and 4. The findings from this chapter were also used to inform some of the exploratory analyses conducted in chapter 4 (i.e., selecting the tests to include in the analyses). These findings demonstrate that GPs approach paediatric diagnostics differently to adults, and their thresholds for testing and referral vary. The previous chapters explored the role of demographic factors on test variation such as gender, age, and socioeconomic deprivation level, however, performing this qualitative study enabled me to explore other potential contributors to testing variation that are not available within the datasets. The subsequent chapters will explore two of the extrinsic drivers of variation that arose from this study, including the role of guidelines and quality of evidence.

## 5.7 Conclusions

This qualitative interview study provides GP perspectives on paediatric testing and the factors that underpin their diagnostic decision-making for children. These findings highlight actionable issues for clinicians, researchers, and policymakers to address, including gaps in education, evidence, and guidance. Tackling these issues can reduce unwarranted differences in test use and improve the quality of healthcare delivered to children in general practice.

## 5.8 Chapter Summary

- To explore GPs' perspectives on paediatric test use in general practice and explore the reasons for variation in test use, I conducted interviews with 18 GPs and 2 trainees. The interviews were transcribed and analysed thematically.
- GPs reflected that their approach to testing in children differed from adults; their threshold to test was higher, and their threshold to refer children to specialists was lower.
- GPs had varying opinions about the utility of specific tests, including objective testing for asthma.
- Perceived drivers of variation in testing included: 1) intrinsic (clinician) factors relating to their risk tolerance and
  experience, and 2) extrinsic factors, including disease prevalence, parental concern and expectations of healthcare, workforce changes leading to fragmentation in care, time constraints, and inconsistencies in guidelines.
- These findings highlight key actions that clinicians, researchers, and policymakers need to take to reduce unwarranted variation in test use and provide equitable care for children in general practice.

## 5.9 References


1. Ruzangi J, Blair M, Cecil E, et al. Trends in healthcare use in children aged less than 15 years: A population-based cohort study in England from 2007 to 2017. BMJ Open. 2020;10(5). doi:10.1136/bmjopen-2019-033761.
2. Hobbs FDR, Bankhead C, Mukhtar T, et al. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007–14. The Lancet. 2016;387(10035). doi:10.1016/S0140-6736(16)00620-6.
3. Foot C, Naylor C, Imison C. The quality of GP diagnosis and referral. London, UK: The King's Fund. Published online 2010.
4. Hiscock H, Perera P, Mclean K, et al. Variation in paediatric clinical practice: A review of care in inpatient, outpatient and emergency department settings. J Paediatr Child Health. 2016;52(7). doi:10.1111/jpc.13212.
5. Public Health England. The 2nd Atlas of Variation in NHS Diagnostic Services in England. NHS RightCare. 2017;(January).
6. Wennberg J, Cooper M. Dartmouth Atlas of Health; 1998.
7. NHS RightCare. NHS Atlas of Variation in Healthcare for Children and Young People; 2012.
8. Malterud K, Siersma VD, Guassora AD. Sample Size in Qualitative Interview Studies: Guided by Information Power. Qual Health Res. 2016;26(13). doi:10.1177/1049732315617444.
9. Clarke V, Braun V. Thematic analysis: A practical guide. London: SAGE; 2021.
10. Braun V, Clarke V. Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. Couns Psychother Res. 2021;21(1). doi:10.1002/capr.12360.
11. Lumivero. NVivo (Version 12), 2018. www.lumivero.com.
12. O'Sullivan JW, Stevens S, Hobbs FDR, et al. Temporal trends in use of tests in UK primary care, 2000-15: Retrospective analysis of 250 million tests. The BMJ. 2018;363. doi:10.1136/bmj.k4666.
13. Watson J, Whiting PF, Salisbury C, Hamilton WT, Banks J. Blood tests in primary care: A qualitative study of communication and decision-making between doctors and patients. Health Expectations. 2022;25(5). doi:10.1111/hex.13564.

14. Harley A, Latour J, Schlapbach LJ. The role of parental concerns in the recognition of sepsis in children: A literature review. Front Pediatr. 2019;7(APR). doi:10.3389/fped.2019.00161.
15. Sever Z, Schlapbach LJ, Gilholm P, et al. Impact of parental and healthcare professional concern on the diagnosis of pediatric sepsis: a diagnostic accuracy study. Front Pediatr. 2023;11. doi:10.3389/fped.2023.1140121.
16. Gill PJ, Hislop J, Mant D, Harnden A. General practitioners views on quality markers for children in UK primary care: A qualitative study. BMC Fam Pract. 2012;13. doi:10.1186/1471-2296-13-92.
17. Watson J, De Salis I, Salisbury C, Hamilton W. 'I'm fishing really'-Inflammatory marker testing in primary care: A qualitative study. British Journal of General Practice. 2016;66(644). doi:10.3399/bjgp16X683857.
18. Hoffmann TC. Patients' expectations of the benefits and harms of treatments, screening, and tests a systematic review. JAMA Intern Med. 2015;175(2). doi:10.1001/jamainternmed.2014.6016.
19. Hoffmann TC, Del Mar C. Clinicians' expectations of the benefits and harms of treatments, screening, and tests: A systematic review. JAMA Intern Med. 2017;177(3). doi:10.1001/jamainternmed.2016.8254.
20. Walker GJ, Chanchlani N, Thomas A, et al. Primary care faecal calprotectin testing in children with suspected inflammatory bowel disease: A diagnostic accuracy study. Arch Dis Child. 2020;105(10). doi:10.1136/archdischild-2019-317823.
21. Wang Z, Pianosi PT, Keogh KA, et al. The Diagnostic Accuracy of Fractional Exhaled Nitric Oxide Testing in Asthma: A Systematic Review and Meta-analyses. Mayo Clin Proc. 2018;93(2). doi:10.1016/j.mayocp.2017.11.012.
22. Tang S, Xie Y, Yuan C, Sun X, Cui Y. Fractional Exhaled Nitric Oxide for the Diagnosis of Childhood Asthma: a Systematic Review and Meta-analysis. Clin Rev Allergy Immunol. 2019;56(2). doi:10.1007/s12016-016-8573-4.
23. Eaton G, Tierney S, Wong G, Oke J, Williams V, Mahtani KR. Understanding the roles and work of paramedics in primary care: A national cross-sectional survey. BMJ Open. 2022;12(12). doi:10.1136/bmjopen-2022-067476.

# Chapter 6. A quality appraisal of paediatric guidelines in general practice.

## 6.1 Publication statement

A manuscript containing components of this chapter has been published in the *Journal of Paediatrics and Child Health* (**Appendix 1.1**).

## 6.2 Introduction

In the previous chapter, GPs described the challenges in making diagnostic decisions for children, which contribute to variation in paediatric test utilisation. One of the drivers of variation that they cited was limitations in existing clinical practice guidelines. Given the critical role of primary care in delivering health care for children, improved diagnostic guidance can standardize care according to the best available evidence, lead to better children's' health outcomes and prevent unnecessary hospital presentations and referrals.

Clinical practice guidelines aim to reduce uncertainty and standardise care pathways for diagnoses and management. Adherence to clinical practice guidelines is used as a metric of the quality of health service provision.[1] A series of population studies in Australia reported adherence to national and international CPG treatment guidance approximately 60% of the time for 17 common childhood conditions; this ranged from 43% for preventative care[2], 54% for fever[3], and 60% for acute gastroenteritis.[4] CPG adherence was generally lowest in the primary care setting when compared with inpatient and emergency care.

One study has assessed the quality of diagnostic testing guidelines for adults presenting to primary care, however no similar study exists for children presenting to primary care.[5] A systematic review determined that the quality of 159 paediatric guidelines from 1990 to 2020 had improved over time but remained sub-optimal.[6] Another review examining the quality of 216 paediatric guidelines published from 2017 to 2019 found that quality was influenced by whether the guidelines were under the responsibility of organizations or groups, if evidence-based methods were used to develop guidelines, and the socioeconomic development index of a country (i.e., countries from less economically developed countries had lower quality scores in four of the six quality domains).[7]

In this chapter, I aimed to determine the methodological quality and reporting of diagnostic guidelines from high-income countries for three of the most prevalent and important childhood conditions in primary care: fever, constipation, and gastroenteritis. These conditions were selected by screening NICE guidelines and undertaking a feasibility count using the Clinical Practice Research Datalink (**Chapter 2, Table 2.1**).

## 6.3 Aims and objectives

This overall aim of this chapter was to assess whether guideline quality could be a potential driver of variation in test use.

The specific objective was:

1. To assess the methodological quality of diagnostic guidelines from high-income countries for three common childhood conditions in primary care: fever, constipation, and gastroenteritis.

## 6.4 Methods

### 6.4.1 Study design

I reviewed and appraised paediatric clinical practice guidelines published between February 2011 and September 2022 using a systematic approach. This study was conducted and reported in line with a modified Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) checklist.[8] Ethical approval was not required.

### 6.4.2 Search strategy and eligibility criteria

An information specialist (NR) searched Medline and Embase for paediatric clinical practice guidelines (CPGs) relating to fever, constipation, and gastroenteritis (see **Chapter 6 Appendix 6.1** for search strategy). I also searched the following guideline repositories: Trip Database, Guidelines International Network, National Guideline Clearinghouse, and the World Health Organization (WHO) guidelines. I included guidelines from high-income countries with similar health systems to the United Kingdom (i.e., Australia, Canada, Ireland, Norway, Denmark, and the Netherlands) where GPs are the focal point for coordinating referrals to specialist paediatric care.[9] I also included guidelines from the United States to avoid missing important guidance that may be used by practitioners in other countries.

Guidelines published or updated between February 2011 and September 2022 were included. If there were multiple versions of a guideline, the most recently updated version was included. Guidelines were eligible if they included 1) children and adolescents; 2) diagnostic testing recommendations; and 3) relevant to general practice. Non-English language publications were excluded unless English translations were available.

Two reviewers (ET, ST) independently assessed all abstracts for potential inclusion, with discrepancies resolved by consensus and a third reviewer (CH) if required.

### 6.4.3 Assessment of guideline quality and reporting using AGREE II

Two reviewers (ET, ST) independently assessed the methodological quality and reporting of eligible guidelines. The assessment was not restricted to specific diagnostic recommendations contained within each guideline, but rather the guideline in its entirety. We used the AGREE II tool, which has been used widely to evaluate the quality of clinical practice guidelines across six domains: scope and purpose, stakeholder involvement, the rigour of development, clarity of presentation, applicability, and editorial independence.[10] An overall quality score was then determined, ranging from a score of 1 for the lowest possible quality, to 7 for highest possible quality. Domain scores were calculated by adding item scores given by both reviewers within a domain and then calculating this as a percentage of the maximum possible domain score. Disagreements were resolved by discussion and a third reviewer (CH) if required. Further details about the AGREE II tool and score calculation can be found in **Chapter 7 Appendix 7.2** and **7.3**.

### 6.4.4 Statistical analysis

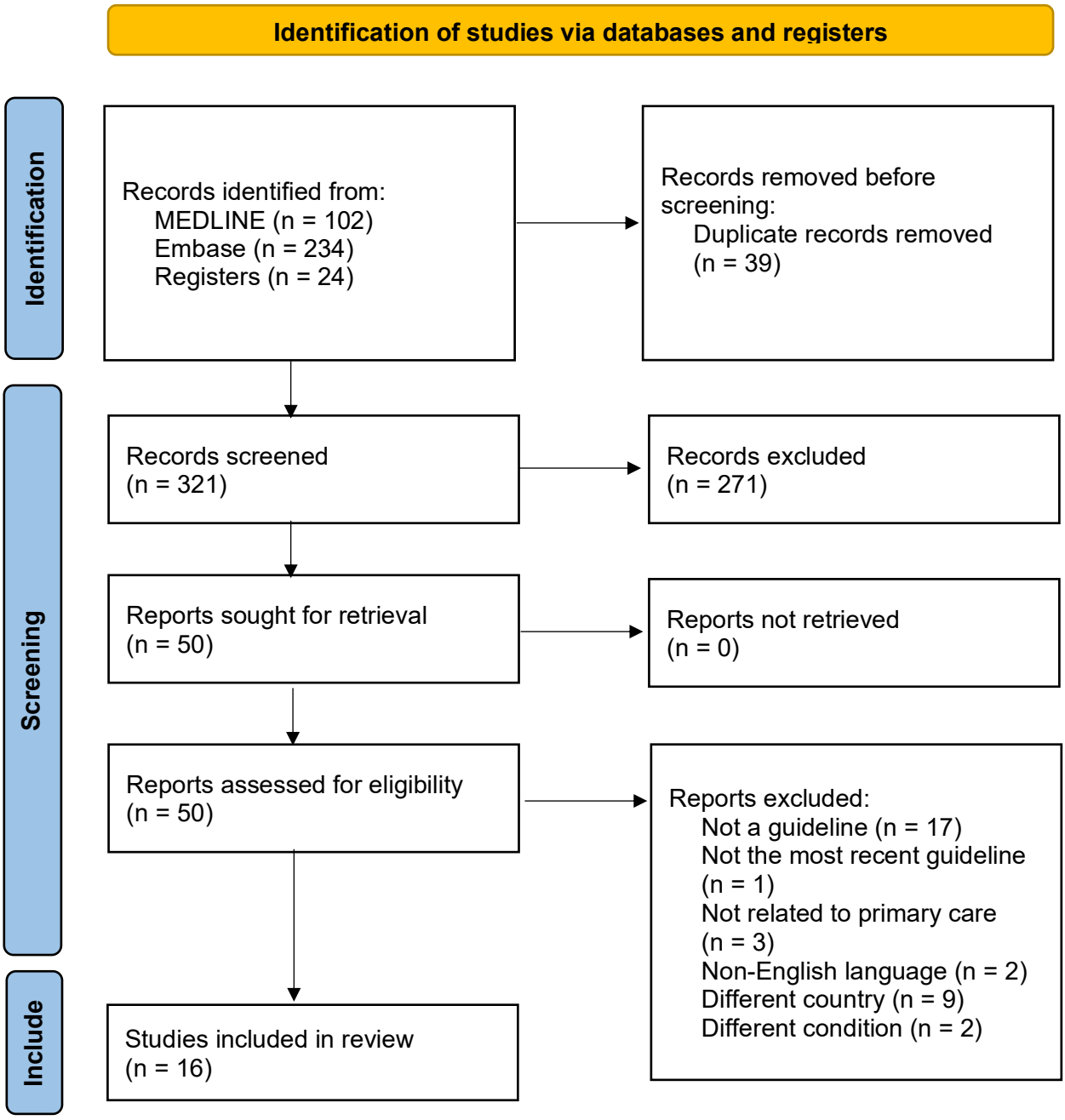
AGREE II scores were reported as medians with corresponding ranges. The results were presented in a grouped box-and-whiskers plot to visualise the variation in scores across domains for each condition. Cohen's kappa was calculated to measure inter-rater reliability.

## 6.5 Results

### 6.5.1 Study selection

The search was conducted on 2nd September 2022 and returned 360 results. After removing 39 duplicates, 321 titles and abstracts were screened for eligibility. The full texts of 50 guidelines were reviewed and 16 guidelines were deemed eligible for inclusion. **Figure 6.1** shows the guideline selection flow diagram.

**Figure 6.1 Guideline selection flow diagram**




### 6.5.2 Characteristics of included guidelines

We included 16 guidelines, comprising seven for fever, four for constipation and five for gastroenteritis (**Table 6.1**). All guidelines were published after or updated since 2013. Of

the fever guidelines, three were published in the United States (American Academy of Paediatrics [AAP], American College of Radiology [ACR] and Cincinnati Children's [CC]), two were published in Australia (Royal Children's Hospital [RCH] and South Australia Ministry of Health [SA]), one was published in the Netherlands (Dutch College of General Practitioners [DCGP]) and one was published in the UK (National Institute for Health and Care Excellence [NICE]). Of the constipation guidelines, one was published in Australia (SA), one was published in the Netherlands (Federatie Medisch Specialisten [FMS]), one was published the UK (NICE) and the guideline published in Switzerland (European Society for Paediatric Gastroenterology Hepatology and Nutrition/ North American Society for Paediatric Gastroenterology Hepatology and Nutrition [ESPGHAN/NASPGHAN]) was published for worldwide use. Finally, the gastroenteritis guidelines, one was published in Australia (SA), one was published in the United States (Infectious Diseases Society of America [IDSA]), one was published in the UK (NICE) and the guidelines (World Gastroenterology Organization [WGO]) and Switzerland (European Society for Paediatric Gastroenterology, Hepatology, and Nutrition/European Society for Paediatric Infectious Diseases [ESPGHAN/ESPID]) were published for worldwide use.

**Table 6.1 Characteristics of included fever, constipation, and gastroenteritis guidelines**

| Condition | Guideline Publisher | Title | Publication year | Year of update | Country of Publication |
|---|---|---|---|---|---|
| Fever | *NICE[11] | Fever in under 5s: assessment and initial management | 2019 | 2022 | UK |
| | AAP[12] | AAP: Evaluation and Management of Well-Appearing Febrile Infants 8 to 60 Days Old | 2021 | - | USA |
| | RCH[13] | Febrile child | - | 2022 | Australia |
| | SA[14] | Fever in children aged 1-2 months | 2014 | | Australia |
| | DCGP[15] | Children with fever | 2016 | 2020 | Netherlands |
| | ACR[16] | ACR Appropriateness Criteria Fever Without Source or Unknown Origin—Child | 2016 | - | USA |
| | CC[17] | Evidence-Based Care Guideline for Management of Infants 0 to 60 days with Fever of Unknown Source Guideline 10 | 1998 | 2019 | USA |
| Constipation | *NICE[18] | Constipation in children and young people: diagnosis and management | 2010 | 2017 | UK |
| | *ESPGHAN/NASPGHAN[19] | Evaluation and Treatment of Functional Constipation in Infants and Children: Evidence-Based Recommendations From ESPGHAN and NASPGHAN | 2014 | - | Switzerland |
| | FMS[20] | Constipation in Children 0-18 Years | - | 2015 | Netherlands |
| | SA[21] | Constipation in Children | 2013 | 2022 | Australia |
| Gastroenteritis | *NICE[22] | Diarrhoea and vomiting caused by gastroenteritis in under 5s: diagnosis and management | 2009 | 2018 | UK |

| | | | | | |
|---|---|---|---|---|---|
| | *WGO[23] | Acute diarrhea in adults and children: a global perspective | 2013 | - | UK |
| | *ESPGHAN/ESPID[24] | ESPGHAN/ESPID Evidence-Based Guidelines for the Management of Acute Gastroenteritis in Children in Europe: update 2014 | 2014 | - | Switzerland |
| | IDSA[25] | 2017 Infectious Diseases Society of America Clinical Practice Guidelines for the Diagnosis and Management of Infectious Diarrhea | 2017 | - | USA |
| | SA[26] | Gastroenteritis in Children | 2013 | 2018 | Australia |

* Intended for worldwide use

**Abbreviations** NICE: National Institute for Care and Clinical Excellence; AAP – American Academy of Paediatrics; RCH – Royal Children's Hospital Melbourne; SA – South Australia Ministry of Health; DCGP – Dutch College of General Practitioners; ACR – American College of Radiology; CC – Cincinnati Children's Guidelines; ESPGHAN/NASPGHAN - European Society for Paediatric Gastroenterology Hepatology and Nutrition/ North American Society for Paediatric Gastroenterology Hepatology and Nutrition; FMS - Federatie Medisch Specialisten; WGO – World Gastroenterology Organization; ESPGHAN/ESPID - European Society for Paediatric Gastroenterology, Hepatology, and Nutrition/European Society for Paediatric Infectious Diseases; IDSA – Infectious Diseases Society of America

The AGREE scores for each domain are shown in **Table 6.2** (fever), **Table 6.3** (constipation) and **Table 6.4** (gastroenteritis). The kappa statistic for inter-rater reliability was 67%, showing substantial agreement. The quality of CPG reporting was highly variable across all three conditions. The overall quality of all the included guidelines was moderate (median 4.8 out of a maximum of 7, range 2.5 to 6.5). The guidelines for constipation had the highest overall quality rating (median 5.3, range 2.5 to 6.0), and the guidelines for gastroenteritis had the lowest overall quality rating (median 4.0, range 2.5 to 6.0). NICE guidelines were consistently rated the highest quality for all three conditions due to their methodological rigour, high standards of reporting, and transparency.

**Table 6.2 AGREE II Scores by domain for childhood fever guidelines**

| Guideline name | Domain 1: Scope and Purpose (%) | Domain 2: Stakeholder Involvement (%) | Domain 3: Rigour of Development (%) | Domain 4: Clarity of Presentation (%) | Domain 5: Applicability (%) | Domain 6: Editorial Independence (%) | Overall Guideline Assessment (out of 7) |
|---|---|---|---|---|---|---|---|
| NICE | 97 | 97 | 98 | 100 | 85 | 58 | 6.5 |
| CC | 94 | 94 | 98 | 86 | 75 | 100 | 6.5 |
| DCGP | 94 | 92 | 88 | 100 | 21 | 33 | 6 |
| ACR | 67 | 81 | 89 | 100 | 15 | 100 | 5 |
| AAP | 89 | 78 | 59 | 100 | 63 | 83 | 4.5 |
| RCH | 14 | 56 | 13 | 100 | 33 | 0 | 3 |
| SA | 61 | 36 | 0 | 75 | 25 | 0 | 2.5 |

Abbreviations: AAP – American Academy of Paediatrics; ACR – American College of Radiology; CC – Cincinnati Children's Guidelines; DCGP – Dutch College of General Practitioners; NICE – National Institute for Care and Clinical Excellence; RCH – Royal Children's Hospital Melbourne; SA – South Australia Ministry of Health.

**Table 6.3 AGREE II Scores by domain for childhood constipation guidelines**

| Guideline name | Domain 1: Scope and Purpose (%) | Domain 2: Stakeholder Involvement (%) | Domain 3: Rigour of Development (%) | Domain 4: Clarity of Presentation (%) | Domain 5: Applicability (%) | Domain 6: Editorial Independence (%) | Overall Guideline Assessment (out of 7) |
|---|---|---|---|---|---|---|---|
| NICE | 100 | 97 | 96 | 100 | 88 | 38 | 6 |
| FMS | 100 | 94 | 98 | 100 | 54 | 54 | 6 |
| ESPGHAN/NASPGHAN | 89 | 39 | 70 | 100 | 25 | 38 | 4.5 |
| SA | 39 | 36 | 0 | 100 | 23 | 4 | 2.5 |

Abbreviations: NICE – National Institute for Health and Care Excellence; FMS – Federatie Medisch Specialisten; ESPGHAN/NASPGHAN – European Society for Paediatric Gastroenterology Hepatology and Nutrition/North American Society for Paediatric Gastroenterology Hepatology and Nutrition; SA – South Australia Ministry of Health.

**Table 6.4 AGREE II Scores by domain for childhood gastroenteritis guidelines**

| Guideline name | Domain 1: Scope and Purpose (%) | Domain 2: Stakeholder Involvement (%) | Domain 3: Rigour of Development (%) | Domain 4: Clarity of Presentation (%) | Domain 5: Applicability (%) | Domain 6: Editorial Independence (%) | Overall Guideline Assessment (out of 7) |
|---|---|---|---|---|---|---|---|
| NICE | 97 | 100 | 99 | 100 | 85 | 33 | 6 |
| IDSA | 94 | 50 | 82 | 100 | 25 | 54 | 5 |
| ESPGHAN/ ESPID | 81 | 39 | 59 | 100 | 40 | 42 | 4 |
| WGO | 33 | 8 | 5 | 56 | 56 | 46 | 2.5 |
| SA | 39 | 36 | 15 | 100 | 23 | 4 | 2.5 |

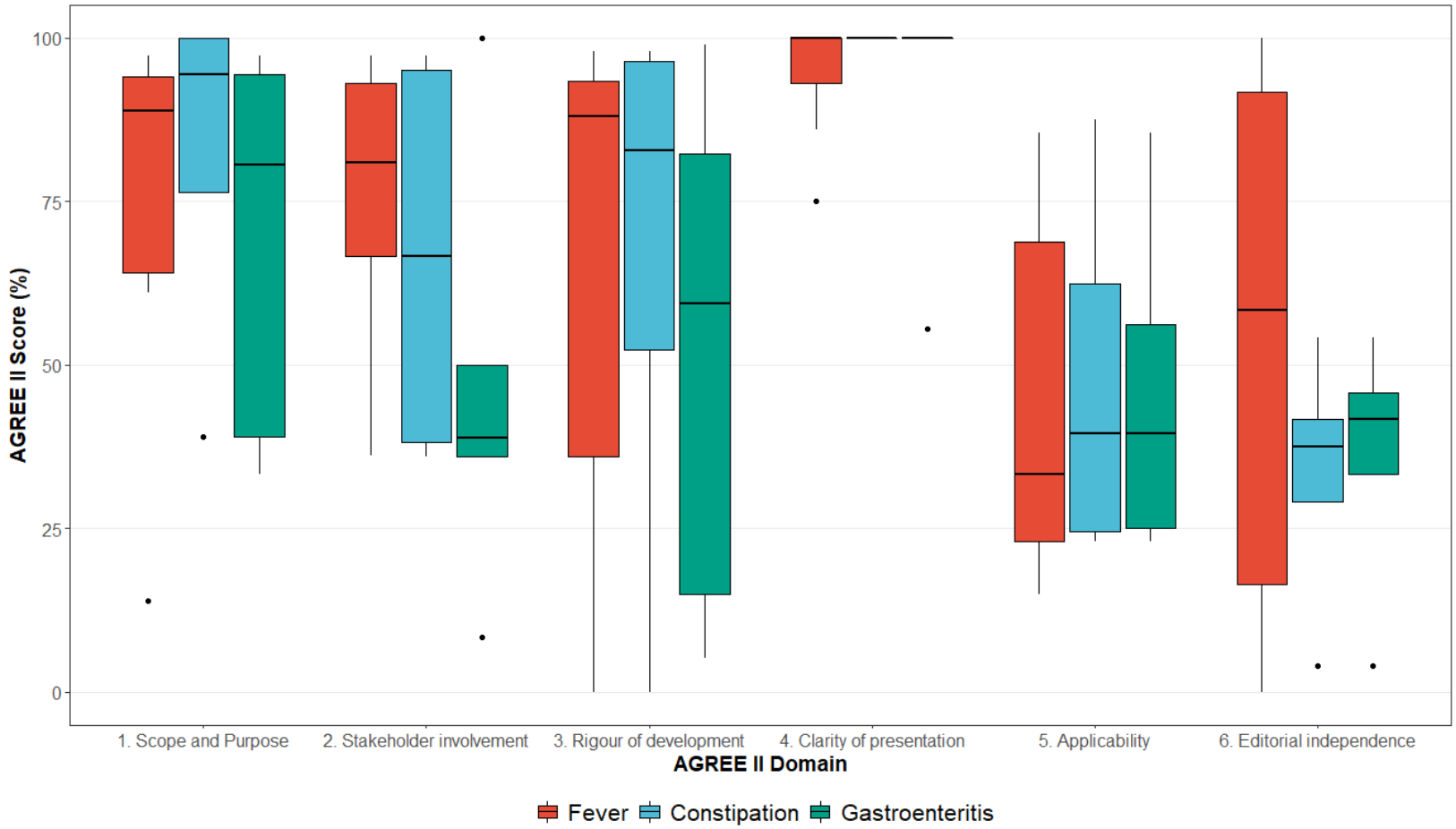
Abbreviations: NICE – National Institute for Health and Care Excellence; IDSA – Infectious Diseases Society of America; FMS – Federatie Medisch Specialisten; ESPGHAN/ESPID – European Society for Paediatric Gastroenterology, Hepatology, and Nutrition/European Society for Paediatric Infectious Diseases; WGO – World Gastroenterology Organization; SA – South Australia Ministry of Health.

### 6.5.3 Evaluation of the methodological quality of guidelines

**Figure 6.2** presents the variation in the AGREE II domain scores. Most guidelines demonstrated excellent clarity in the presentation of their recommendations. For all three conditions, the greatest variation in guideline quality scores occurred in Domain 3, which relates to the rigour of development. This domain assesses a guideline's transparency in reporting search strategies, reasons for selecting the evidence, consideration of health benefits, side effects and risks in giving recommendations and the process for monitoring, auditing, and updating guidelines. Across all conditions, a major weakness was the limited consideration given to the applicability of guidelines in the clinical setting, particularly the facilitators and barriers of guideline implementation. For example, 44% (7/16) of guidelines did not factor in the cost implications of their recommendations.

Eight guidelines (fever [2], constipation [2] and gastroenteritis [4]) guidelines did not explicitly state whether parents or public representatives were involved in the recommendation development process. Editorial independence was also highlighted as a major area of concern for 56% (9/16) of guidelines, with funding sources and conflicts of interest either not declared, or committee members failing to provide adequate explanations of how competing interests were addressed.

**Figure 6.2 Boxplot demonstrating variation in AGREE II domain scores across guidelines for fever, constipation, and gastroenteritis**



## 6.6 Discussion

### 6.6.1 Summary of findings

For three of the most frequently encountered problems in paediatric primary care, the quality of guidelines was highly variable. The major areas of weakness in guideline development were related to guideline applicability and editorial independence. Guidelines for fever were the most variable in quality with overall scores ranging from 2.5 to 6.5 out of 7. For all three conditions, NICE guidelines performed the best in most quality domains compared to the other guidelines. Variable guideline quality and more importantly, limited consideration of how guidelines can be implemented in clinical settings present a major barrier to their uptake in clinical practice, which likely contributes to variation in care health care delivery and outcomes.

### 6.6.2 Strengths and limitations in the context of existing literature

To better understand the perceived barriers to guideline implementation and reasons for poor adherence to guideline recommendations, a group of researchers in the Netherlands conducted a qualitative study of Dutch GPs analysing the barriers to guideline implementation. They reported the most common reason for poor guideline uptake was lack of agreement with the recommendations due to poor applicability, unclear recommendations, and limited evidence.[27] Other factors included organisational constraints, lack of awareness of guideline recommendations as well as patient preferences and behaviours not aligning with recommendations.

This is the first review to systematically search and appraise guidelines for the most common childhood presentations to primary care. I specifically focused on guidelines that included diagnostic testing recommendations, as GPs tend to encounter more diagnostic uncertainty than other clinicians.[28] The findings of this study highlight the need for high quality clinical practice guidelines to aid GPs.

This review included guidelines that are applied in similar settings; I deliberately focused on guidelines from high-income countries. I also included guidelines for specific conditions, rather than grouping guidelines by organ system. This was to allow for fair comparisons between guidelines. It also allows clinicians to identify the highest quality guideline for a particular condition.

These findings are restricted to guidelines relevant to primary care to better understand how GPs diagnose children, and I only focused on three of the most common conditions that children present with. Other reviews have sought to look at the issue more broadly across a wide range of conditions.[6,7] Their findings are consistent with this review's findings, suggesting that guidance for paediatric conditions is generally of suboptimal quality across health care settings.

### 6.6.3 Implications for policy and practice

Good quality guidelines are a valuable resource for synthesising evidence in the context of a growing body of primary research. However, to effectively inform health care decision-making and reduce unwarranted variation, they should adhere to reporting standards, be developed using robust methods and synthesised using the best available evidence. Guidelines should also promote transparency and ensure that contributors declare and address conflicts of interest. To improve their applicability, guideline developers should consider that specific tests are not readily available in primary care (as was described in Chapter 5) and ensure that their diagnostic recommendations are feasible and adaptable to different local contexts.

This chapter's findings also highlight the need for international collaboration for improving the accessibility of CPGs. There are currently multiple sources to search for guidelines including the BIGG International database of GRADE guidelines; Dynamed; ECRI Guidelines Trust; MAGIC authoring and publication platform; NICE; TRIP database; and the U.S. Preventive Services Task Force. The Guidelines International

Network (GIN) was established to promote collaboration and improve the efficiency of guideline development, adaptation, dissemination, and implementation. However, it is not a comprehensive source of all paediatric guidelines – a search of all guidelines in GIN related to children returned 77 results of which only 20 were from the UK. Yet, there are 85 UK NICE guidelines related to children and young people. To improve the retrieval and accessibility of paediatric guidelines, a separate repository should be developed to collate all the available guidance for children by condition. Furthermore, guidelines should be open-access and include quality ratings, to enable GPs to identify and evaluate the best diagnostic and management options for their patients.

Guidelines only play a part when it comes to standardising care for children presenting to a general practice. The paucity of high-quality guidance could necessitate lengthening the duration of GP specialist training to increase paediatric clinical exposure. There are also calls to develop flexible GP models to improve paediatric care delivery, such as child health GP hubs, GPs with extended roles, or increasing GP access to specialist opinions.[29] In Australia, where paediatric experience is already a core requirement of general practice training, GP registrars reported lower confidence in managing mental health and behavioural presentations but more confidence in managing acute presentations, emphasising the importance of outpatient paediatric medicine exposure in GP training programs.[30]

### 6.6.4 How this chapter fits within the wider thesis

In this chapter, I examined the quality of clinical guidelines as one of the potential extrinsic drivers of test variation which was previously highlighted in Chapter 5. I found that the quality and reporting of guidelines for the most commonly presenting paediatric conditions in primary care were highly variable. This is consistent with the perspectives of GPs and GP trainees in Chapter 5 who believed that limited consideration of how guidelines are implemented in clinical settings can impede their uptake in practice. Most interview participants reported referring to NICE guidelines which reassuringly

performed the best across the range of primary care conditions in this review, despite concerns related to the editorial independence of guideline developers.

## 6.7 Conclusions

High quality diagnostic guidelines for children in primary care can improve health care delivery and reduce unwarranted test variation. There is substantial variation in quality of paediatric guidelines for primary care presentations. To improve guidelines, developers should adhere to strict methodological standards of reporting. Streamlining guidelines into one resource would improve GPs' access to the best quality guidance to improve diagnosis in children.

## 6.8 Chapter summary

- Paediatric clinical practice guidelines related to the diagnosis of fever, constipation, and gastroenteritis are of varying quality (median AGREE II score 4.8, range 2.5 to 6.5 out of 7).
- Specific areas of weaknesses across the guidelines include limited consideration of guideline applicability, failure to involve parent representatives in guideline development, and not adequately declaring or addressing competing interests.
- Poor-quality paediatric clinical guidance may contribute to poor guideline adherence and unwarranted variation in care in general practice.

## 6.9 References


1. Milchak JL, Carter BL, James PA, Ardery G. Measuring adherence to practice guidelines for the management of hypertension: An evaluation of the literature. Hypertension. 2004;44(5). doi:10.1161/01.HYP.0000144100.29945.5e.
2. Wiles LK, de Wet C, Dalton C, et al. The quality of preventive care for pre-school aged children in Australian general practice. BMC Med. 2019;17(1). doi:10.1186/s12916-019-1455-x.
3. Holt J, White L, Wheaton GR, et al. Management of fever in Australian children: A population-based sample survey. BMC Pediatr. 2020;20(1). doi:10.1186/s12887-020-1911-y.
4. Sunderland N, Westbrook J, Urwin R, et al. Appropriate management of acute gastroenteritis in Australian children: A population-based study. PLoS One. 2019;14(11). doi:10.1371/journal.pone.0224681.
5. O'Sullivan JW, Albasri A, Koshiaris C, Aronson JK, Heneghan C, Perera R. Diagnostic test guidelines based on high-quality evidence had greater rates of adherence: a meta-epidemiological study. J Clin Epidemiol. 2018;103. doi:10.1016/j.jclinepi.2018.06.013.
6. Zhang S, Wu L, Wang Y, et al. Methodological and reporting quality of pediatric clinical practice guidelines: a systematic review. Ann Transl Med. 2021;9(15). doi:10.21037/atm-21-2686.
7. Liu Y, Zhang Y, Wang S, et al. Quality of pediatric clinical practice guidelines. BMC Pediatr. 2021;21(1). doi:10.1186/s12887-021-02693-1.
8. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. The BMJ. 2021;372. doi:10.1136/bmj.n71.
9. Van Esso D, Del Torso S, Hadjipanayis A, et al. Paediatric primary care in Europe: Variation between countries. Arch Dis Child. 2010;95(10). doi:10.1136/adc.2009.178459.
10. Brouwers M, Kho ME, Browman GP, et al. AGREE Next Steps Consortium. AGREE II: Advancing guideline development, reporting and evaluation in healthcare. Can Med Assoc J. 2010;(182).
11. NICE. Fever in under 5s: assessment and initial management NICE guideline [NG143]. NICE Guidelines Online. 2019;(November 2019).
12. Pantell RH, Roberts KB, Adams WG, et al. Evaluation and management of well-appearing febrile infants 8 to 60 days old. Pediatrics. 2021;148(2). doi:10.1542/peds.2021-052228.

13. The Royal Children's Hospital Melbourne. Febrile child. Published September 2022. Accessed October 4, 2022. https://www.rch.org.au/clinicalguide/guideline_index/Febrile_child/
14. Government of South Australia Department of Health and Wellbeing. Fever in Children aged 1-2 months. Published September 2022. Accessed October 4, 2022. https://www.sahealth.sa.gov.au/wps/wcm/connect/public+content/sa+health+internet/resources/policies/fever+in+children+aged+1-2+months+-+sa+paediatric+clinical+guideline#:~:text=Non%2Dtoxic%20infants%20aged%201,Children%20aged%201%2D2%20months.
15. NHG working group, Berger MY, Berghuis ICA, et al. Children with fever. Published 2016. Accessed September 4, 2022. https://richtlijnen.nhg.org/standaarden/kinderen-met-koorts#volledige-tekst-belangrijkste-wijzigingen.
16. Westra SJ, Karmazyn BK, Alazraki AL, et al. ACR Appropriateness Criteria Fever Without Source or Unknown Origin—Child. Journal of the American College of Radiology. 2016;13(8). doi:10.1016/j.jacr.2016.04.028.
17. Cincinnati Children's. Evidence-Based Care Guideline for Management of Infants 0 to 60 Days with Fever of Unknown Source Guideline 10.; 2019. Accessed September 4, 2022. https://www.cincinnatichildrens.org/-/media/cincinnati%20childrens/home/service/j/anderson-center/evidence-based-care/recommendations/default/fever%20of%20unknown%20source%20guideline%2010.pdf?la=en.
18. NICE. Constipation in children and young people: diagnosis and management. Clinical Guideline. 2010;(May).
19. Tabbers MM, Dilorenzo C, Berger MY, et al. Evaluation and treatment of functional constipation in infants and children: Evidence-based recommendations from ESPGHAN and NASPGHAN. J Pediatr Gastroenterol Nutr. 2014;58(2). doi:10.1097/MPG.0000000000000266.
20. Federatie Medisch Specialisten. Constipation in children aged 0 to 18 years. 2015. Published December 1, 2015. Accessed October 4, 2022. https://richtlijnendatabase.nl/richtlijn/obstipatie_bij_kinderen_van_0_tot_18_jaar/obstipatie_-_startpagina.html.
21. Government of South Australia Department for Health and Wellbeing. Constipation in Children.; 2022. Accessed October 4, 2022. https://www.sahealth.sa.gov.au/wps/wcm/connect/public+content/sa+health+int

ernet/resources/policies/constipation+in+children+-+sa+paediatric+clinical+guideline.
22. NICE Guidelines. Diarrhoea and vomiting caused by gastroenteritis in under 5s: diagnosis and management. NICE Clinical Guidelines. 2009;(April).
23. Farthing M, Salam MA, Lindberg G, et al. Acute diarrhea in adults and children: A global perspective. J Clin Gastroenterol. 2013;47(1). doi:10.1097/MCG.0b013e31826df662.
24. Guarino A, Ashkenazi S, Gendrel D, lo Vecchio A, Shamir R, Szajewska H. ESPGHAN/ESPID Evidence-Based Guidelines for the Management of Acute Gastroenteritis in Children in Europe: Update 2014. J Pediatr Gastroenterol Nutr. 2014;59(1).
25. Shane AL, Mody RK, Crump JA, et al. 2017 Infectious Diseases Society of America Clinical Practice Guidelines for the Diagnosis and Management of Infectious Diarrhea. Clinical Infectious Diseases. 2017;65(12). doi:10.1093/cid/cix669.
26. Government of South Australia Department for Health and Wellbeing. Gastroenteritis in Children.; 2014. Accessed October 4, 2022. https://www.sahealth.sa.gov.au/wps/wcm/connect/public+content/sa+health+internet/resources/policies/gastroenteritis+in+children+-+sa+paediatric+clinical+guideline.
27. Lugtenberg M, Zegers-van Schaick JM, Westert GP, Burgers JS. Why don't physicians adhere to guideline recommendations in practice? An analysis of barriers among Dutch general practitioners. Implementation Science. 2009;4(1):54. doi:10.1186/1748-5908-4-54.
28. Alam R, Cheraghi-Sohi S, Panagioti M, Esmail A, Campbell S, Panagopoulou E. Managing diagnostic uncertainty in primary care: A systematic critical review. BMC Fam Pract. 2017;18(1). doi:10.1186/s12875-017-0650-0.
29. Ridd MJ, Thompson MJ. Would primary care paediatricians improve UK child health outcomes? No. British Journal of General Practice. 2020;70(693). doi:10.3399/bjgp20x709289.
30. Williames S, Temple-Smith M, Chondros P, et al. Are we preparing Victorian general practice registrars to be confident in all aspects of primary care paediatrics? Aust J Gen Pract. 2020;49(11). doi:10.31128/AJGP-08-19-5028.

# Chapter 7. The quality of paediatric asthma guidelines: evidence underpinning diagnostic test recommendations from a meta-epidemiological study

## 7.1 Publication statement

This chapter is based on my publication in *Family Practice* (**Appendix 1.1**).

## 7.2 Introduction

In Chapter 6, I assessed the quality of guidelines for common childhood conditions as one of the proposed extrinsic drivers of testing variation. The utility of guidelines, however, also depends on the quality of the underlying evidence supporting the individual recommendations. I therefore sought to extend my investigation of guideline quality as a source of test variation by assessing the strength of evidence supporting diagnostic test recommendations. I chose to examine guidelines for paediatric asthma, which was identified as one of the most prevalent paediatric primary care conditions in the Introduction**,** asthma (**Chapter 2, Table 2.1**).

Asthma is the most prevalent chronic respiratory condition in children worldwide. Childhood asthma rates based on general practice clinical records vary between countries, at 8% in the United Kingdom, 11% in the Netherlands, and 14% in Australia.[1–3] Diagnosis of childhood asthma is challenging, highlighting the need for high quality clinical practice guidelines (CPGs) which 'translate the best available evidence into the best practice'.[4] An Australian childhood asthma population study showed CPG adherence ranged from 54% among GPs to 80% in a hospital in-patient setting.[5] Measuring CPG adherence as a metric for health care quality can be problematic, as individual patient characteristics and preferences can warrant variation in clinical care. CPG development may be influenced by interested third parties i.e., pharmaceutical companies. [4,6] They also tend to be of varying quality with

recommendations often based on expert opinion (Level 5 evidence) rather than high-quality Level 1 evidence from randomised controlled trials.[7]

Diagnosing childhood asthma is difficult for a few reasons. First, there is no single objective gold-standard diagnostic test. Younger children cannot often perform objective testing and so a diagnosis is made on clinical symptoms or response to a treatment trial. Second, asthma is a dynamic illness with phenotypic heterogeneity. Asthma can also mimic other conditions, including viral-associated wheeze, allergic rhinitis, vocal cord dysfunction and gastro-oesophageal disease. Test results can vary over time, making it necessary to carry out repeat tests to reduce uncertainty. This is intrinsically problematic in primary care where access to respiratory specialists, and training in the proper use and interpretation of spirometry and other tests in the paediatric population can vary significantly as described by interview participants in Chapter 5. CPGs aim to reduce diagnostic uncertainty, by identifying children with asthma who would derive benefit from therapy and avoiding the harms of unnecessary or ineffective interventions in children without asthma.[6]

No prior studies have assessed the quality of evidence and reporting of diagnostic guidelines for childhood asthma. One study evaluated diagnostic guidelines in an adult primary care setting and found that recommendations based on high-quality evidence had greater rates of clinician adherence.[7] Two reviews compared childhood asthma recommendations across international guidelines, however they focused on management and did not examine diagnostic recommendations. [8,9] While conducting primary research in asthma is challenging due to the absence of a gold-standard test, the strength of evidence underpinning current diagnostic recommendations for paediatric asthma remains unknown.

## 7.3 Aims and objectives

The chapter aims to explore how the quality of diagnostic guidance and strength of evidence supporting diagnostic recommendations for paediatric asthma may contribute to variation in test use among general practitioners.

The objectives were:

1. To assess the methodological quality and reporting of paediatric asthma guidelines for primary care.
2. To determine the strength of evidence supporting diagnostic test recommendations for childhood asthma.

## 7.4 Methods

### 7.4.1 Study design

I conducted a meta-epidemiological study of paediatric asthma CPGs published between February 2011 and September 2022 using a systematic approach. There are no formal reporting standards for meta-epidemiological studies; therefore, this study followed a modified Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) checklist.[11]

Two reviewers (ET, ST) assessed the quality and reporting of CPGs using the Appraisal of Guidelines for Research and Evaluation (AGREE II) tool.[12] We evaluated the quality of evidence underpinning diagnostic testing recommendations using Grading of Recommendations Assessment, Development and Evaluation (GRADE).[13]

### 7.4.2 Search strategy and eligibility criteria

An information specialist (NR) searched Medline and Embase for paediatric asthma CPGs (see **Chapter 7 Appendix 7.1** for the search strategy). I also searched the following guideline repositories: Trip Database, Guidelines International Network,

National Guideline Clearinghouse and WHO guidelines. Following the same approach as Chapter 6, I included guidelines from high-income countries with similar health systems to the UK, specifically Australia, Canada, Ireland, Norway, Denmark, and the Netherlands, where GPs act as the gatekeepers to specialist paediatric care.[14]

I included guidelines published or updated between February 2011 and September 2022. I also considered strategy documents as guidelines if they included diagnostic recommendations for clinicians. The most recently published guideline was used if multiple versions of a guideline existed. CPGs which 1) focused on children and adolescents with suspected asthma, 2) related to primary care and 3) included diagnostic testing recommendations were eligible. Non-English language publications were excluded. Two reviewers (ET, ST) independently assessed all guidelines for potential inclusion, with discrepancies resolved by consensus and a third reviewer (CH) if required.

### 7.4.3 Data extraction

Two reviewers (ET, ST) independently extracted data on the country of origin, date of guideline publication and update. We also extracted the diagnostic tests included in each guideline and associated recommendations.

### 7.4.4 Assessment of guideline quality and reporting

Two reviewers (ET, ST) independently evaluated the methodological quality and reporting of included CPGs using the AGREE II tool.[12] The guideline quality was assessed across six domains; scope and purpose, stakeholder involvement, the rigour of development, clarity of presentation, applicability, and editorial independence. A global quality score was given ranging from 1 (lowest quality) to 7 (highest quality). Disagreements were resolved by discussion and a third author (CH) if required. **Chapter 7 Appendix** demonstrates the use of the AGREE II tool (**7.2**) and score calculation (**7.3**).

### 7.4.5 Selection of diagnostic tests related to primary care

A list of included diagnostic tests across all relevant CPGs was provided to two GPs (JT, BN) who work within the primary author's research department and practise in the UK. They independently identified the tests that could be requested in primary care from those that are only accessible in specialist care. The Cohen's kappa for interrater agreement between the two GPs was 76%. A third GP (CH) resolved any disagreements. The preliminary list of included and excluded diagnostic tests **(see Chapter 7 Appendix 7.4)** was sent to four other GPs (GB, SB, SD, JL) who practice in different parts of the UK. They unanimously agreed on its validity.

### 7.4.6 Assessment of quality of evidence

We examined the evidence underpinning recommendations for the diagnostic tests identified above. Supporting references could include other guidelines, evidence syntheses, and primary studies. The full texts of referenced studies supporting each guideline recommendation were identified from Medline and Embase searches. If no studies supported a guideline recommendation and it was not explicitly stated that the recommendation was based on expert opinion, I contacted the author organisation for clarification.

Two reviewers (ET, ST) independently appraised the supporting studies for each recommendation and assessed their quality using GRADE, the most widely adopted tool for grading the quality of evidence supporting guideline recommendations.[13] Disagreements were resolved by discussion and a third author (CH) if required. While GRADE assessments are typically tailored towards intervention outcomes, we used the guidance for performing GRADE assessments for test accuracy outcomes.[15,16] The assessment encompasses five categories: risk of bias, imprecision, inconsistency, indirectness, and publication bias, which culminate in an overall GRADE score of high, moderate, low, or very low quality. If one of the categories was deemed to have serious

limitations, the quality was downgraded by one, e.g., from high to moderate. The Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool was used to assess the risk of bias of diagnostic accuracy studies. The QUADAS-2 comprises four domains: patient selection, conduct of the index test, reference test, and patient flow and timing.[17] Expert opinion was considered “very low quality” evidence.

### 7.4.7 Statistical analysis

I compared the AGREE II score (guideline quality) and GRADE rating (for each test recommendation) for all included diagnostic test recommendations. The median and range of the overall quality AGREE II scores were calculated. I presented the AGREE II scores using box-and-whisker plots to demonstrate the median and variation for each domain across the included guidelines. For each diagnostic test, I collated the number and proportion of recommendations based on high, moderate, low, or very low-quality evidence.

To investigate potential associations between the quality of guideline recommendations and age group, tests, and recommendation type, I employed Fisher's exact test, which can be used to detect associations between two categorical variables. Factors were categorised as follows: age groups (younger vs. older, defined as < 5 or 6 years and ≥ 5 or 6 years as specified by the guideline), diagnostic test type (imaging, blood test, lung function test, or other), and the type of recommendation ('do' or 'do not do'). Statistical significance was set at a P value of less than 0.05. All analyses were conducted using R version 4.1.3.[18]

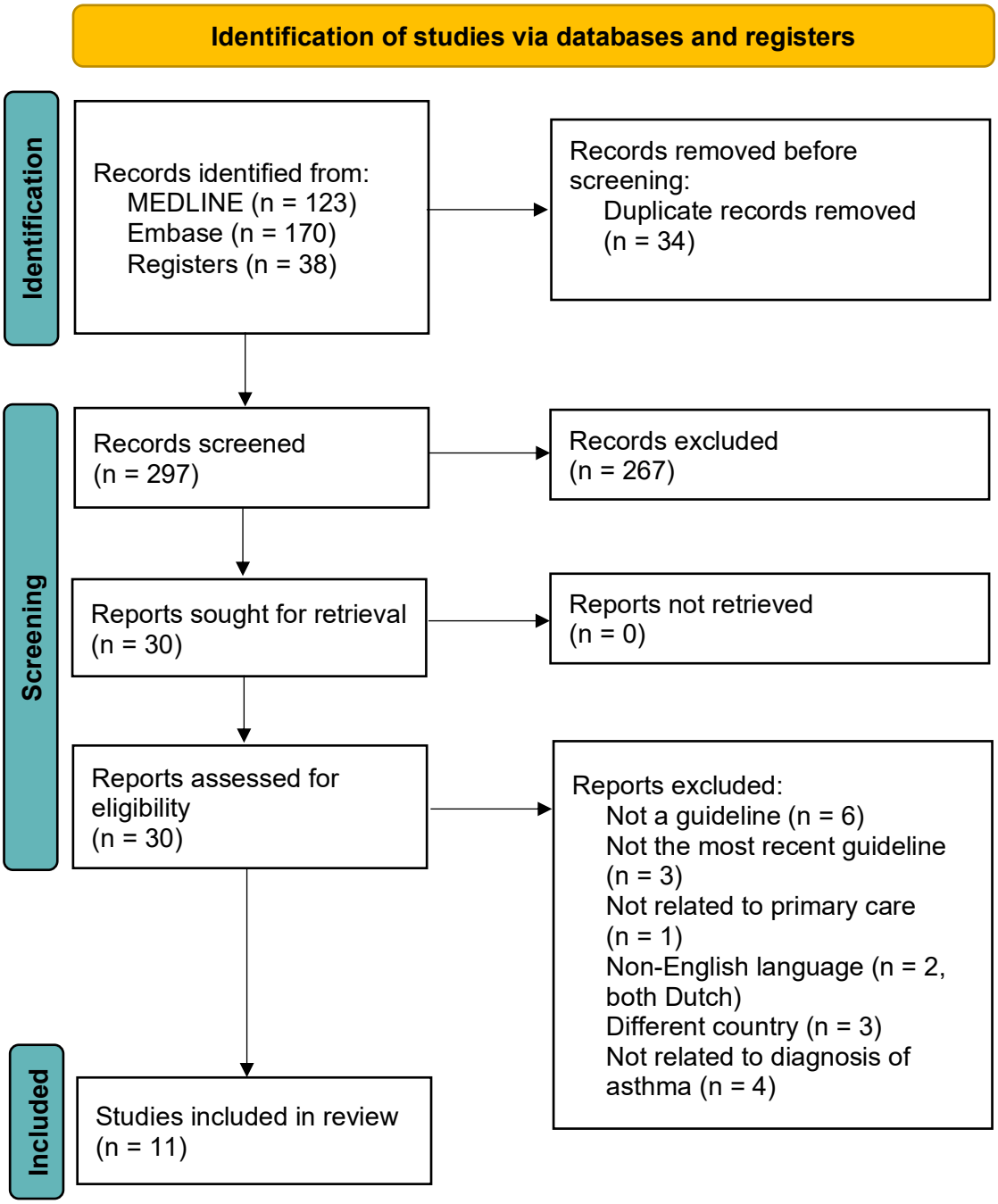
## 7.5 Results

### 7.5.1 Study selection

The search was conducted on 2nd September 2022, with 293 results. An additional 38 results were identified through manual searches. After removing 34 duplicates, 297

titles and abstracts were screened for eligibility. The full texts of 30 guidelines were retrieved, and eleven guidelines were deemed eligible for inclusion. **Figure 7.1** shows the guideline selection flow diagram.

**Figure 7.1 Guideline selection flow diagram**



### 7.5.2 Study characteristics

Of the eleven included guidelines, four were published for use worldwide (Global Initiative for Asthma [GINA][19], International consensus on paediatric asthma [ICON][20], National Institute for Clinical and Care Excellence [NICE][21] and National Heart Lung and Blood Institute [NHLBI][22,23]). There were additional guidelines from the Canadian Thoracic Society (CTS)[24], British Columbia (BC)[25], the British Thoracic Society/Scottish Intercollegiate Guideline Network (BTS/SIGN)[26], the Irish College of General Practitioners (ICGP)[27], the European Respiratory Society (ERS)[28], Australian National Asthma Council (NAC)[29] and Kaiser Permanente (KP)[30]. All were published or updated between 2012 and 2022. The most recent NHLBI guidance[23] provided six priority updates to its 2007 guideline[22], instead of reviewing all the recommendations. ICGP guidance was based on recommendations from GINA before 2020. KP recommendations were adapted from NHLBI, GINA, NICE and BTS/SIGN guidance. ICON was also based on recommendations from several existing guidelines including NAC, GINA, NHLBI and BTS/SIGN, in addition to others. Guideline characteristics are shown in **Table 7.1**.

**Table 7.1 Characteristics of included guidelines**

| Guideline Publisher | Title | Publication year | Year of update | Country of Publication |
|---|---|---|---|---|
| ***NICE**[21] | Asthma: diagnosis and monitoring of asthma in adults, children and young people | 2017 | 2021 | United Kingdom |
| ***ERS** [28] | European Respiratory Society Clinical Practice Guidelines for the diagnosis of asthma in children aged 5 – 16 years | 2021 | 2021 | Switzerland |
| **NAC** [29] | Australian Asthma Handbook | 2014 | 2022 | Australia |
| ***NHLBI** [22,23] | National Asthma Education and Prevention Program | 2007 | 2020 | USA |

| | | | | |
|---|---|---|---|---|
| | Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma | | | |
| **BTS / SIGN**[26] | British guideline on the management of asthma | 2003 | 2019 | Britain/Scotland |
| ***ICON**[20] | International Consensus on Pediatric Asthma | 2012 | - | Italy |
| **CTS**[24] | Canadian Thoracic Society 2021 Guideline update: Diagnosis and management of asthma in preschoolers, children and adults | 2010 | 2021 | Canada |
| ***GINA**[19] | Global strategy for asthma management and prevention | 1995 | 2022 | Australia |
| **ICGP**[27] | Asthma – Diagnosis, Assessment and Management in General Practice | 2020 | - | Ireland |
| **BC**[25] | Asthma in Children – Diagnosis and Management | 2015 | - | Canada |
| **KP**[30] | Asthma Diagnosis and Treatment Guideline | 1999 | 2021 | USA |

* Intended for worldwide use
**Abbreviations: NICE** – National Institute of Clinical Excellence; **ERS** – European Respiratory Society; **NAC –** National Asthma Council Australian Asthma Handbook; **NHLBI –** National Heart, Lung, Blood Institute; **BTS / SIGN**– British Thoracic Society / Scottish Intercollegiate Guideline Network; **ICON** – International Consensus on Pediatric Asthma; **CTS** – Canadian Thoracic Society; **GINA** – Global Initiative for Asthma; **ICGP** – Irish College of General Practitioners; **BC** – British Columbia Guideline; **KP** - Kaiser Permanente

### 7.5.3 AGREE II score

The quality of reporting in clinical guidelines was highly variable. The median AGREE II score rating the quality of the included guidelines was 4.5 out of 7 (range 2 to 6). The scores for each guideline are shown in **Table 7.2**. Most CPGs demonstrated excellent clarity in the presentation of their recommendations. The third and sixth domains saw the most variation in guideline quality (see **Figure 7.2**). The third domain assessed the rigour of guideline development, specifically transparency in reporting search

strategies, reasons for selecting the evidence, consideration of health benefits, side effects, and risks in implementing the recommendation, and finally, the process for monitoring, auditing, and updating guidelines. The sixth domain related to editorial independence. Scores were downgraded if guideline funding sources and conflicts of interest were not declared (CTS, ICGP, BC, KP) or committee members failed to provide adequate explanations of how their competing interests were addressed (CTS, ICGP, BC, KP). NICE guidance included pharmaceutical companies in their stakeholders list without clarifying these companies were not involved in developing the guideline content. In contrast, NAC and ERS disclosed financial support from pharmaceutical companies and stated they did not influence the guideline content. Other shortcomings of guidelines included failure to consider their applicability, specifically, facilitators and barriers to the guideline's implementation as well as the cost implications. Many guidelines did not also mention whether parents or public representatives were involved in the recommendation development process.

**Figure 7.2 Boxplot demonstrating variation in AGREE II domain scores across eleven asthma guidelines**

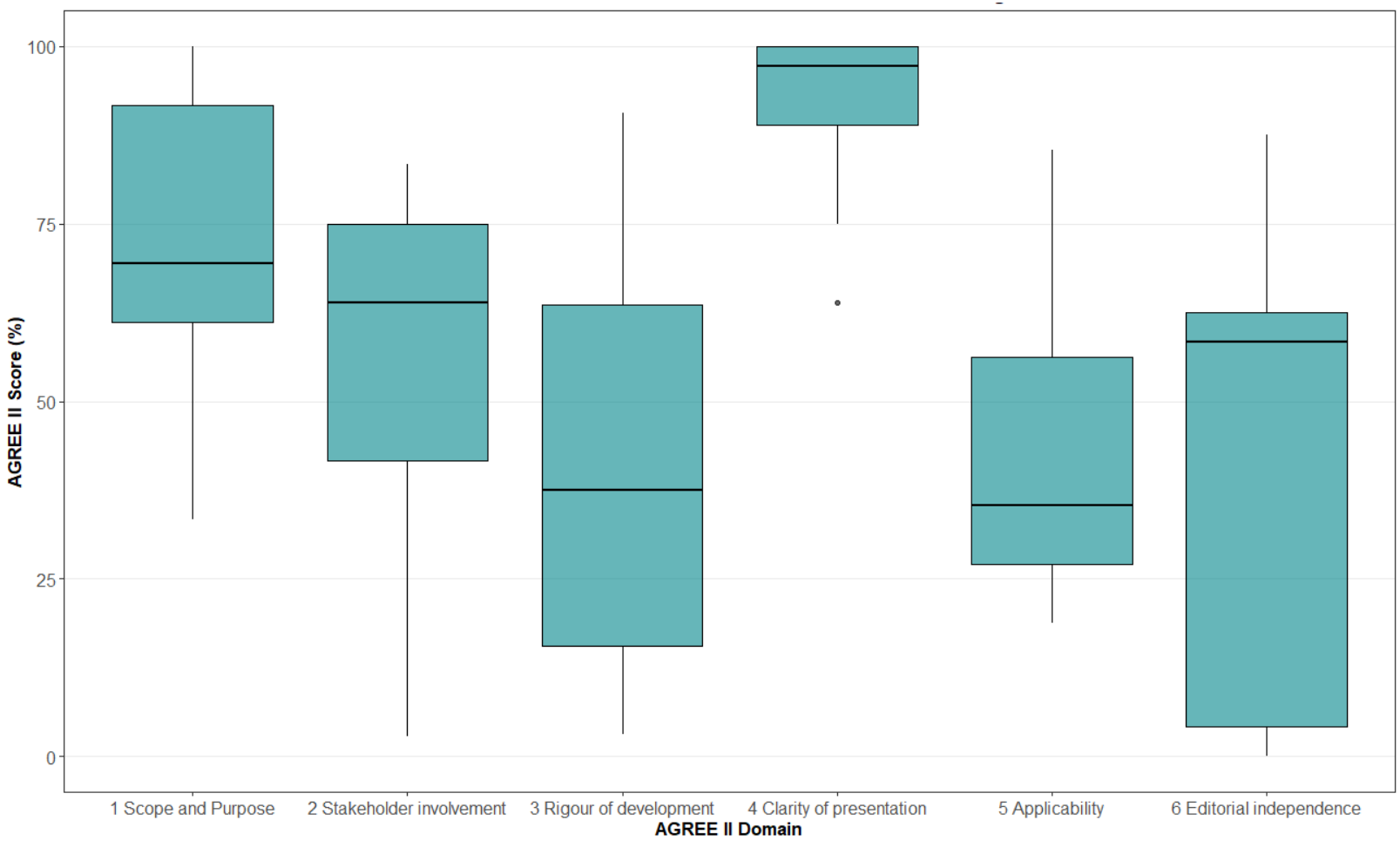


**Table 7.2 AGREE II scores for Asthma Guidelines**

| Guideline | Domain 1: Scope and Purpose | Domain 2: Stakeholder Involvement | Domain 3: Rigour of Development | Domain 4: Clarity of Presentation | Domain 5: Applicability | Domain 6: Editorial Independence | Overall Assessment (out of 7) |
|---|---|---|---|---|---|---|---|
| **NICE** | 94 | 75 | 91 | 100 | 85 | 63 | 6 |
| **ERS** | 100 | 69 | 71 | 100 | 56 | 63 | 6 |
| **NAC** | 89 | 64 | 64 | 97 | 29 | 88 | 6 |
| **NHLBI** | 97 | 94 | 69 | 89 | 56 | 67 | 5 |
| **BTS / SIGN** | 92 | 75 | 38 | 100 | 19 | 75 | 4.5 |
| **GINA** | 39 | 83 | 64 | 94 | 65 | 58 | 4.5 |
| **ICON** | 33 | 28 | 12.5 | 75 | 35 | 46 | 4 |
| **CTS** | 61 | 47 | 21 | 100 | 48 | 4 | 3.5 |
| **ICGP** | 69 | 42 | 16 | 64 | 27 | 0 | 3 |
| **BC** | 61 | 3 | 3 | 89 | 25 | 0 | 3 |
| **KP** | 14 | 15 | 15 | 83 | 0 | 0 | 2 |

**Abbreviations: NICE** – National Institute of Clinical Excellence; **ERS** – European Respiratory Society; **NAC –** National Asthma Council Australian Asthma Handbook **; NHLBI –** National Heart, Lung, Blood Institute; **BTS / SIGN**– British Thoracic Society / Scottish Intercollegiate Guideline Network; **ICON** – International Consensus on Pediatric Asthma; **CTS** – Canadian Thoracic Society; **GINA** – Global Initiative for Asthma; **ICGP** – Irish College of General Practitioners; **BC** – British Columbia Guideline; **KP** - Kaiser Permanente

### 7.5.4 Diagnostic recommendations from each guideline

Across the eleven included guidelines, there were recommendations relating to the use of 17 diagnostic tests (**Table 7.3**). Of these, GPs identified seven tests which were relevant to general practice. There were 50 recommendations for the seven diagnostic tests. **Table 7.4** presents the final tests from each guideline, the associated recommendations, and their GRADE ratings.

Across the guidelines, spirometry and bronchodilator reversibility testing was recommended for children aged over five or six years old for suspected childhood asthma. There was also agreement among guidelines to avoid testing for blood eosinophilia to diagnose asthma. Sputum microbiology was only mentioned by one guideline (NAC) which recommended against its use.

For the three other tests (peak expiratory flow, serum immunoglobulin, and chest X-ray) the guideline recommendations were inconsistent. Notably, NICE guidance conditionally supported the use of peak flow measurement, but the ERS guidance conditionally recommended against the use of peak flow measurement based on evidence from the same primary study. [31]

Despite guideline consensus to perform spirometry, there was disagreement regarding the appropriate thresholds for diagnosing airflow obstruction. GINA, NICE, BTS/SIGN, NAC, ERS, NHLBI, KP and CTS guidelines defined spirometry-based airflow obstruction as a ratio of forced expiratory volume in 1 second (FEV1) to forced vital capacity (FVC), or FEV1/FVC less than the lower limit of normal (usually <90% in children according to GINA, <80% according to ERS, <85% according to NHBLI, and KP, and <80-90% according to CTS). NICE provided an alternative threshold of FEV1/FVC <70% if the lower limit of normal was not available, which was the same definition used by ICGP. The ICON statement did not use either definition; it provided a threshold of FEV1 less than 80% of predicted.

### 7.5.5 GRADE (strength of evidence) supporting recommendations

The GRADE scores for the evidence supporting 50 guideline recommendations are shown in **Table 7.4**. All recommendations were based on either low (n=8) or very low (n=42) quality evidence. 18 of the 50 recommendations had no supporting primary studies and were solely based on expert opinion. Nine recommendations were based on non-primary studies, such as review articles or other guidelines. The median number of cited studies per recommendation was 1 (range 0 – 12, IQR 3.5).

#### *7.5.5.1 Recommendations for younger children*

Of all the diagnostic test recommendations, only four pertained to younger children aged less than 5 or 6, see **Table 7.4**. GINA and ICGP recommended against using spirometry to diagnose asthma, and no primary studies were cited to support this recommendation. GINA and NAC recommended consideration of the use of serum immunoglobulins as an adjunct test to support asthma diagnosis. NAC cited a secondary review article[32] which in turn referenced a primary study[33] reporting the association between serum IgE in infants and the number of wheezing episodes. GINA cited a review article[34] and one research letter[35] reporting a cohort study of infants at high genetic risk for asthma who had measurements collected at various intervals throughout childhood. However, the authors did not specify if serum IgE was even measured. These studies highlight the scarcity of high-quality evidence in this age group.

#### *7.5.5.2 Recommendations for older children*

Half of all the recommendations (25 out of 50) specifically focused on children above the age of 5 or 6 years. Of all the 61 studies that were cited to support these 25 recommendations, only one of the cited primary studies[36] directly related to their corresponding guideline recommendations in ERS. The study authors employed an

appropriate study design to evaluate the diagnostic accuracy of spirometry, bronchodilator reversibility and serum IgE (in addition to other tests), in 111 consecutively sampled patients with suspected asthma.[36] These were compared against a reference standard of physician diagnosis based on history, examination, allergy tests, spirometry and FeNO measurement. They provided diagnostic accuracy measures with 95% confidence intervals for the following FEV1/FVC thresholds:

- <70%
    - Sensitivity = 8% (3 – 16%)
    - Specificity = 99% (88 – 99%)
- <80%
    - Sensitivity = 46% (35 – 58%)
    - Specificity = 93% (78 – 99%)
- <90%
    - Sensitivity = 83% (72 – 90%)
    - Specificity = 27% (12 – 46%)

The low negative predictive value of the test suggests that a negative test does not exclude asthma, though a positive test is highly suggestive.

**Table 7.3 List of all diagnostic tests included in the eleven asthma guidelines**

| Test | | Number of guidelines including test | Recommendations in favour of test use | Recommendations against test use | Recommendations for conditional use* |
|---|---|---|---|---|---|
| **Lung function tests** | Spirometry | 13 | 10 | 2 | |
| | Bronchodilator reversibility testing | 11 | 11 | | |
| | Fractional Exhaled Nitric Oxide - FeNO | 10 | 1 | 3 | 6 |
| | Indirect bronchoprovocation testing (exercise or mannitol challenge test) | 10 | 1 | 1 | 8 |
| | Direct bronchoprovocation testing (methacholine or histamine challenge test) | 9 | 1 | 1 | 7 |
| | Peak expiratory flow | 8 | 1 | 3 | 4 |
| | Residual volume measurements | 2 | | 1 | 1 |
| | Specific airways resistance (sRaw) | 1 | | 1 | |
| | Impulse oscillometry | 1 | | | 1 |
| **Blood tests** | Serum immunoglobulins (including serum IgE/Radioallergosorbent testing) | 7 | 1 | 3 | 3 |
| | Blood eosinophilia | 4 | | 3 | 1 |
| **Imaging** | Chest X-Ray | 5 | | 1 | 4 |
| **Miscellaneous** | Skin allergy tests | 9 | 1 | 2 | 6 |
| | Sputum eosinophils | 2 | | 1 | 1 |
| | Microbiological/sputum testing | 1 | | 1 | |
| | Airway wall biopsy | 1 | | 1 | |
| | Bronchoalveolar lavage | 1 | | 1 | |

*or uncertain on whether test should be used

**Table 7.4 Diagnostic test recommendations from international guidelines for children with suspected asthma and their associated GRADE rating.**
The number of primary studies cited for each recommendation is shown in brackets.

| Test* | Age/subgroup (if noted) | Test recommendation and GRADE rating<br>Red = Not recommended; Orange = Conditional recommendation against;<br>Yellow = Conditional recommendation for; Green = Recommended | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | BC | BTS/SIGN | CTS | ERS | GINA | ICGP | ICON | KP | NAC | NHLBI | NICE |
| **Spirometry** | ≥6 years | Low (6) | | Low (1) | Low[1] (3) | Very Low (3) | Very Low (0) | Very Low [2] (1) | Very Low [3] (4) | Very Low (1) | Very Low [3] (9) | Very Low [3] (1) |
| | <6 years | | | | | Very Low (0) | Very Low (2) | | | | | |
| | Not specified | | Very Low (3) | | | | | | | | | |
| **Reversibility testing** | ≥6 years | Very Low (1) | | Low (0) | | Very Low (3) | Very Low (0) | Very Low [2] (0) | Very Low [3] (4) | Very Low (0) | Very Low [3] (12) | Very Low [1] (0) |
| | All children with obstructive spirometry[4] | | | | Low (5) | | | | | | | |
| | Not specified | | Very Low (5) | | | | | | | | | |
| **Peak expiratory flow** | ≥6 years | | | | Low[1] (1) | | Very Low (0) | Very Low (0) | | Very Low (0) | Very Low (9) | Low (2) |
| | Not specified | Very Low (0) | Very Low (1) | | | Very Low[5] (1) | | | | | | |
| **Serum immunoglobulin** | ≤5 years | | | | | Very Low[6] (2) | | | | Very Low (2) | | |
| | Not specified | | Low (4) | | Very Low (4) | | | Very Low (2) | | | Very Low (0) | Very Low (0) |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Blood eosinophilia** | Not specified | | Very Low (1) | | | | | | | Very Low (3) | Very Low (0) | Very Low (10) |
| **Chest X-Ray** | Not specified | Very Low (0) | Very Low (1) | | | Very Low (0) | | | | Very Low (0) | Very Low (0) | |
| **Sputum microbiology** | Not specified | | | | | | | | | Very Low (0) | | |

***** Seven general practitioners identified these tests as relevant to primary care

[1] 5-16 years

[2] ≥ 5-7 years

[3] ≥5 years

[4] all children with FEV1 <LLN or <80% predicted and/or FEV1/FVC <LLN or <80%

[5] once obstructive defect confirmed

[6] 1-5 years

### 7.5.6 Subgroup analysis

There were only four recommendations specifically for children under five (or six) and 24 for children aged five years (or six years) and over, with no association between the age group and quality of evidence (p= 0.15, Fisher's exact test for count data). There were 24 recommendations 'for' doing the test and 14 'against' with no statistically significant association with the strength of supporting evidence (low vs very low, p=0.36). Test type (e.g., blood test or imaging) was not associated with evidence quality (p=0.54). These subgroup analyses were, however, exploratory and under-powered to detect a true association.

## 7.6 Discussion

### 7.6.1 Summary of evidence

Guidelines aim to reflect the appropriate standards of clinical care. This study demonstrated the variable quality of methodology and reporting across international guidelines for childhood asthma. Few asthma guidelines employed rigorous methods to search for the best available evidence, considered barriers and facilitators in implementing guideline recommendations, appropriately involved all relevant stakeholders (including parents or public representatives), or had all guideline contributors declare and appropriately address conflicts of interest. The quality of evidence supporting childhood asthma diagnostic recommendations was also generally poor. Exploratory analyses showed that the quality of evidence did not vary based on age group, test type, or if the recommendation was 'for' or 'against' performing the test.

There were discrepancies in guidelines' recommendations about whether to perform a test based on different interpretations of the supporting evidence. Even though guidelines agreed that spirometry should be utilised for diagnosing asthma in children older than five years, the recommended diagnostic test thresholds differed. This would

lead to significant uncertainty around testing and diagnosing asthma in children and unreliable disease prevalence estimates.

The poor quality of evidence supporting recommendations and inconsistencies in guideline recommendations for a given diagnostic test would translate to variation in testing in clinical practice. This could mean some children receive tests unnecessarily, and others do not get tested, missing out on the appropriate diagnosis and treatment. This also has systemic implications relating to equity and fair distribution of resources.

### 7.6.2 Strengths and Limitations

This meta-epidemiological study has several strengths. First, it provides a comprehensive overview and comparison of international guidelines and their recommendations for childhood asthma diagnosis. This is the first study to appraise the quality of childhood asthma guidelines and supporting evidence for diagnostic recommendations. Knowing the quality of evidence underpinning asthma tests can inform clinical decision-making and it also highlights evidence gaps for future research.

The AGREE II tool assesses the quality of reporting of guidelines, however, this does not necessarily correlate with a guideline's applicability in clinical practice. NICE performed a feasibility evaluation of its asthma guidance and found that of the 33 people diagnosed with asthma (including children and adults), only nine participants had an obstructive pattern on spirometry and 24 people had a normal result.[21] Some of the highlighted issues in the evaluation included the poor applicability of the diagnostic recommendations to younger children (under the age of eight) given their difficulties in performing spirometry.

Due to resource constraints, my study was limited by the exclusion of non-English guidelines. Given the widespread adoption of English-language guidelines by local guideline bodies and clinicians from non-English-speaking countries, it was deemed improbable that their guidelines would incorporate recommendations based on high-

quality evidence, given the established scarcity of research in this area. It is, however, possible that their guidelines had better reporting quality compared with the English-language guidelines included in our study. This study was also limited to guidelines published from high-income countries with similar primary care systems for paediatric care. I included guidelines from the US because it was considered plausible that clinicians/local guideline bodies from other countries may adopt recommendations from the US. Overall, it is unlikely that guidelines published in the English language that used more robust methodologies and incorporated high-quality evidence were missed.

The scope of this study was limited to diagnostic tests which GPs could regularly use in the UK, which I based on seven UK GPs' perspectives. I acknowledge, however, that test availability varies regionally in primary care, and tests may have been missed. For example, FeNO testing, though not readily accessible in all primary care practices is used in some (as evidenced by their high practice variability in Chapter 4), with recently published studies evaluating their feasibility in this setting.[37] Further research is required to establish the accuracy and value of FeNO for paediatric asthma diagnosis in primary care. Additionally, we did not evaluate other diagnostic methods such as a trial of salbutamol treatment, which may be an effective diagnostic strategy for those who are unable to perform spirometry.

### 7.6.3 Implications for future research and practice

Poor quality evidence or the lack of evidence limits the ability of guideline developers to develop robust CPGs that inform practice. Guidelines could include specific research recommendations to address evidence gaps. The very low quality of included evidence for diagnostic recommendations highlights the need for research collaboration to produce high-quality systematic reviews and diagnostic accuracy studies to provide evidence for tests in a primary care setting. In younger children especially, there is a critical need for new research to evaluate existing diagnostic methods (such as trial of treatment) or to develop newer diagnostic strategies. Guideline committees should also

adhere to strict methodological standards so that important studies that could influence guideline recommendations are not missed.

There are several proposed drivers of poor clinician adherence to practice guidelines, and among these, a lack of clarity and credibility in the evidence has been cited.[38] Other proposed barriers for the implementation of CPGs include financial, personnel and time constraints, lack of clinician confidence or knowledge about the guideline, and the sociocultural beliefs of patients. Future research should explore these potential barriers in the context of childhood asthma.

For the time being, practitioners need to be aware of the variation in CPGs for asthma and the limitations of existing evidence for asthma tests. NICE guidance had the highest reporting standards, despite some limitations related to their contributors' competing interests, and applicability for children in primary care. Until there is a new gold standard for asthma diagnosis, clinicians should utilise spirometry with the lower limit of normal diagnostic threshold and bronchodilator reversibility if childhood asthma is clinically suspected. Most importantly, the diagnosis should be regularly reviewed and updated.

The practice of evidence-based medicine integrates clinical experience with the best available evidence; taken together they can improve the targeting of diagnostics to improve asthma outcomes.

### 7.6.4 How this chapter fits in within the wider thesis

In the previous chapter, I identified that the quality of diagnostic guidelines relating to children in primary care was highly variable. In this chapter, I delved deeper into the role of guidelines as a potential driver of variation by looking at the specific diagnostic recommendations for asthma and the quality of evidence supporting them. While all guidelines generally agreed that spirometry should be used to confirm a diagnosis, the

thresholds for diagnosis differed. Interestingly, spirometry was noted to be one of the tests that have decreased over time in general practice and subject to higher-than-average variability between practices in Chapter 4. This chapter also identified inconsistencies across guidelines on which other tests should or should not be used. The evidence supporting asthma diagnostic test recommendations were either of poor or very poor quality; this may be correlated with their declining use and variation in practice.

## 7.7 Conclusions

Good-quality evidence is lacking to support the use of diagnostic tests in childhood asthma. Discrepant guideline recommendations based on the same evidence may contribute to poor clinician adherence to asthma guidelines and variation in testing practices. Therefore, guideline bodies should prioritise evidence to address the essential gaps and employ rigorous methods to identify all the best available evidence when making diagnostic recommendations for children with asthma.

## 7.8 Chapter summary

- A previous study demonstrated poor GP adherence to childhood asthma guidelines. In this chapter, I showed the quality of these diagnostic guidelines was highly variable.
- The evidence supporting diagnostic test recommendations was generally of very poor quality.
- There was discrepant advice on the recommended diagnostic tests and diagnostic thresholds for asthma across international guidelines.
- Guideline bodies should prioritise evidence when developing recommendations and improve the quality of their reporting.

- High quality evidence is needed to evaluate existing and emerging diagnostic tests.
- Clinicians should be aware of the discrepancies in testing recommendations and limitations of the current evidence base when making a paediatric asthma diagnosis.

## 7.9 References


1. Pearce N, Aït-Khaled N, Beasley R, et al. Worldwide trends in the prevalence of asthma symptoms: Phase III of the International Study of Asthma and Allergies in Childhood (ISAAC). Thorax 2007;62. doi:10.1136/thx.2006.070169.
2. World Health Organization Europe. Prevalence of asthma and allergies in children. 2007.
3. Rosenfeld RM, Shiffman RN. Clinical practice guideline development manual: A quality-driven approach for translating evidence into action. Otolaryngology - Head and Neck Surgery 2009;140. doi:10.1016/j.otohns.2009.04.015.
4. Woolf SH, Grol R, Hutchinson A, et al. Clinical guidelines. Potential benefits, limitations, and harms of clinical guidelines. Br Med J. 1999;318. doi:10.1136/bmj.318.7182.527.
5. Homaira N, Wiles LK, Gardner C, et al. Assessing appropriateness of paediatric asthma management: A population-based sample survey. Respirology 2020;25. doi:10.1111/resp.13611.
6. Ransohoff DF, Pignone M, Sox HC. How to decide whether a clinical practice guideline is trustworthy. JAMA - Journal of the American Medical Association. 2013;309. doi:10.1001/jama.2012.156703.
7. OCEBM Levels of Evidence Working Group. The Oxford Levels of Evidence 2. Oxford Centre for Evidence-Based Medicine: https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence. 2011.
8. O'Sullivan JW, Albasri A, Koshiaris C, et al. Diagnostic test guidelines based on high-quality evidence had greater rates of adherence: a meta-epidemiological study. J Clin Epidemiol 2018;103. doi:10.1016/j.jclinepi.2018.06.013.
9. Martin J, Townshend J, Brodlie M. Diagnosis and management of asthma in children. BMJ Paediatr Open 2022;6. doi:10.1136/bmjpo-2021-001277
10. Cope SF, Ungar WJ, Glazier RH. International Differences in Asthma Guidelines for Children. Int Arch Allergy Immunol 2009;148:265–78. doi:10.1159/000170380.
11. Murad MH, Wang Z. Guidelines for reporting meta-epidemiological methodology research. Evid Based Med 2017;22. doi:10.1136/ebmed-2017-110713.
12. Brouwers MC, Kerkvliet K, Spithof K. The AGREE reporting checklist: A tool to improve reporting of clinical practice guidelines. BMJ (Online) 2016;352. doi:10.1136/bmj.i1152.

13. Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. The BMJ 2004;328.
14. van Esso D, del Torso S, Hadjipanayis A, et al. Paediatric primary care in Europe: Variation between countries. Arch Dis Child 2010;95. doi:10.1136/adc.2009.178459.
15. Schünemann HJ, Mustafa RA, Brozek J, et al. GRADE guidelines: 21 part 1. Study design, risk of bias, and indirectness in rating the certainty across a body of evidence for test accuracy. J Clin Epidemiol. 2020;122. doi:10.1016/j.jclinepi.2019.12.020.
16. Schünemann HJ, Mustafa RA, Brozek J, et al. GRADE guidelines: 21 part 2. Test accuracy: inconsistency, imprecision, publication bias, and other domains for rating the certainty of evidence and presenting it in evidence profiles and summary of findings tables. J Clin Epidemiol. 2020;122. doi:10.1016/j.jclinepi.2019.12.021.
17. Whiting PF, Rutjes AWS, Westwood ME, et al. Quadas-2: A revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8). doi:10.7326/0003-4819-155-8-201110180-00009.
18. R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
19. Reddel HK, Bacharier LB, Bateman ED, et al. Global Initiative for Asthma Strategy 2021. Respirology 2022;27. doi:10.1111/resp.14174.
20. Papadopoulos NG, Arakawa H, Carlsen KH, et al. International consensus on (ICON) pediatric asthma. Allergy: European Journal of Allergy and Clinical Immunology 2012;67. doi:10.1111/j.1398-9995.2012.02865.x.
21. NICE. Overview | Asthma: diagnosis, monitoring and chronic asthma management | Guidance | NICE. National Institute for Health and Care Excellence 2017.
22. National Heart Lung and Blood Institute, National Asthma Education and Prevention Program. Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma Full Report 2007. U.S. Department of Health and Human Services National Institutes of Health.
23. National Heart Lung and Blood Institute. 2020 Focused Updates to the Asthma Management Guidelines: A Report from the National Asthma Education and Prevention Program Coordinating Committee Expert Panel Working Group. U.S. Department of Health and Human Services National Institutes of Health.

24. Yang CL, Hicks EA, Mitchell P, et al. Canadian Thoracic Society 2021 Guideline update: Diagnosis and management of asthma in preschoolers, children and adults. Canadian Journal of Respiratory, Critical Care, and Sleep Medicine 2021;5. doi:10.1080/24745332.2021.1945887.
25. British Columbia Guidelines. Asthma in Children - Diagnosis and Management. https://www2.gov.bc.ca/gov/content/health/practitioner-professional-resources/bc-guidelines/asthma-children#:~:text=Diagnosis%20and%20Management.-,Key%20Recommendations,medication%20daily%20and%20not%20intermittently. 2015.
26. British Thoracic Society, NHS Scotland. SIGN158: British guideline on the management of asthma. 2019.
27. Irish College of General Practitioners. Asthma - Diagnosis, Assessment and Management in General Practice. 2020.
28. Gaillard EA, Kuehni CE, Turner S, et al. European respiratory society clinical practice guidelines for the diagnosis of asthma in children aged. European Respiratory Journal 2021;58. doi:10.1183/13993003.04173-2020.
29. National Asthma Council. Australian Asthma Handbook. 2022.
30. Kaiser Permanente. Asthma Diagnosis and Treatment Guideline. https://wa.kaiserpermanente.org/static/pdf/public/guidelines/asthma.pdf.
31. Brouwer AFJ, Visser CAN, Duiverman EJ, oorda R, Brand Pet al. Is home spirometry useful in diagnosing asthma in children with nonspecific respiratory symptoms? Pediatr Pulmonol. 2010;45(4):326–332. doi:10.1002/ppul.21183.
32. Brand PLP, Baraldi E, Bisgaard H, Boner AL, Castro-Rodriguez JA, Custovic A, de Blic J, de Jongste JC, Eber E, Everard ML, et al. Definition, assessment and treatment of wheezing disorders in preschool children: an evidence-based approach. Eur Respir J. 2008;32:1096–1110.
33. Rusconi F, Patria MF, Cislaghi GU, Sideri S, Gagliardi L, et al. Total serum IgE and outcome in infants with recurrent wheezing. Arch Dis Child. 2001;85(1):23–25. doi:10.1136/adc.85.1.23.
34. Ahlstedt S, Murray CS. In vitro diagnosis of allergy: how to interpret IgE antibody results in clinical practice. Prim Care Respir J. 2006;15:228–236.
35. Azad MB, Chan-Yeung M, Chan ES, Dytnerski AM, Kozyrskyj AL, Ramsey C, Becker AB. Wheezing patterns in early childhood and the risk of respiratory and allergic disease in adolescence. JAMA Pediatr. 2016;170:393. doi:10.1001/jamapediatrics.2015.4127.

36. De Jong CCM, Pedersen ESL, Mozun R, Goutaki M, Trachsel D, Barben J, Kuehni C, et al. Diagnosis of asthma in children: the contribution of a detailed history and test results. Eur Respir J. 2019;54 (6):1901326. doi:10.1183/13993003.01326-2019.
37. Lo D, Beardsmore C, Roland D, Richardson M, Yang Y, Danvers L, Wilson A, Gaillard EA. Spirometry and FeNO testing for asthma in children in UK primary care: a prospective observational cohort study of feasibility and acceptability. Br J Gen Pract. 2020;70:e809–e816.
38. Correa VC, Lugo-Agudelo LH, Aguirre-Acevedo DC, Contreras J, Borrero A, Patiño-Lugo D, Valencia D, et al. Individual, health system, and contextual barriers and facilitators for the implementation of clinical practice guidelines: a systematic metareview. Health Res Policy Syst. 2020;18(1):74. doi:10.1186/s12961-020-00588-8.

# Chapter 8. Discussion and conclusions

## 8.1 Introduction

I was introduced to the concepts of overdiagnosis, over-testing, and over-treatment during my undergraduate medical studies. I first worked on a systematic review which examined breast cancer overdiagnosis, a contentious topic at a time when breast cancer screening was, and continues to be, heavily promoted. As part of my Doctor of Medicine Thesis, I also conducted a study which explored natural age-related decline in lung function to prevent overdiagnosis of chronic airways disease in the elderly. When I attended the Preventing Overdiagnosis conference 2016 to present my paper on breast cancer overdiagnosis, it struck me how pervasive the problem of potential overdiagnosis was across every healthcare discipline. This realisation challenged my idealistic views of medicine as purely a force for good, to think that patients could be entangled in a cycle of unnecessary appointments, tests, treatments, and surveillance which would not necessarily equate to improved long-term health outcomes. Health "care" delivered by well-meaning practitioners could also be wasteful and harmful.

When I started working in clinical practice as a junior doctor, I reflected on how often my colleagues and I would request tests needlessly. This was part of the hospital culture. As junior doctors, we practised "defensive" medicine; in a fast-paced and highly critical environment, ordering more tests assuaged some of that fear and anxiety of "not doing enough" or missing a diagnosis. I also experienced the problem firsthand when my sister, at the age of six, was diagnosed with asthma in the emergency department after an episode of wheezing. The registrar who examined her incidentally found a heart murmur that led to an echocardiogram performed at the State's Children's Hospital one-and-a-half hours away, which showed that the murmur had no pathological features. I wondered at the time if a different clinician would have also

referred her based on the same examination findings. This prompted me to think about variation in paediatric testing.

Variation in paediatric healthcare has been researched previously. The first known study of variation in healthcare by J Alison Glover in 1938 revealed widespread differences in children's tonsillectomies and highlighted the procedure as "preference-sensitive", as described in **Introduction Chapter 2.4**. The NHS Atlas of Variation published an edition related to the health of children and young people in 2012[1] which measured healthcare use for pre-specified indicators (e.g., admission rates, health expenditure, mortality, immunisation coverage) that included very few diagnostic test indicators. An Atlas of NHS Diagnostic Services was published in 2017 but included only two indicators related to children and young people: admission rate for endoscopy and the proportion of newborns with a newborn blood spot test recorded.[2] I did not identify other sources that specifically examined diagnostic test variation for children, specifically testing related to primary care. Test use carries potential implications for children and young people and their families, staff, and the health system at large, therefore, it is essential to investigate. This body of work therefore sought to quantify and analyse variation in paediatric test use as well as drivers of variation in testing within primary care.

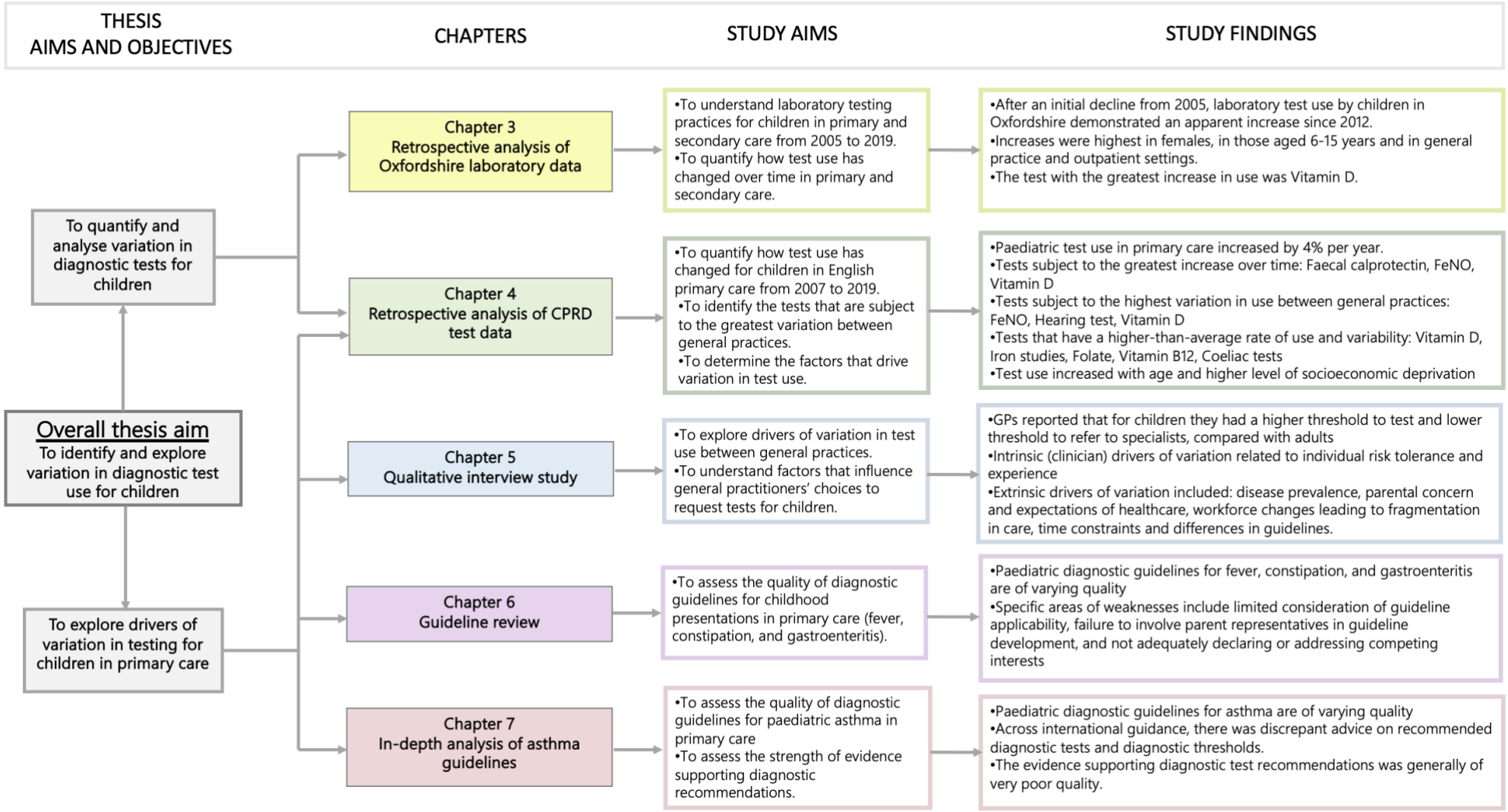
## 8.2 Summary of findings

I conducted five studies to address the overall aims of my thesis as well as identify the evidence gaps. The aims, methods and findings of each chapter are summarised in **Figure 8.1.**

Chapters 3 and 4 focused on quantifying and analysing variation in diagnostic tests.

Chapter 3 was a retrospective analysis of laboratory data from Oxford University Hospitals NHS Trust Laboratories, which aimed to examine the landscape of diagnostic tests for children across primary and secondary care and investigate how test use has changed in Oxfordshire over a 15-year-period. 1.7 million laboratory tests were conducted during the study period, and over two thirds (71%) of tests were conducted in the inpatient setting. Tests in general practice comprised 17% of tests, where test use increased by 4% per year during the study period. Overall test use demonstrated an initial decline until 2012 but increased thereafter. The greatest increases occurred for females aged 6-15 years, in outpatient and general practice settings. The most frequently requested blood tests overall were identified. Vitamin D testing increased by the greatest margin overall, with a 27% increase per year from 0.9 tests per 1,000 child-years to 28.9 tests per 1,000 child-years. This was followed by parathyroid hormone, which increased 10% per year from 1.3 tests per 1,000 child-years to 4.9 tests per 1,000 child-years, then iron studies which increased 9% per year from 9.2 tests per 1,000 child-years to 33.3 tests per 1,000 child-years. Vitamin $B_{12}$, folate, and coeliac tests also increased. These trends were not consistent across settings and age groups.

**Figure 8.1 Overall summary of thesis**



Chapter 4 was a retrospective analysis of diagnostic test data from the Clinical Practice Research Datalink (CPRD) Aurum database. I aimed to analyse temporal variation from 2007 to 2019, practice variability of test use in 2019, and determined how demographic factors influence testing rates. Overall test use increased by 4% per year, from 399 tests per 1,000 child-years to 2007 to 608 tests per 1,000 child-years in 2019. This appeared to be largely driven by an increase in blood tests (which comprised half of the total requested tests); these increased by 8% per year from 145 tests per 1,000 child-years at baseline to 357 tests per 1,000 child-years at the end of the study. Testing rates were generally higher in females compared with males, and this gender-difference was most evident in the 11–15-year age group. Children from the most socioeconomically deprived group also received more tests. Faecal calprotectin tests and fractional exhaled nitric oxide (FeNO) tests for asthma were subject to the greatest temporal increases. This was because they were rolled out between 2009 and 2010 in general practice, increasing by 105% and 40% per year, respectively, from a baseline of 0 tests per 1,000 child-years in 2007. These were followed by vitamin D, folate, vitamin $B_{12}$, coeliac testing, helicobacter testing and iron studies. Tests subject to the greatest practice variation were FeNO, hearing test, vitamin D, helicobacter and monospot testing. Tests subject to high rate of use and high practice variability included Vitamin D, coeliac test, folate, vitamin $B_{12}$, and iron studies, which indicates that these tests were potentially overused.

Chapter 5 to 7 explored the potential drivers of testing variation in primary care.

Chapter 5 described a qualitative interview study with 18 general practitioners and two general practitioner trainees, exploring their views on using tests for children in primary care. I sought to understand factors that underpin their diagnostic decision-making in

children and identify perceived drivers of variation in testing practices. Based on these interviews, the following themes were conceptualised: 1) how decisions to test in children differed from adults; 2) the perceived utility of tests, and 3) the perceived drivers of variation, which included clinician-specific (intrinsic) factors, and other factors related to the child, parent, test, colleague testing behaviours, work environment, resources, and local policies and procedures as well as clinical practice guidelines (extrinsic).

Chapter 6 specifically examined the quality of diagnostic guidelines relevant to children presenting to primary care to assess whether this could be a potential driver of variation. 16 guidelines which included diagnostic recommendations for fever, constipation, and gastroenteritis were included. The methodological quality and reporting of these guidelines (assessed using AGREE II) was found to be highly variable, with major weaknesses relating to limited consideration of guideline applicability, failure to involve parent representatives in guideline development, and inadequately declaring or addressing competing interests.

Chapter 7 provided an in-depth analysis of guidelines and focused on asthma, a condition that is prone to diagnostic uncertainty. I performed a meta-epidemiological study that, appraised the quality of 11 international guidelines using AGREE II, examined the differences in diagnostic test recommendations across guidelines, and assessed the strength of evidence supporting these diagnostic recommendations. I found that the quality of asthma guidelines varied, and they provided discrepant advice on which tests to use to diagnose childhood asthma. The quality of evidence underpinning all 50 diagnostic test recommendations for asthma was either Low or Very Low, and 36% of recommendations were based on expert opinion (where no studies were cited to inform the recommendation). Therefore, the varying quality of

guidelines, differences in diagnostic recommendations, and poor quality of evidence were identified as factors that are likely to contribute to variation in test use.

## 8.3 Findings in the context of the wider literature

**Section Summary**

- Previous research has shown increased test use and variation in adults in primary care; however, this is the first known study of paediatric testing trends in this setting.
- Other work has aimed to describe variation in paediatric clinical care, but these have been focused on specific conditions or presentations.
- Other studies have sought to understand the drivers of variation in paediatric clinical care which identified similar drivers of variation related to clinicians, parents and families, and systems.
- *Helicobacter pylori* testing provides a case example of a test that is subject to increasing use and practice variation. In the context of the guideline recommendations, and the low risk of developing serious disease, there is a potential that *Helicobacter pylori* testing is overused.

### 8.3.1 Comparing paediatric testing trends with adults in primary care

This is the first known study of paediatric testing trends in general practice. The problems of increasing test use and variation has been well researched and established in adults, however, this thesis highlights that the problem extends to children. There are additional considerations to be made in using tests in children given their differing physiology, vulnerability, the higher degree of diagnostic uncertainty, and the greater potential for harm compared with adults.

O'Sullivan used similar methods to measure temporal trends and practice variation in test use by adults in primary care. Test rates in adults increased by 8.5% per year from 2000 to 2015[3], which is even higher than the 3.6% per year temporal increase

observed in children up to 2019. In both children and adults, vitamin D tests were identified as tests subject to high temporal increases as well as high practice variability.[4] Substantial increases and practice variability in test use across English general practice reflects inconsistency in the guidance and evidence base for diagnostic tests.

### 8.3.2 Examining paediatric test variation in other settings

My research focused on variation in paediatric diagnostic testing in general practice, which has not been studied before. Previous research examined variations in paediatric clinical care across inpatient, outpatient, and emergency settings internationally. Only a few studies evaluated variations in diagnostic testing, and they were limited to a specific presentation or condition.[5–7] For example, a cross-sectional analysis of a national hospital dataset in the United States found widespread variation in rates of diagnostic lumbar punctures when newborns presented with suspected sepsis.[8] The odds of performing a lumbar puncture were higher in children who had Medicaid coverage, and those born in urban or teaching hospitals.[8] Another retrospective chart review in the US reported variation in the proportion of children with acute exacerbations of asthma receiving ancillary testing (such as Chest X-Rays and blood tests) in emergency departments, and examined the factors associated with test use.[9] Chest X-Rays were more frequently ordered in patients with fever. Lower rates of testing occurred when (1) the physician had subspecialty training in paediatric emergency medicine, (2) the child was treated at a children's hospital, (3) had normal oxygen saturation, and (4) had a disposition to home.[9] In contrast, my quantitative analyses (Chapter 3 and Chapter 4) were disease-agnostic. Factors related to specialty training and patient clinical features were not explored in my thesis due to the unavailability of such data, but future analyses should seek to include this.

### 8.3.3 Factors driving paediatric clinical variation

A previous evidence review from Australia examined the factors driving variation in paediatric clinical care[10,11], however, no studies have taken place in primary care or examined the use of diagnostics. The studies that explored factors driving paediatric care variation were mostly based in outpatient settings. They proposed similar drivers to the factors I described in Chapters 5, 6 and 7. These included:

- Clinician-factors: lack of awareness of clinical practice guidelines, poorly written clinical practice guidelines, lack of familiarity with recommended practices, clinical uncertainty, clinicians' preconceived views, varying degrees of clinician experience/competency, physician preferences, physicians following adult-based care, increased pressure placed on clinicians.
- Patient/family factors: case complexity, patient characteristics, health literacy of family members, improper or lack of adequate information provided to family members, patient/family preferences and parental enthusiasm.
- Systemic factors: differences in educational materials provided to patients, lack of standardised practices/recommendations, access to care setting and cost of transportation, supply factors and institutional restraints including the availability of physicians, resources, or beds, lack of time, differences in local and regional primary care systems.

Interestingly, the Australian review also identified factors associated with reduced variation in effective paediatric hospital care. These included[10,11]:

- Children's hospitals compared with general hospitals.
- Hospital-based doctors compared with community-based doctors.
- Younger clinicians compared with older clinicians.
- Adherence to clinical practice guidelines, and

- Use of electronic order sets within electronic medical records that can help guide and standardise practice.

The results of reduced variation being associated with younger clinicians appeared to contrast with empirical observations by paediatricians in Chapter 3.6 and the results presented in Chapter 5, where GPs believed that older, more experienced doctors may be more likely to request fewer tests compared with their junior colleagues. This has not been studied prospectively so there is no high-quality evidence to support either hypothesis. Further research is needed to investigate this claim.

### 8.3.4 A closer look: examining variation in paediatric testing for *Helicobacter pylori* infection

A few GPs that I interviewed for the qualitative study observed variation in testing for *Helicobacter pylori* across GP practices and hence I included this in my Chapter 4 analysis. While there were low testing rates overall, the quantitative analysis confirmed high temporal and practice variability. Arguably, evaluating test variation and appropriateness in this space is important particularly if testing results in further invasive testing and treatment which can cause net patient harm and incurs high healthcare costs. Non-invasive *H. pylori* testing encompasses *H. pylori* serology, stool testing and urease breath testing. This may lead to invasive gold standard testing, which is a gastric or duodenal biopsy, acquired via gastroduodenoscopy.

GP access to gold standard testing, and gastroenterology specialists vary geographically.[2] It is relevant to mention that the non-invasive *H. pylori* diagnostic test performance in children is highly variable and age dependent. For example, a meta-analysis showed that the urea breath test was less accurate for the diagnosis of *H. pylori* infection in children aged 6 or younger, where sensitivity estimates ranged from 50 to 100% across the studies; the overall summary sensitivity estimates were 89%

(95% confidence interval 80 – 95%) when using a tracer dose based on body weight and 93% (95% CI 91 – 97%) with a unique tracer dose. In contrast, the sensitivity estimate in children aged older than 6 years was very high, at 97% (95% CI 95 – 98%).[12] The authors concluded that to improve its accuracy in younger children, the cut-off values for diagnosis and urea dose need to be adjusted.[12]

The global prevalence of *H. pylori* infection in children is 32.3% (95% confidence interval 27.3 – 37.8%) though rates in high income countries are lower at 21.7% (95% CI 16.9 – 27.4%).[13] The recognized complication of *H. pylori* infection is the development of peptic ulcers in adulthood, which affects 10 – 15% of infected children.[14] The European Society for Paediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN) and their North American counterpart (NASPGHAN) jointly published guidance[11] recommending against testing and treating children with *H. pylori* infection in the absence of peptic ulcer disease. This is likely to the fact that infection rarely gives rise to other symptoms. Whilst this was a strong recommendation, the quality of evidence underpinning this recommendation was low because of its indirectness, though there was 100% agreement from the consensus group.

The indications for *H. pylori* eradication treatment are also limited. There would be an indication to treat in the presence of active peptic ulcer disease[14] and possible indication to treat if *H. pylori* test positive and the child also has a family history of peptic ulcer disease. The joint European and North American Societies of Paediatric Gastroenterology, Hepatology and Nutrition guidance recommends against testing in children (1) with functional abdominal pain disorders (High quality of evidence, agreement 100%), (2) in the initial workup for iron deficiency anaemia (Moderate quality of evidence, agreement 93%) or (3) investigating causes of short stature (Moderate quality of evidence, agreement 79%).[15]

This evidence highlights that *H. pylori* testing in children may be over-utilised and unwarranted in UK general practice. Understanding whether testing was appropriate with more certainty (defined as guideline concordant testing or discordant testing) involves comparing guideline recommendations with the clinical indications for each *H. pylori* test. However, this approach has its own fallibilities; guidelines may not be the best measure of appropriate care if the evidence base underpinning their recommendations are limited or low quality.

## 8.4 Strengths and limitations

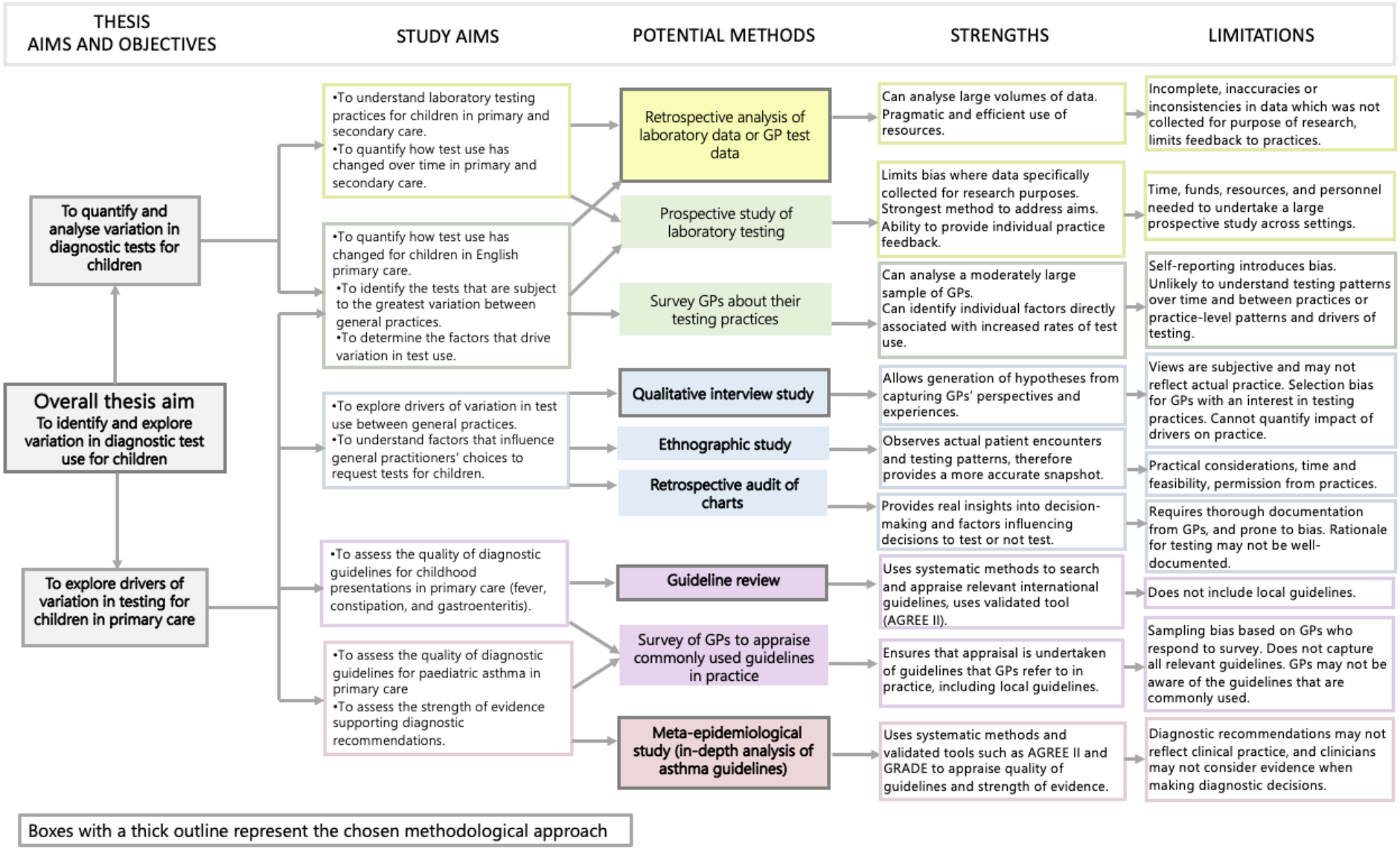
**Section Summary**

- A key strength of this thesis is the diversity of methodological approaches (quantitative, qualitative, evidence appraisal) used to summarise the landscape of paediatric diagnostic testing, using both local and national data sources.
- The thesis was strengthened by the involvement of parent representatives who provided their personal perspectives and advice and informed the research process.
- Limitations included the inability to provide feedback to individual practices on their testing patterns, possibility of bias due to the retrospective nature of data collection, and its unknown external validity and relevance in the post-pandemic era.

This is the first known study of paediatric testing trends in general practice based on a comprehensive review of the literature. This thesis provides a starting point for measuring whether tests are being performed in an effective, evidence-based, and equitable manner. Future research should aim to provide greater clarity on the observed trends and investigate testing patterns in other healthcare settings, for example in specialty outpatient settings where testing was noted to increase by the highest margin in Chapter 3.

The considered benefits and limitations of each methodological approach in addressing the thesis aims and objectives are briefly outlined in **Figure 8.2**.

**Figure 8.2 Strengths and limitations of various methodological approaches to address the thesis aims and objectives**



A key strength of this thesis is the diversity of methodological approaches used to address the overarching research aims. I used validated statistical methods to analyse laboratory data in Oxfordshire (Chapter 3) and CPRD data (Chapter 4). The qualitative methods used in Chapter 5 sought to complement the findings of the preceding chapters, and to provide meaning and context to the quantitative findings. The study themes expanded on current knowledge of testing issues in children. Interviewing GPs proved to be essential to building an understanding about how testing is implemented in general practice. I also used systematic methods to identify relevant international guidelines (Chapter 6 and 7), and applied evidence synthesis skills to examine the evidence base for diagnostic recommendations contained within guidelines and appraise the quality of the evidence using GRADE (Chapter 7). Four of these studies have been peer-reviewed and published in academic journals: Archives of Disease in Childhood (Chapter 3), the British Journal of General Practice (In Press, Chapter 5), the Journal of Paediatrics and Child Health (Chapter 6), and Family Practice (Chapter 7).

Though my thesis had a focus on primary care due to the inherent challenges of diagnosing children in general practice, I also examined test use more broadly across general practice, inpatient (including emergency) and outpatient care in Chapter 3. This was conducted to establish where general practice fits into the overall landscape of diagnostic testing for children and understand how test use has changed in general practice compared with other settings.

This thesis utilises both local and national data. This allowed for a granular and aerial understanding of testing patterns and trends in the UK. Moreover, I sought to examine some of the drivers of testing variation on a global level by analysing international guidelines from high-income countries including the US as well as countries with a similar paediatric primary care system to the UK. Therefore, the findings of this thesis

have global implications for improving the value of diagnostics and the quality of care for children.

One of the major strengths of this thesis is the involvement of parent representatives who formed a Patient and Public Involvement (PPI) advisory panel for Chapters 4 and 5. The three PPI representatives were from different parts of the UK and were each a parent of three to four children with differing experiences of health care. We held several meetings to discuss the planned research and gather their perspectives and experiences of their children having tests. Their perspectives were invaluable and shaped the qualitative study, especially regarding future directions of this strand of research.

Early on in my DPhil journey, I learned about the principles of Open Science and the importance of ensuring research is transparent and reproducible. I have applied these to my research, by registering my protocols on the Open Science Framework, sharing my web scraping and analytic code on GitHub, and ensuring that my studies have all been published in Open Access journals. While the data used for Chapters 3 and 4 are not available openly, I provided instructions on how I applied for these data and shared the code that details my data cleaning, manipulation, management, and analysis.

This thesis has several limitations. First, it does not provide feedback to individual practices on their testing behaviours as the data were de-identified on a practice and patient level. This prevented me from generating a heat map of testing rates across different practices (like the NHS Atlases of Variation), which can be a valuable way to identify which areas are high or low testers. This information would have enabled me to identify, approach, and potentially interview GPs from practices that requested more higher rates and lower rates of testing compared with other practices on average for Chapter 5, to elucidate some of the key differences in testing behaviours that drive variation. To address this limitation, I contacted the consultant chemical pathologist at

the Oxford University Hospitals NHS Trust clinical laboratory who advertised the qualitative study recruitment material to GPs who requested the most tests in the county prior to the onset of the pandemic. Three of these GPs participated in the study.

There are also inherent limitations associated with using retrospective data in Chapter 3 and 4. The collected data were not designed for the purpose of my studies; therefore, data may be incomplete, inaccurate, or inconsistent in how they were recorded. I primarily used this approach for pragmatic reasons; performing prospective studies of such a large scale would have required a significant amount of time and funds beyond the scope of a three-year DPhil.

The study period in the quantitative chapters ended prior to the onset of the COVID-19 pandemic. While this was specifically chosen to exclude pandemic effects, there are uncertainties related to its external validity and its relevance to the current context of testing for children in primary care. Testing rates may not have recovered to pre-pandemic levels, and therefore estimates and predictions about potential rates of testing in the future may be flawed. Foley and colleagues performed a longitudinal trends analysis which showed that primary care consultations fell by 50% or more in children aged 1 to 14 from March to June 2020 compared to the preceding five years; rates across all children and young people recovered but remained at around 18% below pre-pandemic levels by October 2020.[16] However, a longer-term study that also used CPRD Aurum data analysed consultation rates by age and deprivation, found that consultation rates in 2021-2022 had, in fact, recovered to higher than pre-pandemic levels for all patients except those aged 0-4 and 75 years or older.[17] Increases in overall consultation rates were driven by an increase in remote consultations in 2021-2022.[17] It is unclear whether the mode of consultation had any influence on the rates of testing. During the interviews, some GPs revealed that their testing behaviours as a practice had changed since the pandemic, and that since then, their practice had

dramatically reduced the number of tests they performed. Future analyses can explore the impact of the pandemic by performing an interrupted time series analysis and should also explore the influence of consultation mode (face-to-face, remote) on testing rates.

This thesis is limited to general practitioners; however, other clinical personnel in primary care also request tests. A recently published study found that 40% of test requests in primary care originate from clinicians who are not GPs or GP registrars, including requests from secondary care (5%), nurse practitioners (5%), nurses (4%), health care assistants (2%), pharmacists (1%), physician associates (1%), paramedics (0.04%).[18] This was previously identified as a limitation within Chapter 5.6; however, it also pertains to other chapters. GPs were also solely involved in the process of confirming the diagnostic guidelines that were applicable to primary care in Chapter 2 as well as the selecting asthma tests that were appropriate for inclusion in Chapter 7. Despite this, GPs within the qualitative study stated that due to higher clinical uncertainty in paediatric patients amongst their colleagues, GPs who had a special interest in paediatrics or previous paediatric experience were frequently referred children who needed tests from their fellow primary care colleagues. Therefore, it is unlikely that limiting the studies to GPs impacted the results significantly, given that most test requests for children likely originated from GPs.

## 8.5 Implications for research

**Section Summary**

- Future research should aim to:
  - evaluate tests for their impact on patient outcomes, population outcomes, cost, and resource use prior to their implementation in practice.
  - assess the appropriateness of testing by developing testing quality indicators or 'appropriate use criteria' to benchmark the quality and appropriateness of tests. A proposed research agenda is introduced.
  - understand the implications of test variation on the number of diagnoses, testing cascades, treatment, flow on referrals and hospital activity, resource utilisation, and cost.

### 8.5.1 Tests should be fully evaluated prior to implementation

New tests and technologies are incorporated into clinical practice without a thorough evaluation of the limitations, benefits, and pitfalls, both in paediatric and adult care settings. Once they become a part of routine care, these tests become difficult to de-implement. Tests therefore need to be evaluated to ensure they are fit for purpose in the intended population and improve patient outcomes before being adopted into clinical practice. Test evaluations should not be focused on test accuracy alone. Numerous models and frameworks for the evaluation of medical tests have been proposed since 1978; one review identified 19 different models that had been published prior to 2009[19], and there have been several others published since. [20–22] The different frameworks encompass the following aspects: technical accuracy, diagnostic accuracy; diagnostic thinking (how the test alters diagnostic decision-making); therapeutic efficacy, patient outcome, and societal outcome.[19] Ferrante Di Ruffano and colleagues proposed that the evaluation of tests should occur in the context of its place in the test-treatment pathway, considering the delivery of the test,

the results, its influence on diagnostic decision-making, treatment decision-making and treatment implementation.[20] The different components of this pathway include:

1) Test delivery
    a. Timing of test
    b. Feasibility – acceptability, clinical contraindications, technical failure rates
    c. Test process – procedural harms or benefits, placebo effect (if patients feel more assured about the test's thoroughness)
2) Test result
    a. Interpretability
    b. Accuracy (does the test correctly identify the target condition in the population?)
    c. Timing of results (does it require specialist interpretation?)
3) Diagnostic decision
    a. Timing of diagnosis
    b. Diagnostic yield (is there more weight applied to the findings of one test compared with another?)
    c. Diagnostic confidence (doctors' and patients' confidence in diagnosis)
4) Treatment decision
    a. Therapeutic yield (does it lead to appropriate management?)
    b. Therapeutic confidence (doctors' and patients' confidence in treatment choice)
5) Treatment implementation
    a. Time to treatment
    b. Efficacy of treatment (does it lead to improvement in patient outcomes?)
    c. Does it impact treatment adherence?

The authors provided a helpful case example of a randomised controlled trial evaluating positron emission tomography (PET) imaging and its role in the reducing the rate of thoracotomies in patients with Stage I and II non-small-cell lung cancer compared with conventional staging methods.[23] The trial reported no difference in the number of thoracotomies between both treatment groups. The reason for this was not related to test accuracy but related to treatment effects; the test did not change management decisions due to surgeon's policies of operating on patients with completely resectable stage IIIA disease and there were strong preferences for the existing management (to perform thoracotomies).[23] Other models go beyond assessing the patient outcome from a clinical standpoint alone, incorporating the emotional effects, social effects, cognitive effects, and behavioural impacts of testing.[22] Economic evaluations are also vital. Basatemur and colleagues found that the cost of supplementation outweighs the cost of testing for vitamin D, arguing that it is not a cost-effective strategy.[24] All these outcomes should be incorporated into studies evaluating the performance of diagnostics and included in guidance to enable clinicians and patients to make better-informed decisions relating to diagnostic tests.

### 8.5.2 Test appropriateness

My thesis was designed to understand the landscape of testing, but it does not directly assess whether variation was warranted (appropriate) or unwarranted (inappropriate). Instead, the presented chapters identified that there is a dearth of evidence on paediatric diagnostic tests in primary care and that consequently, it is difficult to ascertain the 'appropriate' level of testing. Without first defining what constitutes appropriate testing, it is not possible to identify underuse or overuse of tests with certainty. While there are indirect measures of inappropriate use, including: 1) comparing testing indications with guideline recommendations and 2) identifying the proportion of tests that return abnormal results, these carry limitations and fail to account for the complexities and uncertainties associated with testing and their

implications. This is a nascent area of research; therefore, high quality, patient-centred research is needed to improve the quality of testing for children in primary care that addresses some of these complexities and uncertainties.

One of the main challenges of identifying inappropriate testing is the absence of benchmarks or quality indicators for appropriate testing, sometimes called “appropriate use criteria”.[18] Similar quality indicators already exist in the context of prescribing, such as outpatient antimicrobial prescribing indicators that aim reduce unnecessary antibiotic use.[25,26] Quality indicators for paediatric testing in primary care need to be developed and applied to optimise the delivery of diagnostic services and reduce unwarranted variation. Similar initiatives in the hospital setting, such as the Choosing Wisely campaign found that in Canada, the implementation of recommendations had a moderate effect on reducing low value care, for example chest X-Rays for children with asthma and bronchiolitis decreased by 21 and 32% respectively.[27] I present a proposed research agenda that outlines the steps to be taken to develop these testing quality indicators.

### 8.5.3 A proposed approach to develop testing indicators

The research agenda include the following steps which are detailed below.

1. Identifying the tests and/or conditions for further research.
2. Evidence review and identifying research gaps
3. Prioritising research questions
4. Addressing the research gaps
5. Development of appropriate use test criteria or test quality indicators
6. Piloting phase: test indicators
7. Using test quality indicators to change practice

1. *Identifying the tests and/or conditions for further research.*

The findings presented in this thesis identified tests that require further research and potential intervention. The tests that were identified as high-test rate – high practice variability (**Figure 4.6**) - vitamin D, iron studies, vitamin $B_{12}$, folate, and coeliac testing - could provide a starting point. A more comprehensive strategy would be to present the distilled summary of this thesis (**Table 8.1**) to stakeholders which include clinicians, patients, caregivers, local health service researchers, and managers to discuss and identify which tests should be prioritised. They could alternatively select the conditions or presentations that encompass these tests, for example, asthma, or fatigue (which may include testing for haematinic or vitamin deficiencies), which would offer more clinically oriented approach.

Decisions on which tests or conditions to prioritise should be based on the potential health, social, emotional, and economic impacts of the test or condition. This discussion should also consider whether testing variation is likely to reflect local systemic processes (for example, restrictions on whether GPs can request MRIs) and whether variation is likely to have persisted post-pandemic. Prioritisation could take place using a nominal group technique method to ensure the process is systematic and centred around partnership with patients and caregivers. The following process of selecting and refining the paediatric diagnostics research priorities are adapted from James Lind Alliance Priority Setting Partnership, which has been used to identify Top 10 lists of research priorities in many areas of healthcare.[28] Gill et al, for example, used this method to develop a list of unanswered research questions in paediatric hospital medicine from the perspectives of children and young people, parents and caregivers, and health care professionals.[29]

**Table 8.1 Tests ranked by practice variability, temporal increase (since 2007) and highest use in 2019**

| Rank | Practice variation: 2019 | Temporal increase: 2007 to 2019 | Most frequently used tests: 2019 |
|---|---|---|---|
| 1 | Fractional exhaled nitric oxide | Faecal calprotectin | Full blood count |
| 2 | Hearing test | Fractional exhaled nitric oxide | Urinalysis |
| 3 | Vitamin D | Vitamin D | Urine microscopy, culture, and sensitives |
| 4 | *Helicobacter* test | Folate | Peak flow |
| 5 | Monospot test | Vitamin $B_{12}$ | Urea and electrolytes |
| 6 | Skin / wound microscopy, culture and/or sensitivities | Coeliac test | Liver function test |
| 7 | Enteric virus screen | *Helicobacter* test | Iron studies |
| 8 | MRI brain | Iron studies | Thyroid function test |
| 9 | Immunoglobulins | HbA1c | Bone profile |
| 10 | Stool ova, cysts, parasites | Immunoglobulins | C reactive protein |
| 11 | Folate | C reactive protein | Vitamin $B_{12}$ |
| 12 | Spirometry | MRI brain | Folate |
| 13 | Coeliac test | Bone profile | Glucose |
| 14 | Iron studies | Allergen-specific IgE | HbA1c |
| 15 | Vitamin $B_{12}$ | ECG | Erythrocyte sedimentation rate |
| 16 | Chest X-Ray | Liver function test | Vitamin D |
| 17 | Allergen specific IgE | Thyroid function test | Coeliac test |
| 18 | CT head | Urea and electrolytes | Stool microscopy, culture, and sensitivities |
| 19 | Glucose | Abdominal ultrasound | Electrocardiogram |
| 20 | Bond profile | Full blood count | Wound/Skin MCS |
| 21 | Renal ultrasound | Skin / wound microscopy, culture and/or sensitivities | Spirometry |
| 22 | Abdominal ultrasound | Erythrocyte sedimentation rate | Hearing test |
| 23 | Peak flow | Urinalysis | Immunoglobulins (IgG, IgA, IgM) |
| 24 | HbA1c | Enteric virus screen | Enteric virus screen |
| 25 | Full blood count | Glucose | Chest X-ray |

| | | | |
|---|---|---|---|
| 26 | ECG | Stool ova, cysts, parasites | Stool ova, cysts, parasites |
| 27 | Thyroid function test | Stool microscopy, culture, and sensitivities | Allergen Specific IgE |
| 28 | Erythrocyte sedimentation rate | Chest X-ray | *Helicobacter* test |
| 29 | Liver function test | Urine microscopy, culture, and sensitives | Calprotectin |
| 30 | Urinalysis | Hearing test | Monospot |
| 31 | Faecal calprotectin | Spirometry | Abdominal ultrasound |
| 32 | Urea and electrolytes | CT head | Renal ultrasound |
| 33 | Stool microscopy, culture, and sensitivities | Peak flow | MRI brain |
| 34 | Urine microscopy, culture, and sensitives | Renal ultrasound | Fractional exhaled nitric oxide |
| 35 | C reactive protein | Monospot | CT head |

1. *Evidence review to identify research gaps.*

After the tests have been selected and prioritised, it is pertinent to examine and summarise the existing evidence for the test's use and identify evidence gaps to direct future research. Surveys can be sent to clinicians, researchers, patients and caregivers and policymakers to submit research questions that can address research gaps for the chosen tests.

Examples of questions that clinicians may include:

- How does the test perform in children from ethnically diverse backgrounds?
- What are the risk factors, clinical features or lab tests that indicate the child is likely to have the target condition?
- How does the test perform in primary care in improving patients' physical, emotional, or social outcomes?
- Of children with the specified presentation in primary care, what proportion of the requested tests are likely to return an abnormal result for an underlying condition that would benefit from intervention?

2. *Prioritising research questions*

Once the list of research questions has been collated and screened (to ensure they have not been already been addressed), a second survey could be sent to the stakeholders to score and prioritise the research questions based on importance, relevance, and potential impact, similar to Gill et al.[29] These scores can then be used to generate an aggregate score for each research question and a list of the top 10 questions to put forward into the next phase of research.

3. *Addressing the research gaps*

Once the prioritised list of questions is finalised, the next phase is to design and plan the appropriate studies to address the research questions. Studies should be co-designed with clinicians, patients, parents, and caregivers so that the outcomes are patient centered. The appropriate study designs to address the research questions include randomised controlled trials, feasibility and acceptability studies, analyses using large electronic health record datasets, prospective cohort studies, and qualitative studies utilising ethnography or interviews.

4. *Development of appropriate use test criteria or test quality indicators*

When evidence has been generated to better understand the best use of the relevant test or tests (for a condition), quality indicators for their best use should be drafted. Where evidence is equivocal, this should be clearly stated. The process of developing NICE guideline recommendations can be applied to the process of developing test quality indicators.[30] Indicators should incorporate the best available evidence, emphasise the quality of evidence used, the benefits and harms of testing, considerations for its implementation in practice, size of effect and potential impact on public health. Other considerations include their applicability and implementation in

more disadvantaged groups, and as well as addressing insufficient evidence (to remove the indication entirely or explain the limitations of evidence).

The wording of indicators is essential to their successful implementation. They should be specific and use patient-centred, clear, and concise language that can be interpreted by clinicians, policy makers and health service managers alike. This process will require multiple rounds of revision and discussion between all stakeholders to ensure the indicators appropriately reflect the evidence, are valid, feasible, and achieves consensus among the stakeholder group.

5. *Piloting phase: test indicators*

Following the development of evidence-based testing indicators, pilot testing needs to occur to evaluate their external validity and whether they improve the quality of diagnoses and outcomes for children in primary care. This could take place within several local integrated care systems to assess its applicability in different localities. Pilot testing of these indicators in these areas should ask GPs and community-based paediatricians to assess whether these indicators are valid, feasible, and important. Pilot testing will further refine these indicators.

6. *Using test quality indicators to change practice*

After the final evidence-based test indicators or criteria have been developed and piloted, researchers, practices, and health services can use these to identify areas of low-value care as well as instances where more testing is needed to improve patient and population outcomes. A systematic review by Cliff et al identified an increased likelihood of success in reducing low value care when health systems adopted guidance to reduce low-value care and targeted multicomponent interventions towards clinicians.[31] Development of indicators alone will not improve the quality of testing for

children in primary care, but the success of testing interventions can be measured using these indicators. Follow up studies will be needed to assess the impact of these interventions on testing practices and measure their impact on individual patient outcomes as well as downstream activity including further tests, treatments, referrals, and the overall cost.

### 8.5.4 Implications of variation in paediatric diagnostic test use

My thesis did not directly address the consequences of testing variation. These implications include: the number of diagnoses, cascades of further testing, treatment, flow on referrals and hospital activity, resource utilisation, cost, and the impact on children and their families. A US-based, multicentre, cross-sectional study using the Pediatric Health Information System quantified and analysed variation in stool testing rates in children diagnosed with acute gastrointestinal infections.[32] Rates of stool PCR testing ranged from 0 to 65% for children in hospital, but there were no statistically significant differences in length of stay, costs, or revisit rates across the different institutions. [32] Another retrospective multicentre study analysed variation in laboratory testing for hyperbilirubinaemia in hospitalised infants; testing bilirubin levels was found to be associated with longer hospital stays, suggesting that additional testing unnecessarily increased resource use.[33] Leyenaar et al conducted a retrospective cohort study of children hospitalised with uncomplicated pneumonia across hospitals in the US that observed similar rates of chest X-rays, blood cultures and viral respiratory testing among admitted children but variability in tests for acute-phase reactants (erythrocyte sedimentation rate, C-reactive protein). The authors reported that performing blood tests or tests for respiratory viruses was significantly associated with increased length of stays.[34]

The other important consideration that my thesis did not examine is the cost implications of test use. A previous study provided an approximation of the average cost of a laboratory test at £6, an imaging test at £29 and a miscellaneous test at £53.[3] Applying these costs to the findings from Chapter 4, the estimated expenditure on diagnostic tests for children in primary care in 2019 was £184 million (£27 million for laboratory tests, £15 million for imaging and £142 million for miscellaneous tests). These are rough estimates, and future economic analyses of the costs associated with tests in children should utilise the unit costings from the different NHS laboratory trusts.

A sustainability-focused assessment, called a "triple-bottom line" assessment, to examine the (1) financial, (2) health, and (3) environmental impact of increased test use can be conducted to understand test implications. This triple-bottom line assessment was conducted for vitamin D testing in Australia (across all ages), and the authors estimated that of the 4.4 million vitamin D tests performed in 2020 (based on Medicare data), 77% provided no health benefit.[35] The total cost of unnecessary vitamin D testing was $87 million AUD, and the carbon footprint ranged between 29,000- 42,000 kg of $CO_2$ emissions.[35] This case example broadly illustrates the potential environmental, health, and financial benefits of reducing low-value care.

## 8.6 Implications for patients, parents, and caregivers

**Section Summary**

- This thesis highlights the need for improved public understanding about the benefits and harms of tests.
- Patients should be encouraged to participate in these discussions with their clinicians, including potential false positives and negatives.
- Tests should be accompanied by information provided in leaflets or test safety profiles that include information on the benefits and harms for patients.

Research has shown that patients tend to overestimate the benefits of tests and underestimate the harms.[36] Previous work has described the uncertainty that patients experience with respect to diagnostic tests, particularly how they make sense of the test results in the context of their current symptoms, and how test results impact on their management. [37] Difficulties in interpreting tests may be due to differences in health literacy, numeracy, ambiguities in the results, and conflicting opinions.[37] This sheds light on the need for improved public understanding about the benefits and harms of tests, the latter of which include the likelihoods of false positive or negative results or the need for further testing. To achieve this, all tests could be accompanied by a test safety profile or information leaflet for both clinicians and patients that details the potential benefits and risks as well as special considerations. Clinicians should engage in discussions with children and their parents or caregivers about reasons for performing the test(s) as well as the potential risks of testing, as these can lead to improved understanding of diagnostic decision-making and uncertainty.[38]

## 8.7 Implications for practice

**Section Summary**

- Inappropriate test use can be addressed using electronic clinical decision support tools, reminders, and audit/feedback strategies.
- GPs may benefit from reviewing their own testing practices in a similar way to prescribing.
- Clinicians should be aware of the variation and uncertainties of test results and reference ranges while work is ongoing to improve their representativeness and accuracy.

### 8.7.1 Identifying and mitigating inappropriate testing in clinical practice

My thesis highlighted some of the challenges of researching variation in testing. Paediatric diagnostic guidelines are flawed and vary in their quality of reporting, and

test recommendations are underpinned by poor quality evidence. Some tests, such as vitamin D, do, however, have indications for their use, including suspected rickets or osteomalacia, abnormal serum bone profile or X-Ray findings, or known osteoporosis.[39] From a practice standpoint, some areas have implemented electronic clinical decision support tools for clinicians to ensure that tests are used appropriately and not over-used or under-used. A scoping review that explored the impact of electronic clinical decision support tools on workload and workflow implications for health professionals in general practice found equivocal evidence on whether they increased or decreased the duration of consultations and workload.[40] A systematic review studied a variety of de-implementation strategies and interventions that aimed to reduce the overuse of tests in primary care.[41] Strategies that resulted in greater reductions in the use of low-value tests included using reminders (22% vs 14%), audit/feedback (22% vs 13%) and patient-targeted interventions (51% vs 17%).[41] Another systematic review[42] that assessed behavioural interventions to reduce carbon emissions in health care activity identified two studies[43,44] which successfully implemented behaviour change techniques to reduce unnecessary test ordering. The behavioural techniques included: goal setting (e.g., setting targets to reduce testing), feedback on the outcome of behaviour (e.g., report on individual and/or practice testing rates), social support (e.g., encouragement or counselling on testing practices), instructions on how to perform the behaviour (e.g., skills training), salience of consequences (e.g., using a graphics designer to highlight consequences in a way that elicits an emotional response), providing information about social and environmental consequences, prompts/cues, credible source (e.g., having the policy approved by the president of the college), incentives, and adding objects to the environment (e.g., educational posters).[42]

In clinical practice, GPs would benefit from the opportunity to review and analyse their testing patterns, in a similar way that is already occurring for prescribing. OpenPrescribing.net is such a tool that was developed by the Bennett Institute for Applied Data Science that allows clinicians, researchers, and members of the public to assess the performance of practices and NHS organisations against their peers as well as standard prescribing measures.[45] It also analyses prescribing trends and expenditure over time and allows users to run their own analyses to get prescribing data in near-real time. OpenPathology.net was established to achieve similar aims in the realm of diagnostic tests, however, further funding is needed to develop the tool to achieve these objectives.[46] In this thesis, I found that paediatric testing rates vary by age and gender, therefore, testing analyses using these platforms should also be stratified by age and gender to gain a more nuanced understanding of testing trends.

### 8.7.2 Addressing variation in results and reference ranges

Practitioners should be aware of the uncertainties relating to results, reference ranges, and understand that there is variation in the reference intervals used in laboratories.[47] Future research should aim to address these evidence gaps to improve the interpretation, standards, and quality of paediatric diagnostic testing.

#### *Variation in test results*

Every laboratory test result carries a level of uncertainty which originates from several sources. Firstly, every individual has a degree of within-person variability, or biological variation, resulting from a combination of human physiology, environment, diet, activity, stress, mood, weather, and climate factors.[48] The degree of variation depends on human physiological factors, and certain measurements are subject to more variation than others. Serum cholesterol, for example, is subject to variation that is 10-fold higher than sodium measurements, and 25-hydroxy-vitamin D is subject to 24-fold higher variation than serum sodium measurements.[48] The second source of uncertainty

comes from pre-analytic variation, which encompasses issues related to specimen collection, handling, shipping, and storage. These can lead to random error or systematic bias.[48] Third, analytic variation relates to the error within the laboratory, specifically operator or equipment-related issues, used for analysing the specimen. The analytical variation can be as high as 11 to 20% for a single measurement of vitamin $B_{12}$, alkaline phosphatase (a liver enzyme requested as part of a liver function test), or rheumatoid factor. When compounded with the potential biologic variation, measurements can be subject to variation as high as >41-50% for serum total bilirubin or triglycerides, meaning that for these two values, two serial measurements need to vary by >50% for there to be "true" statistical difference with 95% confidence. This value is called the reference change value.[48,49] Biologic variation is the largest contributor to the reference change value. There are also post-analytical errors that can occur from incorrectly transcribing the results. The uncertainty related to each result needs to be considered in the context of the reference ranges, which also vary.

*Variation in reference ranges*

Reference ranges are statistical measures for detecting outliers in a healthy, normal population. Reference ranges are defined by the inclusion of 95% of the values of a healthy population, so in 5% of cases, people will have a value that is outside (either above or below) the reference range. Each person's set point is different, which is known as "between-subject" variation. If, as is commonly the case, a panel is ordered which contains several different measurements, the chance of one value falling outside of the reference range is: $1 - 0.95^k$, where k equals the number of measurements, assuming the measurements are independent of each other. Therefore, if eight different measurements are taken, the probability of a result outside of the reference range is: $1 - 0.95^8$ = 0.34 or 34%. However, it is unlikely this would indicate a clinically significant abnormality. In the ideal scenario, reference ranges and disease thresholds are derived from threshold of clinical significance, the point at which intervention would

improve outcomes. However, this would require prospective studies to understand the short-term and long-term implications of different laboratory measurements.

The other consideration is that different laboratories use different reference ranges to interpret results. Some thresholds are defined by consensus guidelines, for example, NICE guidelines use a vitamin D threshold as greater than 50nmol/L to suggest sufficiency. In other cases, there are age- and sex- specific thresholds, such as the cut-offs used to diagnose anaemia. The thresholds to diagnose anaemia were derived from a predominantly white adult population in 1968, with amendments made in 2000 for children and young people based on data from the US National Health and Nutrition Examination Survey 2.[50] However, reference ranges for each test component differs according to gender, age, and ethnicity, proving this to be a challenge when interpreting results. The partitioned age-based reference range are also problematic; a change in age by one day could abruptly change a result from normal to abnormal (or vice versa), which does not reflect the continuous variation of physiology with age. Laboratories are required to generate their own reference intervals based on their local populations using their analysers, but this does not always occur. The Pathology Harmony initiative found that three major laboratories in the city of Birmingham had identical biochemistry analysers but despite using the same methods for the same population group, the laboratories used different reference ranges. 37% of the 153 reference ranges were sourced from the manufacturer, 14% were based on the literature, 18% were based on in-house measurements and 35% of reference ranges had an unknown source.[51] Analyser manufacturers may even derive their reference ranges from measurements in the population or from the literature, but these may differ substantially, and it is not known which is more accurate unless they are validated in the local healthy population. The result is that a value that is flagged as normal in one area may be abnormal in another. When laboratories change equipment, (which was

the case in the Oxfordshire laboratory [Chapter 3]) they generally adopt the manufacturer's reference ranges. This decision of which reference range to apply will directly impact rates of abnormal results being returned. The Pathology Harmony Group has attempted to standardise the reference ranges for several of analytes[51], however, there remains a vast majority of reference ranges which are not calibrated and standardised, making formal comparisons between different areas and practices a challenge. The paediatric equivalent of the Pathology Harmony Group, called the Canadian Laboratory Initiative on Pediatric Reference Intervals (CALIPER), developed age and sex-specific continuous paediatric reference intervals for more than 200 laboratory tests based on a dataset containing 13,000 children.[52–54] These reference ranges could be applied to future analyses examining test results. Clinicians should be aware of the variability of results, as well as local laboratory reference intervals and their limitations when interpreting diagnostic test results in current practice.[52]

## 8.8 Implications for policy

**Section Summary**

- Tests should be subject to regulation to ensure they are fit for purpose in the intended testing population.
- Gaps in the current GP training program for providing paediatric care need to be addressed to improve GP confidence and certainty when caring for children.
- Integrated care systems may prove to be an effective model of child health delivery in the community.
- Paediatric guidance needs to be high quality, rooted in evidence, and accessible to all GPs.

### 8.8.1 Tests need increased regulation

The Medicines and Healthcare products Regulatory Agency (MHRA) in the UK are responsible for ensuring that tests are subject to strict regulatory standards.[55]

Independent health technology assessments of tests should be undertaken to establish the clinical and cost-effectiveness of tests in comparison with existing standards. Test safety should consider not only physical harms, but the potential number of false positive and false negative results. Post-marketing surveillance is also needed to evaluate both the intended and unintended consequences of testing.[55]

If tests have not undergone rigorous evaluation, are proven to be ineffective, or cause net harm for the population, then these should be de-implemented. Current regulatory gaps in testing have resulted in the rise of direct-to-consumer testing that have not undergone rigorous evaluations.[56] This gap needs to be urgently addressed, as these ineffective tests can lead to undue anxiety and worry for patients and cost the NHS.[56]

### 8.8.2 Gaps in general practice training

The findings of this thesis support growing calls for increased paediatric exposure within general practice specialty training. There appeared to be substantial variation in GP confidence relating to paediatric presentations which depended on the length of experience they had working as a GP, whether they had inpatient paediatric rotations during their specialist training program, whether they were parents themselves or have worked abroad in resource-poor settings which helped develop confidence in identifying sick children. One way to standardise the quality of paediatric care within primary care would be to increase exposure to paediatrics within GP training. Sweden, for example, has a primary care model where all family medicine specialists have had a minimum of five years of specialist training with compulsory rotations in internal medicine, psychiatry, paediatrics, gynaecology, and family medicine. Applying a similar model in the UK or even including paediatrics as a core term within GP training is likely to enhance GP confidence in managing paediatric presentation.

### 8.8.3 Moving to enhanced paediatric primary care delivery through integrated care models

An alternative model to deliver better quality paediatric primary care is to implement integrated models of health care, such as GP child health hubs for children and young people.[57–59] The Connecting Care for Child Health model recently emerged in the UK.[57,59] This consists of a group of local GP practices which work with a local hospital paediatrician using one practice as the centre of care to meet the needs of the local population. The GP calls the paediatrician for advice and guidance, which results in one of four outcomes: (1) the child being managed directly by the GP, (2) referral to the specialist outreach hub clinic, (3) discussion at a multidisciplinary team meeting and then managed by the GP, or (4) a referral to a specialty outpatient clinic.[57] The core features of this model aim to develop the paediatric capabilities of GPs by promoting multidisciplinary team discussions and education, improving patient access to specialists, and enhancing parent and child self-confidence in how to manage their health and access health services appropriately. A service evaluation of this model found that these hubs reduced secondary care use, including 39% of new patient hospital appointments, 42% of appointments being shifted from hospital to GP practices and a 20% reduction in sub-specialty referrals, admissions, and emergency department attendees. They also received high patient satisfaction ratings, and professionals fed back that it provided helpful learning and development. [60] Other models of integrated care, include the Cluster Clinic, which involves a paediatrician and GP with an established special interest or experience in child health working together to provide advice and triage referrals in a cluster of GP practices, and the Wessex model, which employs the multidisciplinary team format to review patients with the paediatrician and GP and other specialists including allied health if required.[59] All these models of integrated care may provide better access to specialty care, reduce healthcare costs, improve patient quality of life[58] and reduce unnecessary tests and referrals. A pragmatic randomised controlled trial compared integrated care systems

with usual GP care over three years. While integrated care systems did not result in a difference in the primary outcomes including non-elective hospital admissions or paediatric quality of life, there was an observed improvement in care quality for children with asthma[61], suggesting that large-scale health services research is needed to study this effect long term. To be rolled out successfully, it needs to be co-designed and co-produced by both primary and secondary care, there needs to be buy-in from all stakeholders including funders, as well as quality improvement cycles to continuously assess and evaluate the performance of these models over a long period of time.

### 8.8.4 Improved quality and access to paediatric guidance

To deliver standardised care to children in primary care, the quality of paediatric guidance needs to be improved and GPs need to have improved accessibility and awareness of paediatric guidelines. In Chapters 6 and 7, I proposed a central paediatric repository for clinical practice guidelines. These guidelines should reflect the needs and resource constraints of local health systems. Guideline developers should formulate recommendations based on a systematic review of the existing evidence and clearly state where the recommendations are based on expert opinion.

## 8.9 Conclusions

Increasing financial and service pressures on health systems and worsening inequities in child health outcomes highlight the urgent need to measure the quality of care that is delivered to children. Most children encounter the health system through their GP; however, GPs face a high degree of diagnostic uncertainty and can choose from a variety of diagnostic strategies to assist them, which includes performing tests. To date, no research has examined diagnostic test use for children in primary care. My thesis utilised both quantitative and qualitative methods to examine the use of tests in

children and explored some of factors that contribute to variation, such the varying quality of clinical guidance and evidence for paediatric diagnostic tests.

My thesis builds on existing knowledge of paediatric diagnostics by describing trends in test use and highlighting the tests that are subject to high variation in use. These findings should be investigated further using a structured approach that aims to prioritise tests and/or conditions, identify evidence gaps, and generate evidence in partnership with clinicians, patients, parents and caregivers and health service stakeholders to ultimately develop testing quality indicators. The drivers of variation in paediatric test use that were identified in this thesis highlight the urgent need for clinician-level and policy-level intervention to improve testing practices in general practice. Addressing unwarranted variation in paediatric test use can reduce health care waste, ensure resource allocation is equitable, and improve the quality and value of health care delivered to children in primary care.

## 8.10 References


1. NHS RightCare. NHS Atlas of Variation in Healthcare for Children and Young People; 2012.
2. Public Health England. The 2nd Atlas of Variation in NHS Diagnostic Services in England. NHS RightCare. 2017;(January).
3. O'Sullivan JW, Stevens S, Hobbs FDR, et al. Temporal trends in use of tests in UK primary care, 2000-15: Retrospective analysis of 250 million tests. The BMJ. 2018;363. doi:10.1136/bmj.k4666.
4. O'Sullivan JW, Stevens S, Oke J, et al. Practice variation in the use of tests in UK primary care: A retrospective analysis of 16 million tests performed over 3.3 million patient years in 2015/16. BMC Med. 2018;16(1). doi:10.1186/s12916-018-1217-1.
5. Lee LK, Monuteaux MC, Hudgins JD, et al. Variation in the evaluation of testicular conditions across United States pediatric emergency departments. American Journal of Emergency Medicine. 2018;36(2):208-212. doi:10.1016/j.ajem.2017.07.078.
6. Rogers AJ, Kuppermann N, Anders J, et al. Practice Variation in the Evaluation and Disposition of Febrile Infants ≤60 Days of Age. Journal of Emergency Medicine. 2019;56(6):583-591. doi:10.1016/j.jemermed.2019.03.003.
7. Florin TA, French B, Zorc JJ, Alpern ER, Shah SS. Variation in emergency department diagnostic testing and disposition outcomes in pneumonia. Pediatrics. 2013;132(2). doi:10.1542/peds.2013-0179.
8. Patrick SW, Schumacher RE, Davis MM. Variation in lumbar punctures for early onset neonatal sepsis: a nationally representative serial cross-sectional analysis, 2003-2009. BMC Pediatr. 2012;12. doi:10.1186/1471-2431-12-134.
9. Stanley RM, Teach SJ, Mann NC, et al. Variation in Ancillary Testing among Pediatric Asthma Patients Seen in Emergency Departments. Academic Emergency Medicine. 2007;14(6). doi:10.1197/j.aem.2007.01.016.
10. Hiscock H, Perera P, Mclean K, et al. Variation in paediatric clinical practice: A review of care in inpatient, outpatient and emergency department settings. J Paediatr Child Health. 2016;52(7):691-693. doi:10.1111/jpc.13212.
11. Hiscock H, Perera P, McLean K, Roberts G. Evidence Check: Variation in Paediatric Clinical Practice; 2015.
12. Leal YA, Flores LL, Fuentes-Pananá EM, Cedillo-Rivera R, Torres J. 13C-Urea Breath Test for the Diagnosis of Helicobacter pylori Infection in Children:

A Systematic Review and Meta-Analysis. Helicobacter. 2011;16(4). doi:10.1111/j.1523-5378.2011.00863.x.

13. Yuan C, Adeloye D, Luk TT, et al. The global prevalence of and factors associated with Helicobacter pylori infection in children: a systematic review and meta-analysis. Lancet Child Adolesc Health. 2022;6(3). doi:10.1016/S2352-4642(21)00400-4.
14. Rowland M, Imrie C, Bourke B, Drumm B. How should Helicobacter pylori infected children be managed? In: Gut. Vol 45. ; 1999. doi:10.1136/gut.45.2008.i36.
15. Jones NL, Koletzko S, Goodman K, et al. Joint ESPGHAN/NASPGHAN Guidelines for the Management of Helicobacter pylori in Children and Adolescents (Update 2016). J Pediatr Gastroenterol Nutr. 2017;64(6). doi:10.1097/MPG.0000000000001594.
16. Foley KA, Maile EJ, Bottle A, et al. Impact of COVID-19 on primary care contacts with children and young people in England: longitudinal trends study 2015-2020. Br J Gen Pract. 2022;72(720):e464-e471. doi:10.3399/BJGP.2021.0643.
17. Vestesson EM, De Corte KLA, Crellin E, Ledger J, Bakhai M, Clarke GM. Consultation Rate and Mode by Deprivation in English General Practice from 2018 to 2022: Population-Based Study. JMIR Public Health Surveill. 2023;9. doi:10.2196/44944.
18. Watson J, Burrell A, Duncan P, et al. Exploration of reasons for primary care testing (the Why Test study): a UK-wide audit using the Primary care Academic CollaboraTive. British Journal of General Practice. Published online October 2, 2023. doi:10.3399/BJGP.2023.0191.
19. Lijmer JG, Leeflang M, Bossuyt PMM. Proposals for a phased evaluation of medical tests. Medical Decision Making. 2009;29(5). doi:10.1177/0272989X09336144.
20. Ferrante Di Ruffano L, Hyde CJ, McCaffery KJ, Bossuyt PMM, Deeks JJ. Assessing the value of diagnostic tests: A framework for designing and evaluating trials. BMJ (Online). 2012;344(7847). doi:10.1136/bmj.e686.
21. Thompson M, Weigl B, Fitzpatrick A, Ide N. More Than Just Accuracy: A Novel Method to Incorporate Multiple Test Attributes in Evaluating Diagnostic Tests Including Point of Care Tests. IEEE J Transl Eng Health Med. 2016;4. doi:10.1109/JTEHM.2016.2570222.

22. Bossuyt PMM, McCaffery K. Additional patient outcomes and pathways in evaluations of testing. Medical Decision Making. 2009;29(5). doi:10.1177/0272989X09347013.
23. Viney RC, Boyer MJ, King MT, et al. Randomized controlled trial of the role of positron emission tomography in the management of stage I and II non-small-cell lung cancer. Journal of Clinical Oncology. 2004;22(12). doi:10.1200/JCO.2004.04.126.
24. Basatemur E, Hunter R, Horsfall L, Sutcliffe A, Rait G. Costs of vitamin D testing and prescribing among children in primary care. Eur J Pediatr 2017; 176. DOI:10.1007/s00431-017-2986-9.
25. Adriaenssens N, Coenen S, Tonkin-Crine S, Verheij TJM, Little P, Goossens H. European Surveillance of Antimicrobial Consumption (ESAC): Disease-specific quality indicators for outpatient antibiotic prescribing. BMJ Qual Saf. 2011;20(9). doi:10.1136/bmjqs.2010.049049
26. Williams MR, Greene G, Naik G, Hughes K, Butler CC, Hay AD. Antibiotic prescribing quality for children in primary care: An observational study. British Journal of General Practice. 2018;68(667). doi:10.3399/bjgp18X694409.
27. Reyes MA, Etinger V, Hall M, et al. Impact of the choosing wisely® campaign recommendations for hospitalized children on clinical practice: Trends from 2008 to 2017. J Hosp Med. 2020;15(2). doi:10.12788/jhm.3291.
28. The James Lind Alliance. The James Lind Alliance Priority Setting Partnerships. Published 2022. Accessed December 1, 2023. https://www.jla.nihr.ac.uk/priority-setting-partnerships/.
29. Gill PJ, Bayliss A, Sozer A, et al. Patient, Caregiver, and Clinician Participation in Prioritization of Research Questions in Pediatric Hospital Medicine. JAMA Netw Open. 2022;5(4):e229085-e229085. doi:10.1001/jamanetworkopen.2022.9085.
30. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual. Process and methods guides. 2014;(October).
31. Cliff BQ, Avanceña ALV, Hirth RA, Lee SYD. The Impact of Choosing Wisely Interventions on Low-Value Medical Services: A Systematic Review. Milbank Quarterly. 2021;99(4). doi:10.1111/1468-0009.12531.
32. Markham JL, Hall M, Collins ME, et al. Variation in stool testing for children with acute gastrointestinal infections. J Hosp Med. 2023;18(6):473-482. doi:10.1002/jhm.13087.

33. DePorre AG, Hall M, Puls HT, et al. Variation in care and clinical outcomes among infants hospitalized with hyperbilirubinemia. Hosp Pediatr. 2020;10(10):844-850. doi:10.1542/hpeds.2020-0161.
34. Leyenaar JK, Lagu T, Shieh MS, Pekow PS, Lindenauer PK. Variation in resource utilization for the management of uncomplicated community-acquired pneumonia across community and children's hospitals. Journal of Pediatrics. 2014;165(3). doi:10.1016/j.jpeds.2014.04.062.
35. Breth-Petersen M, Bell K, Pickles K, McGain F, McAlister S, Barratt A. Health, financial and environmental impacts of unnecessary vitamin D testing: A triple bottom line assessment adapted for healthcare. BMJ Open. 2022;12(8). doi:10.1136/bmjopen-2021-056997.
36. Hoffmann TC. Patients' expectations of the benefits and harms of treatments, screening, and tests a systematic review. JAMA Intern Med. 2015;175(2). doi:10.1001/jamainternmed.2014.6016.
37. Meyer AND, Giardina TD, Khawaja L, Singh H. Patient and clinician experiences of uncertainty in the diagnostic process: Current understanding and future directions. Patient Educ Couns. 2021;104(11). doi:10.1016/j.pec.2021.07.028.
38. Kljakovic M. Patients and tests: A study into patient understanding of blood tests ordered by their doctor. Aust Fam Physician. 2012;41(4).
39. National Institute for Health and Care Excellence. Vitamin D deficiency in children. Clinical Knowledge Summaries.
40. Fletcher E, Burns A, Wiering B, et al. Workload and workflow implications associated with the use of electronic clinical decision support tools used by health professionals in general practice: a scoping review. BMC Primary Care. 2023;24(1). doi:10.1186/s12875-023-01973-2.
41. Takada T, Heus P, Doorn S van, et al. Strategies to reduce the use of low-value medical tests in primary care: a systematic review. British Journal of General Practice. 2020;70(701):e858-e865. doi:10.3399/BJGP20X713693
42. Batcup C, Breth-Petersen M, Dakin T, et al. Behavioural change interventions encouraging clinicians to reduce carbon emissions in clinical activity: a systematic review. BMC Health Serv Res. 2023;23(1). doi:10.1186/s12913-023-09370-2.
43. Regan W, Hothi D, Jones K. Sustainable approach to reducing unnecessary combined biochemistry tests on a paediatric cardiology ward. BMJ Open Qual. 2018;7(4). doi:10.1136/bmjoq-2018-000372.

44. McAlister S, Smyth B, Koprivic I, et al. Carbon emissions and hospital pathology stewardship: a retrospective cohort analysis. Intern Med J. 2023;53(4). doi:10.1111/imj.15622.
45. Bennett Institute for Applied Data Science. OpenPrescribing.net. University of Oxford.
46. Bennett Institute for Applied Data Science. OpenPathology.net . University of Oxford.
47. Hoq M, Matthews S, Donath S, Carlin J, Ignjatovic V, Monagle P. Paediatric Reference Intervals: Current Status, Gaps, Challenges and Future Considerations. Clinical Biochemist Reviews. 2020;41(2). doi:10.33176/AACB-19-00036.
48. McCormack JP, Holmes DT. Your results may vary: The imprecision of medical measurements. The BMJ. 2020;368. doi:10.1136/bmj.m149
49. Fraser CG. Reference change values. Clin Chem Lab Med. 2012;50(5). doi:10.1515/CCLM.2011.733.
50. Sachdev HS, Porwal A, Acharya R, et al. Haemoglobin thresholds to define anaemia in a national sample of healthy children and adolescents aged 1–19 years in India: a population-based study. Lancet Glob Health. 2021;9(6). doi:10.1016/S2214-109X(21)00077-2.
51. Berg J. The approach to pathology Harmony in the UK. Clinical Biochemist Reviews. 2012;33(3).
52. Adeli K, Higgins V, Trajcevski K, White-Al Habeeb N. The Canadian laboratory initiative on pediatric reference intervals: A CALIPER white paper. Crit Rev Clin Lab Sci. 2017;57(2). doi:10.1080/10408363.2017.1379945.
53. Karbasy K, Ariadne P, Gaglione S, Nieuwesteeg M, Adeli K. Advances in pediatric reference intervals for biochemical markers: Establishment of the caliper database in healthy children and adolescents. J Med Biochem. 2014;34(1). doi:10.2478/jomb-2014-0063.
54. Asgari S, Higgins V, McCudden C, Adeli K. Continuous reference intervals for 38 biochemical markers in healthy children and adolescents: Comparisons to traditionally partitioned reference intervals. Clin Biochem. 2019;73. doi:10.1016/j.clinbiochem.2019.08.010.
55. Royal Statistical Society. Royal Statistical Society Diagnostic Tests Working Group Report.; 2021. Accessed December 2, 2023. https://rss.org.uk/RSS/media/File-library/Policy/2021/RSS-Diagnostic-tests-report-FINAL.pdf.

56. McCartney M, Watson J, Finney B, Salisbury C. Why blood testing companies need effective regulation. The BMJ. Published online 2022. doi:10.1136/bmj.o2517.
57. The King's Fund. Imperial Child Health General Practice Hubs.; 2014. Accessed November 13, 2023. https://www.kingsfund.org.uk/sites/default/files/media/imperial-child-health-general-practice-hubs-kingsfund-oct14.pdf
58. Wolfe I, Satherley RM, Scotney E, Newham J, Lingam R. Integrated Care Models and Child Health: A Meta-analysis. Pediatrics. 2020;145(1). doi:10.1542/peds.2018-3747.
59. Watson M, Struthers S, Turner SW. Lessons learnt (so far) from establishing models of integrated clinical care for children and young people. Arch Dis Child. Published online November 10, 2023:archdischild-2023-325818. doi:10.1136/archdischild-2023-325818.
60. Montgomery-Taylor S, Watson M, Klaber R. Child Health General Practice Hubs: A service evaluation. Arch Dis Child. 2016;101(4). doi:10.1136/archdischild-2015-308910.
61. Wolfe I, Forman J, Cecil E, et al. Effect of the Children and Young People's Health Partnership model of paediatric integrated care on health service use and child health outcomes: a pragmatic two-arm cluster randomised controlled trial. Lancet Child Adolesc Health. 2023;7(12):830-843. doi:10.1016/S2352-4642(23)00216-X.

# Chapter Appendices

## Chapter 1 Appendices

## Appendix 1.1 List of publications

Publications arising from the DPhil

1. **Thomas ET**, Thomas ST, Perera R, Gill PJ, Moloney S, Heneghan CJ. The quality of paediatric asthma guidelines: evidence underpinning diagnostic test recommendations from a meta-epidemiological study. *Fam Pract*. Published online May 17, 2023. doi:10.1093/fampra/cmad052
2. **Thomas ET**, Thomas ST, Perera R, Gill PJ, Moloney S, Heneghan C. The quality of diagnostic guidelines for children in primary care: A meta-epidemiological study. *J Paediatr Child Health*. 2023;59(9):1053-1060. doi:10.1111/jpc.16454
3. **Thomas ET**, Withrow DR, Shine B, Gill P, Perera R, Heneghan C. Trends in diagnostic tests ordered for children: a retrospective analysis of 1.7 million laboratory test requests in Oxfordshire, UK from 2005 to 2019. *Arch Dis Child*. Published online November 10, 2023:archdischild-2023-325550. doi:10.1136/archdischild-2023-325550
4. **Thomas ET**, Glogowska M, Hayward G, Gill P, Perera R, Heneghan C. General practitioners' perspectives on diagnostic tests for children: a qualitative interview study. *Br J Gen Pract (In Press).*

Other publications

5. Dernie F, France HS, **Thomas ET**, et al. Preventable deaths involving opioids in England and Wales, 2013–2022: a systematic case series of coroners' reports. *J Public Health (Bangkok)*. Published online November 29, 2023. doi:10.1093/pubmed/fdad147
6. Money NM, Wolf ER, Marin JR, Liang D, **Thomas ET**, Ho T. 2023 Update on Pediatric Medical Overuse. *Pediatrics*. 2023;152(4). doi:10.1542/peds.2023-062650
7. Brueckle MS, **Thomas ET**, Seide SE, et al. Amitriptyline's anticholinergic adverse drug reactions–A systematic multiple-indication review and meta-analysis. *PLoS One*. 2023;18(4):e0284168. doi:10.1371/journal.pone.0284168
8. Wu AD, Lindson N, Hartmann-Boyce J, **et al.** Smoking cessation for secondary prevention of cardiovascular disease. *Cochrane Database of Systematic Reviews*. 2022;2022(8). doi:10.1002/14651858.CD014936.pub2
9. Nunan D, Cai T, Gardener AD, **et al**. Physical activity for treatment of irritable bowel syndrome. *Cochrane Database of Systematic Reviews*. 2022;2022(6). doi:10.1002/14651858.CD011497.pub2
10. Fundaun J, **Thomas ET**, Schmid AB, Baskozos G. The power of integrating data: Advancing pain research using meta-analysis. *Pain Rep*. 2022;7(6). doi:10.1097/PR9.0000000000001038
11. Gill PJ, **Thomas ET**, Van den Bruel A. Managing paediatric gastroenteritis in primary care: is there a role for ondansetron? *British Journal of General Practice*. 2021;71(711):440-441. doi:10.3399/bjgp21X717089
12. **Thomas ET**, Richards GC. Diclofenac in adolescents: Diagnosing and treating gastrointestinal adverse drug reactions can prevent future deaths. *BMJ Evid Based Med*. 2021;26(4). doi:10.1136/bmjebm-2020-111640
13. **Thomas ET**, Heneghan C. Catalogue of bias: Selective outcome reporting bias. *BMJ Evid Based Med*. Published online 2022. doi:10.1136/bmjebm-2021-111845

14. Thomas ST, **Thomas ET**, McLean M, Titus TT. Paving the way to achieving the United Nations Sustainable Development Goals for women from Indigenous communities: lessons from Attappady, India. *Discover Sustainability*. 2021;2(1):4. doi:10.1007/s43621-021-00009-y

15.

## Appendix 1.2 National Institute for Health and Care Research (NIHR) School of Primary Care Research Grant Award Letter

**GRANT AWARD LETTER (Individual Research Project)**

NIHR | School for Primary Care Research

School for Primary Care Research
School of Primary Care, Community and Social Care
School of Medicine
David Weatherall Building
Keele University
Staffordshire
ST5 5BG

Web: www.spcr.nihr.ac.uk

Elizabeth Thomas
Radcliffe Primary Care Building
Radcliffe Observatory Quarter
Woodstock Road
Oxford
OX2 6GG

28 June 2022

Dear Elizabeth,

**RE: Member Application to the National Institute for Health Research School for Primary Care Research (NIHR SPCR)**
**Grant Reference Number**: 624
**Research Project Title:** Diagnostic Testing for Children in Primary Care

The National Institute for Health and Care Research ("NIHR") has, agreed to provide funds, reference: NIHR SPCR-2021-2026 (the "Grant") for the National Institute for Health and Care Research School for Primary Care Research ("NIHR SPCR") to be used to increase the evidence base for primary care practice in England by commissioning and conducting world class research. The Grant is to be administered by Keele under the terms of an agreement between Keele and Secretary of State for Health (the "Authority") dated with effect from April 2021 (the "Main contract").

I am delighted to inform you that your recent Project Application to the SPCR Board for funding for the Research Project named above has been successful. The NIHR SPCR will provide funding of no more than £23,085.00 over 18 months for the project described in your Project Application.

By signing this Agreement, you agree to carry out the Research Project in accordance with:

(a) the work plan submitted as part of the Project Application, in the form in which it was approved by the SPCR Board; and

(b) the terms and conditions of a Partnership Agreement to which your institution is a party dated with effect from April 1 2021; and

(c) the following additional conditions specified by the SPCR Board:

(i) Notification to the SPCR of any third party rights that exist in relation to Background IP;

(ii) Notification to the SPCR Board if Foreground IP will not vest with you.

You will ensure that any publication, including patent applications, or resulting from research carried out under this Award shall acknowledge the NIHR's financial support and shall carry a disclaimer as per web link: https://www.nihr.ac.uk/documents/nihr-research-outputs-and-publications-guidance/12250#Types_of_research_output

Details of the financial award are set forth in the document headed 'Financial Arrangements' attached to this letter.

Costs may vary from the award breakdown stipulated. Transfer of funds between sub-headings is permitted without prior permission. However, funds cannot be transferred between Non-Staff and Staff without prior written approval from Keele. Total costs must not exceed the total award amount stipulated for this project. While approval does not need to be sought from Keele for transfer of funds, Keele reserves the right to query any expenditure outlined in the financial report which has not been incurred in line with the Award Breakdown.

You shall invoice Keele quarterly in arrears on the basis of actual expenditure as set out in the quarterly financial report and Keele shall pay such invoice within thirty (30) days of receipt of a correct tax invoice, subject always to receipt of funds from the Authority.

The quarterly financial report and invoices should be sent to: SPCR Finance and Contracts Officer. Failure to submit your financial report will result in a delay in payment.

I would be grateful if you would organise for an authorised signatory for your institution to sign this letter and return a copy to Jess Nye, Programme Manager jess.nye@ucl.ac.uk confirming your acceptance of the award which will be governed under the terms and conditions of the Partnership Agreement. Please note that funds cannot be released until a signed copy of the letter has been returned.

Signed on behalf of Keele University

Name: Dr Georgina Fletcher

Title: SPCR Assistant Director

Date: 28 June 2022

Signed on behalf of University of Oxford

Name: Nicola Small

Title: Head of Administration and Finance

Date: 28 July 2022

**FINANCIAL ARRANGEMENTS**

**NIHR School for Primary Care Research**

Award issued to the University of Oxford is as follows:

**Start Date of Individual Project** 01/10/2022

**End Date of Individual Project** 31/03/2024

**Award Breakdown**

| | £ |
|---|---|
| **STAFF** | |
| Research | |
| Academic & Related | |
| Support | |
| **Staff sub-total** | **0.00** |
| **NON-STAFF** | |
| Consultancy | |
| Consumables | |
| Dissemination Activities | 500.00 |
| Equipment | |
| Patient & Public Involvement | 1,475.00 |
| Recruitment Costs | 1,250.00 |
| Travel & Subsistence | 700.00 |
| Other | 19,160.00 |
| **Non-staff sub-total** | **23,085.00** |
| **OVERHEADS** | |
| **TOTAL COSTS** | **23,085.00** |

## Chapter 3 Appendices

### Appendix Table 3.1 Number of tests requested by age group and setting

| Age group | Setting | Number of children | % |
|---|---|---|---|
| <1 year | Total | 580,636 | 100·0 |
| | General Practice | 4,157 | 0·7 |
| | Inpatient | 558,716 | 96·2 |
| | Outpatient | 17,763 | 3·1 |
| 1-5 years | Total | 439,770 | 100·0 |
| | General Practice | 49,123 | 11·2 |
| | Inpatient | 315,321 | 71·7 |
| | Outpatient | 75,326 | 17·1 |
| 6-10 years | Total | 319,387 | 100·0 |
| | General Practice | 85,081 | 26·6 |
| | Inpatient | 170,158 | 53·3 |
| | Outpatient | 64,148 | 20·1 |
| 11-15 years | Total | 409,632 | 100·0 |
| | General Practice | 155,145 | 37·9 |
| | Inpatient | 188,361 | 46·0 |
| | Outpatient | 66,126 | 16·1 |

**Appendix Table 3.2 Standardised rates of test use, annual percentage change and average annual percentage change by setting, sex, and age from 2005 to 2019**

| | | | | Test rate* (per 1,000 child years) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Start | End | Start | End | APC (%) | 95% CI Lower Limit | 95% CI Upper Limit | p value | AAPC (%) 95% CI | p value |
| Overall | | 2005 | 2012 | 878.4 | 781.9 | -2.2 | -3.8 | -0.6 | 0.01 | 1.5 | 0.2 |
| | | 2012 | 2015 | 781.9 | 981.9 | 9 | -3 | 22.5 | 0.1 | (-0.8 to 3.9) | |
| | | 2015 | 2019 | 981.9 | 1107.8 | 2.8 | -0.5 | 6.3 | 0.1 | | |
| Setting | General Practice | 2005 | 2011 | 123.5 | 125.9 | -3.3 | -8.3 | 1.9 | 0.2 | 4.2 | 0.002 |
| | | 2011 | 2019 | 125.9 | 259.9 | 10.1 | 6.5 | 14 | < 0.001 | (1.5 to 6.9) | |
| | Inpatient | 2005 | 2015 | 620.5 | 699.4 | 1.0 | -0.2 | 2.3 | 0.1 | -0.6 | 0.4 |
| | | 2015 | 2019 | 699.4 | 564.5 | -4.5 | -9.3 | 0.5 | 0.07 | (-2.1 to 0.9) | |
| | Outpatient | 2005 | 2011 | 134.4 | 38.3 | -21.6 | -28.3 | -14.2 | < 0.001 | 6.6 | 0.005 |
| | | 2011 | 2019 | 38.3 | 283.5 | 34.1 | 26.6 | 42.2 | < 0.001 | (1.9 to 11.5) | |
| Sex | Female | 2005 | 2012 | 790.7 | 698.9 | -1.6 | -3.3 | 0.1 | 0.05 | 2.3 | |
| | | 2012 | 2019 | 698.9 | 1098.1 | 6.4 | 4.7 | 8 | < 0.001 | (1.3 to 3.3) | <0.001 |
| | Male | 2005 | 2012 | 949.0 | 857.1 | -2.2 | -3.9 | -0.5 | 0.02 | 1.2 | 0.3 |
| | | 2012 | 2015 | 857.1 | 1077.2 | 10 | -3 | 24.7 | 0.1 | (-1.3 to 3.8) | |
| | | 2015 | 2019 | 1077.2 | 1114.8 | 1 | -2.6 | 4.7 | 0.5 | | |
| Age group | <1 | 2005 | 2019 | 301.5 | 232.5 | -1.2 | -2.2 | -0.2 | 0.02 | -1.2 | 0.02 |
| | | | | | | | | | | (-2.2 to -0.2) | |
| | 1-5 | 2005 | 2010 | 203.0 | 192.9 | -1.8 | -6 | 2.6 | 0.4 | 1.9 | 0.2 |
| | | 2010 | 2015 | 192.9 | 257.4 | 6.9 | 0.5 | 13.8 | 0.04 | (-0.8 to 4.6) | |
| | | 2015 | 2019 | 257.4 | 269.2 | 0.4 | -5.7 | 6.8 | 0.9 | | |
| | 6-10 | 2005 | 2010 | 153.0 | 117.4 | -6.3 | -9 | -3.6 | 0.001 | 4.4 | <0.001 |
| | | 2010 | 2015 | 117.4 | 236.8 | 16.6 | 12 | 21.4 | < 0.001 | (2.6 to 6.3) | |
| | | 2015 | 2019 | 236.8 | 265.3 | 4.2 | 0.1 | 8.5 | 0.048 | | |
| | 11-15 | 2005 | 2013 | 216.9 | 164.3 | -6.1 | -9.8 | -2.2 | 0.006 | 2.0 | 0.2 |
| | | 2013 | 2020 | 164.3 | 336.2 | 13.8 | 6.9 | 21.1 | 0.001 | (-1.1 to 5.2) | |

*Observed rates; Abbreviations: APC Annual percentage change; AAPC Average annual percentage change

**Appendix Table 3.3 Proportion of children in Oxfordshire who had at least one test, 2005 to 2019.**

| Year | Proportion of population aged below 16 (%) | | | |
|---|---|---|---|---|
| | Overall | General Practice | Inpatient | Outpatient |
| 2005 | 8.8 | 2.8 | 4.7 | 2.2 |
| 2006 | 8.8 | 2.8 | 4.6 | 2.3 |
| 2007 | 9.0 | 3.1 | 5.3 | 1.6 |
| 2008 | 8.3 | 2.4 | 5.5 | 1.0 |
| 2009 | 8.0 | 2.1 | 5.7 | 0.8 |
| 2010 | 7.6 | 2.1 | 5.3 | 0.7 |
| 2011 | 8.0 | 2.1 | 5.8 | 0.7 |
| 2012 | 7.9 | 2.1 | 5.6 | 0.7 |
| 2013 | 8.3 | 2.4 | 5.8 | 0.9 |
| 2014 | 9.6 | 2.5 | 6.7 | 1.6 |
| 2015 | 10.3 | 2.7 | 7.0 | 1.8 |
| 2016 | 10.9 | 3.1 | 7.2 | 1.9 |
| 2017 | 11.0 | 3.3 | 5.9 | 3.5 |
| 2018 | 11.8 | 3.7 | 6.0 | 4.0 |
| 2019 | 12.3 | 4.0 | 6.1 | 4.0 |

* Denominator based on ONS estimates of children aged 0 to 15 in Oxfordshire from 2005 to 2019

**Appendix Table 3.4 Age-standardised rates of test use and annual percentage change, 2005 to 2019**

| Test name | Start | End | Rate (per 1,000 child years) Start | End | APC (%) | 95% Confidence Interval Lower Limit | Upper Limit | p value | AAPC (%) | 95% Confidence Interval Lower Limit | Upper Limit | p value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Amylase | 2005 | 2010 | 4.8 | 3.0 | -7.3 | -15 | 1.2 | 0.082 | 0.6 | -2.6 | 3.9 | 0.715 |
| | 2010 | 2019 | 3.0 | 5.7 | 5.3 | 2 | 8.6 | 0.005 | | | | |
| Blood gas | 2005 | 2010 | 3.5 | 1.7 | -17.9 | -48 | 29.6 | 0.341 | -11.4 | -28.5 | 9.7 | 0.265 |
| | 2010 | 2015 | 1.7 | 35.9 | 98.1 | 39.5 | 181.3 | 0.002 | | | | |
| | 2015 | 2019 | 35.9 | 1.1 | -64.4 | -79.3 | -38.7 | 0.003 | | | | |
| C reactive protein | 2005 | 2010 | 3.5 | 68.6 | -7.7 | -10 | -5.4 | <0.001 | 0.9 | -0.6 | 2.4 | 0.224 |
| | 2010 | 2015 | 68.6 | 105.4 | 9.6 | 5.8 | 13.5 | <0.001 | | | | |
| | 2015 | 2019 | 105.4 | 113.4 | 1.8 | -1.3 | 5 | 0.212 | | | | |
| Calcium, magnesium, phosphate | 2005 | 2014 | 94.5 | 67.0 | -3.8 | -5.3 | -2.3 | <0.001 | -1.1 | -2.6 | 0.4 | 0.156 |
| | 2014 | 2019 | 67.0 | 76.6 | 4 | -0.1 | 8.2 | 0.055 | | | | |
| Coagulation profile | 2005 | 2012 | 29.5 | 20.2 | -5.6 | -8.1 | -3 | 0.001 | -2.9 | -6.8 | 1.2 | 0.162 |
| | 2012 | 2015 | 20.2 | 22.6 | 6.9 | -12.9 | 31.2 | 0.468 | | | | |
| | 2015 | 2019 | 22.6 | 19.7 | -5.1 | -11.1 | 1.4 | 0.103 | | | | |
| Coeliac testing | 2005 | 2010 | 7.9 | 9.1 | 1.1 | -4.7 | 7.2 | 0.686 | 7.7 | 5.5 | 10 | <0.001 |
| | 2010 | 2019 | 9.1 | 23.5 | 11.6 | 9.7 | 13.5 | <0.001 | | | | |
| Creatine Kinase | 2005 | 2011 | 3.1 | 2.7 | -2.1 | -6.4 | 2.3 | 0.291 | 4.6 | 1 | 8.4 | 0.011 |
| | 2011 | 2015 | 2.7 | 4.8 | 14.9 | 2.2 | 29.3 | 0.027 | | | | |
| | 2015 | 2019 | 4.8 | 6.1 | 5.3 | -0.6 | 11.6 | 0.073 | | | | |
| Erythrocyte sedimentation rate | 2005 | 2009 | 24.4 | 15.9 | -8.4 | -14.7 | -1.7 | 0.02 | -2.1 | -4.1 | 0 | 0.054 |
| | 2009 | 2019 | 15.9 | 17.2 | 0.6 | -1.3 | 2.5 | 0.495 | | | | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Folate | 2005 | 2012 | 3.4 | 3.4 | -2.4 | -7.4 | 2.9 | 0.326 | 8.4 | 5.3 | 11.5 | <0.001 |
| | 2012 | 2019 | 3.4 | 12.2 | 20.3 | 15.9 | 24.9 | <0.001 | | | | |
| Full blood count | 2005 | 2012 | 163.6 | 145.1 | -2.4 | -3.8 | -0.9 | 0.008 | 1.1 | -1.1 | 3.3 | 0.344 |
| | 2012 | 2015 | 145.1 | 176.4 | 7.8 | -3.4 | 20.4 | 0.149 | | | | |
| | 2015 | 2019 | 176.4 | 194.6 | 2.2 | -1 | 5.5 | 0.146 | | | | |
| Glucose | 2005 | 2012 | 15.5 | 11.8 | -7.3 | -10.8 | -3.6 | 0.002 | 8 | 2.2 | 14.1 | 0.006 |
| | 2012 | 2015 | 11.8 | 27.6 | 47 | 10.3 | 96.1 | 0.016 | | | | |
| | 2015 | 2019 | 27.6 | 43.9 | 11.8 | 6.5 | 17.4 | 0.001 | | | | |
| HbA1c | 2005 | 2013 | 7.1 | 10.4 | 1.5 | -2.5 | 5.7 | 0.419 | 8 | 5.3 | 10.8 | <0.001 |
| | 2013 | 2019 | 10.4 | 25 | 17.3 | 12.4 | 22.4 | <0.001 | | | | |
| IgA | 2005 | 2010 | 8.7 | 9.8 | 0.5 | -4.7 | 5.9 | 0.839 | 7.9 | 5.2 | 10.7 | <0.001 |
| | 2010 | 2015 | 9.8 | 20.9 | 16.8 | 10.3 | 23.8 | <0.001 | | | | |
| | 2015 | 2019 | 20.9 | 26.8 | 6.8 | 2.4 | 11.5 | 0.008 | | | | |
| IgG | 2005 | 2015 | 3.1 | 5.9 | 6.8 | 5 | 8.8 | <0.001 | 2.9 | 0.9 | 4.9 | 0.004 |
| | 2015 | 2019 | 5.9 | 4.4 | -6.4 | -12.1 | -0.4 | 0.04 | | | | |
| IgM | 2005 | 2015 | 3.1 | 5.9 | 6.9 | 4.9 | 8.8 | <0.001 | 2.9 | 0.9 | 4.9 | 0.005 |
| | 2015 | 2019 | 5.9 | 4.4 | -6.4 | -12.2 | -0.3 | 0.042 | | | | |
| Iron studies | 2005 | 2011 | 9.2 | 10.3 | -1.2 | -5.3 | 3.2 | 0.563 | 9.3 | 7.3 | 11.4 | <0.001 |
| | 2011 | 2019 | 10.3 | 33.3 | 17.9 | 15.7 | 20.1 | <0.001 | | | | |
| Lipids | 2005 | 2009 | 4.4 | 4.6 | 0.1 | -8.5 | 9.5 | 0.983 | 3.9 | -2.5 | 10.7 | 0.243 |
| | 2009 | 2012 | 4.6 | 3.2 | -15.2 | -38.9 | 17.8 | 0.274 | | | | |
| | 2012 | 2019 | 3.2 | 8.7 | 15.7 | 11.8 | 19.8 | <0.001 | | | | |
| Liver function test | 2005 | 2012 | 135.3 | 113.3 | -2.5 | -4.1 | -0.9 | 0.006 | 0.1 | -0.9 | 1.1 | 0.785 |
| | 2012 | 2019 | 113.3 | 136.1 | 2.9 | 1.2 | 4.5 | 0.003 | | | | |
| Medication level | 2005 | 2019 | 2.9 | 9.8 | 2.9 | 0 | 5.9 | 0.051 | 2.9 | 0 | 5.9 | 0.051 |
| Monospot test | 2005 | 2019 | 5.9 | 2.2 | -8.8 | -11.4 | -6.1 | <0.001 | -8.8 | -11.4 | -6.1 | <0.001 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Parathyroid hormone | 2005 | 2011 | 1.3 | 3.9 | 17.6 | 10.3 | 25.3 | <0.001 | 9.8 | 6.8 | 12.9 | <0.001 |
| | 2011 | 2019 | 3.9 | 4.9 | 4.3 | 1.5 | 7.2 | 0.006 | | | | |
| Thyroid function test | 2005 | 2011 | 23.9 | 23 | -2.9 | -7 | 1.4 | 0.159 | 3.8 | 1.7 | 5.9 | <0.001 |
| | 2011 | 2019 | 23 | 43.8 | 9.1 | 6.6 | 11.7 | <0.001 | | | | |
| Urea and electrolytes | 2005 | 2012 | 153.9 | 142.1 | -1.6 | -3.4 | 0.1 | 0.065 | 0.6 | -1.9 | 3.2 | 0.64 |
| | 2012 | 2015 | 142.1 | 174.9 | 7.9 | -5.1 | 22.7 | 0.206 | | | | |
| | 2015 | 2019 | 174.9 | 173.4 | -0.6 | -4.4 | 3.2 | 0.703 | | | | |
| Urine creatinine | 2005 | 2012 | 7 | 6.1 | -2.8 | -5 | -0.5 | 0.024 | 0.6 | -2.7 | 4.1 | 0.717 |
| | 2012 | 2015 | 6.1 | 7.9 | 9.2 | -7.9 | 29.6 | 0.261 | | | | |
| | 2015 | 2019 | 7.9 | 7.8 | 0.5 | -4.2 | 5.4 | 0.822 | | | | |
| Vitamin $B_{12}$ | 2005 | 2012 | 3.6 | 3.8 | -1.7 | -6.9 | 3.7 | 0.486 | 8.4 | 5.3 | 11.6 | <0.001 |
| | 2012 | 2019 | 3.8 | 12.9 | 19.5 | 15.1 | 24.1 | <0.001 | | | | |
| Vitamin D | 2005 | 2019 | 0.9 | 28.9 | 26.5 | 23.7 | 29.3 | <0.001 | 26.5 | 23.7 | 29.3 | <0.001 |

Abbreviations: APC, annual percentage change; AAPC, average annual percentage change

## Chapter 4 Appendices

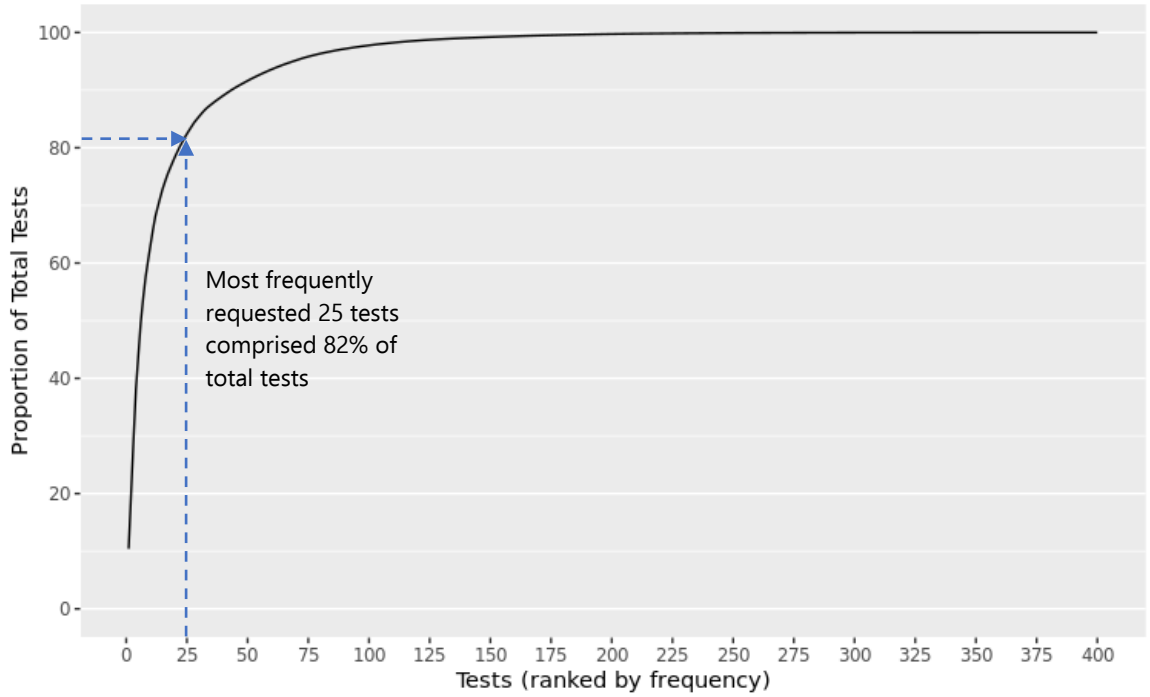
### Appendix Table 4.1 Sources of included tests for this chapter

| | Rank | Test name | Number of tests |
|---|---|---|---|
| 25 most frequently recorded tests* | 1 | Urine MCS | 1391785 |
| | 2 | Urinalysis | 1261359 |
| | 3 | Full blood count | 1249322 |
| | 4 | Peak flow | 1184109 |
| | 5 | Urea and electrolytes | 790364 |
| | 6 | Liver function test | 760041 |
| | 8 | Thyroid function tests | 560244 |
| | 9 | Iron studies | 489989 |
| | 10 | Bone profile | 375914 |
| | 11 | Glucose | 361960 |
| | 12 | C reactive protein | 349276 |
| | 13 | Erythrocyte sedimentation rate | 314392 |
| | 14 | Vitamin $B_{12}$ | 217619 |
| | 15 | Stool MCS | 214476 |
| | 16 | Folate | 204739 |
| | 17 | Vitamin D | 168323 |
| | 18 | HbA1c | 162022 |
| | 19 | Wound/Skin MCS | 138975 |
| | 20 | Spirometry | 131770 |
| | 21 | Hearing test | 130290 |
| | 22 | Coeliac test | 120337 |
| | 23 | Enteric virus screen | 117247 |
| | 25 | ECG | 105247 |
| | 26 | Chest X-Ray | 104604 |
| | 29 | Stool OCP | 92839 |
| Tests included from other chapters | 33 | Immunoglobulins (IgG, IgA, IgM) | 70549 |
| | 34 | Monospot | 68121 |
| | 41 | Allergen Specific IgE | 42642 |
| | 49 | Abdominal ultrasound | 34195 |
| | 54 | *Helicobacter* test | 30655 |
| | 57 | Renal ultrasound | 27606 |
| | 80 | Calprotectin | 14637 |
| | 84 | MRI head | 12936 |
| | 115 | CT head | 5437 |
| | 197 | Fractional exhaled nitric oxide | 977 |

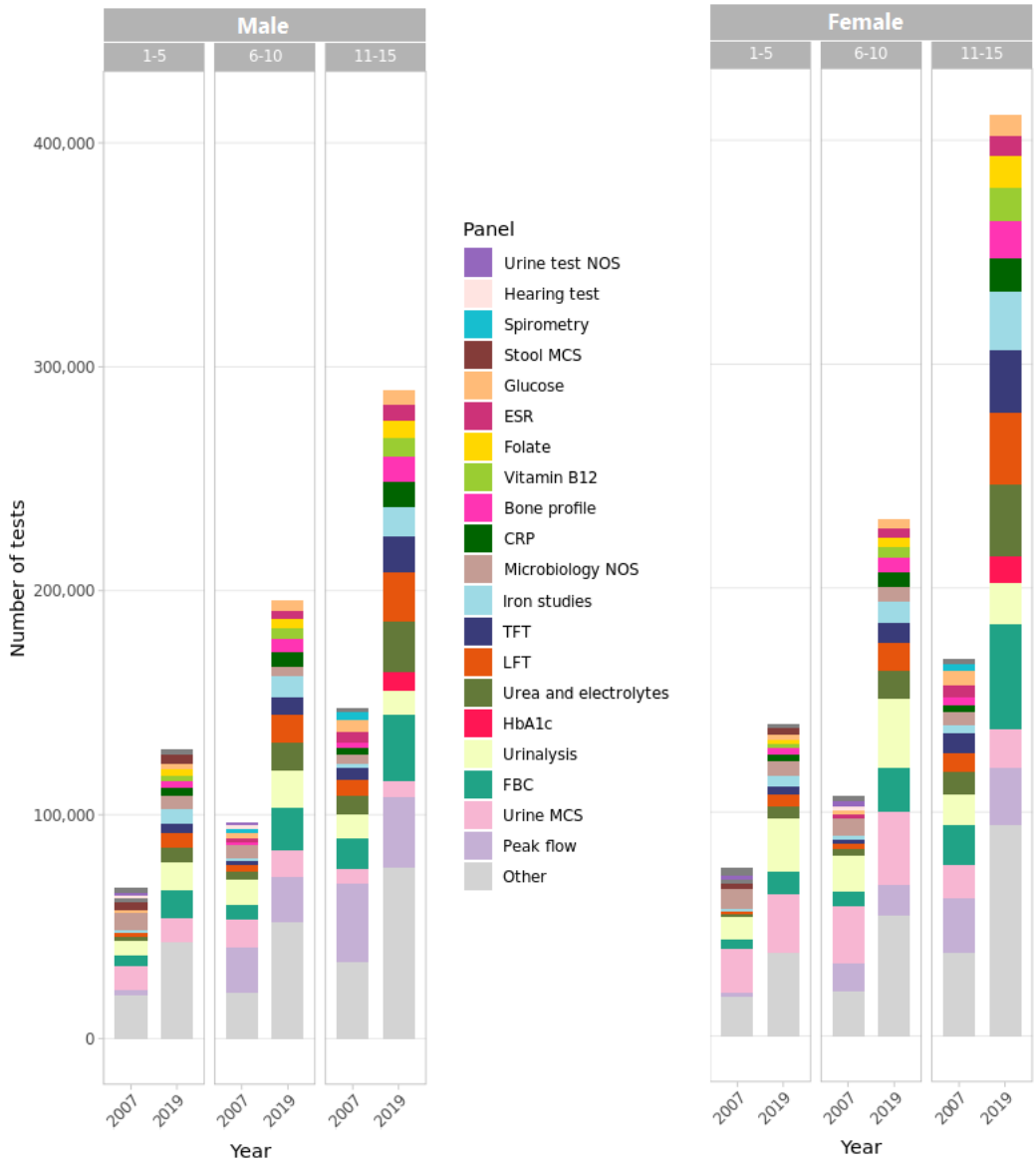
*Excluding unspecified tests
Abbreviations: CT - Computed tomography; MCS – Microscopy, culture, sensitivities; OCP – Ova/cysts/parasites

**Appendix Figure 4.1 Cumulative frequency of the 400 tests by rank from highest to lowest frequency**

Excluding the following tests: Microbiology NOS; Urine test NOS, US NOS, Diagnostic test NOS, which comprised 8.2% of all tests.

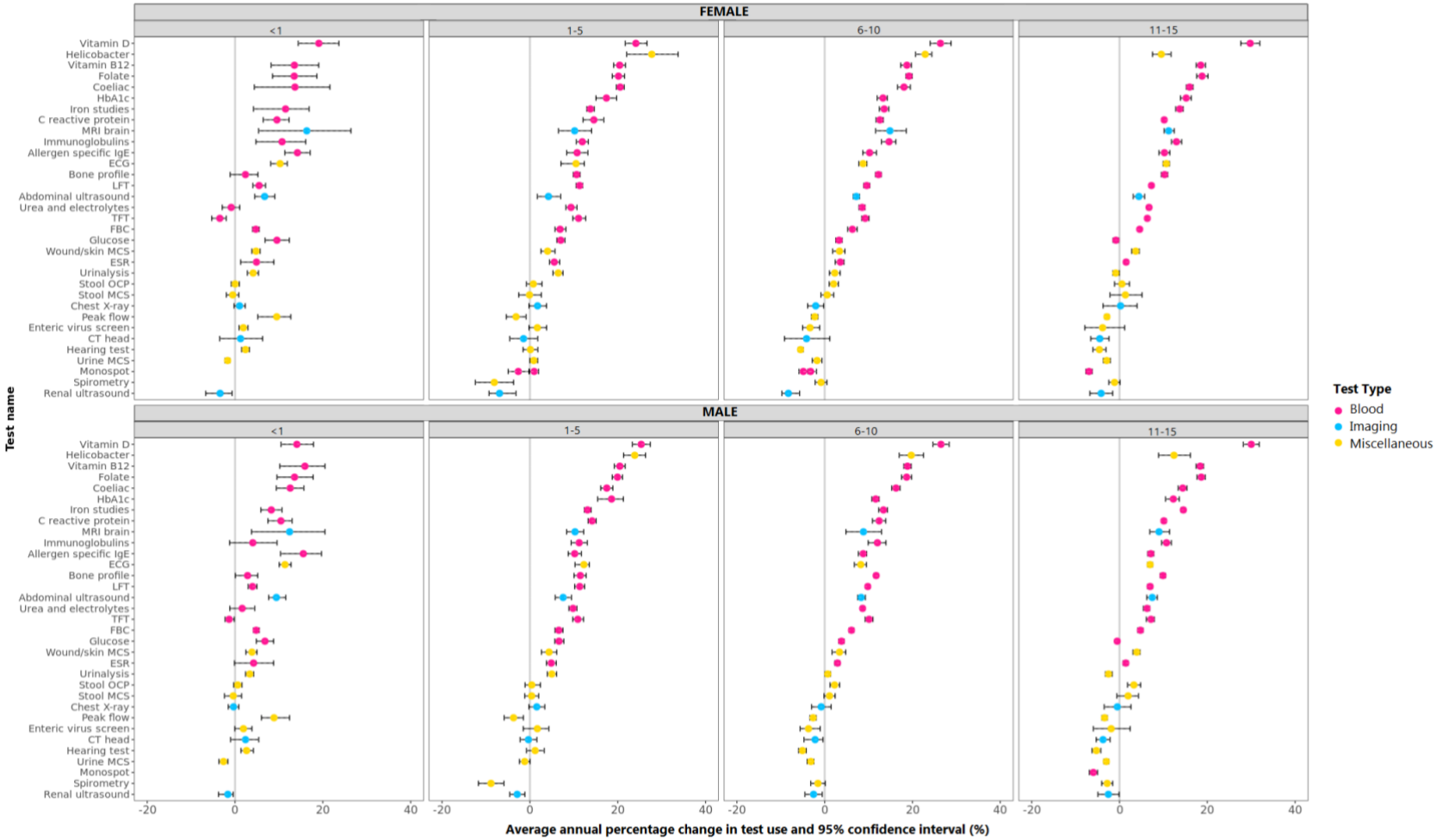


**Appendix Figure 4.2 Changes in absolute numbers of tests by gender and age group from 2007 to 2019**



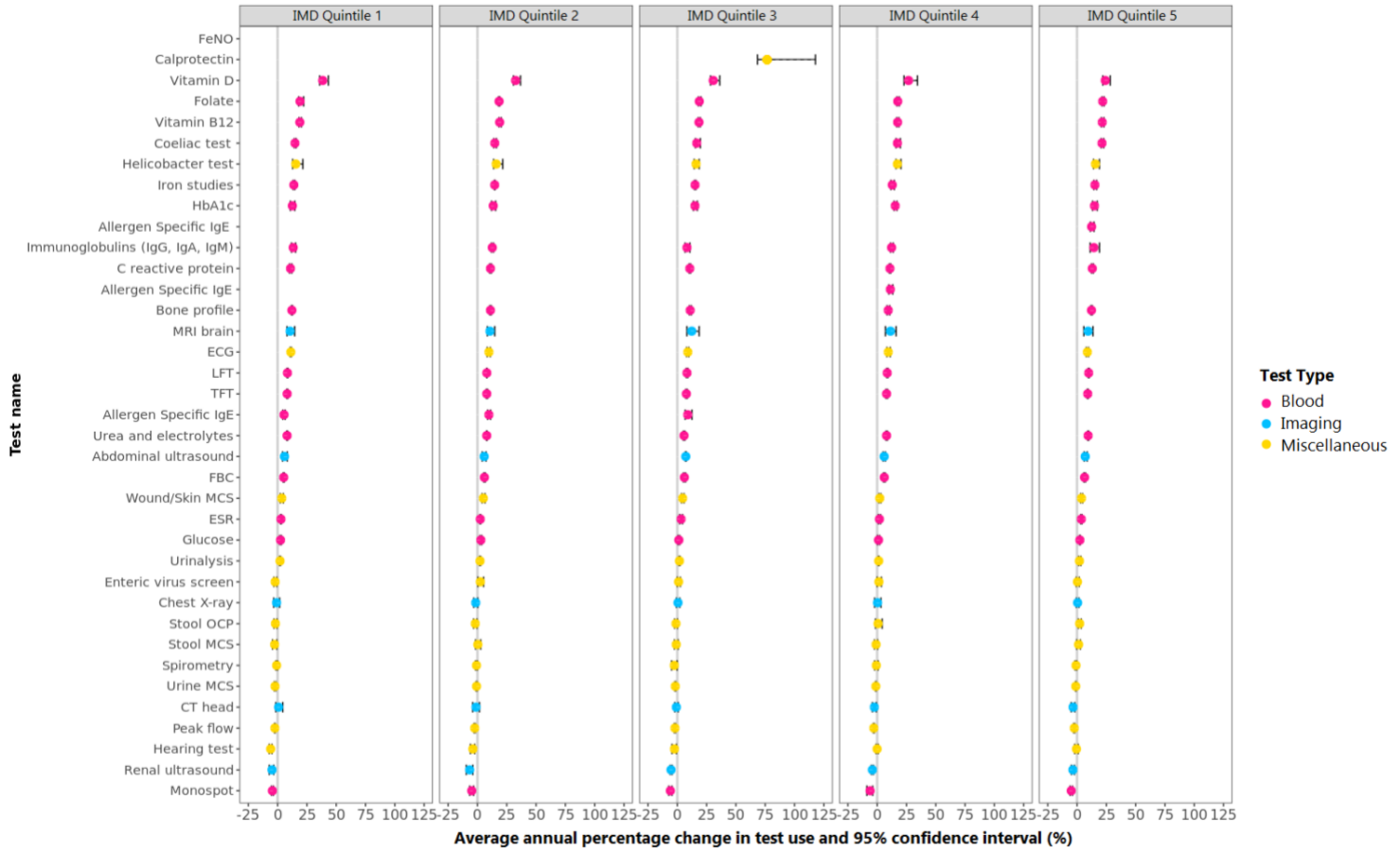

CRP – C reactive protein; ESR – Erythrocyte sedimentation rate; FBC – Full blood count; LFT – Liver function test; MCS – Microscopy, culture, sensitivities; NOS – Not otherwise specified; TFT – Thyroid function test

**Appendix Figure 4.3 Temporal changes in specific tests for children aged 0 to 15 from 2007 to 2019; stratified by gender and age.**




CRP – C reactive protein; ECG – Electrocardiogram; ESR – Erythrocyte sedimentation rate; FBC – Full blood count; LFT – Liver function test; MCS – Microscopy, culture, sensitivities; NOS – Not otherwise specified; OCP – Ova/cysts/parasites; TFT – Thyroid function test; US – Ultrasound. Tests were excluded if there were too few tests to perform a meaningful analysis, or the test was not technically possible to perform, leaving the outcome blank in the graph (e.g., spirometry in <1 year old).

**Appendix Figure 4.4 Temporal changes in specific tests for children aged 0 to 15 from 2007 to 2019; stratified by Index of Multiple Deprivation Quintile.**




CRP – C reactive protein; ECG – Electrocardiogram; ESR – Erythrocyte sedimentation rate; FBC – Full blood count; FeNO – Fractional exhaled nitric oxide; LFT – Liver function test; MCS – Microscopy, culture, sensitivities; NOS – Not otherwise specified; OCP – Ova/cysts/parasites; TFT – Thyroid function test; US – Ultrasound.
Tests were excluded if there were too few tests to perform a meaningful analysis, leaving the outcome blank in the graph.

**Appendix Table 4.2 Adjusted test rates and coefficient of variation (CoV) with corresponding 95% confidence intervals in 2019**

| Test name | Test type | Adjusted mean rate (tests/1,000 child-years) | Adjusted CoV (%) | 95% CI lower limit (%) | 95% CI upper limit (%) | Rate-Variability* |
|---|---|---|---|---|---|---|
| FBC | Blood | 58.7 | 13.0 | 12.9 | 13.1 | High test rate - Low CoV |
| Urine MCS | Miscellaneous | 48.4 | 3.8 | 3.7 | 3.9 | High test rate - Low CoV |
| Urinalysis | Miscellaneous | 47.9 | 10.5 | 10.3 | 10.7 | High test rate - Low CoV |
| Peak flow | Miscellaneous | 39.4 | 13.0 | 12.8 | 13.3 | High test rate - Low CoV |
| Urea and electrolytes | Blood | 39.0 | 10.1 | 10.0 | 10.2 | High test rate - Low CoV |
| LFT | Blood | 38.2 | 10.5 | 10.4 | 10.7 | High test rate - Low CoV |
| Iron studies | Blood | 29.7 | 18.8 | 18.7 | 19.0 | High test rate - High CoV |
| TFT | Blood | 28.6 | 11.5 | 11.3 | 11.6 | High test rate - Low CoV |
| Bone profile | Blood | 19.8 | 15.5 | 15.4 | 15.7 | High test rate - Low CoV |
| CRP | Blood | 18.8 | 6.7 | 6.5 | 6.8 | High test rate - Low CoV |
| Vitamin B12 | Blood | 15.1 | 18.4 | 18.3 | 18.6 | High test rate - High CoV |
| Folate | Blood | 14.6 | 19.7 | 19.6 | 19.9 | High test rate - High CoV |
| Glucose | Blood | 12.9 | 15.8 | 15.7 | 15.9 | High test rate - Low CoV |
| HbA1c | Blood | 12.3 | 13.0 | 12.9 | 13.1 | High test rate - Low CoV |
| ESR | Blood | 11.6 | 11.2 | 11.1 | 11.4 | High test rate - Low CoV |
| Vitamin D | Blood | 8.7 | 38.1 | 38.0 | 38.3 | High test rate - High CoV |
| Coeliac | Blood | 7.4 | 18.9 | 18.8 | 19.0 | High test rate - High CoV |
| Stool MCS | Miscellaneous | 6.9 | 9.3 | 9.1 | 9.4 | High test rate - Low CoV |
| ECG | Miscellaneous | 6.3 | 12.7 | 12.5 | 12.8 | Low test rate - Low CoV |
| Wound/Skin MCS | Miscellaneous | 5.0 | 27.1 | 26.9 | 27.2 | Low test rate - High CoV |
| Spirometry | Miscellaneous | 4.6 | 19.0 | 18.9 | 19.1 | Low test rate - High CoV |
| Hearing test | Miscellaneous | 4.2 | 51.6 | 51.4 | 51.7 | Low test rate - High CoV |
| Immunoglobulins | Blood | 4.0 | 22.2 | 22.1 | 22.3 | Low test rate - High CoV |
| CXR | Imaging | 3.8 | 17.9 | 17.8 | 18.1 | Low test rate - High CoV |
| Enteric virus screen | Miscellaneous | 3.8 | 26.2 | 26.1 | 26.4 | Low test rate - High CoV |
| Stool OCP | Miscellaneous | 2.4 | 20.4 | 20.3 | 20.5 | Low test rate - High CoV |
| Specific IgE | Blood | 2.4 | 16.5 | 16.4 | 16.7 | Low test rate - High CoV |
| *Helicobacter* pylori | Miscellaneous | 2.1 | 37.0 | 36.8 | 37.1 | Low test rate - High CoV |

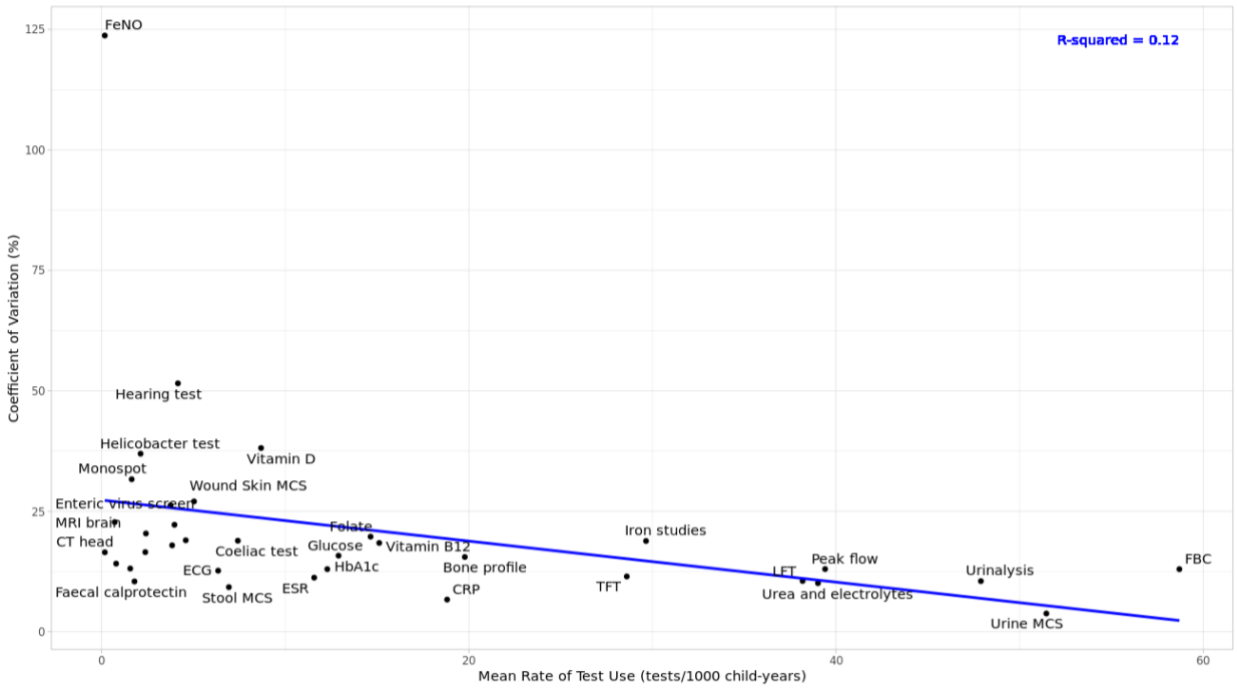
| | | | | | | |
|---|---|---|---|---|---|---|
| Faecal calprotectin | Miscellaneous | 1.8 | 10.4 | 10.3 | 10.6 | Low test rate - Low CoV |
| Monospot | Blood | 1.6 | 31.7 | 31.5 | 31.8 | Low test rate - High CoV |
| US abdomen | Imaging | 1.6 | 13.1 | 13.0 | 13.3 | Low test rate - Low CoV |
| US renal | Imaging | 0.8 | 14.1 | 14.0 | 14.3 | Low test rate - Low CoV |
| MRI head | Imaging | 0.7 | 22.8 | 22.6 | 22.9 | Low test rate - High CoV |
| CT head | Imaging | 0.2 | 16.5 | 16.4 | 16.6 | Low test rate - High CoV |
| FeNO | Miscellaneous | 0.2 | 123.7 | 123.6 | 123.9 | Low test rate - High CoV |

*High and Low in relation to test rate and variability are relative to the median test rate of 6.9 tests/1,000 child-years and median CoV of 16.5%
Abbreviations: CRP – C reactive protein; CT – computed tomography; CXR – Chest X-ray; ECG – Electrocardiogram; ESR – Erythrocyte sedimentation rate; FBC – Full blood count; FeNO – Fractional exhaled nitric oxide; LFT – Liver function test; MCS – Microscopy, culture, sensitivities; MRI – Magnetic resonance imaging; NOS – Not otherwise specified; OCP – Ova/cysts/parasites; TFT – Thyroid function test; US – Ultrasound.

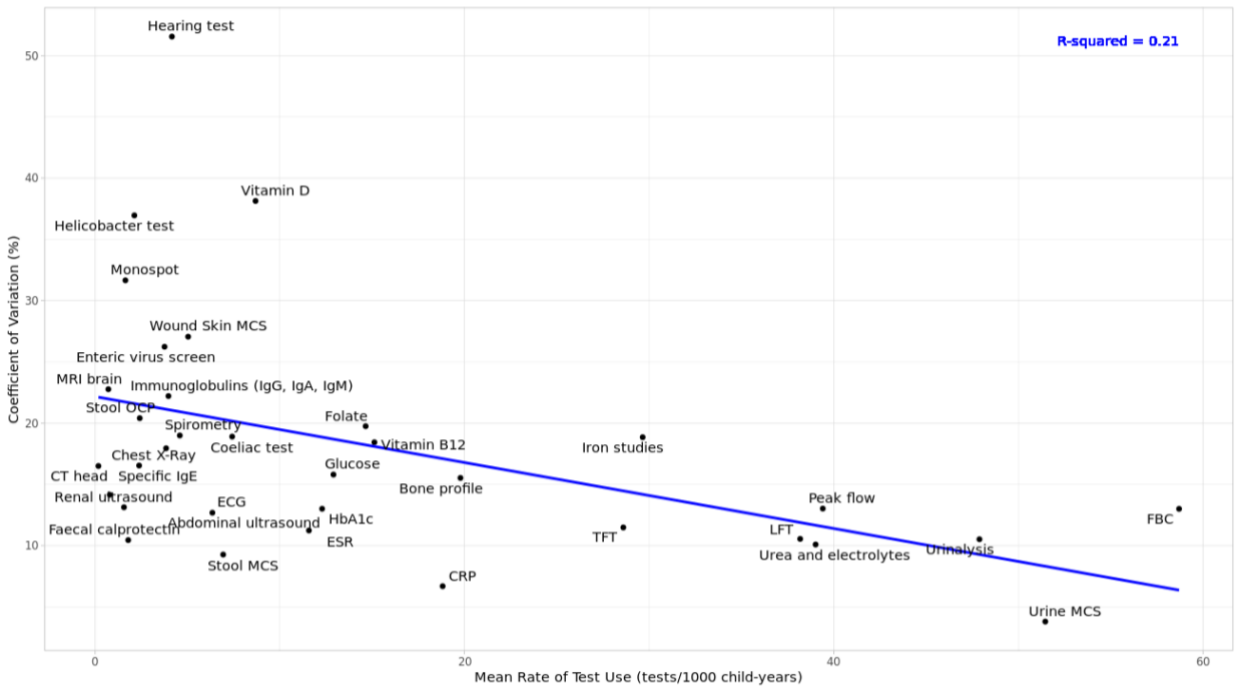
**Appendix Figure 4.5 Exploring potential linear associations between test utilisation rate (tests per 1,000 child-years) and practice variability (coefficient of variation [CoV]).**

A. Including all tests; B. Excluding FeNO.

**A.**



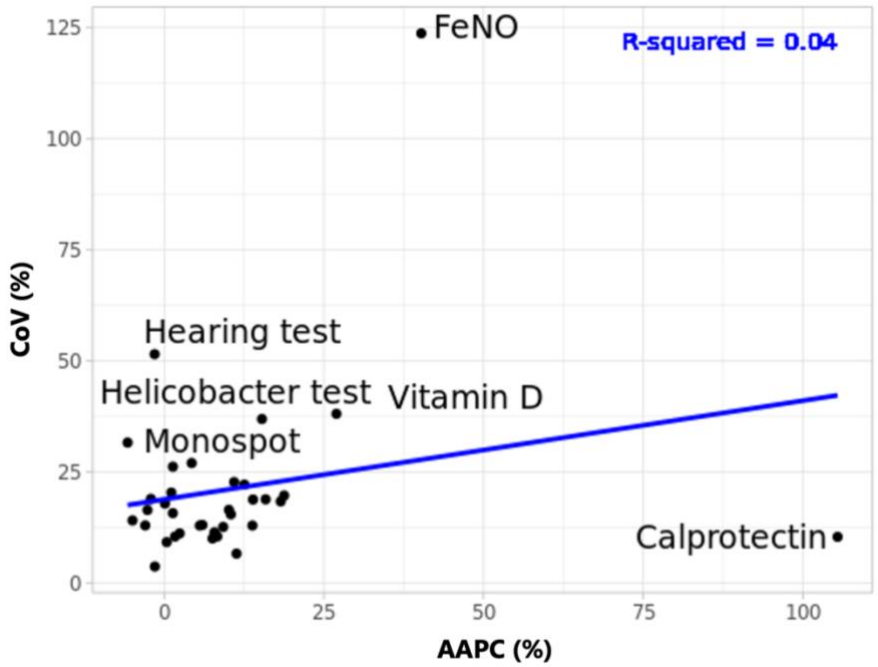

**B.**



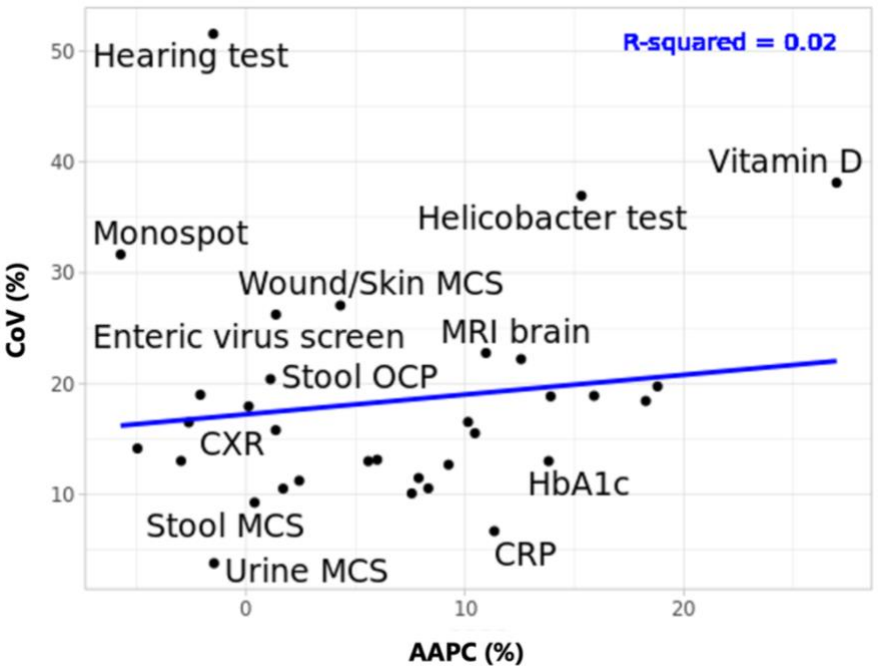
**Appendix Figure 4.6 Exploring potential linear associations between temporal change (average annual percentage change [AAPC] and practice variability (coefficient of variation [CoV]) of tests.**

A. Including all tests; B. Excluding calprotectin; C. Excluding FeNO; D; Excluding FeNO and calprotectin.

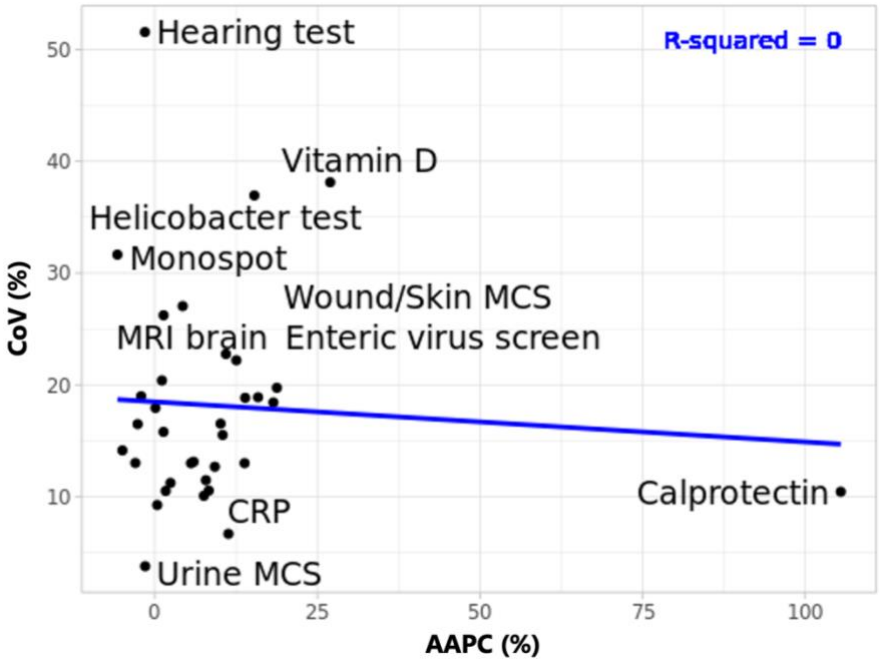
**A.**



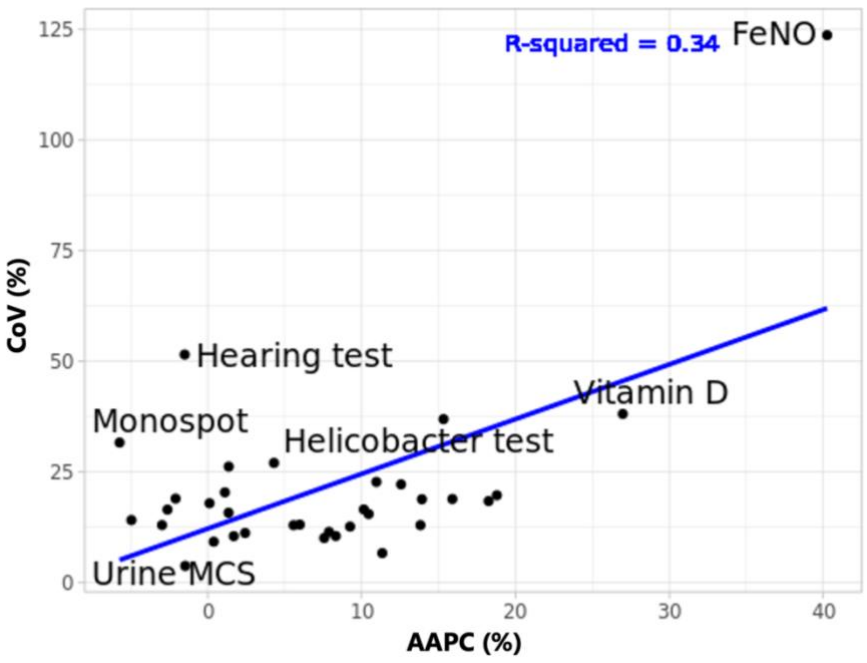

**B.**



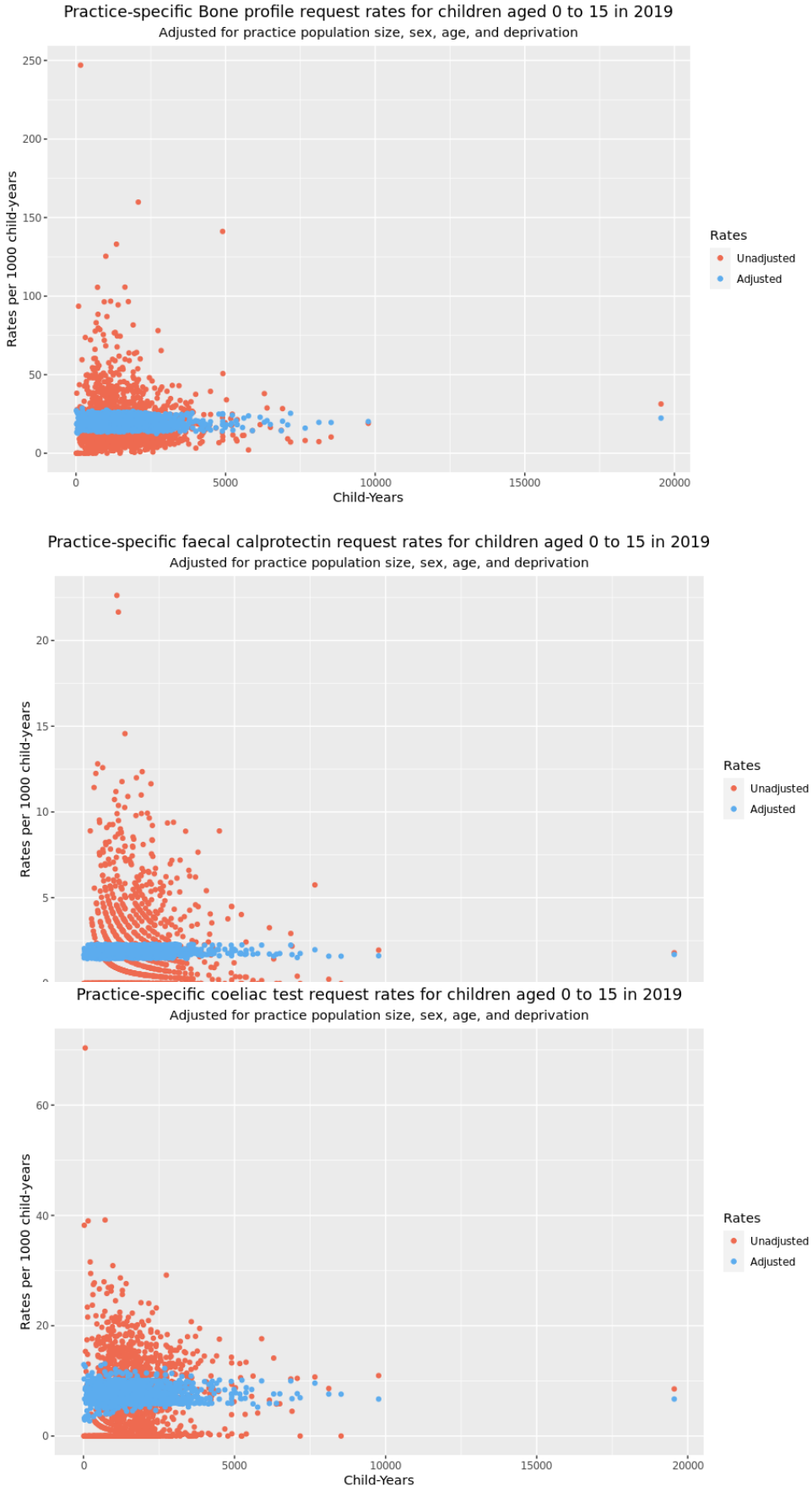
C.

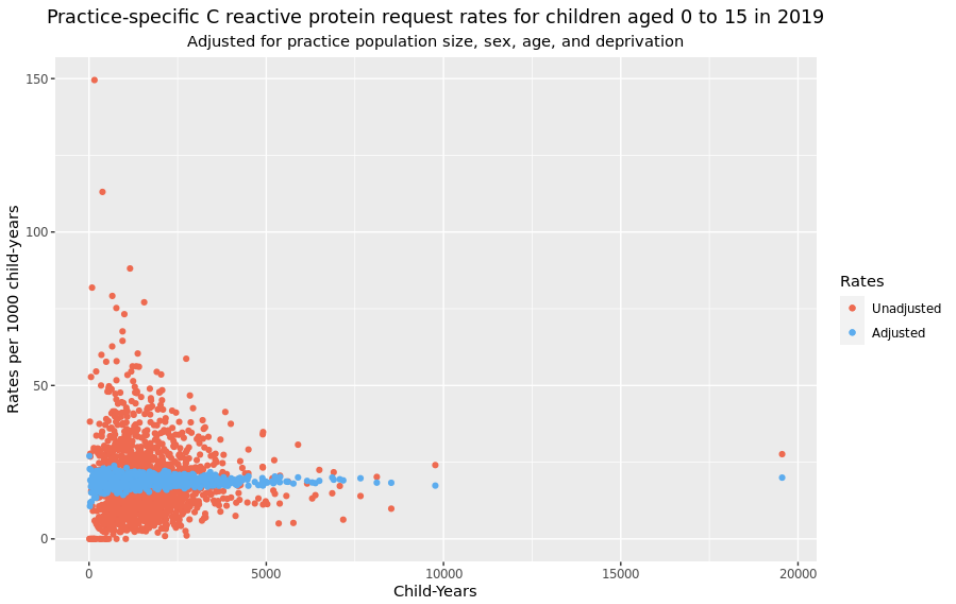

Hearing test
R-squared = 0
Vitamin D
Helicobacter test
Monospot
Wound/Skin MCS
MRI brain
Enteric virus screen
Calprotectin
CRP
Urine MCS
CoV (%)
AAPC (%)
50
40
30
20
10
0
25
50
75
100

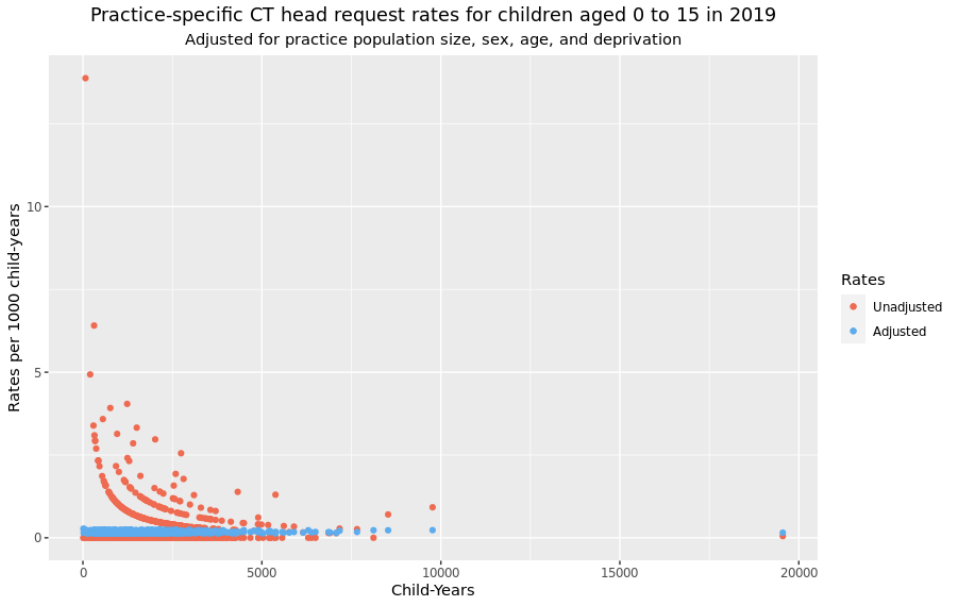
D.

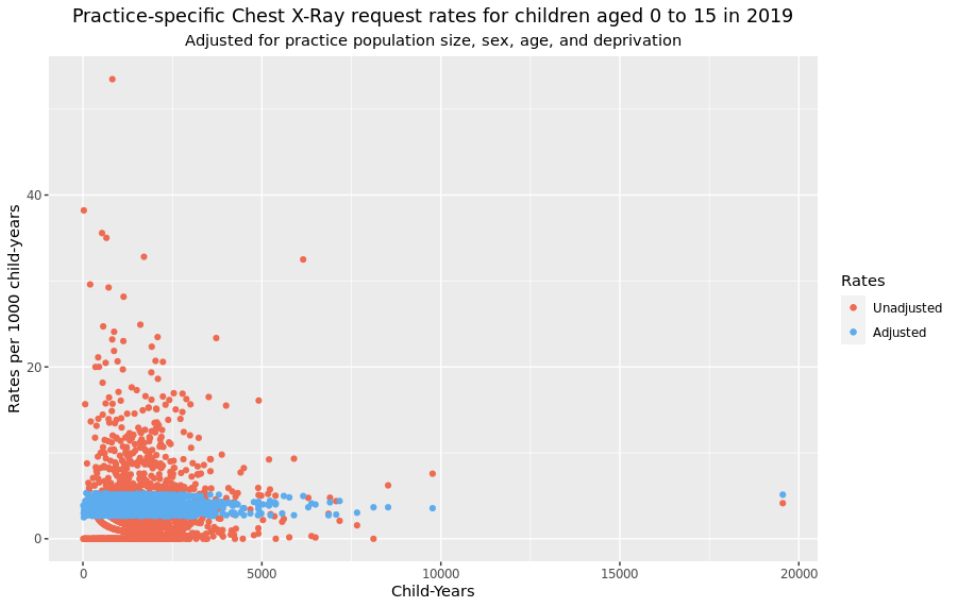

R-squared = 0.34
FeNO
Hearing test
Vitamin D
Monospot
Helicobacter test
Urine MCS
CoV (%)
AAPC (%)
125
100
75
50
25
0
0
10
20
30
40

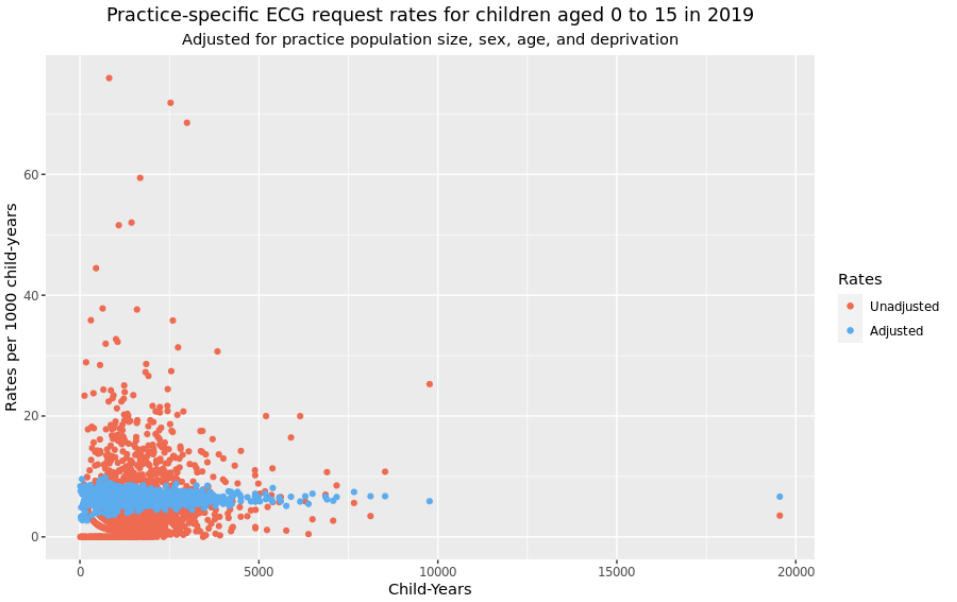
**Appendix Figure 4.7 Unadjusted and adjusted paediatric test rates from each practice for specific tests in 2019.**

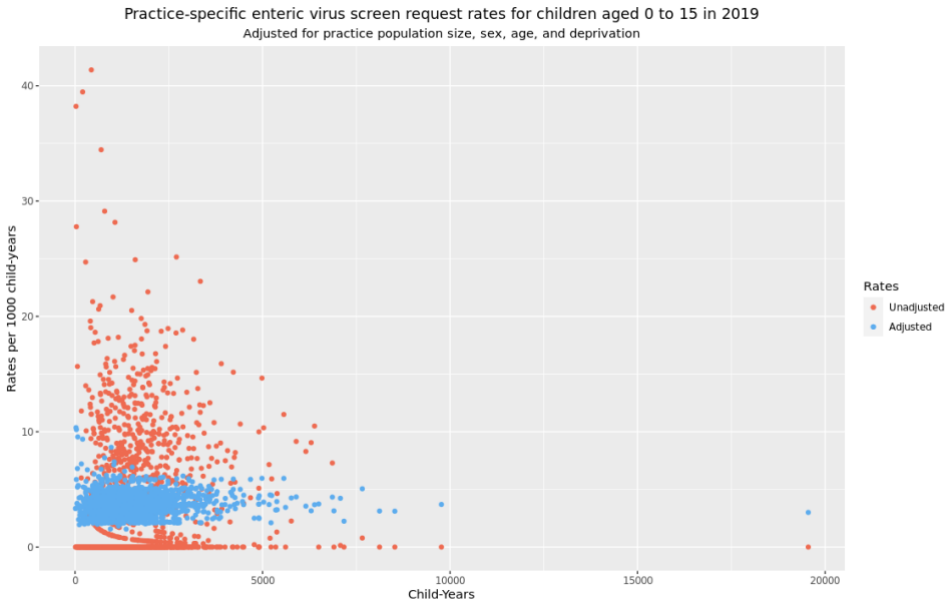


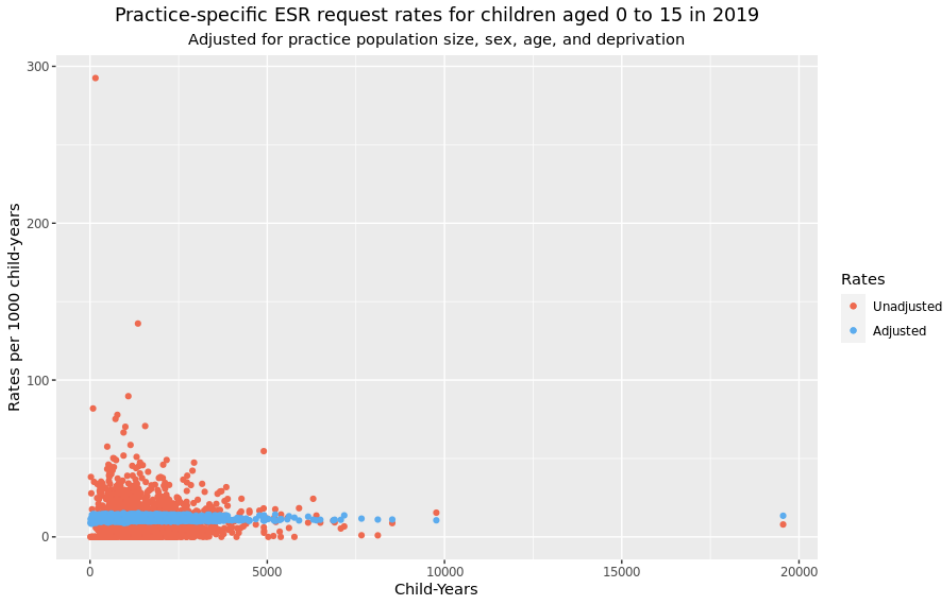

Practice-specific C reactive protein request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
150
100
50
0
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted

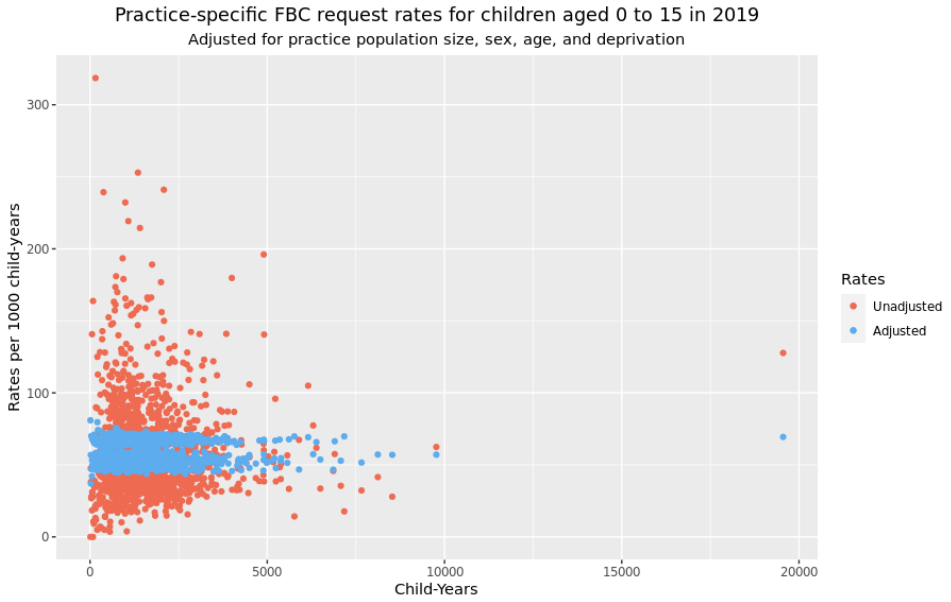

Practice-specific CT head request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
10
5
0
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted

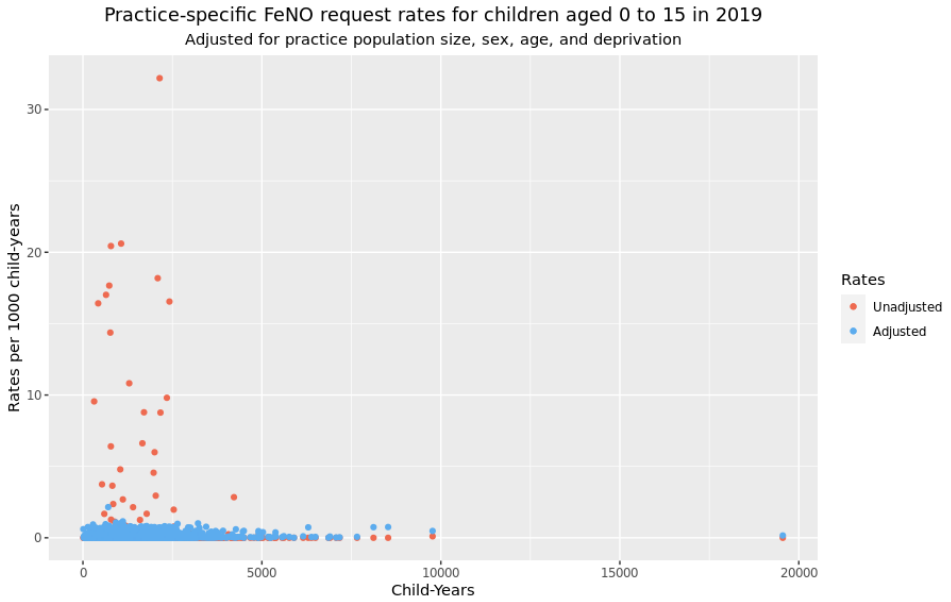

Practice-specific Chest X-Ray request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
40
20
0
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted

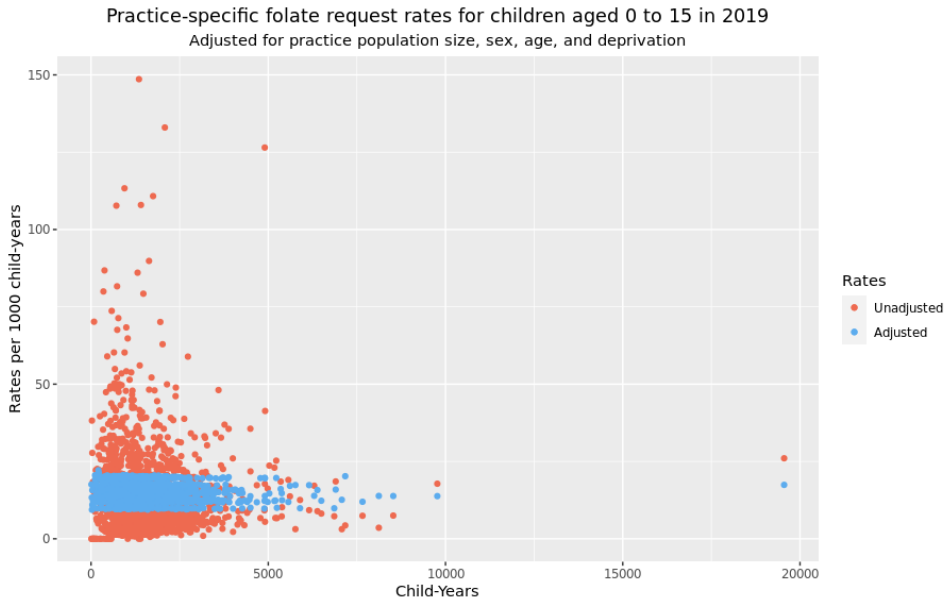

Practice-specific ECG request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
0
20
40
60
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted

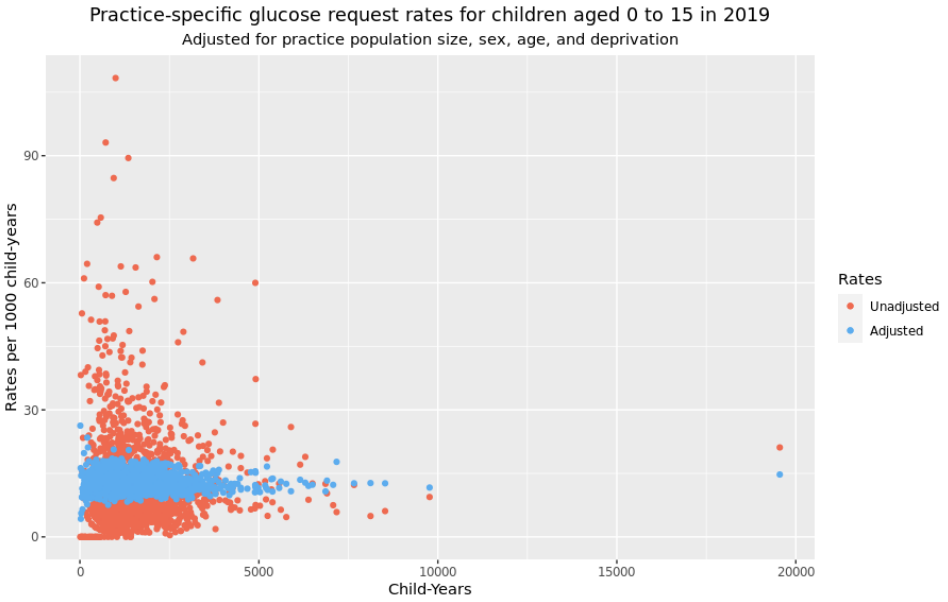

Practice-specific enteric virus screen request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
0
10
20
30
40
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted

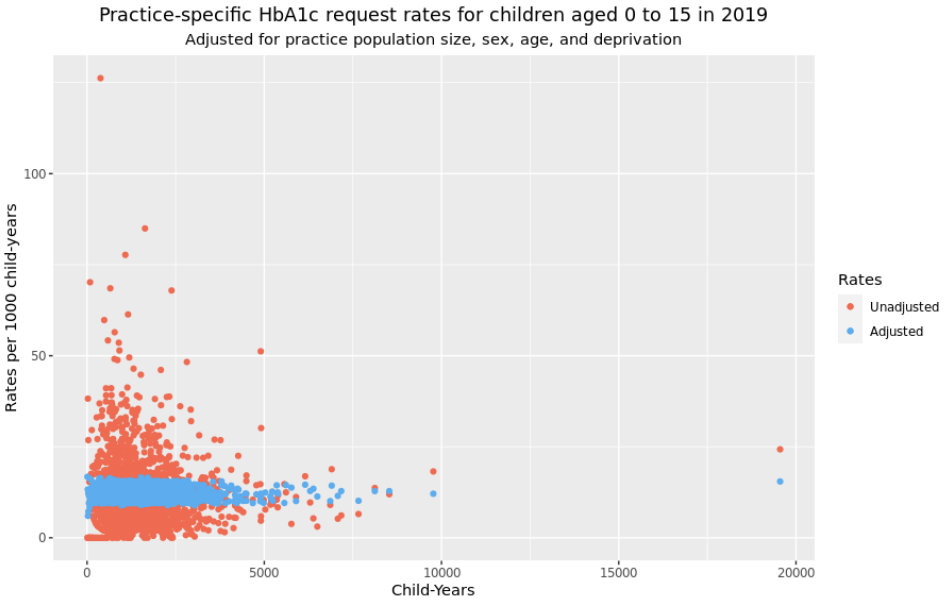

Practice-specific ESR request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
0
100
200
300
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted

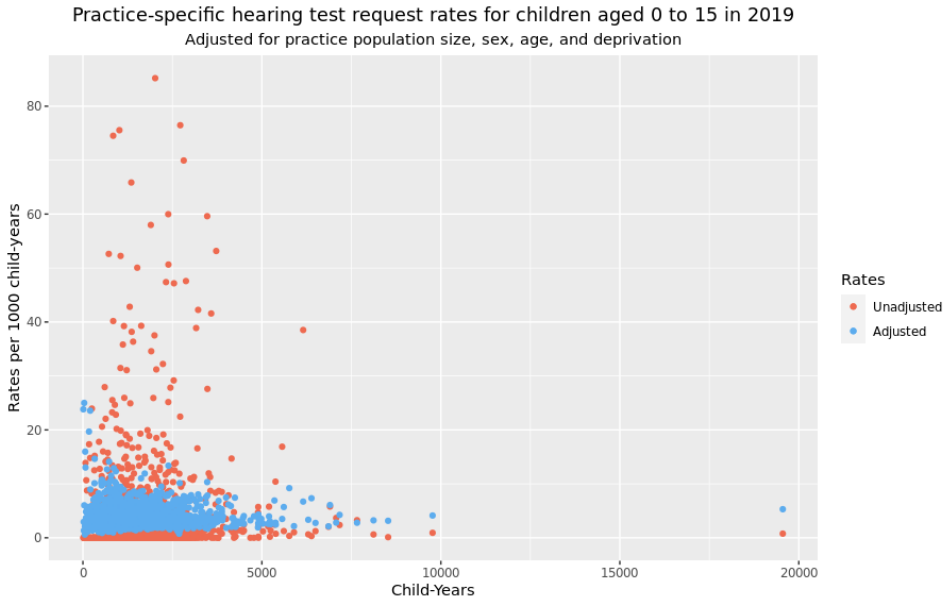

Practice-specific FBC request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
300
200
100
0
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted

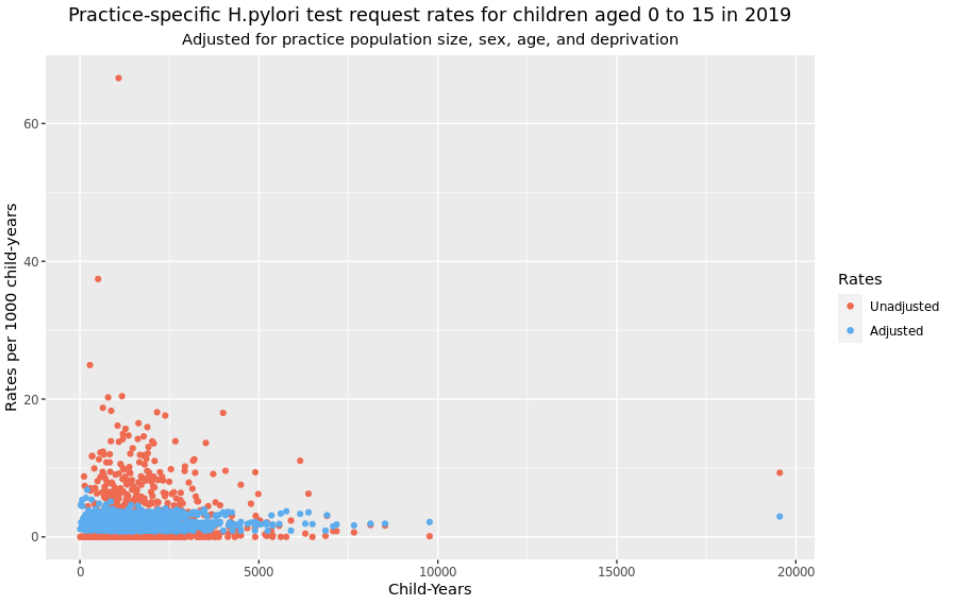

Practice-specific FeNO request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
30
20
10
0
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted

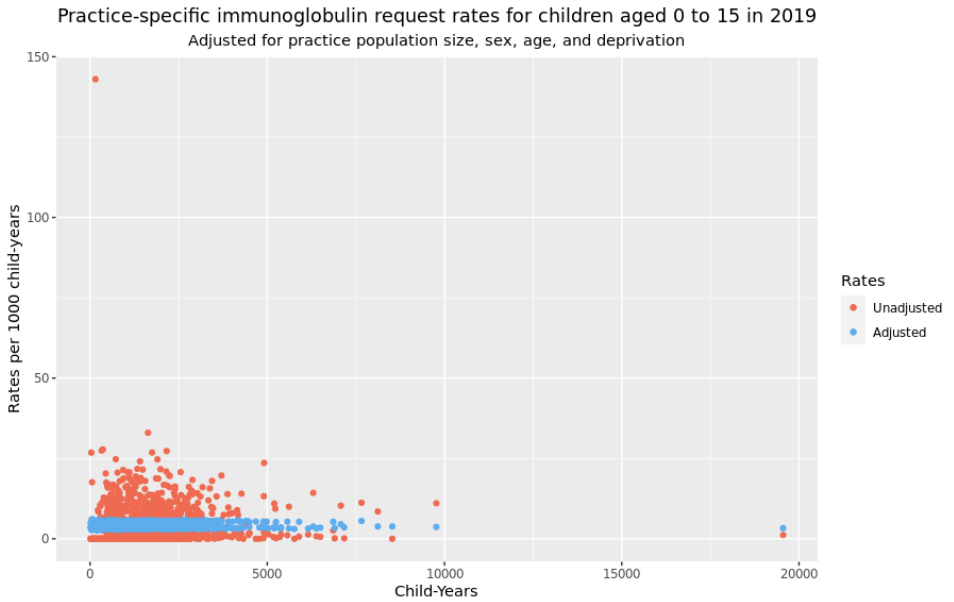

Practice-specific folate request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
150
100
50
0
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted

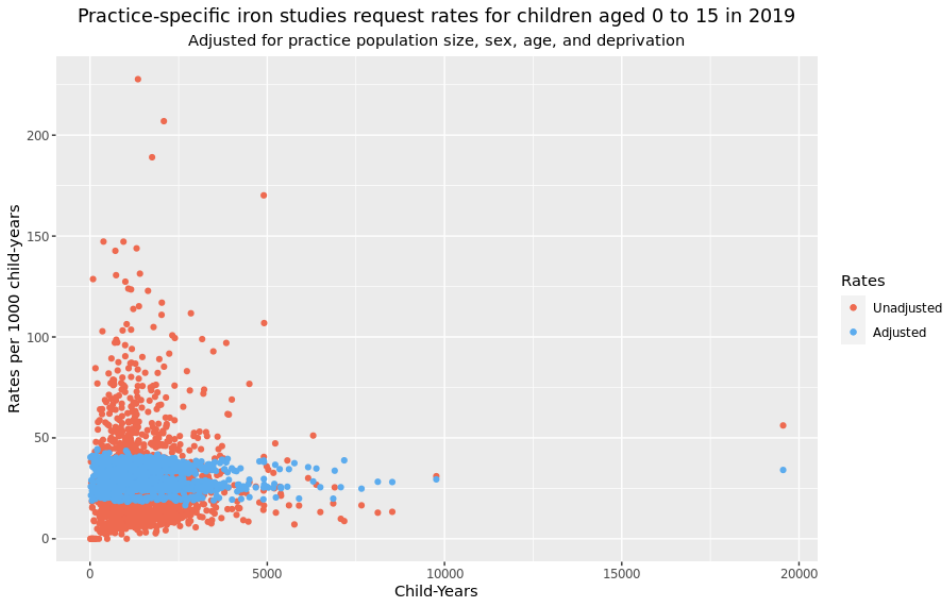

Practice-specific glucose request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
0
30
60
90
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted

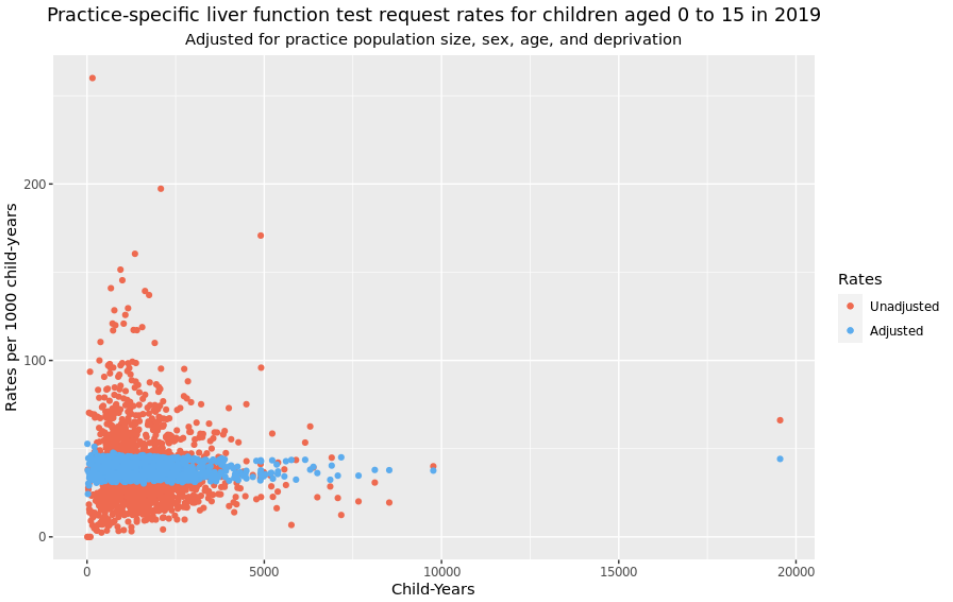

Practice-specific HbA1c request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
0
50
100
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted

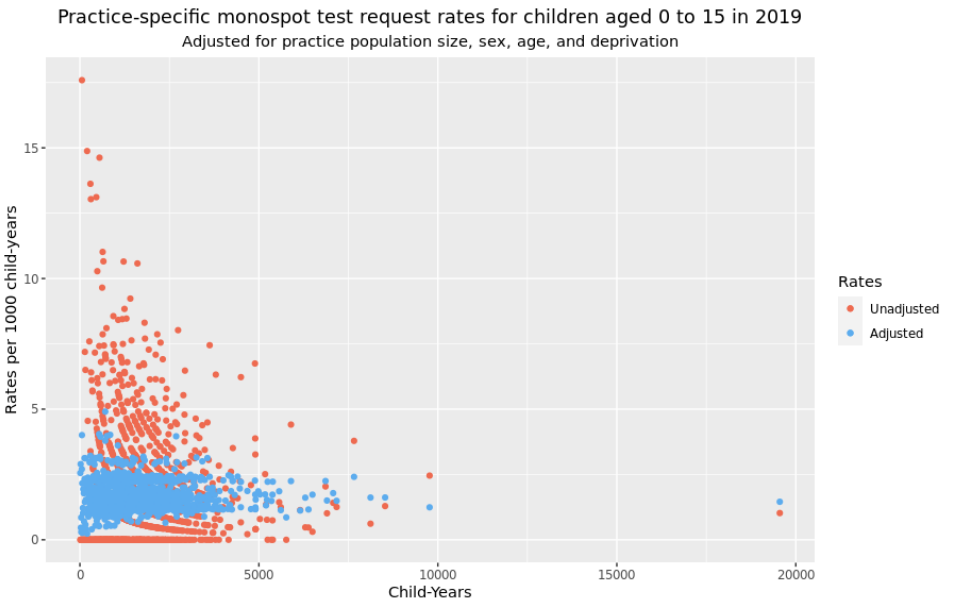

Practice-specific hearing test request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
0
20
40
60
80
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted

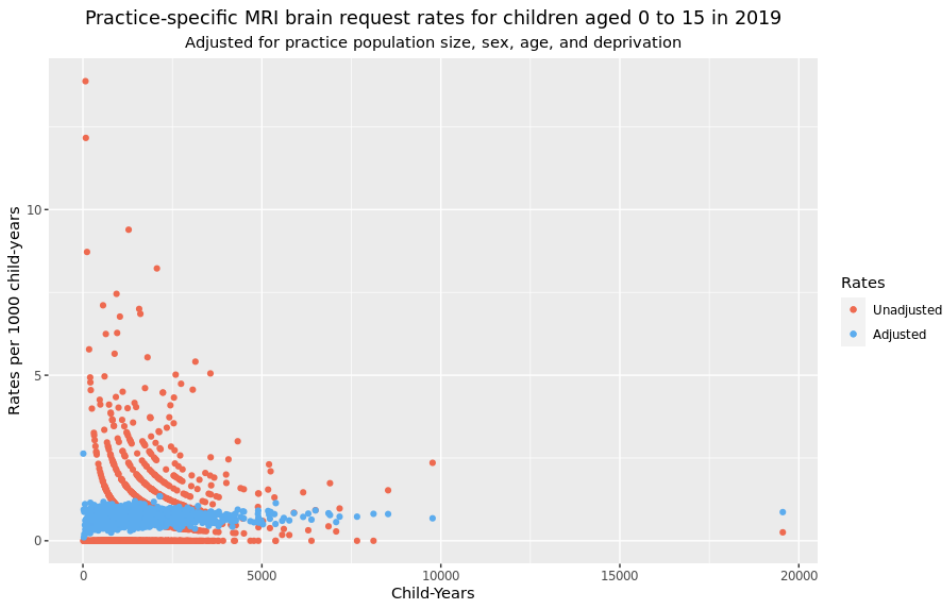

Practice-specific H.pylori test request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
0
20
40
60
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted

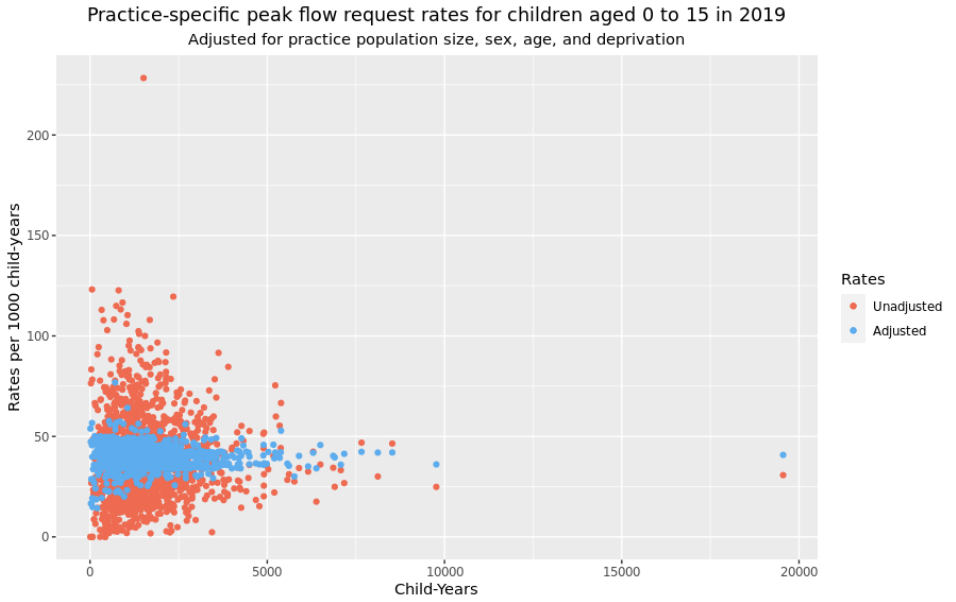

Practice-specific immunoglobulin request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
0
50
100
150
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted

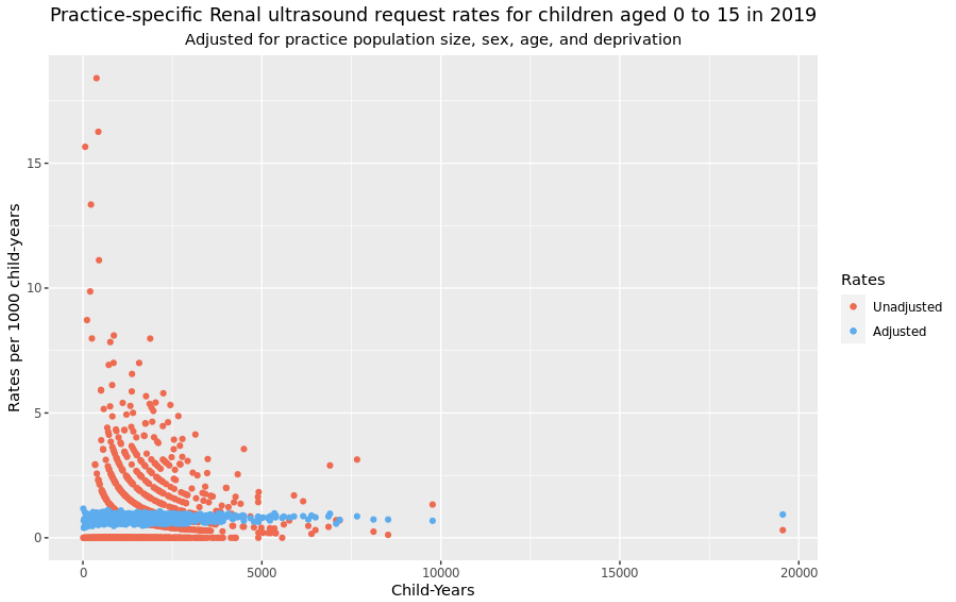

Practice-specific iron studies request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
0
50
100
150
200
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted

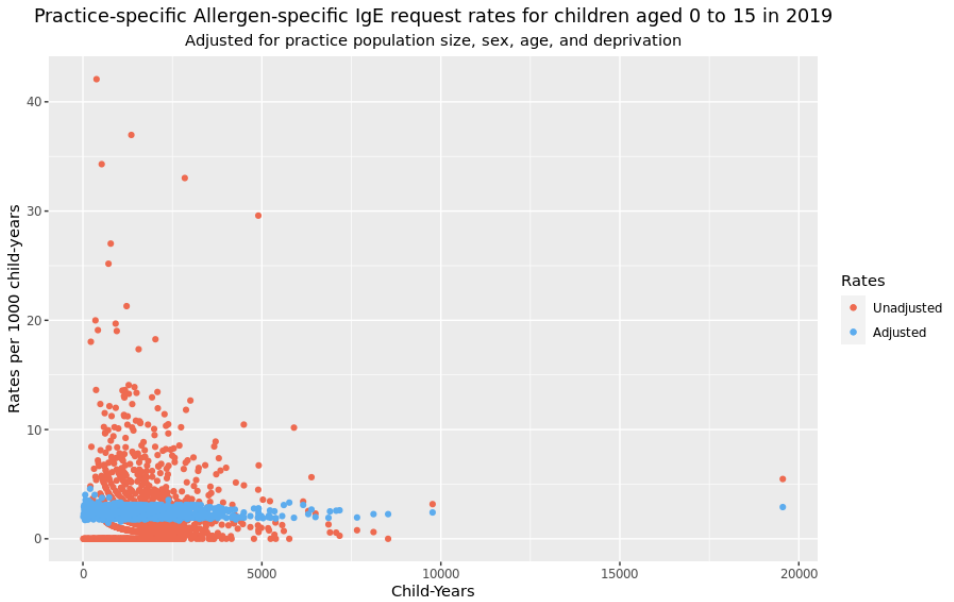

Practice-specific liver function test request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
200
100
0
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted

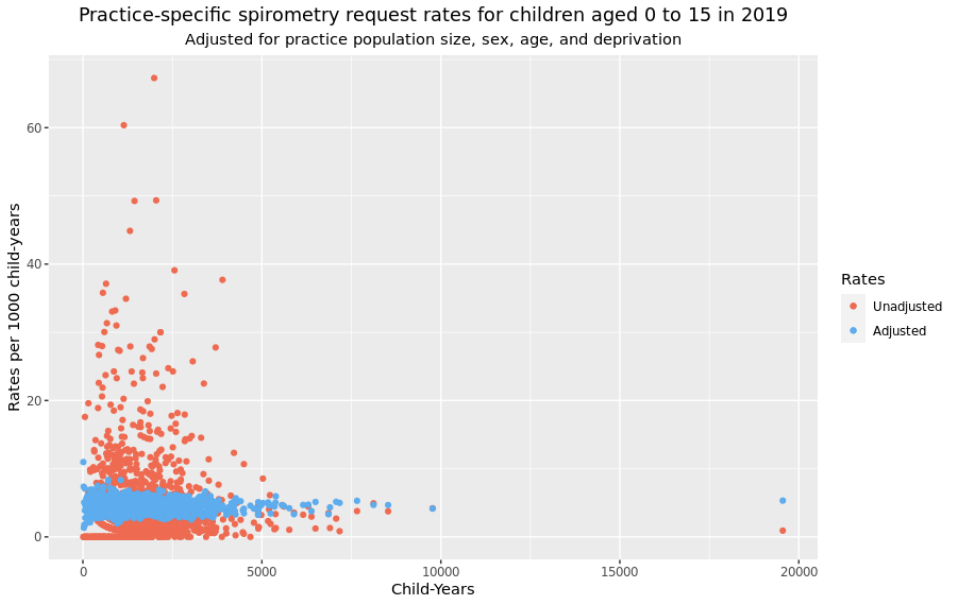

Practice-specific monospot test request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
15
10
5
0
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted

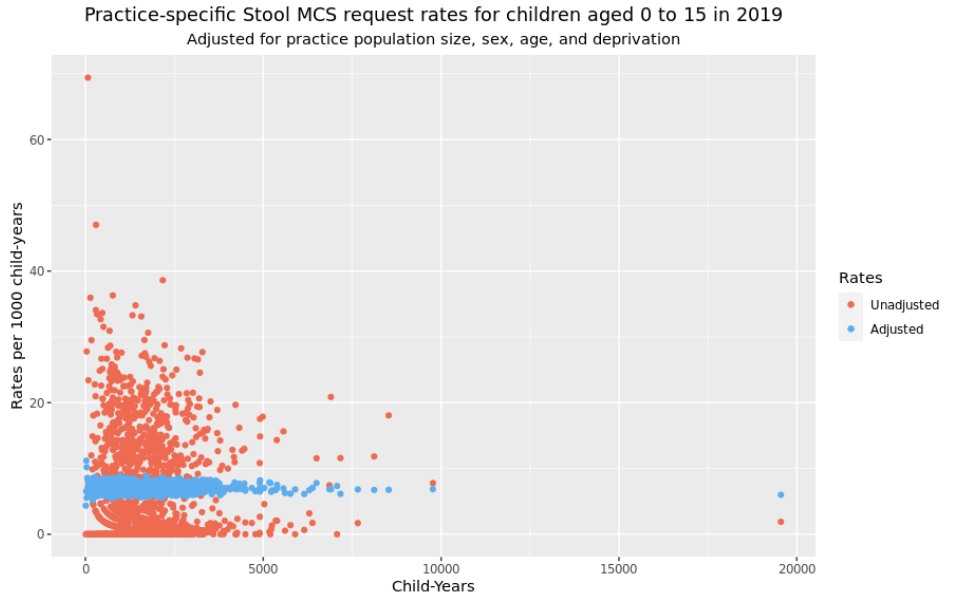

Practice-specific MRI brain request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
10
5
0
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted

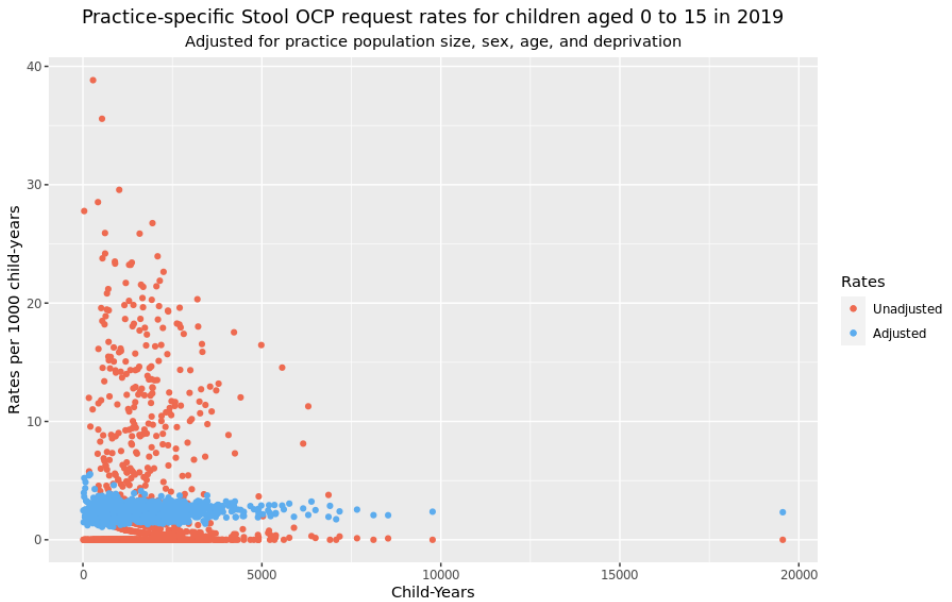

Practice-specific peak flow request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
Child-Years
Rates
Unadjusted
Adjusted

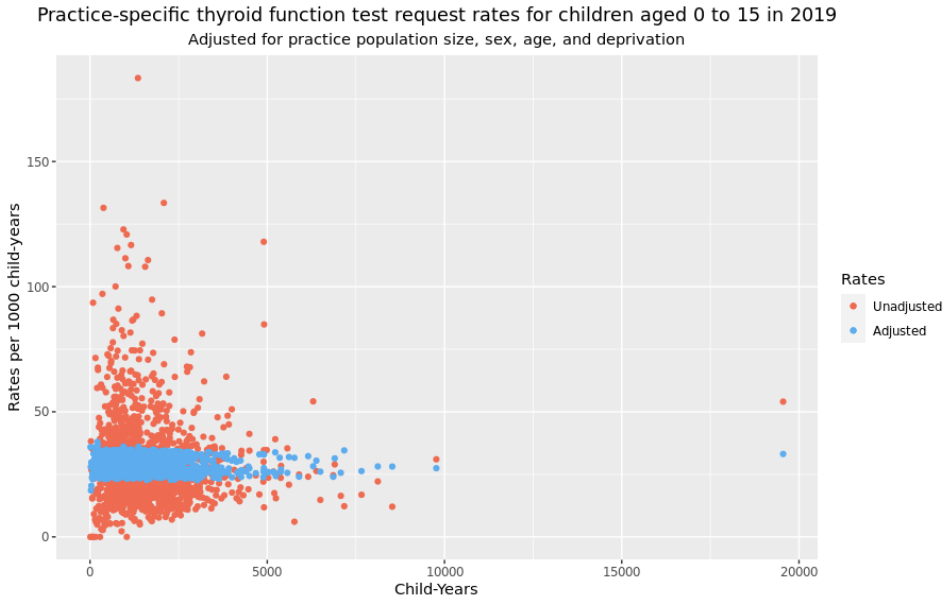

Practice-specific Renal ultrasound request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
Child-Years
Rates
Unadjusted
Adjusted

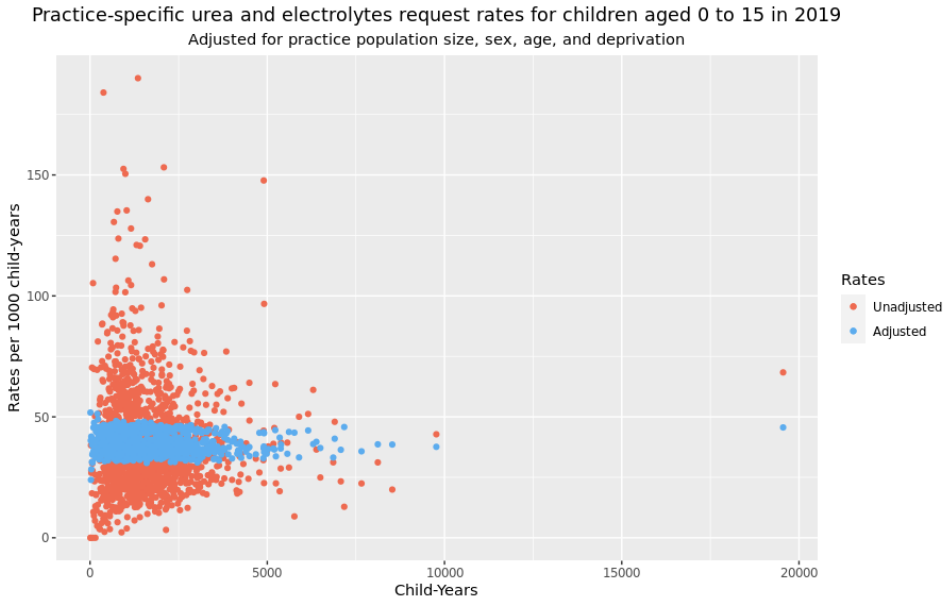

Practice-specific Allergen-specific IgE request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
Child-Years
Rates
Unadjusted
Adjusted

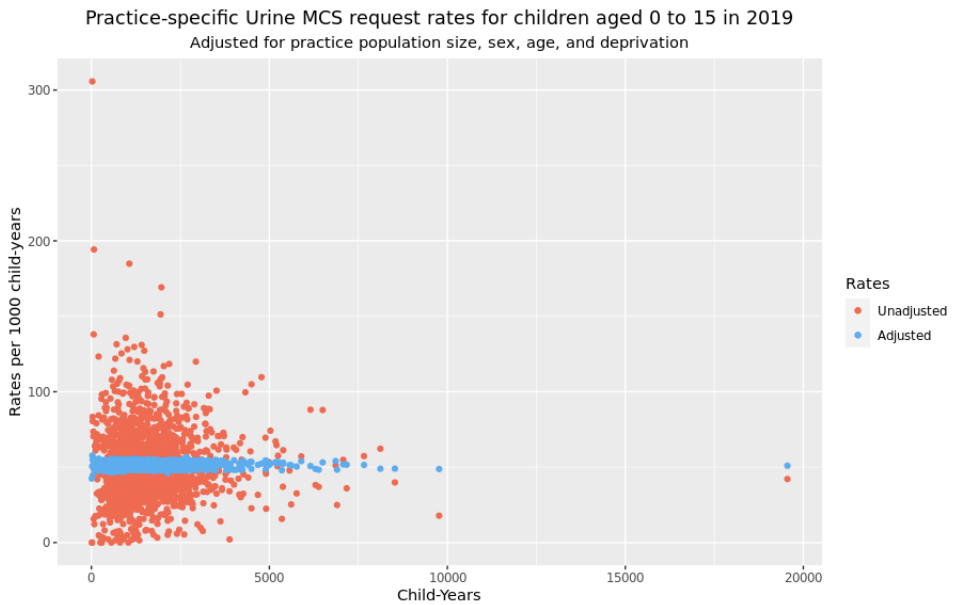

Practice-specific spirometry request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
0
20
40
60
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted

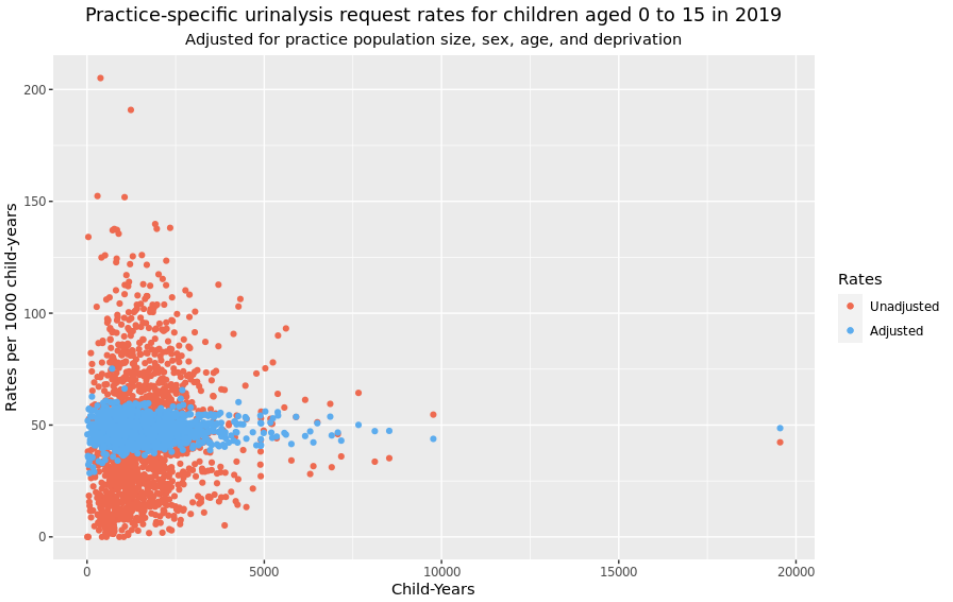

Practice-specific Stool MCS request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
0
20
40
60
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted

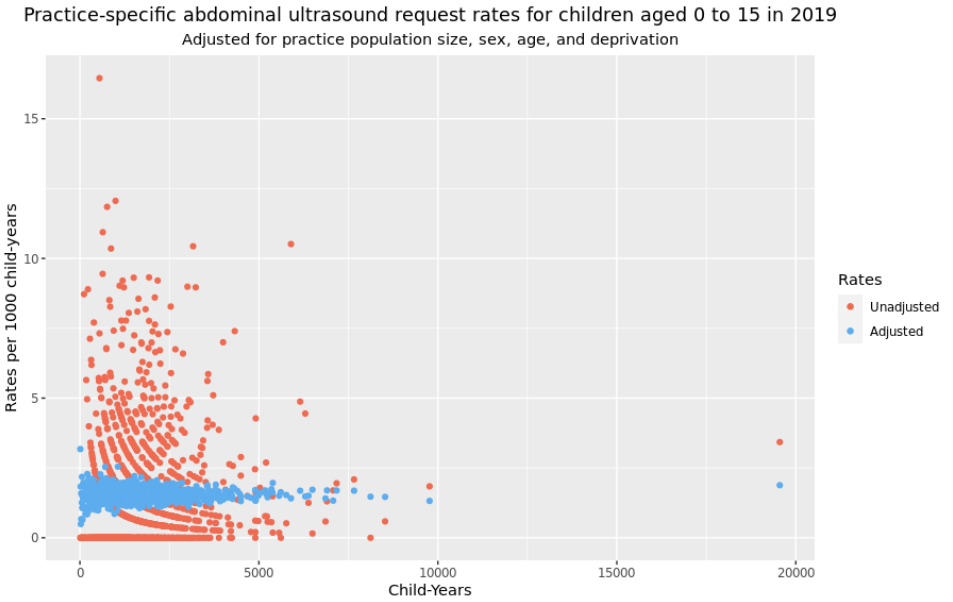

Practice-specific Stool OCP request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
0
10
20
30
40
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted

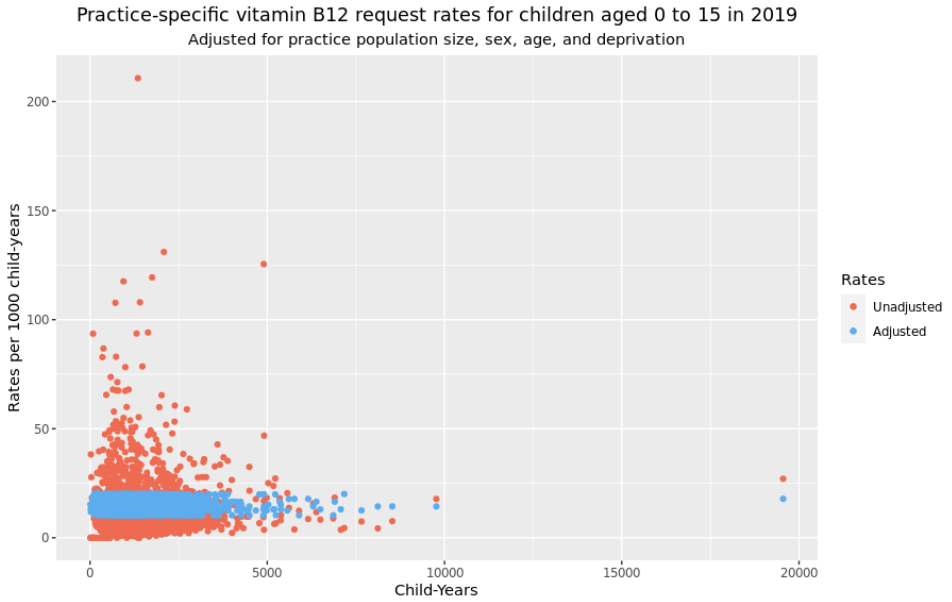

Practice-specific thyroid function test request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
0
50
100
150
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted

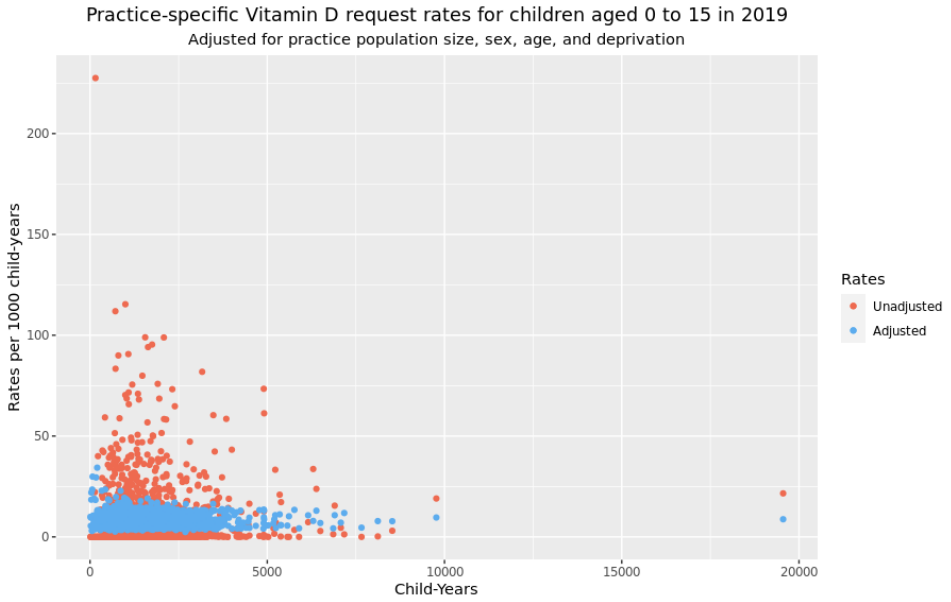

Practice-specific urea and electrolytes request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
0
50
100
150
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted

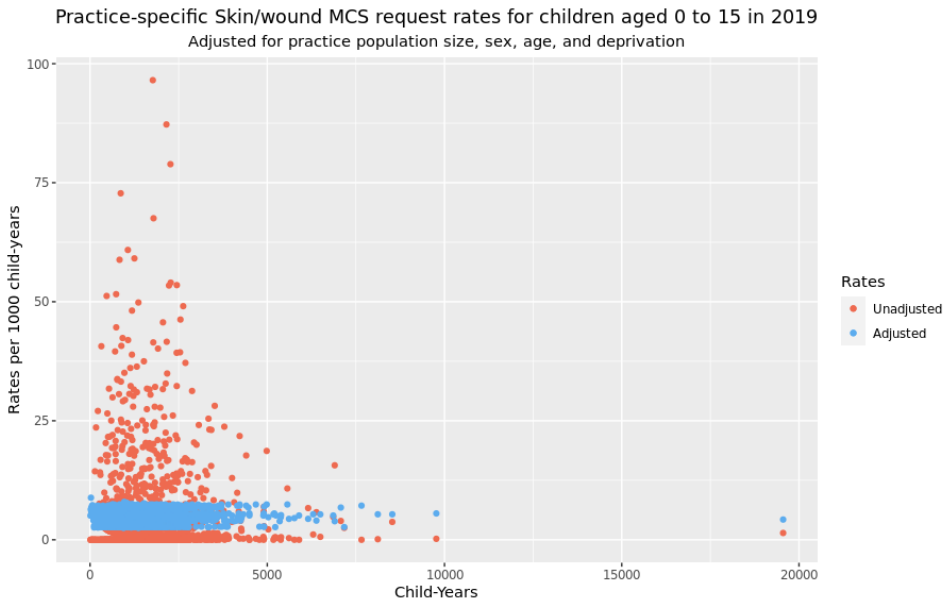

Practice-specific Urine MCS request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
0
100
200
300
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted


Practice-specific urinalysis request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
Child-Years
Rates
Unadjusted
Adjusted


Practice-specific abdominal ultrasound request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
Child-Years
Rates
Unadjusted
Adjusted


Practice-specific vitamin B12 request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
Child-Years
Rates
Unadjusted
Adjusted


Practice-specific Vitamin D request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
200
150
100
50
0
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted


Practice-specific Skin/wound MCS request rates for children aged 0 to 15 in 2019
Adjusted for practice population size, sex, age, and deprivation
Rates per 1000 child-years
100
75
50
25
0
0
5000
10000
15000
20000
Child-Years
Rates
Unadjusted
Adjusted

## Chapter 5 Appendix

## Appendix 5.1 Interview Topic guide

**Descriptive/demographic questions**

- Are you a GP partner/salaried GP or locum?
- On average, how many days a week do you work?
- Locum – areas practised?
- How would you describe the deprivation level of the population your practice serves? Low (least deprived), Low-moderate, moderate, moderate-high, high (most deprived)
- What proportion of your patients come from ethnic minority backgrounds? (<10%, 10-19%, 20-29%, 30% or more)
- On average how many patients might you see per day?
- <30
- 30 to 49
- >50
- How many GPs (including you) work at your practice?
- How many nurses work at your practice?

**Understanding the paediatric workload in general practice**

- Can you tell me about the children you might see in a typical working week?
- How many children do you see and what types of conditions do they have?
- Of the proportion of children you see, how often would you consider requesting a test? Urine test, blood test, swab, imaging, other types of special tests
- What types of tests would you most commonly order?
- The process of requesting tests tend to differ by practice, could you talk me through how you request a diagnostic test for a child and how it works in your practice?

**Understanding the factors that influence testing**

- What factors determine your choice to request a diagnostic test for a child rather than not request the test?
- What factors determine your choice to request certain tests in primary care compared to referring a child to secondary care?
- How do you think GPs in general feel about requesting tests for children?
- (If uncertainty exists, why do you think there is uncertainty amongst general practitioners in requesting tests for children?)
- How do the testing behaviours of other GPs in your practice affect your own testing practices? i.e., do you perceive there a “culture” of doing tests/not performing tests within a practice?
- Did you do a paediatrics term during your training? How did this impact on your confidence in managing childhood presentations?
- Structures – local pathways, guidance, proformas, types of consultations? How do these impact your testing decisions?

**Exploring the drivers of variation**

- Do you think testing trends in general practice have changed over time? Do you have any thoughts on why / why not?
- Has your own diagnostic approach in children changed over the course of your time in general practice?
- Have there been things that have happened to you / your practice / elsewhere that have influenced these changes?
- Are there certain tests that you believe to have more variation amongst general practitioners, why do you believe there is more variation with these tests?
- What do you perceive to be the potential harms of inappropriately requesting a test?

## Chapter 6 Appendix

## Appendix 6.1 Search strategy

| | Search Strategy | Results |
|---|---|---|
| 1 | adolescent/ or exp child/ or exp infant/ | 3884889 |
| 2 | Pediatrics/ | 57642 |
| 3 | (infan* or newborn? or neonate? or baby or babies or child* or preschool* or pre-school* or toddler* or schoolchild* or "school age" or boys or girls or adolescen* or teen* or p?ediatric* or pe?diatric* or youth?).ti,ab,kw. | 2527407 |
| 4 | 1 or 2 or 3 | 4565052 |
| 5 | Constipation/ | 15561 |
| 6 | (constipat* or stool? or (bowel adj2 (movement? or problem? or disorder?))).ti. | 14065 |
| 7 | gastroenteritis/ or enteritis/ or gastritis/ | 43608 |
| 8 | exp diarrhea/ or vomiting/ | 78953 |
| 9 | (gastroenteritis or gastro-enteritis or gastritis or enteritis or diarrh?ea* or vomit*).ti. | 63685 |
| 10 | fever/ or "fever of unknown origin"/ | 46661 |
| 11 | exp Sepsis/ | 137512 |
| 12 | (fever or febrile or high temperature?).ti. | 81346 |
| 13 | (sepsis or septic?emia or septic shock or bacter?emia or bacterial infection?).ti. | 72991 |
| 14 | 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 | 411426 |
| 15 | 4 and 14 | 117257 |
| 16 | afghanistan/ or exp africa/ or albania/ or andorra/ or antarctic regions/ or argentina/ or exp asia, central/ or exp asia, northern/ or exp asia, southeastern/ or exp atlantic islands/ or bahrain/ or bangladesh/ or bhutan/ or bolivia/ or borneo/ or "bosnia and herzegovina"/ or brazil/ or bulgaria/ or exp central america/ or exp china/ or colombia/ or "commonwealth of independent states"/ or croatia/ or "democratic people's republic of korea"/ or ecuador/ or gibraltar/ or guyana/ or exp india/ or indonesia/ or iran/ or iraq/ or jordan/ or kosovo/ or kuwait/ or lebanon/ or liechtenstein/ or macau/ or "macedonia (republic)"/ or exp melanesia/ or moldova/ or monaco/ or mongolia/ or montenegro/ or nepal/ or netherlands antilles/ or new guinea/ or oman/ or pakistan/ or paraguay/ or peru/ or philippines/ or qatar/ or "republic of belarus"/ or romania/ or exp russia/ or saudi arabia/ or serbia/ or sri lanka/ or suriname/ or syria/ or taiwan/ or exp transcaucasia/ or ukraine/ or uruguay/ or united arab emirates/ or exp ussr/ or venezuela/ or yemen/ | 1291482 |
| 17 | organisation for economic co-operation and development/ | 471 |
| 18 | australasia/ or exp australia/ or austria/ or exp baltic states/ or belgium/ or exp canada/ or chile/ or czech republic/ or europe/ or exp france/ or exp germany/ or greece/ or hungary/ or ireland/ or israel/ or exp italy/ or exp japan/ or korea/ or luxembourg/ or mexico/ or netherlands/ or new zealand/ or north america/ or poland/ or portugal/ or exp "republic of korea"/ or exp "scandinavian and nordic countries"/ or slovakia/ or slovenia/ or spain/ or switzerland/ or turkey/ or exp united kingdom/ or exp united states/ | 3422984 |
| 19 | european union/ | 17364 |
| 20 | developed countries/ | 21220 |
| 21 | 17 or 18 or 19 or 20 | 3438594 |
| 22 | 16 not 21 | 1200998 |
| 23 | 15 not 22 | 98142 |
| 24 | limit 23 to (guideline or practice guideline) | 236 |
| 25 | limit 24 to yr="2011 -Current" | 102 |

## Chapter 7 Appendices

## Appendix 7.1 Search strategy

| | Search strategy | Results |
|---|---|---|
| 1 | adolescent/ or exp child/ or exp infant/ | 3880783 |
| 2 | Pediatrics/ | 57617 |
| 3 | (infan* or newborn? or neonate? or baby or babies or child* or preschool* or pre-school* or toddler* or schoolchild* or "school age" or boys or girls or adolescen* or teen* or p?ediatric* or pe?diatric* or youth?).ti,ab,kw. | 2523247 |
| 4 | 1 or 2 or 3 | 4560292 |
| 5 | asthma/ or asthma, exercise-induced/ or status asthmaticus/ | 137847 |
| 6 | asthma*.ti. | 103557 |
| 7 | 5 or 6 | 147595 |
| 8 | 4 and 7 | 59492 |
| 9 | afghanistan/ or exp africa/ or albania/ or andorra/ or antarctic regions/ or argentina/ or exp asia, central/ or exp asia, northern/ or exp asia, southeastern/ or exp atlantic islands/ or bahrain/ or bangladesh/ or bhutan/ or bolivia/ or borneo/ or "bosnia and herzegovina"/ or brazil/ or bulgaria/ or exp central america/ or exp china/ or colombia/ or "commonwealth of independent states"/ or croatia/ or "democratic people's republic of korea"/ or ecuador/ or gibraltar/ or guyana/ or exp india/ or indonesia/ or iran/ or iraq/ or jordan/ or kosovo/ or kuwait/ or lebanon/ or liechtenstein/ or macau/ or "macedonia (republic)"/ or exp melanesia/ or moldova/ or monaco/ or mongolia/ or montenegro/ or nepal/ or netherlands antilles/ or new guinea/ or oman/ or pakistan/ or paraguay/ or peru/ or philippines/ or qatar/ or "republic of belarus"/ or romania/ or exp russia/ or saudi arabia/ or serbia/ or sri lanka/ or suriname/ or syria/ or taiwan/ or exp transcaucasia/ or ukraine/ or uruguay/ or united arab emirates/ or exp ussr/ or venezuela/ or yemen/ | 1289305 |
| 10 | "organisation for economic co-operation and development"/ | 466 |
| 11 | australasia/ or exp australia/ or austria/ or exp baltic states/ or belgium/ or exp canada/ or chile/ or czech republic/ or europe/ or exp france/ or exp germany/ or greece/ or hungary/ or ireland/ or israel/ or exp italy/ or exp japan/ or korea/ or luxembourg/ or mexico/ or netherlands/ or new zealand/ or north america/ or poland/ or portugal/ or exp "republic of korea"/ or exp "scandinavian and nordic countries"/ or slovakia/ or slovenia/ or spain/ or switzerland/ or turkey/ or exp united kingdom/ or exp united states/ | 3420247 |
| 12 | european union/ | 17347 |
| 13 | developed countries/ | 21214 |
| 14 | 10 or 11 or 12 or 13 | 34535838 |
| 15 | 9 not 14 | 1198859 |
| 16 | 8 not 15 | 56272 |
| 17 | limit 16 to (guideline or practice guideline) | 170 |
| 18 | limit 17 to yr="2011 -Current" | 44 |

## Appendix 7.2 AGREE II tool

**NICE guideline: Diagnosis of Asthma**

| Domain | Item | AGREE II Rating | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 *Strongly Disagree* | 2 | 3 | 4 | 5 | 6 | 7 *Strongly Agree* |
| Scope and purpose | 1. The overall objective(s) of the guideline is (are) specifically described. | | | | | | | ✓ |
| | 2. The health question(s) covered by the guideline is (are) specifically described. | | | | | | ✓ | |
| | 3. The population (patients, public, etc.) to whom the guideline is meant to apply is specifically described. | | | | | | | ✓ |
| Stakeholder involvement | 4. The guideline development group includes individuals from all the relevant professional groups. | | | | | ✓ | | |
| | 5. The views and preferences of the target population (patients, public, etc.) have been sought. | | | ✓ | | | | |
| | 6. The target users of the guideline are clearly defined. | | | | | | | ✓ |
| Rigor of development | 7. Systematic methods were used to search for evidence. | | | | | | | ✓ |
| | 8. The criteria for selecting the evidence are clearly described. | | | | | | ✓ | |
| | 9. The strengths and limitations of the body of evidence are clearly described. | | | | | | | ✓ |
| | 10. The methods for formulating the recommendations are clearly described. | | | | | | ✓ | |
| | 11. The health benefits, side effects and risks have been considered in formulating the recommendations. | | | | | | | ✓ |
| | 12. There is an explicit link between the recommendations and the supporting evidence. | | | | | | | ✓ |
| | 13. The guideline has been externally reviewed by experts prior to its publication. | | | ✓ | | | | |
| | 14. A procedure for updating the guideline is provided. | | | | | | | ✓ |
| Clarity of presentation | 15. The recommendations are specific and unambiguous. | | | | | | | ✓ |
| | 16. The different options for management of the condition or health issue are clearly presented. | | | | | | | NA |
| | 17. Key recommendations are easily identifiable. | | | | | | | ✓ |
| Applicability | 18. The guideline describes facilitators and barriers to its application. | | | | | | ✓ | |
| | 19. The guideline provides advice and/or tools on how the recommendations can be put into practice. | | | | | ✓ | | |
| | 20. The potential resource implications of applying the recommendations have been considered. | | | | | | | ✓ |
| | 21. The guideline presents monitoring and/ or auditing criteria. | | | | | | ✓ | |
| Editorial independence | 22. The views of the funding body have not influenced the content of the guideline. | | | ✓ | | | | |
| | 23. Competing interests of guideline development group members have been recorded and addressed. | | | | | | ✓ | |
| Overall Guideline Assessment | 1. Rate the overall quality of this guideline. | 1 *Lowest possible quality* | 2 | 3 | 4 | 5 | (6) | 7 *Highest possible quality* |
| Overall Guideline Assessment | 2. I would recommend this guideline for use. | *Yes* | *Yes, with modifications* | | | | | *No* |

## Appendix 7.3 Calculation of AGREE II Aggregate Score

As outlined in the AGREE II tool, domain scores are calculated by cumulating the scores for individual items in a domain and scaling the total as a percentage of the maximum possible score for that domain.

For example, using the NICE guideline for asthma.

**Domain 5 (Applicability) Scores:**

| | Item 1 (out of 7) | Item 2 (out of 7) | Item 3 (out of 7) | Item 4 (out of 7) | **Total** |
|---|---|---|---|---|---|
| Appraiser 1 | 6 | 5 | 7 | 6 | **24** |
| Appraiser 2 | 5 | 7 | 7 | 6 | **25** |
| **Total** | **11** | **12** | **14** | **12** | **49** |

Maximum possible score = 7 x 4 (items) x 2 (appraisers) = 56

Minimum possible score = 1 x 4 (items) x 2 (appraisers) = 8

Scaled domain score = obtained score - minimum possible score

$$\boldsymbol{Scaled\ domain\ score} = \frac{\boldsymbol{obtained\ score - minimum\ possible\ score}}{\boldsymbol{maximum\ possible\ score - minimum\ possible\ score}} \text{ x } 100$$

$$\boldsymbol{Scaled\ domain\ score} = \frac{\boldsymbol{49 - 8}}{\boldsymbol{56 - 8}} \text{ x } 100$$

Scaled domain score = 85.4%

## Appendix 7.4 Selection of asthma diagnostic tests relevant to children in primary care

### Tests that can be routinely requested or performed in primary care

1. Spirometry
2. Bronchodilator reversibility testing
3. Peak expiratory flow
4. Serum immunoglobulins
5. Blood eosinophilia
6. Chest XR
7. Sputum microbiology

### Tests that are more likely to be requested or performed in secondary care/specialist settings i.e., wouldn't routinely be requested or performed by a GP

1. Impulse oscillometry
2. Residual volume measurements
3. Skin allergy tests
4. Sputum eosinophils
5. Fractional exhaled nitric oxide
6. Chest CT
7. Direct bronchoprovocation testing (methacholine or histamine challenge test)
8. Indirect bronchoprovocation testing (exercise or mannitol challenge test)
9. Specific airways resistance
10. Airway wall biopsy
11. Bronchoalveolar lavage
12. Sweat test
13. Genotyping for cystic fibrosis
14. Nasal biopsy