PhD Thesis

Division of Clinical Medicine

School of Medicine and Population Health

Faculty of Health

Name of Candidate:

Neil Richard Lawrence

Title:

Development of a clinical decision support tool for monitoring and treatment of children with Congenital Adrenal Hyperplasia

Year of submission:

2025

Name of Supervisor:	**Name of Co-Supervisors:**
Professor Nils Krone	Jeremy Dawson / Zi Qiang Lang / Gary Collins (Birmingham)

Contents

Acknowledgements

I would like to take this opportunity to thank my supervisory team, my co-authors and the examiners who have volunteered to assess this thesis for their engagement with this project and overall support.

I thank the patients and families who have contributed data and guidance through public involvement in this research, without whom it would have not been possible.

Most importantly I thank my family, for their never-ending support.

Publications and Contribution of others

A large proportion of this thesis has been published in international peer reviewed journal articles. The author has also published and contributed to other peer reviewed journal articles within the PhD funding period that are not reproduced within this thesis.

All of the work presented has been originally produced and written by the author. Published chapters have received feedback and advice from the supervisory team and co-authors, that has led to alterations in wording and explanation. Further details about the contribution of co-authors to each manuscript is contained within the relevant chapters.

This thesis is not a 'PhD by publication', but is a 'publication format thesis', in line with the guidance in section 1.4 of 'Guide to the thesis examination process for candidates of research degree programmes', University of Sheffield Research, Partnerships & Innovation (1). Each chapter that has been published also contains a 'thesis contribution' section that is not contained within the journal article publications, to help inform the reader about the contribution of each chapter to the overall aims and objectives.

Thesis Summary

Introduction

Congenital adrenal hyperplasia (CAH) is an autosomal recessive condition caused in over 90% of cases by 21-hydroxylase deficiency (21OHD), that leads to cortisol deficiency. This thesis applies longitudinal statistical modelling to gain insights into clinical trial and international registry data from patients with CAH, to inform and develop a prototype clinical decision support tool that supports clinicians managing children and young people with CAH.

Methods

Contemporary models that predict height or weight were systematically reviewed. A nationwide UK service evaluation quantified the satisfaction of services provided for children with CAH. Longitudinal multilevel modelling was used to assess biomarkers, blood pressure and growth in patients with CAH. A web application was developed in a browser interface to exhibit a prototype clinical decision support tool. All analysis was carried out in *R* (https://www.R-project.org/).

Results

Models developed (n=180) or validated (n=61) to predict height or weight in 2022 were at significant risk of bias. Questionnaires from 229 respondents showed satisfaction with services for CAH in the UK is high, although families would appreciate further education about the condition. Spline modelling of 477 24-hour profiles of 17-Hydroxyprogesterone from 122 patients show they follow a similar pattern to those of androstenedione, with adults taking modified-release hydrocortisone exhibiting reduced variability in these markers throughout the day. Registry data from 554 children with 21OHD showed higher BP was common at younger ages, but the clinical significance of this remains uncertain. Analysis of the growth of 573 children with CAH showed an early adiposity rebound and blunted latest peak height velocity.

Conclusion

The prototype clinical decision support tool (https://endocrinology.shinyapps.io/cah_data_visualiser/) demonstrates how data visualisation and automated calculations of clinical relevance could be incorporated into registry data entry platforms, to make the process of data entry more attractive and useful for frontline clinicians.

Chapter 1 – Thesis Introduction

1.01 Congenital Adrenal Hyperplasia

Congenital adrenal hyperplasia (CAH) is an autosomal recessive condition caused by enzyme deficiencies that lead to glucocorticoid deficiency. Deficiency in the most prominent glucocorticoid, cortisol, results in overstimulation of the adrenal cortex and increased production of androgens such as testosterone. Excess androgens result in virilisation of affected females and ambiguous genitalia at birth. Different degrees of enzyme deficiency lead to a spectrum of CAH severity that extends towards life-threatening adrenal crisis and salt wasting. It is known that treatment is best tailored to the individual patient (2), but despite international guidelines available since 2002, the best way to titrate glucocorticoid replacement and to monitor this disease is unknown (3-5).

Approximately 1 in 15,000 people in the UK are affected by CAH, the commonest inherited cause of adrenal insufficiency (6), equating to 4000 people in the UK and half a million people worldwide. Overtreatment in CAH is associated with reduced adult height, fertility issues in both sexes and adverse metabolic, cardiovascular and bone health (2, 7, 8). Clinicians use blood tests to monitor CAH without precise guidance on how to interpret the results (5). Under-replacement of glucocorticoids in CAH can cause hypoglycaemic or hyponatraemic seizures (6), incentivising clinicians to treat patients conservatively with relatively high doses, despite the knowledge of longer-term consequences caused by overtreatment (2). There is no web application or software that can provide clinicians with appropriate individual patient data calculations or data visualisations that can assist management decisions. The International Congenital Adrenal Hyperplasia Registry (I-CAH Registry) provides a user-friendly web interface for the upload of patient data, but does not allow users to display individual patient data across different clinic visits to support clinical decision making.

International guidance involves reviewing patients with CAH every three to four months after diagnosis and throughout infancy, with reviews six monthly or yearly in older children. At each assessment, it is recommended to record height, weight, blood pressure, physical examination, and timings of the medication schedule. Bone age assessment is recommended annually, and serum biomarker measurement is dependent upon clinical judgement (5), shown in preliminary work to be recorded frequently, with 2/3 of visits in the I-CAH Registry cohort analysed reporting either androstenedione or 17-Hydroxyprogesterone (17OHP) (2, 9).

Alterations in growth trajectory and bone maturation have been studied previously in CAH. Reduction in adult height has been quantified through meta-analysis of over 1000 patients with CAH as averaging 10cm (10). The pubertal growth spurt has been shown as suppressed in CAH (11), and doses of glucocorticoids during the pubertal growth spurt negatively correlate with final adult height (11, 12). Bone age is used to detect abnormal advancement in relation to chronological age indicating ongoing exposure to adrenal androgens and undertreatment (2), with bone age advancement at eight years of age in children with CAH associated with adult height less than two standard deviations below the national average (11). However, these studies have focused on quantifying average differences in these metrics in CAH, often lacking in describing interindividual differences or the association of important covariates of treatment with those differences.

Studies in the UK and Sweden have shown an increased prevalence of obesity in adults with CAH (13-15), whilst adults with CAH in the United States of America (USA) and France have shown similar Body Mass Index (BMI) to the general population (16). A French study in children showed poor disease control in CAH associated with a higher BMI in adulthood (17). Children in the USA were more likely to be obese if they had CAH, and Glucocorticoid dose was shown to correlate with obesity (18). Obesity therefore occurs earlier in those living with CAH, likely contributing to their increased cardiovascular risk profile in later life (19). Whilst undertreatment with glucocorticoids and androgen excess in CAH can cause increased height velocity before puberty and reduced adult height (5), one can see a similar effect due to obesity in otherwise healthy children (20). The extent obesity contributes to the disordered height seen in CAH patients in relation to the effect of the disease itself is not well described.

Blood pressure in children with CAH is contentious, with several small studies reporting normal blood pressure values, other studies finding elevated values in systolic or diastolic readings, and others reporting that high blood pressure readings are a transient problem in early childhood that resolves (21). High doses of fludrocortisone in CAH to replace salt loss can cause hypertension, as can high doses of glucocorticoids (2, 22). The optimum dose of fludrocortisone has not been studied, and the extent to which hypertension in children with CAH correlates with adverse outcomes or co-morbidities is unknown (5).

The optimum level of serum biomarkers to aim for in CAH is disputed, culminating in international guideline advising not to normalise the serum cortisol precursor 17OHP as this indicates overtreatment, yet falling short of providing a target reference range. Recommended replacement of hydrocortisone is 10-15mg/m^2/day, with agreement that

different doses will be required between patients to maintain the same level of control (5). Preliminary work presented internationally has highlighted significant variation in the treatment of CAH, and large variability in serum biomarkers between different centres (9, 23).

In conclusion, glucocorticoid replacement in CAH requires titration of dosing to the individual patient, clinicians considering multiple variables that covary. This has resulted in international guidance that lacks precise advice as to how to adjust medications in children. The I-CAH Registry provides a platform for collation of multicentre data for research, but does not allow visualisation of individual patient data across multiple visits that could provide clinical decision support.

1.02 Prediction of growth

High glucocorticoid dosing in CAH has been shown to cause reduction in height velocity by a randomised controlled trial in 1997 that compared 15mg/m^2 hydrocortisone total daily dosing to 25mg/m^2 hydrocortisone total daily dosing (24). Retrospective analysis of data from 92 patients in Germany showed deleterious effects on final height at doses over 17mg/m^2/day (25). However, one must be careful making direct comparisons between groups formed by the dichotomisation of continuous variables within real-world registry data, as a lack of randomisation can lead to heterogeneity between groups (26). This problem has been addressed in relation to growth hormone deficiency by using multivariable regression models, of which there are multiple iterations (27-29). At the funding panel review for this fellowship, the increasing popularity and flexibility of machine learning techniques was highlighted as a valuable area of possible methodologies that could support the overarching aims and objectives of the research contained within this thesis. However, with the increasing popularity of machine learning models, there have been increasing concerns about their inappropriate application and risk of providing unexplainable and bias results (30). Prediction models of growth had not, as of the beginning of the funding period for this research (1/1/2022), been systematically reviewed and assessed objectively for bias. The first objective was therefore tailored around understanding the current standard of research into prediction modelling in this domain.

1.03 Patient and Public Involvement

There is increasing appreciation that research should be conducted ‘with’ and ‘by’ patients and the public that are directly affected by medical conditions, rather than ‘on’ or ‘for’ them (31). To inform the development of the project

plan of this thesis, two events were held at Sheffield Children's Hospital, the first with four children (aged seven to ten years) and their parents, the second with seven young people aged 13 to 22. These were funded by a £500 PPI grant from the Research and Design Service Yorkshire and Humber Public Involvement in Grant Applications Funding Award (Call 10), with top up funding from Sheffield Children's NHS Foundation trust.

These events investigated the challenges children face when living with CAH, and used their opinions to directly inform the research funding proposal. A variety of games were used to investigate their thoughts about the variation of treatment between hospitals, followed by telephone interviews with four parents to investigate parental concerns about the disease, pseudonymised information sharing and long-term outcomes. The activities employed differed depending on patient age groups. The approaches employed were presented at a national conference, the author of this thesis as senior author (32). A focus group at the CAH support group in Bristol, 2021 was also used to inform the objectives of the work presented in this thesis.

All children who attended the PPI event disliked blood tests, but there was minimal dissatisfaction with other measures of disease control, including assessing height and weight, and using x-rays to quantify bone age. Most children thought their care should be the same at different hospitals, and carers volunteered advantages of using technology when interacting with clinicians. All participants were in favour of digital tools that support clinicians and improve the process of adding data to registries that can be used for further research.

The focus groups conducted at the CAH support group showed there is enthusiasm for data sharing amongst parents of children with CAH. However, there is also awareness of the importance of good data governance, to ensure data is appropriately pseudonymised and that their children are not identifiable from published research. An area of keen discussion was based around education about the disease, and the importance of being able to equip families with the knowledge to be able to educate those around them about the importance of recognising and treating adrenal crises.

The early PPI experience encouraged the investigation of how real-world data can be best analysed and visualised to be of pertinent clinical relevance to clinicians and families that are managing chronic disease. It also provided valuable insight into the importance of understanding the management of chronic disease from not only the clinician perspective, but the patient and family perspective. The second objective was therefore designed to get a diverse

range of opinions from families of children with CAH through a national service evaluation to help contextualise the overall research by providing a contemporary snapshot of current UK practice and the satisfaction associated with it.

1.04 Biomarkers of disease control in Congenital Adrenal Hyperplasia

The insights into the detriment of high glucocorticoid doses from the randomised controlled trial that compared total daily glucocorticoid replacement of 15mg/m^2 to 25mg/m^2 led clinicians to infer that low concentrations of 17OHP were associated with patients on the higher dose that was detrimental to height velocity (24). This led to an assumption that 'normalising' 17OHP to concentrations seen in healthy patients was not to be recommended, and an indicator of overtreatment (33, 34).

Cortisol follows a diurnal rhythm, rising as humans wake and falling throughout the afternoon (35). This natural physiology is mimicked by providing patients who require cortisol replacement with larger doses in the morning, and increasingly smaller doses throughout the day (5). However, such regimes still result in inappropriate 'spikes' in cortisol that occur around dose administration. Efmody is a modified release hydrocortisone, developed under the name 'Chronocort' to better mimic the natural diurnal rhythm of the hormone (35). The phase 3 'DIUR-005' trial in adult patients with CAH was setup under the premise that target levels of 17OHP should not be low, culminating in the primary efficacy endpoint being a complicated metric of the change from baseline of the natural logarithm of the mean over a 24-hour period of the SDS profile for 17OHP (36). This study randomised participants to standard replacement or a hydrocortisone equivalent dose of the novel Chronocort. Despite showing improvement in disease control by a variety of other metrics, this trial missed its primary efficacy endpoint. Application was later made to the European Medicines Agency (EMA) for extra post hoc analysis of the data from this study to inform consideration of licensing of the drug. The better control of morning 17OHP concentrations was cited as the justification for the EMA to approve the drug (37).

Despite extra post hoc analysis and scrutiny by the EMA, there remained significant limitations on the analysis performed on the data from this trial. Patients underwent comprehensive 2-hourly measurements 17OHP and androstenedione, to produce 24-hour profiles. However, these profiles were not modelled within patients, and the biomarkers were not comprehensively regressed against one another in the original analysis (36). The third objective

of this research thesis was therefore set to further investigate these biomarkers and the relationship between them by conducting a retrospective exploratory interrogation of the data from this study.

1.05 The International Congenital Adrenal Hyperplasia Registry (I-CAH Registry)

The I-CAH Registry is the largest international database of pseudonymised information on patients with CAH in the world. This is a rich dataset with information compiled from over 600 patients with more than 10,000 assessments over the last 20 years. Data entry into the registry is increasing year on year and has already provided novel insights into CAH treatment and monitoring (9, 38).

1.06 Blood pressure in Congenital Adrenal Hyperplasia

Elevated blood pressure contributes significantly to the burden of cardiovascular disease (39), but the impact within patients with CAH is poorly understood. Studies in adults with CAH have shown elevated blood pressure (40) and higher ambulatory blood pressure (41). A cross-sectional analysis of 199 adults showed an increased risk of hypertension associated with later diagnosis, leading the authors to hypothesise that increased androgen exposure in childhood may predispose adverse cardiovascular modelling. Research into the longitudinal blood pressure changes in children with CAH is likely to improve the understanding of its prevalence and relevance in this disease.

Existing studies investigating blood pressure in children with CAH are conflicting (21). Many studies feature a small number of patients, but in those studies with data from more than 100 patients, researchers have found higher proportions of hypertension than in normative populations (19, 42-45). However, the association between high blood pressure and different dosing or markers of disease control is inconsistent, with variable associations between diastolic and systolic readings. However, none of the existing studies appropriately employed joint modelling to investigate the paired metrics of diastolic or systolic blood pressure. The fourth objective of this thesis was therefore constructed around understanding more about blood pressure within children living with CAH, by interrogation of real-world data from the I-CAH Registry.

1.07 Growth in children with Congenital Adrenal Hyperplasia

In healthy children, body mass index increases after birth before falling after approximately 1 year and then increasing in early childhood after approximately 6 years. Adiposity rebound is defined as the last nadir in BMI in early childhood, that corresponds to a decline in fat mass in relation to height. Earlier adiposity rebound in otherwise healthy children is associated with obesity in later life and cardiometabolic diseases in adulthood (46).

Several studies have shown that patients with CAH have an earlier adiposity peak and rebound compared to healthy children. However, this age differs between studies and ethnicity, from as early as 1.7 years in a UK study to 3.8 years in a study from the USA (47, 48). The reason for the early timing of adiposity rebound in CAH is poorly understood and the factors influencing this difference in body composition remain elusive. Dose of glucocorticoids or mineralocorticoids have not been shown to affect the adiposity rebound, although prior to 2022 this had only been studied in a small population of patients that had received high dose induction therapy of glucocorticoids (49). However, a study from an American cohort had shown that increased disease suppression in younger children had been associated with earlier adiposity rebound suggesting there may be in inadequately described relationship between glucocorticoid dosing and adiposity rebound. The studies into this phenomenon had quantified adiposity rebound within different groups of patients with CAH, but failed to appropriately quantify the variability of adiposity rebound between patients. Whilst there was an appreciation that it was occurring earlier, the fifth objective of this project was designed to use more appropriate statistical modelling techniques that would be able to compare the variability of this non-linear pattern between patients, rather than just the average pattern amongst groups of patients (26).

SuperImposition by Translation and Rotation is a non-linear multilevel modelling technique that estimates an average trajectory of growth across a sample population by applying an overall ‘fixed effects’ spline model, and then adjusting that model to each individual patient fit by three different ‘random effect’ transformations to provide individual patient level fits (50). The transformations include a shift on the y axis consistent with ‘size’, a shift on the x axis consistent with ‘timing’ and a stretching or compression akin to a rotation in the plane of the graph consistent with ‘intensity’ of growth. Such transformations are excellent at quantifying the pubertal growth spurt in an explainable fashion (51). The fifth objective of this research was tailored to use this modelling technique on data from the I-CAH

Registry to quantify the average and between-participant variability in both the pubertal growth spurt and adiposity rebound seen in CAH, and thus gain novel insights.

1.08 Data visualisation and automated calculations to improve clinical decision support for CAH

After coordinating the collation of data for the national service evaluation, as well as the blood pressure and adiposity rebound studies across a multitude of different centres, the author of this thesis understood how there were increasing demands on the time of clinicians, often combined with reduced administrative support. Enquiries to centres about biologically implausible data points or inconsistencies flagged during data analysis resulted in improved data quality. However, this liaison was also combined with complaints about the time it can take to reassess mistakes in data previously entered into the registry. These challenges led the author to investigate how data entry into the I-CAH Registry could be combined with visualisation and disease appropriate calculations that would support clinical decisions, rather than be restricted to research outputs that inevitably take years to come to fruition.

As part of the NIHR doctoral research fellowship undertaken by the author, one year was spent as a part-time clinical fellow in the Medicines and Healthcare products Regulatory Agency (MHRA) medical software team, to gain insights into the process of the regulation of software as a medical device. Specific focus was given to the importance of assessing the ergonomics and human factors around software as a medical device development and deployment (52). This fellowship provided valuable insight into the necessary initial investment as well as ongoing maintenance costs that are required to develop and sustain software as a medical device, increasing as the class of that device goes up from basic class I devices that can provide low risk algorithmic support to patient management, through class II devices that include those that provide advice around dosing of medication (53). This fellowship helped guide the final objective of this thesis towards the development of a prototype clinical decision support tool that is not software as a medical device, but could be integrated into the I-CAH Registry data entry platform to help advocate for more utility for clinicians during data entry.

1.09 Research Plan

1.09.01 Research Question

Can real world data from the International Congenital Adrenal Hyperplasia Registry be used to develop a prototype clinical decision support tool that supports clinicians managing children and young people with CAH.

1.09.02 Aim

Use real world data from the International Congenital Adrenal Hyperplasia registry to develop a prototype clinical decision support tool that supports clinicians managing children and young people with CAH.

1.09.03 Research Objectives

The objectives of this thesis have evolved throughout the course of this work. Working as a clinical fellow for the medicines and healthcare products regulatory agency, the author gained significant insight into the economic challenges, importance of ergonomics, and ongoing maintenance costs that are associated with developing software as a medical device (52). The author received specific training in causal inference, but tempered the application of these methods to real world data analysis after international presentations, peer review and the core and advisory supervisory teams highlighting the difficulties in inferring causality from clinical data outside of the context of controlled clinical trials. Objectives were thus directed towards robust description of available data sets using appropriate repeated measures methods that appropriately quantify associations, to gain novel insights into CAH.

Thorough description can appropriately inform clinical practice by contextualising estimates of how important variables measured in CAH vary across sample populations. However, care must be taken before using real world data to predict outcomes for patients outside of the dataset in a disease that exhibits significant heterogeneity. This work has been published in several peer review journals throughout the PhD, and thus the final objective was focused on producing a visualisation and calculation tool relevant to CAH that has the potential to increase data entry and engagement with the I-CAH Registry, and therefore increase the likelihood of more detailed research insights into CAH in the future.

1.09.03.01 Objective 1: Systematic Review into models predicting height and weight in fetus' and children

Systematically review the academic literature to assess the quality of prediction modelling published related to growth prediction in fetus' and children using the Prediction model Risk Of Bias Assessment Tool (PROBAST) (54).

1.09.03.02 Objective 2: Patient and Public Involvement: Service evaluation into the provision of care of children with CAH in the UK

Carry out a national service evaluation to review the variability of and satisfaction with care provision for children with CAH in the UK.

1.09.03.03 Objective 3: Analysis of data from a phase 3 clinical trial into the use of modified-release hydrocortisone in adults with CAH

Assess the variability of 17OHP and androstenedione between and within patients with CAH in detail by retrospective analysis of 24-hour profiles measured as part of a phase 3 clinical trial investigating modified-release hydrocortisone (55).

1.09.03.04 Objective 4: Analysis of blood pressure data from the I-CAH Registry

Analyse data from the I-CAH Registry to investigate the variability of blood pressure in patients with CAH. The author of this thesis was primary investigator for this study, study ID 202107_NL (https://sdmregistries.org/studies/).

1.09.03.05 Objective 5: Analysis of growth data from the I-CAH Registry

Analyse data from the I-CAH Registry to investigate the age and associations of the adiposity rebound and pubertal growth spurt in patients with CAH using non-linear spline modelling. The author of this thesis was primary investigator for this study, study ID 202305_NL (https://sdmregistries.org/studies/).

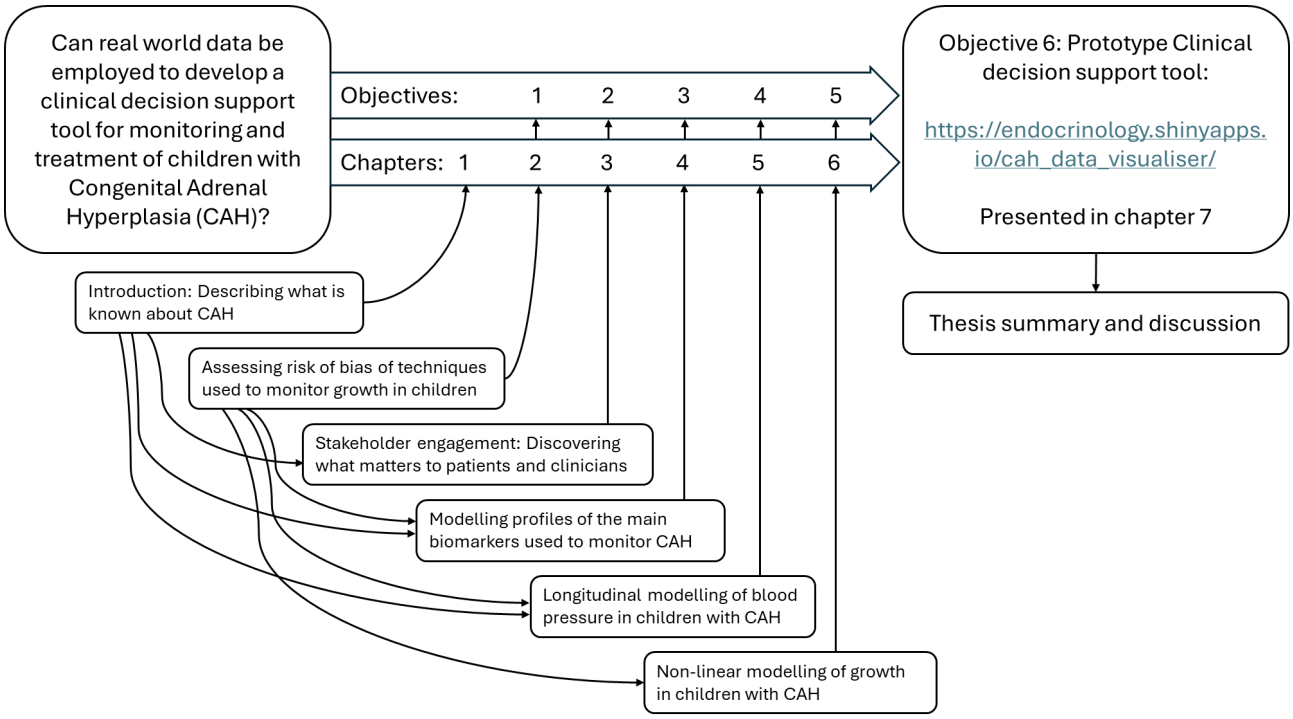
1.09.03.06 Objective 6: Development of a prototype clinical decision support tool to increase utility of data entry into the I-CAH Registry

Produce a web application that facilitates rapid visualisation and automated calculations relevant to the management of CAH. This prototype aims to demonstrate how data entry processes can be improved to provide utility to clinicians entering information into disease specific registries.

1.09.04 Thesis flow chart

These research aims and objectives work chronologically to provide insight to later research carried out throughout the thesis towards the final objective, as mapped by figure 1.1

Figure 1.1 – Graphical thesis flow chart



1.09.05 Software employed

Analysis has been conducted in *R*, a language and Environment for statistical computing (56). The prototype clinical decision support tool has been developed using the package Shiny (57). Microsoft Office (Microsoft corporation) has been used to draft this thesis and associated publications, with visualisations being collated using the ggplot library in *R* (58), and within Microsoft PowerPoint (Microsoft corporation) or Adobe Illustrator (Adobe Systems Incorporated).

1.10 Ethics

The I-CAH Registry is approved by the National Research Ethics Service in the United Kingdom as a research database that collects information from routine clinical care (19/WS/0131). Using real-world data requires adherence to appropriate ethical principles to the same extent as a traditional clinical trial, but with differing considerations. Responsibility for ethics and governance of this project lies with Neil Lawrence, but all supervisors have received training in research and data governance. Data has been processed in accordance with the General Data Protection Regulation 2018 (59).

Chapter 2 – Risk of bias in machine learning and statistical models to predict height or weight: a systematic review in fetal and paediatric medicine

2.01 Publication and contribution of others

Original submission to the *Journal of Clinical Epidemiology* on 9/4/2024 was unsuccessful. However, submission to the journal *Diagnostic and Prognostic Research* was successful on 15/12/2025 and can be accessed at (https://doi.org/10.1186/s41512-025-00215-6) (60). This article was published with joint first authorship as all articles were reviewed by both Neil Lawrence and Irina Bacila. The original draft of the manuscript and incorporation of feedback and improvements was done by Neil Lawrence. All authors agreed on the final version of the manuscript (full contributions in section 2.10).

Chapter 2 Co-authors:

Neil R Lawrence*[1,2], Irina Bacila*1,2, Joseph Tonge[1], Anthea Tucker[3], Jeremy Dawson[4,5], Zi-Qiang Lang[6], Nils P Krone[1,2], Paula Dhiman[7], Gary S Collins[7]

* Joint first authors.

Author affiliations:

1: Division of Clinical Medicine, University of Sheffield, Sheffield, S10 2RX, United Kingdom

2: Division of Medicine, Sheffield Children's Hospital, Sheffield, S10 2TH, United Kingdom

3: Health Sciences Library, University of Sheffield, Sheffield, S10 2RX, United Kingdom

4: Division of Population Health, University of Sheffield, Sheffield, S10 2RX, United Kingdom

5: Management School, University of Sheffield, Sheffield, S10 2RX, United Kingdom

6: Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, S1 3JD, United Kingdom

7: Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, OX3 7LD, United Kingdom

2.02 Chapter 2 Abstract

Background

Prediction of suboptimal growth allows early intervention that can improve outcomes for developing fetus' as well as infants and children. We investigate the risk of bias in statistical or machine learning models to predict the height or weight of a fetus, infant or child under 20 years of age to inform the current standard of research and provide insight into why equations developed over 30 years ago are still recommended for use by national professional bodies.

Methods

We systematically searched MEDLINE and EMBASE for peer reviewed original research studies published in 2022. We included studies if they developed or validated a multivariable model to predict height or weight of an individual using two or more variables, excluding studies assessing imaging or using genetics or metabolomics information. Risk of bias was assessed for all prediction models and analyses using the Prediction model Risk Of Bias ASsessment Tool (PROBAST).

Results

Sixty-four studies were included, in which we assessed the development of 180 models and validation of 61 models. Sample size was only considered in 10% of developed models and 13% of validated models. Despite height and weight being continuous variables, 77% of models developed predicted a dichotomised outcome variable.
Models performed best in the outcomes domain of PROBAST, with 61% of models developed and 85% of models validated rated at low risk of bias. However, all models were rated as high risk of bias in the analysis domain, resulting in an overall high risk of bias in all 241 models assessed. Only 29% (53/180) of the developed models were fully reported to allow implementation hindering further research and external validation.

Conclusions

Recent models developed and validated in 2022 to predict height and weight in fetuses and children are at high risk of bias and are unsuitable for application in clinical practice. Research into prediction of growth should focus on model updating using appropriate data sources with transparent reporting to allow for external validation.

Registration

The review was registered on PROSPERO (ID: CRD42023421146), the International prospective register of systematic reviews on 26/4/2023

2.03 Background

Impaired growth in the early years of life is reflective of poor health, and a predictor of poor long-term outcomes (61-63). Identifying pregnancies where the fetus has excessive growth can improve birth outcomes by controlling maternal comorbidities and guiding decisions about caesarean section (64). Monitoring of postnatal and childhood growth is equally important, with adverse growth associated with pathology, malnutrition, and neglect (65, 66).

Prediction of suboptimal growth allows for early intervention that can improve prognosis (67). The application of ultrasound revolutionised antenatal birthweight prediction (68), the Hadlock formulae developed in the 1980s facilitating accurate prediction still widely used today, recommended in the UK, Canada and United States by their respective national obstetric advice bodies (62, 69-72). Statistical tables to predict growth using bone age in children by Bayley and Pinneau were first published in 1946 (73), and seminal work was conducted by Tanner et al to inform predictions using parental height in the 1970s (74, 75). Adult height prediction in children recommended by the UK Royal College of Paediatrics and Child Health is via a prediction model developed in 2011 (76), although the American Academy of Paediatrics recommend employing x-ray quantification of bone age, one of the most widely used being a model first developed in 2009 (77).

The development of prediction models in medicine is increasing exponentially, with a growing proportion employing machine learning methods (78, 79), despite often not demonstrating additional predictive performance over more transparent techniques such as logistic regression (80). Only a fraction of prediction models undergo external validation, and even fewer translate into clinical practice (81). Health inequalities can be perpetuated if prediction models are applied in groups of individuals that are poorly represented in development data, regardless of methodology (82). Understanding the limitations of traditional statistical techniques and machine learning helps ensure appropriate application.

The Prediction model Risk Of Bias ASsessment Tool (PROBAST) has been developed to help assess risk of bias in the development or validation of prediction models (54, 83). To date, there has been no systematic assessment of novel algorithms to predict growth in fetuses or children. We investigate the risk of bias in statistical or machine learning models recently published to predict height or weight of a fetus, infant or child under 20 years of age to provide

insight into why some equations developed over 30 years ago are still recommended for use by national professional bodies (84-86).

2.04 Methods

We systematically reviewed the literature by searching MEDLINE and EMBASE using the OVID library interface on 27/4/2023. This report follows the transparent reporting of multivariable prediction models for individual prognosis or diagnosis for systematic reviews and meta-analyses (TRIPOD-SRMA) (Supplementary Checklists 2.1) (87). The review was registered on PROSPERO (ID: CRD42023421146), the International prospective register of systematic reviews on 26/4/2023 (88).

Search Strategy

Search terms were constructed with the help of an information specialist (AT) using a strategy that combined four groups of terms, all of which would be included in eligible studies, combined with the Boolean operator 'OR'. The four groups included statistical modelling or machine learning methods, model performance terms, terms relating to height or weight, and terms relating to the age of participants. These groups were combined with the Boolean operator 'AND', and restricted to articles published in 2022 to ensure a contemporary sample were assessed. The full search strategy is detailed in Appendix Table S2.1.

Eligibility Criteria

We included peer reviewed original research studies published in English that reported the development or validation of a model to predict height or weight of an individual, restricted to those published in 2022 to provide a contemporary reflection of the studies that are developing and validating these models. Studies were included if models used two or more predictor variables, with continuous or categorical outcomes, using data from any study design. We use the term 'model' to describe any statistical model, equation or machine learning algorithm that produces a predicted outcome designed to apply to an individual.

Included models could employ metrics derived from imaging (e.g. ultrasound), but models predicting results from images were excluded to avoid quantitative medical imaging interpretation studies. Models using composite outcomes that included height or weight alongside other outcomes were excluded to ensure homogenous

comparisons. Models incorporating genomics or metabolomics were excluded to ensure models assessed used predictors that were accessible to clinicians. Studies where the focus was to identify predictors, rather than develop a model to predict an outcome in an individual, were excluded, alongside reviews and conference abstracts to evaluate models at a stage of development suitable for potential application in clinical practice.

Selection Process

Titles and abstracts were screened independently by two authors (Neil Lawrence and Joseph Tonge), with full articles read and models assessed independently by two authors (Neil Lawrence and Irina Bacila). Disagreements were arbitrated by a third author (Paula Dhiman).

Risk of bias assessment

PROBAST was used to assess each prediction model in four domains, each consisting signalling questions answered as 'yes', 'probably yes', 'no information', 'probably no' or 'no'. Questions prompt consideration of risks of bias within each domain, but need not be exclusively answered positively or negatively to define the domain rating, which may be 'low', 'unclear', or 'high'. If any domain is rated 'high', overall assessment of the model is high risk of bias (Appendix Tables S2.2-S2.3 outline PROBAST criteria). Model ratings were discussed between both authors to reach agreement on each screening question and domain rating, any disagreements arbitrated by a third author (Paula Dhiman).

Data Collection and analysis

A data collection form was developed using Excel (Microsoft, Redmond, Washington) (https://tinyurl.com/3aazxyen). We collected whether the studied population was maternal or paediatric, the outcome variable(s) of interest, geographical origin of the data, number of eligible models reported and the following model specific data items:

For each eligible model:

- Type of prediction model study (development or validation).
- Whether sample size was considered.
- Number of participants used in analysis, with number of 'events' for categorical outcomes.
- Whether analysis assessed discrimination, calibration, or both.
- PROBAST signalling question answers and risk of bias rating.

For models developed:

- Number of candidate predictor parameters.
- Modelling methodology employed.
- Whether final model equation or equivalent that would allow external validation of the model was reported.

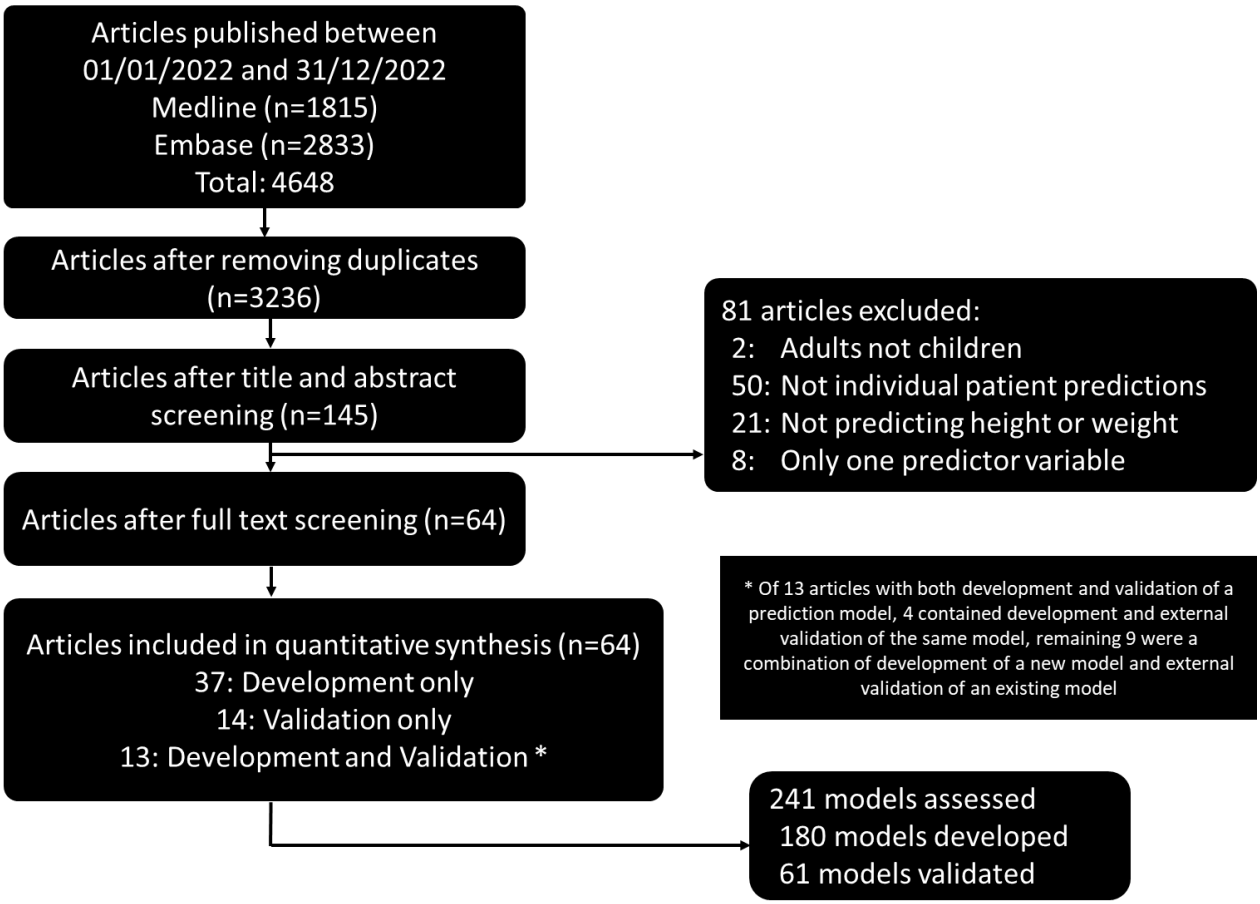
Summary statistics were calculated using *R: A language and environment for statistical computing* (https://www.R-project.org), and reported with lower quartile, median and upper quartile values. Percentages were calculated for the number of models rated in each category. Models were subdivided into type of methodology (regression-based methods, flexible machine learning methods and ensemble machine learning methods) to assess for any difference in the risk of bias between categories.

2.05 Results

Number of studies and models assessed

The search returned 3236 eligible articles after deduplication (Figure 2.1), with abstract screening returning 145 articles. At least one model that satisfied inclusion criteria was found in 64 articles (Appendix Table S2.4). Data used to develop and validate the models originated from participants in 26 different countries, most commonly China (n=20/64, 31%) (Table 2.1, Appendix Table S2.5, Appendix Figure S2.1). Model development was the focus of 37 articles (58%), model validation in 14 articles (22%), and 13 featured both development and validation suitable for rating (20%). A median of 3 models were rated in each article (Q1 to Q3: 1 to 4.5, range: 1 to 25). In all articles with multiple models, the same dataset was used either with different predictors, a different outcome variable, a subset of the sample or a different methodology.

759 Figure 2.1 – PRISMA flow diagram of included studies



760

761 Table 2.1 – Characteristics of studies included in review

	Studies including model development only n = 37	Studies including model validation only n = 14	Studies including both model development and validation n = 13	All studies n = 64
Population studied	**Number of studies (%)**			
Maternal population	20 (54.1%)	11 (78.6%)	8 (61.5%)	39 (60.9%)
Individuals under 18 years	17 (45.9%)	3 (21.4%)	5 (38.5%)	25 (39.1%)
Outcome of interest				
Height	9 (24.3%)	2 (14.3%)	3 (23.1%)	14 (21.9%)
Weight	23 (62.2%)	11 (78.6%)	8 (61.5%)	42 (65.6%)
BMI	4 (10.8%)	0 (0.0%)	1 (7.7%)	5 (7.8%)
Multiple outcomes†	1 (2.7%)	1 (7.1%)	1 (7.7%)	3 (4.7%)
Geographical location				
Africa	2 (5.4%)	1 (7.2%)	0 (0.0%)	3 (4.7%)
Asia	24 (64.9%)	5 (35.7%)	5 (38.4%)	34 (53.1%)
Australia	1 (2.7%)	0 (0.0%)	1 (7.7%)	2 (3.1%)
Europe	6 (16.2%)	4 (28.5%)	4 (30.8%)	14 (21.9%)
North America	3 (8.1%)	3 (21.4%)	3 (23.1%)	9 (14.0%)
Multiple: North America and Europe	1 (2.7%)	0 (0.0%)	0 (0.0%)	1 (1.6%)
Multiple: South America, North America, Europe and Asia	0 (0.0%)	1 (7.2%)	0 (0.0%)	1 (1.6%)

762 †Studies investigating multiple outcomes included height and weight in childhood in one article; height, weight, and

763 BMI in childhood in one article; weight and length in infancy in 1 article

Methods employed

The final model equation or code that would enable external validation of the model was provided for only 53/180 (29%) of the developed models. Logistic regression was most commonly used to develop models (n=63/180, 35%), multiple linear regression second most common (23/180, 13%), with support vector machines and random forest used next most frequently (16/180, 9% each) (Appendix Figure S2.2). Models were categorised as regression-based methods (97/180, 53%), flexible machine learning methods (45/180, 25%) and ensemble methods (35/180, 20%), remaining models (3/180, 2%) developed with unclear methodology (Appendix Table S2.6). Both discrimination and calibration were assessed in 28/180 (16%) models developed, and 10/61 (16%) models validated with 122/180 (68%) models developed and 32/61 (52%) models validated relying on discrimination alone (Appendix Figure S2.3).

Sample size

Sample size was only considered in 10% of developed models and 13% of validated models. Sample size methodology ranged from expert opinion to cited literature but was never consistent with the most up to date guidance (89-93). A median of 1993 participants (Q1 to Q3: 286 to 6363) were used for model development. A continuous outcome was predicted in 42/180, the remaining 138/180 predicting a binary outcome, of which the number of outcome events used to develop 49/138 models was unclear. Of the 89/138 binary outcome models developed that clearly reported the number of outcome events, a median of 130 participants (Q1 to Q3: 78 to 412) had the outcome of interest. Where it could be calculated (84/138), there were a median of 11.7 events per candidate variable (Q1 to Q3: 4.1 to 24.7) for models predicting a binary outcome.

The sample size used to validate 5/61 models was unclear, with 56/61 using data from a median sample size of 374 (Q1 to Q3: 170 to 660). A continuous outcome variable was predicted in 30/61 models. For models predicting a binary outcome, the sample size contained a median of 85 (Q1 to Q3: 68 to 147) events.

Risk of bias ratings

Overall PROBAST ratings

Every model assessed was rated high risk of bias (Table 2.2, Figure 2.2, precise screening question results explaining rationale for domain in Appendix Table S2.7, Appendix Figure S2.4). This comprised a median of 3/4 domains (Q1 to Q3: 2 to 3) rated high risk of bias in models developed, and 2/4 domains (Q1 to Q3: 1 to 3) for models validated. Whilst problematic analysis resulted in high risk of bias across all models, a post-hoc sensitivity analysis showed that 87% of models developed and 74% of models validated would have rated high risk of bias due to at least one other domain, even had analysis been adequate. There was no difference in proportion of models rated high risk of bias dependent upon type of machine learning methodology employed (Appendix Table S2.8), or when only the top-rated model from each article was included to calculate summary statistics (Appendix Tables S2.9-S2.10).

Domain 1: Participants

Participant selection contributed high risk of bias in 54% of developed models, and unclear risk of bias in 19%. Validated models were high risk in 64% and unclear in 7%. The screening question most frequently highlighting problems was 1.2 referring to the suitability of eligibility criteria, rated as 'N/PN' in 51% of developed models and 44% of validated models.

Domain 2: Predictors

Predictor measurement contributed high risk of bias in 44% of developed models, and 26% in models validated. Question 2.1 assessing heterogeneity in the way predictors were defined and assessed across participants most frequently highlighted problems, rated as 'N/PN' in 61% of models developed and 34% of models validated.

Domain 3: Outcome

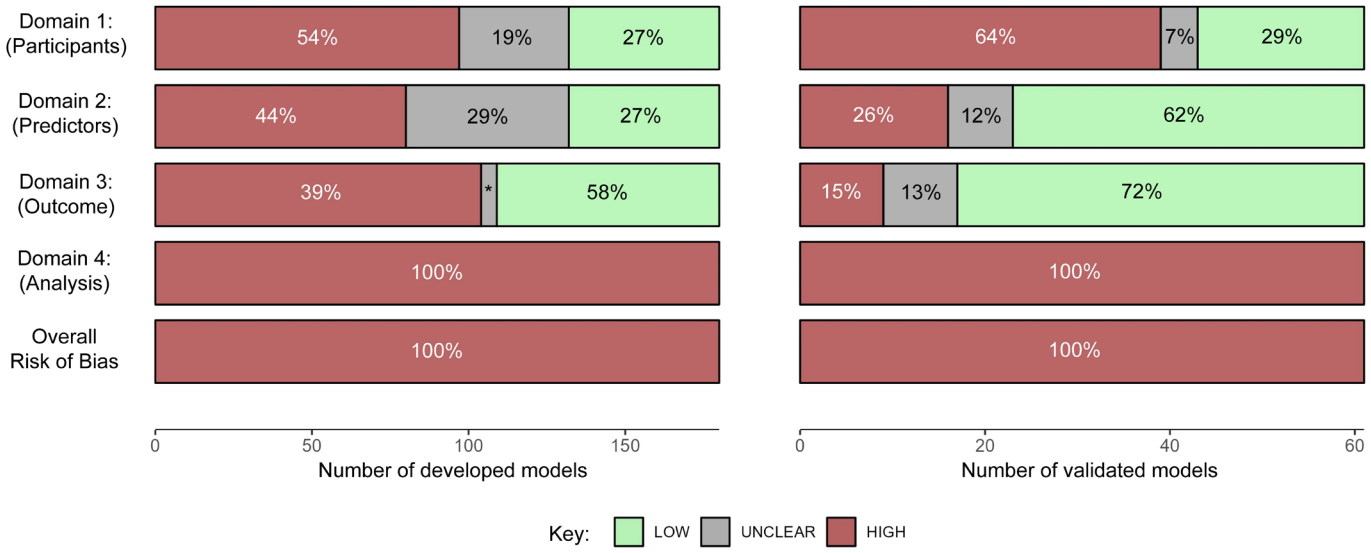
Outcome measurement contributed high risk of bias in 58% of developed models, and 15% in models validated. Question 3.1 assessing appropriate determination of the outcome most frequently highlighted problems, 43% of models developed and 10% of models validated rated 'N/PN'. The second most frequent was 3.6 pertaining to appropriate time between predictor and outcome assessment, where 39% of developed models and 16% of validated models were rated 'N/PN'. Heterogeneity in the assessment of the outcome was a problem in 28% of models developed.

Domain 4: Analysis

Every model developed and validated was rated as high risk of bias due to problematic analysis. Question 4.7 regarding model performance measures used to evaluate models was most frequently highlighted, inappropriate in 88% of models developed and 93% of those validated. Next most frequently flagged questions were 4.3 pertaining to all enrolled participants being included in the analysis, and 4.4 whether missing data was handled appropriately. In these questions 80% and 74% were rated 'N/PN' for developed models, and 61% and 62% for validated models respectively. Across all questions in domain 4, there was a median of 7.5/9 (Q1 to Q3: 6 to 8) questions answered 'N/PN/NI' in developed models and a median of 4/6 (Q1 to Q3: 4 to 5) in models validated.

Figure 2.2 – Risk of bias ratings

Bar chart showing risk of bias ratings by specific domain and overall for developed (n=180) and validated (n=61) models



* = 3% unclear risk of bias in domain 3 of developed models.

831 Table 2.2 – Screening question results

	Title of domain / screening question	Developed models (n=180) n, %			Validated models (n=61) n, %		
Overall risk of bias		**LOW**	**HIGH**	**UNCLEAR**	**LOW**	**HIGH**	**UNCLEAR**
		n = 0, 0%	n = 180, 100%	n = 0, 0%	n = 0, 0%	n = 61, 100%	n = 0, 0%
Domain 1	**Participants**	**LOW**	**HIGH**	**UNCLEAR**	**LOW**	**HIGH**	**UNCLEAR**
		n = 48, 27%	n = 97, 54%	n = 35, 19%	n = 18, 29%	n = 39, 64%	n = 4, 7%
Screening questions:		**Yes / probably yes**	**No / probably no**	**No information**	**Yes / probably yes**	**No / probably no**	**No information**
1.1	Were appropriate data sources used, e.g., cohort, randomized controlled trial, or nested case-control study data?	n = 83, 46%	n = 79, 44%	n = 18, 10%	n = 43, 70%	n = 18, 30%	n = 0, 0%
1.2	Were all inclusions and exclusions of participants appropriate?	n = 62, 34%	n = 91, 51%	n = 27, 15%	n = 31, 51%	n = 27, 44%	n = 3, 5%
Domain 2	**Predictors**	**LOW**	**HIGH**	**UNCLEAR**	**LOW**	**HIGH**	**UNCLEAR**
		n = 48, 27%	n = 80, 44%	n = 52, 29%	n = 38, 62%	n = 16, 26%	n = 7, 12%
Screening questions:		**Yes / probably yes**	**No / probably no**	**No information**	**Yes / probably yes**	**No / probably no**	**No information**
2.1	Were predictors defined and assessed in a similar way for all participants?	n = 52, 29%	n = 110, 61%	n = 18, 10%	n = 37, 61%	n = 21, 34%	n = 3, 5%
2.2	Were predictor assessments made without knowledge of outcome data?	n = 123, 68%	n = 42, 23%	n = 15, 9%	n = 52, 85%	n = 2, 3%	n = 7, 12%
2.3	Are all predictors available at the time the model is intended to be used?	n = 168, 93%	n = 10, 6%	n = 2, 1%	n = 61, 100%	n = 0, 0%	n = 0, 0%
Domain 3	**Outcome**	**LOW**	**HIGH**	**UNCLEAR**	**LOW**	**HIGH**	**UNCLEAR**
		n = 71, 39%	n = 104, 58%	n = 5, 3%	n = 44, 72%	n = 9, 15%	n = 8, 13%
Screening questions:		**Yes / probably yes**	**No / probably no**	**No information**	**Yes / probably yes**	**No / probably no**	**No information**
3.1	Was the outcome determined appropriately?	n = 92, 51%	n = 77, 43%	n = 11, 6%	n = 54, 88%	n = 6, 10%	n = 1, 2%
3.2	Was a prespecified or standard outcome definition used?	n = 169, 94%	n = 11, 6%	n = 0, 0%	n = 57, 93%	n = 3, 5%	n = 1, 2%
3.3	Were predictors excluded from the outcome definition?	n = 168, 93%	n = 12, 7%	n = 0, 0%	n = 58, 95%	n = 3, 5%	n = 0, 0%
3.3	Were predictors excluded from the outcome definition?	n = 168, 93%	n = 12, 7%	n = 0, 0%	n = 58, 95%	n = 3, 5%	n = 0, 0%
3.4	Was the outcome defined and determined in a similar way for all participants?	n = 111, 62%	n = 50, 28%	n = 19, 10%	n = 53, 87%	n = 8, 13%	n = 0, 0%

	Title of domain / screening question	Developed models (n=180) n, %			Validated models (n=61) n, %		
Domain 3	**Outcome**	**Yes / probably yes**	**No / probably no**	**No information**	**Yes / probably yes**	**No / probably no**	**No information**
3.5	Was the outcome determined without knowledge of predictor information?	n = 76, 42%	n = 48, 27%	n = 56, 31%	n = 38, 62%	n = 4, 7%	n = 19, 31%
3.6	Was the time interval between predictor assessment and outcome determination appropriate?	n = 91, 51%	n = 71, 39%	n = 18, 10%	n = 49, 80%	n = 10, 17%	n = 2, 3%
		LOW	**HIGH**	**UNCLEAR**	**LOW**	**HIGH**	**UNCLEAR**
Domain 4	**Analysis**	n = 0, 0%	n = 180, 100%	n = 0, 0%	n = 0, 0%	n = 61, 100%	n = 0, 0%
Screening questions:		**Yes / probably yes**	**No / probably no**	**No information**	**Yes / probably yes**	**No / probably no**	**No information**
4.1	Were there a reasonable number of participants with the outcome?	n = 87, 48%	n = 62, 35%	n = 31, 17%	n = 42, 69%	n = 19, 31%	n = 0, 0%
4.2	Were continuous and categorical predictors handled appropriately?	n = 38, 21%	n = 122, 68%	n = 20, 11%	n = 53, 87%	n = 7, 11%	n = 1, 2%
4.3	Were all enrolled participants included in the analysis?	n = 0, 0%	n = 144, 80%	n = 36, 20%	n = 7, 11%	n = 37, 61%	n = 17, 28%
4.4	Were participants with missing data handled appropriately?	n = 13, 7%	n = 133, 74%	n = 34, 19%	n = 2, 3%	n = 38, 62%	n = 21, 35%
4.5	Was selection of predictors based on univariable analysis avoided?	n = 91, 51%	n = 62, 34%	n = 27, 15%			
4.6	Were complexities in the data (e.g. censoring, competing risks, sampling of control participants) accounted for appropriately?	n = 38, 21%	n = 117, 65%	n = 25, 14%	n = 8, 13%	n = 47, 77%	n = 6, 10%
4.7	Were relevant model performance measures evaluated appropriately?	n = 21, 12%	n = 159, 88%	n = 0, 0%	n = 4, 7%	n = 57, 93%	n = 0, 0%
4.8	Were model overfitting and optimism in model performance accounted for?	n = 27, 15%	n = 114, 63%	n = 39, 22%			
4.9	Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?	n = 31, 17%	n = 17, 10%	n = 132, 73%			
Sensitivity analysis:		**LOW**	**HIGH**	**UNCLEAR**	**LOW**	**HIGH**	**UNCLEAR**
Risk of bias rating not including analysis domain:		*n = 19, 11%*	*n = 157, 87%*	*n = 4, 2%*	*n = 13, 21%*	*n = 45, 74%*	*n = 3, 5%*

Individual question risk of bias ratings grouped as positive, negative or no information for developed and validated models

2.06 Discussion

We conducted a systematic review assessing the risk of bias within the development and validation of models to predict height and weight in fetal or paediatric medicine, and found all models assessed to be at high risk of bias, all having problematic analysis. Less than a third of models were reported to facilitate external validation. This highlights the importance of tools to assess the risk of bias of prediction models prior to application in clinical practice, and the necessity to improve understanding in prediction modelling to increase research impact.

Selection and measurement of variables

Both height and weight are continuous outcomes, yet the most common model development method employed was logistic regression, alongside a variety of machine learning methods that predict categorical outcomes. Mishandling of continuous and categorical predictors was seen in over two-thirds of models developed, with inappropriate preferences to categorise continuous predictors, reducing power and introducing bias. This reproduces recent findings about logistic regression prediction models (94), despite open access guidelines advising against such practice (95, 96). The assessment of predictors or outcomes was frequently heterogenous, particularly in multicentre or lengthy longitudinal studies. In over a third of models developed, there were limitations identified in the interval chosen between predictors and outcome of interest being too short or too long for clinical utility, and in 6% of models developed the predictors would not be available at the time the model was designed to be used.

Sample size

Sample size consideration was rare, in keeping with broader reviews showing consistently inappropriate samples employed to develop prediction models (97). Due to routine inclusion of multiple interaction terms and categorisation of predictors, machine learning models require larger samples to avoid overfitting (93). Future research can benefit from published advice and software packages to assist calculating appropriate sample sizes (98).

Analysis

The most frequent problems in analysis included using inappropriate performance measures, not accounting for all participants studied, and failures in handling missing data. Missing data mechanisms should be considered before carrying out complete case analysis. The exclusion of participants because of information only available at a stage *after* the prediction model is employed is inappropriate (99). Multiple imputation offers the opportunity to account

for missing values by using auxiliary variables and reduces the likelihood of bias, yet was rare, and when employed was often on a small subsection following listwise deletion (54, 100).

Only 16% of models assessed reported calibration alongside discrimination. Exclusively using calibration to assess predictions of continuous measurements may be appropriate, but over half reported only discrimination. Calibration plots to assess the accuracy of predictions across the spectrum of outcomes were rarely employed. Receiver operated characteristics curves featured frequently, providing little insight beyond their area under the curve.

A lack of reporting of the final prediction model, as either an equation, code or online calculator for over 2/3 of models assessed calls into question the value of publishing such models. This was highlighted when assessing whether weights of variables in multivariable analysis were consistent with weights applied within the reported prediction model, a question that cannot be answered without transparent reporting of the final model itself.

Notwithstanding data analysis, 87% of developed models and 74% of models validated rated a high risk of bias based on the scoring of other domains. There would therefore remain significant problems with the underlying data available to the authors of most of these studies that means any model redevelopment, even with gold standard analysis, would remain at risk of bias. Instead, incorporating prediction model intentions within study designs and well defined protocols that promote transparency should be the focus of future research, to increase the likelihood of models that have the potential for clinical application (101).

Clinical and Research Implications

We have shown a high risk of bias in predicting fundamental metrics used for clinical assessment in obstetrics and paediatrics, similar to that recently seen in infectious diseases (102), oncology (103), intensive care (104) and surgery (105), as well as across supervised learning methods in general (30). Existing models most frequently used for comparison within studies related to obstetrics were the Hadlock formulae for fetal weight prediction developed in the 1980's (69), formulae that have recently also shown superiority in meta-analysis of prediction for estimated fetal weight (106, 107). Whilst failure to adopt novel models into clinical practice may in some cases reflect inertia within healthcare, our work alongside these meta-analyses highlights that despite exponential development of prediction models, there is a lack of high quality models developed that are suitable for clinical practice. In terms of paediatric studies predicting height, Bayley-Pinneau and Tanner's formulae from the 1970's and earlier were frequently used for

comparison (73, 75), rather than the more recent 2011 UK model that adjusts for regression to the mean (76), or more modern open-source software that predicts height using bone age (77, 108). When comparing newly developed models in future research, it is important that authors compare models to their most contemporaneous equivalents that are used within clinical practice. A lack of recent meta-analysis robustly comparing prediction models of adult height in children is an area for future research, although the heterogeneity in the predictors used for development of such models within our assessments highlights some of the challenges in direct model comparison. To maximise the likelihood of clinical application, we advocate for the application of high quality, robust prediction model strategies rather than the current trend of simply a greater quantity.

Strengths and limitations

Most studies that validated an external model did not report external validation as their primary aim. Nonetheless, when analysis is used to conclude about the validity of the application of a model, or to advocate for another newly developed model, it is vital it is done robustly. Clinicians should critically analyse prediction models of height and weight using the updated PROBAST+AI (109), as we have seen significant limitations within model validation, that can result in model accuracy being interpreted incorrectly.

This review has provided a contemporary assessment of the risk of bias in models designed to predict height and weight, but is limited to those published in 2022, and thus some models produced more recently may be at lower risk of bias. We have not directly assessed the risk of bias that may exist within the original model development of those that are currently used in clinical practice, and have not included models published within the grey literature which limits the scope of this research. The studies included were heterogenous, some assessing growth over months or years, some assessing point estimates of height or weight using proxy variables. Whilst PROBAST and the more recent PROBAST+AI are well-structured tools that direct robust assessment of models, there remains an element of subjectivity in specific ratings. However, the independent rating by two authors and arbitration of disagreements by a medical statistician, alongside the large number of screening questions and domains that indicated a high risk of bias reassures that the overall risk of bias within all of the models assessed is high, and that their application in clinical practice cannot be recommended.

2.07 Conclusions

Recent models developed and validated in 2022 to predict height and weight in fetuses and children are at high risk of bias and unsuitable for clinical practice. Until the standard of prediction modelling is improved, clinicians should continue to use traditional growth charts and simple formulae recommended by national and international bodies (70, 84-86). The current focus on developing brand new models and use of opaque machine learning methodologies that lead to difficulties in interpretation should be redirected into external validation of existing models and incremental model updating (80, 110, 111).

2.08 Availability of data and materials

All raw data collected about each study and model reported in this study is available via the following link (tinyurl.com/442f4bdz).

2.09 Competing interests

No specific funding was received for this study. The work is part of N.R.L.'s NIHR Doctoral Research Fellowship PhD project registered at the University of Sheffield, United Kingdom. G.S.C. is editor-in-chief of Diagnostic and Prognostic Research. G.S.C. is also a co-author of PROBAST, which is used in this study, and a NIHR Senior Investigator. The views expressed in this article are those of the authors and not necessarily those of the NIHR, or the Department of Health and Social Care. I.B., J.T., A.T., J.D., Z.L., N.P.K. and P.D. have no competing interests.

2.10 Authors' contributions

N.R.L, I.B. and G.S.C. conceived the study. N.R.L, I.B. and G.S.C. designed the study. N.R.L., P.D. and A.T. developed the search strategy. N.R.L. and J.T. carried out the screening. N.R.L, I.B. and P.D. carried out the data extraction of all items from all articles. N.R.L. performed the analysis. N.R.L. and I.B. drafted the first draft. All authors critically reviewed, edited the article and approved the final manuscript.

2.11 Thesis contribution of chapter 2

A systematic review has been completed and published in *Diagnostic and Prognostic Research*, showing the high risk of bias in models recently developed and validated to predict height and weight to achieve objective 1 (60). Critiquing published statistical models using the rigorous approach outlined within the PROBAST guideline and its related documents (54, 83, 109) directly informed the methods employed to analyse data from the national service evaluation, the Chronocort study and the I-CAH registry in subsequent chapters of this thesis (112-115).

Chapter 3 – National service evaluation of the quality of care for children and young people with congenital adrenal hyperplasia in the United Kingdom: survey responses from patients and clinicians

3.01 Publication and contribution of others

This chapter was published in *Hormone Research in Paediatrics* and can be accessed at (https://doi.org/10.1159/000537978) (115). Co-authors outside of the supervisory team contributed through registering the quality improvement project locally and distributing the electronic questionnaire, alongside review and feedback of the manuscript. The original draft of the manuscript and incorporation of feedback and improvements was done by Neil Lawrence. All authors agreed on the final version of the manuscript (full contributions in section 3.09).

Chapter 3 Co-authors:

Neil R Lawrence[1,2], Irina Bacila[1,2], Gary Collins[3,4], Jeremy Dawson[5,6], Zi-Qiang Lang[7], Xiaochen Ji[8], S Faisal Ahmed[9,10], Sabah Alvi[11], Louise Bath[12], Jo Blair[13], Tim Cheetham[14,15], Liz Crowne[16], Justin H Davies[17,18], Mehul Dattani[19,20], Evelien F Gevers[21,22], Ruth Krone[23], Leena Patel[24,25], Ajay Thankamony[26,27], Tabitha Randell[28], Fiona Ryan[29], Sue Elford[30], Sallyann Blackett[30], Nils P Krone[1,2]

Chapter 3 Author affiliations:

1: Division of Clinical Medicine, School of Medicine and Population Health, University of Sheffield, Sheffield, S10 2RX, United Kingdom
2: Department of Paediatric Endocrinology, Sheffield Children's Hospital, Sheffield, S10 2TH, United Kingdom
3: Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, OX3 7LD, United Kingdom
4: NIHR Oxford Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom
5: School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, S10 2RX, United Kingdom
6: Management School, University of Sheffield, Sheffield, S10 2RX, United Kingdom
7: Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, S1 3JD, United Kingdom
8: Department of Endocrinology and Metabolism, the Second Affiliated Hospital of Dalian Medical University, Dalian, China
9: Office for Rare Conditions, Royal Hospital for Children & Queen Elizabeth University Hospital, Glasgow, UK
10: Developmental Endocrinology Research Group, University of Glasgow, Glasgow, United Kingdom
11: Department of Paediatric Endocrinology, Leeds Children's Hospital, Leeds, LS2 3AX, United Kingdom
12: Paediatric Endocrinology, Royal Hospital for Sick Children, Edinburgh, EH16 4TJ, United Kingdom
13: Department of Endocrinology, Alder Hey Children's NHS Foundation Trust, Liverpool, L14 5AB, United Kingdom
14: Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, International Centre for Life, Newcastle upon Tyne, NE1 3BZ, United Kingdom
15: Department of Paediatric Endocrinology, Great North Children's Hospital, Newcastle upon Tyne Hospitals NHS Foundation Trust, NE1 4LP, United Kingdom
16: Bristol Royal Hospital for Children, University Hospitals Bristol and Weston Foundation Trust, Bristol, United Kingdom
17: NIHR Southampton Biomedical Research Centre, University of Southampton and University Hospital Southampton NHS Foundation Trust, Southampton, United Kingdom
18: Paediatric Endocrinology, University Hospital Southampton NHS Foundation Trust, Southampton, UK.
19: Institute of Child Health, University College London, 30 Guildford Street, London, WC1N 1EH
20: Department of Endocrinology, Great Ormond Street Hospital, London, United Kingdom
21: Centre for Endocrinology, William Harvey Research Institute, Queen Mary University of London, London, EC1M 6BQ, United Kingdom
22: Department of Paediatric Endocrinology and Diabetes, The Royal London Hospital, Barts Health NHS Trust, London, E1 1FR, United Kingdom
23: Department of Endocrinology, Birmingham Women's & Children's Hospital, Birmingham, United Kingdom
24: Paediatric Endocrine Service, Royal Manchester Children's Hospital, Manchester University NHS Foundation Trust, Manchester, United Kingdom
25: Division of Medical Education, University of Manchester, Oxford Road, Manchester, M13 9PL, United Kingdom
26: Department of Paediatrics, University of Cambridge, Cambridge United Kingdom
27: Children's Services, Addenbrooke's Hospital, Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom
28: Department of Paediatric Endocrinology, Nottingham Children's Hospital, Nottingham NG15 1PB, United Kingdom
29: Department of Paediatric Endocrinology, Oxford Children's Hospital, Oxford University Hospitals NHS Foundation Trust, Oxford, OX3 9DU, United Kingdom
30: Living with CAH, CAH Support Group, Tamworth, United Kingdom

3.02 Abstract

Aim

To quantify differences in service provision for children and young people (CYP) living with CAH across the United Kingdom.

Methods

A national service evaluation using online questionnaires circulated to patients and clinicians from secondary and tertiary UK centres managing CYP with CAH, and via the "Living with CAH" support group mailing list.

Results

Total of 195 responses relating to patients aged 0-20 years (43 patients, 152 carers), as well as 34 clinicians from 33 hospitals. Only 12% of clinicians were 'completely satisfied' with the service provided, compared to 68% of carers and 76% of patients. Whilst 94% of clinicians reported providing formal training to families with CAH, over 80% of both patients and carers reported not attending what they considered formal training. Appetite for further training was higher in carers (86%) than patients (55%), although further 'unsure' responses suggested formal training sessions would likely be well attended. Biochemical monitoring of treatment was broadly in keeping with international guidelines, with 67% of clinicians reporting regular use of dried blood spots, and 12% regular urinary steroid metabolites.

Conclusions

While there is overall good satisfaction with care provision among patients and carers with CAH in the UK, extra resources addressing the psychological and educational needs about the disease and its management would benefit patients and carers. Improved access to allied health professionals and psychologists will help support families and improve patient outcomes.

3.03 Introduction

Congenital Adrenal Hyperplasia (CAH) is caused in over 90% of cases by 21-hydroxylase deficiency, resulting in glucocorticoid deficiency and androgen excess. Commonly manifesting as ambiguous genitalia in girls, boys may present in the first few days of life with an adrenal or salt losing crisis. However, manifestations occur along a spectrum and may present later with a more insidious picture of glucocorticoid deficiency (5).

International guidance has evolved in the last two decades in attempts to help standardise treatment regimens and improve outcomes for patients with CAH. Nonetheless, the CAH-UK initiative has recently reported on current practice in 107 patients from 14 centres (15) and reveals a heterogeneous approach to managing the condition. A comprehensive analysis of patient reported measures, as part of the CAH-UK project, reported large variability in the quality of life of these patients, poor quality of life related to unhealthy body habitus and levels of hormonal control consistent with overtreatment (116).

In the UK rare diseases framework survey, patients have reported that access to specialist care, coordination of that care and awareness of healthcare professionals are three of the top four challenges linked to living with a rare disease (117). The UK National Diabetes Audit has been shown to allow objective assessment of the national diabetes prevention programme, and provide valuable feedback to ensure diabetes interventions are well targeted (118). We therefore set out to assess the current service provision of CAH across the UK by conducting a questionnaire-based multicentre national service evaluation, to quantify patient, carer and clinician satisfaction with services, and to help identify key elements for best practice.

3.04 Methods

Survey Development

Three questionnaires using similar questions designed for anonymous completion by clinicians, carers and parents were developed with a multidisciplinary team comprising general paediatricians, paediatric endocrinologists, clinical nurse specialists, carers of children with CAH and children living with the condition. The living with CAH support group and British Society for Paediatric Endocrinology Adrenal Special Interest Group approved the final version after pilot testing with a patient and public involvement group. A total of 48 questions were asked to clinicians, and up to 40 questions to patients and carers, with conditional branching employed to ensure redundant questions were not displayed, and a reduced number of questions with simpler wording for younger respondents. A full copy of each survey is available in the research repository (https://github.com/neilxlawrence/CAH_service_evaluation).

Project registration and data collection

The project was approved and registered as a multicentre service evaluation project with Sheffield Children's Hospital Quality and Governance Department (registration number SE1661). Clinicians participating at local centres followed their local service evaluation procedures. The questions were formatted into an online form using Microsoft Forms (Redmond, Washington, USA) with responses collected within the Sheffield Children's hospital secure NHS mail service (Sheffield, UK). The questionnaire was available for completion at hospital trusts between 10/12/2021 and 30/06/2023 and was disseminated to members of the Living with CAH support group on 14/12/2021. A paper version of each questionnaire was also available to minimise digital exclusion, with results scanned and transcribed into the online form after completion.

Respondent characteristics

Patient and carer questionnaires collected age, sex, ethnicity, and name of the centre. Clinician questionnaires collected the name of centre, job role and years spent managing children with CAH.

Categories of questions

Both carers and clinicians were asked about the frequency of aspects assessed and discussed at clinic appointments. Level of confidence in managing the condition was asked to patients and carers, as well as questions about education

and aspects of advice they had received. Clinicians were similarly asked about the education provided, as well as their access to ancillary services and members of the multidisciplinary team.

Statistical Analysis

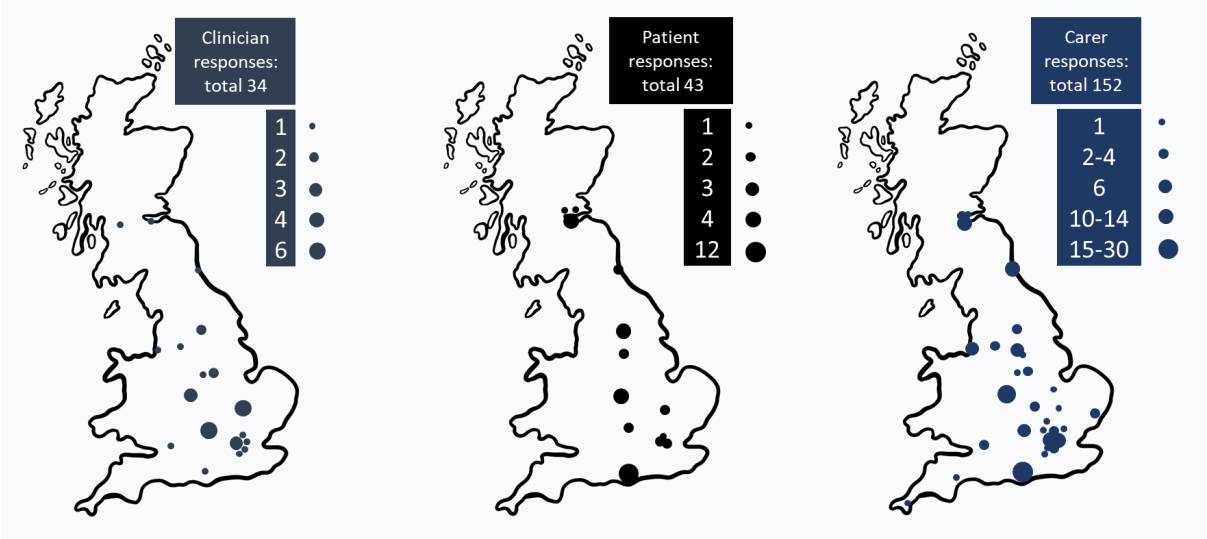
Data analysis was carried out in *R: A Language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria)*, packages employed listed in Appendix Table S3.1. Geographical spread of responses was quantified in comparison to the overall catchment population of the hospital from which the response was registered (119). A sensitivity analysis weighting responses by the size of the catchment population was conducted to assess for any significant change in proportions of responses from centres that had a proportionally larger number of responses to the survey (full details within Appendix Table S3.2). Thematic analysis of free text comments was conducted to summarise these responses and quantify the most commonly emerging topics.

3.05 Results

A total of 229 responses from 33 hospitals pertaining to children with CAH were received (full summary in Appendix Table S3.2). Thirty-four clinicians from 18 hospitals responded, alongside 43 patients and 152 carers reporting their last appointment from 34 different hospitals (Table 3.1). Geographical spread is depicted in Figure 3.1, with the sensitivity analysis of results weighting results by number of clinician responses and hospital catchment area (Appendix Table S3.2) showing no significant difference between the raw proportions of answers and weighted proportions.

Figure 3.1 – Geographical distribution of responses to all surveys

Number of responses from each hospital depicted by size of point on the map.



Demographics of respondents

Patients were aged between 6 and 19 years, with 58% of white ethnicity and 26% Asian. They had most commonly (65%) been treated at the hospital for over ten years, reporting a mode of two visits in the last 12 months (39%). Carer responses pertained on average to slightly younger patients, 30% caring for those 1-5 years old and 9% to those under 12 months of age, 78% white and 13% Asian. Carers most frequently reported 6-10 years at the hospital (27%), and a similar mode of two visits in the last year (28%).

Clinicians were most commonly consultants (82%), alongside 12% nurses and 6% doctors in training. Most had cared for children with CAH for over 10 years (56%). They reported having treated less than 10 patients in the last 12 months in 50% of cases, 11-20 patients in 27% and over 20 in 23%.

1106 Table 3.1 – Demographics and job role of respondents

Patients Percentage of responses (n/total responses)	**Parents** Percentage of responses (n/total responses)	**Clinicians** Percentage of responses (n/total responses)
Number of respondents:		
43	152	34
Number of centres:		
27 (1 unknown)	34 (2 unknown)	18 (0 unknown)
How would you describe the sex of the patient?		
Male: 44.2% (19/43) Female: 55.8% (24/43) Prefer not to say: 0.0% (0/43)	Male: 46.7% (71/152) Female: 52.6% (80/152) Prefer not to say: 0.7% (1/152)	
How long have you been treated at this hospital for CAH?		**How long have you been treating children with CAH?**
1 visit: 3.2% (1/31) < 12 months: 0.0% (0/31) 1-2 years: 6.5% (2/31) 3-5 years: 12.9% (4/31) 6-10 years: 12.9% (4/31) >10 years: 64.5% (20/31)	1 visit: 2.0% (3/152) < 12 months: 11.2% (17/152) 1-2 years: 15.1% (23/152) 3-5 years: 24.3% (37/152) 6-10 years: 27.0% (41/152) >10 years: 20.4% (31/152)	1-2 years: 5.9% (2/34) 3-5 years: 14.7% (5/34) 6-10 years: 23.5% (8/34) >10 years: 55.9% (19/34)
Overall, are you satisfied with the service provided for CAH?		
Completely dissatisfied: 2.4% (1/42) Somewhat dissatisfied: 0.0% (0/42) Neutral: 2.4% (1/42) Somewhat satisfied: 19.0% (8/42) Completely satisfied: 76.2% (32/42)	Completely dissatisfied: 0.7% (1/151) Somewhat dissatisfied: 1.3% (2/151) Neutral: 2.6% (4/151) Somewhat satisfied: 27.8% (42/151) Completely satisfied: 67.5% (102/151)	Completely dissatisfied: 0.0% (0/34) Somewhat dissatisfied: 0.0% (0/34) Neutral: 11.8% (4/34) Somewhat satisfied: 76.5% (26/34) Completely satisfied: 11.8% (4/34)

1107

Service satisfaction and confidence in management

Both patients and carers were satisfied with services in 95% of cases. Patients reported being 'very confident' managing the condition in 71% of cases, with carers 'very confident' managing CAH day to day in 85% of cases. Frequency of carers reporting 'very confident' decreased to 56% during illness, and 40% in an emergency, although only 10% 'not confident' in an emergency. A slight majority of patients (57%) and carers (52%) reported that they were happy with the amount of knowledge they possessed regarding the medication used to treat CAH, the remainder either unsure or wanting more information.

Frequency of follow up was deemed appropriate by 93% of patients and 78% of carers. Most patients (76%) and carers (48%) had not spoken to anyone else with CAH. 90% of patients who had discussed the condition reported they had received the same advice as others, whereas 74% of carers who had spoken to other families in the last 12 months reported that they had received conflicting advice. Clinicians also reported that patients may often or always receive conflicting advice in 59% of cases. Only 30% felt their patients would be managed the same if treated at another hospital, and 54% of carers were unsure if their management would be the same at a different hospital.

Access to the multidisciplinary team

All clinicians reported good access to clinical nurse specialists in endocrinology, but more difficulty with other allied health specialties. Psychologists were either impossible or difficult to access in 44% of cases, dietitians in 71% and physiotherapists in 68%. Access to genital surgery was possible in 77% of responses, but analysis of the free text comments showed that whilst available in some cases, surgery was not currently being offered due to differences of opinion of its application to differences of sexual development in the UK. Clinicians wanted better access to ancillary services in 21% of cases, more frequent patient appointments in 14% and the ability to do more tests in 12%. In contrast, patients and carers reported that they had the right number of appointments in 93% and 78% of cases respectively (Appendix Figure S3.1).

Education about CAH

Over 94% of clinicians reported they delivered formal education about CAH, but 81% of patients and 83% of carers reported that they had not received formal education. Only 19% of patients and 1% of carers reported that they would not attend a formal education course if it were available (Table 3.2, Appendix Figure S3.2). Exploring conflicting advice about CAH, 68% of patients and 73% of carers reported that they had not had varying advice from professionals

about CAH, although 24% of patients and 27% of carers reported this had happened in the last 12 months. There were 11 of 56 free text comments pertaining to education written by carers, all calling for better education about the condition (Appendix Table S3.3).

Clinicians universally reported nurses being involved in education, along with doctors in 72% of cases and psychologists in only 16%. In 32% of clinician responses the education lasts less than an hour, and the remaining 68% less than four hours. Only 47% of clinicians reported always ensuring school is contacted directly for staff education about CAH.

Monitoring of patients with CAH

Figure 3.2 shows the frequency of clinician reported metrics assessed in clinic alongside carer reports of what was measured at last appointment. Auxology and blood pressure are measured commonly, with blood tests measured at least once a year by 61% of clinicians, more frequently than bone age assessed by 42%. Urinary steroids were measured least frequently, reported as 'never measured' by 41% of clinicians. Use of dried blood spot 17-Hydroxyprogesterone was reported by 49% of patients and 66% of clinicians. Saliva sampling was not specifically asked but featured in free text comment answers pertaining to three hospitals as a method of monitoring employed.

Figure 3.3 shows the frequency of topics discussed in clinics, with goals of management and sick day rules being discussed reported most frequently by both carers and clinicians. Carers reported they had less frequent discussions about the psychological effects (51%) and physical effects (63%) of CAH. Fertility and heart disease were the least reported topics of discussion by both clinicians and carers.

1159 Table 3.2 – Education and confidence in managing CAH

Patients Percentage of responses (n/total responses)	**Parents** Percentage of responses (n/total responses)	**Clinicians** Percentage of responses (n/total responses)
Have you attended formal training about CAH?		**Do you provide formal training for CAH?**
No: 80.6% (25/31) Unsure: 9.7% (3/31) Yes: 9.7% (3/31)	No: 82.8% (125/151) Unsure: 3.3% (5/151) Yes: 13.9% (21/151)	No: 0.0% (0/34) Not sure: 5.9% (2/34) Yes: 94.1% (32/34)
Would you attend formal training about CAH if it were available?		**How long is the training that you provide?**
Yes: 54.8% (17/31) No: 19.4% (6/31) Unsure: 25.8% (8/31)	Yes: 85.5% (130/152) No: 1.3% (2/152) Unsure: 13.2% (20/152)	<1 hour: 32.3% (10/31) 1-4 hours: 67.7% (21/31)
Do you think you know enough about the medications you use for CAH?		**What professionals perform the training?**
I don't take medication: 0.0% (0/42) No, I want to know more: 7.1% (3/42) Unsure: 11.9% (5/42) Yes, but want to know more: 23.8% (10/42) Yes: 57.1% (24/42)	Doesn't take medication: 0.7% (1/152) No, I want to know more: 4.6% (7/152) Unsure: 0.0% (0/152) Yes, but want to know more: 42.8% (65/152) Yes: 52.0% (79/152)	Doctor: 71.9% (23/32) Nurse: 100.0% (32/32) Youth worker: 3.1% (1/32) Pharmacist: 3.1% (1/32) Psychologist: 15.6% (5/32) Dietitian: 3.1% (1/32)
Do you feel confident managing your CAH?	**Do you feel confident managing your child's CAH day-to-day?***	**Do you think patients and families are confident managing CAH?**
No: 0.0% (0/31) Somewhat confident: 29.0% (9/31) Very confident: 71.0% (22/31)	No: 0.0% (0/151) Somewhat confident: 14.6% (22/151) Very confident: 85.4% (129/151)	No: 0.0% (0/34) To some extent: 88.2% (30/34) Yes, definitely: 11.8% (4/34)

1160 *Carers were also asked if they felt confident when the child was unwell (No: 1.5% (2/135), somewhat confident:
1161 42.2% (57/135), very confident: 56.3% (76/135)) and in an emergency (No: 10.1% (14/135), somewhat confident:
1162 49.6% (69/135), very confident: 40.3% (56/135))

Figure 3.2 – Frequency of assessment of patient metrics

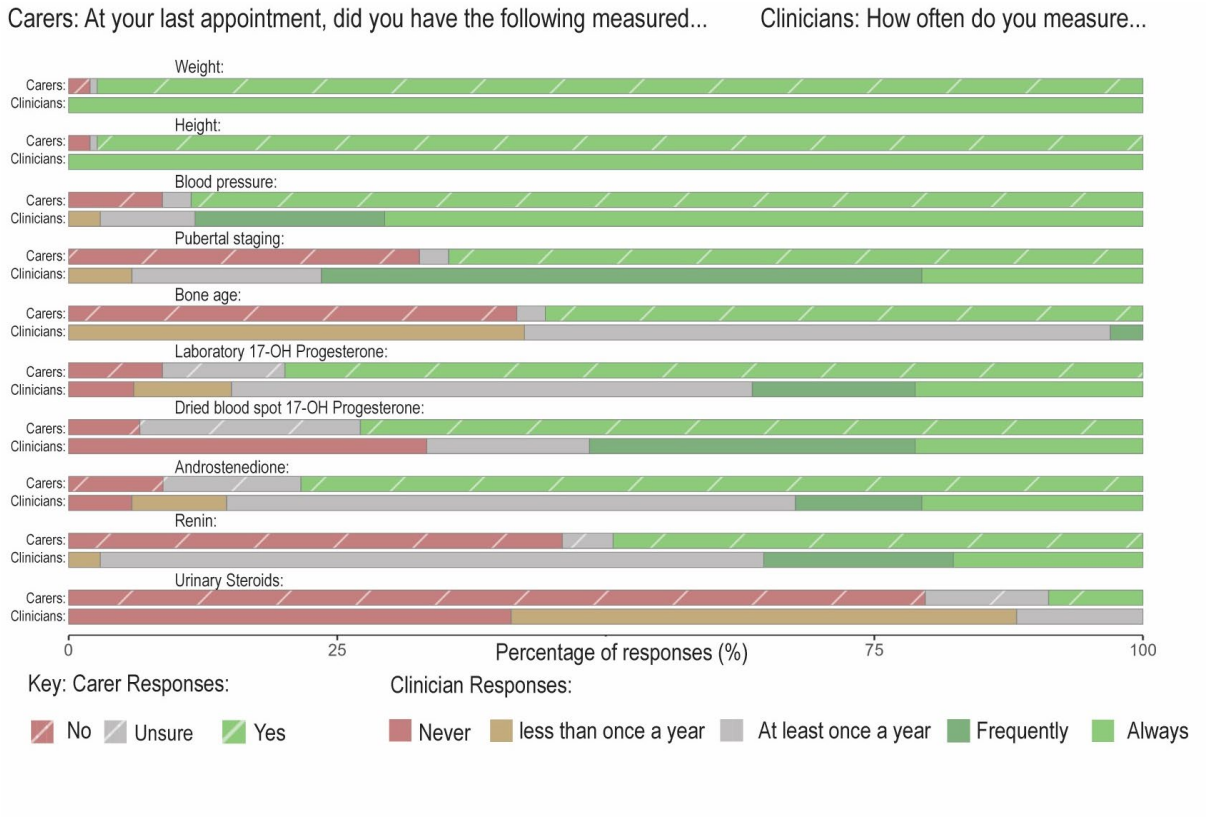
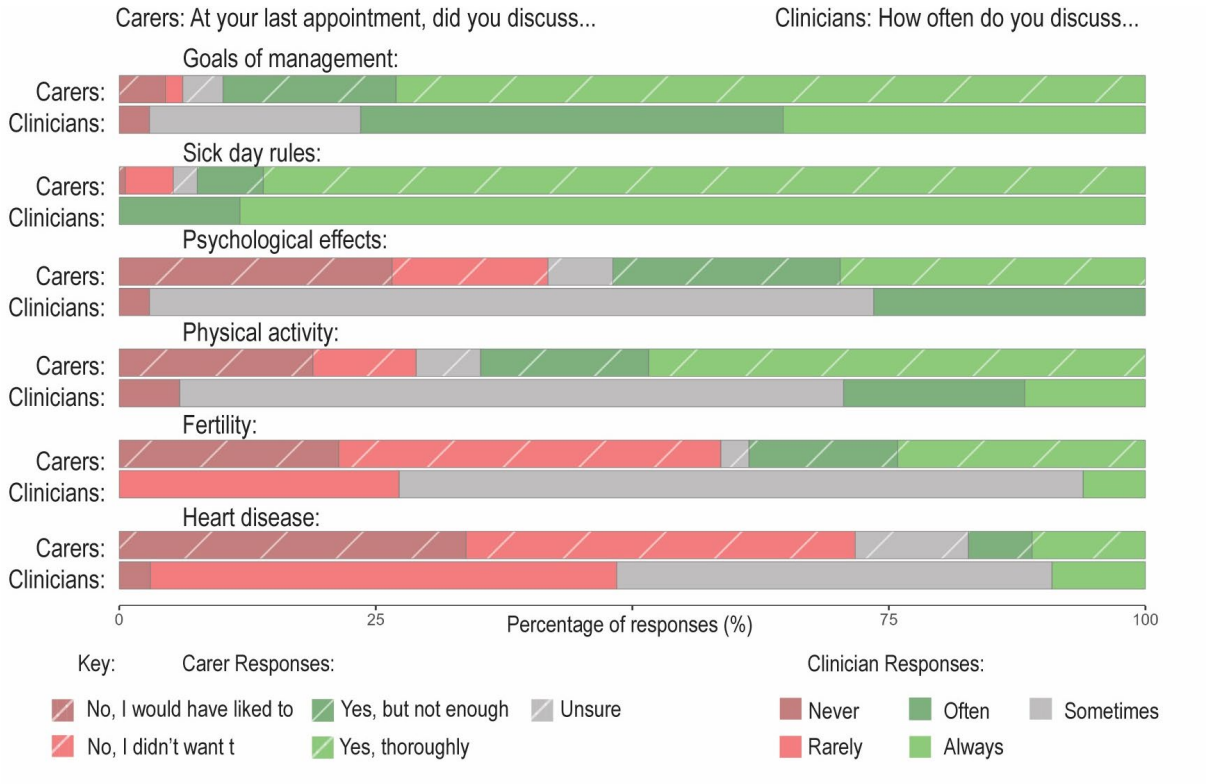


Figure 3.3 – Frequency of topics discussed in clinic



Thematic analysis of the free text comments

The most common type of comment related to the standard of care, featuring in 40% of patient comments, 46% of carer comments and 76% of clinician comments. Carer comments about standard of care spanned the entire spectrum from 21% that were overtly positive about the standard of care to 16% that were overtly negative. Education was discussed next most frequently, in 20% of patient and carer comments but only 4% of clinician comments. A request to have access to 24-hour blood profiles measured was reported by four respondents, and access to modified-release hydrocortisone in one comment. Pertinent quotes from comments are in Appendix Table S3.3.

3.06 Discussion

We conducted a national service evaluation into the care of children with CAH in the UK. Overall satisfaction with services was very good, although there is significant appetite for better education about the condition and greater psychological support for patients and their families.

International consensus guidelines, updated in 2018, contains recommendations for follow up and management of children with CAH (5). Responses received in relation to patient assessment are broadly consistent with this guidance, both patients and parents reporting regular measurements of auxology and blood pressure. Bone age was reported as measured less frequently than the recommended once a year by over 40% of clinicians. The greatest variation in monitoring was seen in the frequency of use of blood biomarkers. However, it must be noted that those reporting 'never' using laboratory 17OHP did report using either dried blood spot 17OHP or salivary 17OHP, showing that measurement of 17OHP is certainly the preferred biochemical marker of disease control.

Whilst access to specialist nursing was good across all clinicians surveyed, there is difficulty in accessing important allied health professionals including psychologists, dietitians, and physiotherapists. Appropriate multidisciplinary working is vital for the proper management of paediatric CAH and has been shown to improve medication compliance (120-122). Whilst clinicians did report overall satisfaction with services, almost half wanted to be able to offer more, be that more tests, more appointments or more allied services. The need to improve care provision also came through strongly in free text comments with heed paid to psychological support, including the quote 'psychology [services are] not adequate anywhere'. Recent findings from the CAH-UK initiative that assessed the quality of life of patients with CAH showed increasingly impaired QoL with more detrimental measures of body habitus, highlighting the importance of giving children with CAH the appropriate support to improve their psychological and physical health (15, 116). Lack of access to surgical services also comprised two free text comments from patients and carers, although the difficulties and controversies around genital surgery were highlighted by three comments from clinicians, reinforcing the importance of further research into understanding both whether and when to offer genital surgery to patients with CAH.

There was a significant discrepancy between what patients and families consider to be formal training in the management of the disease, and what clinicians report as formal training. All the free text comments about education from carers were describing an appetite for more or better education, and the vast majority of both patients and carers reported that they had not received what they would consider formal training. Whilst clinicians report formal training almost universally, they admit to this being either less than an hour, or less than four hours in duration. The requirement to administer emergency steroid injections is rare, and thus regular repeated education sessions at different stages throughout childhood may be more appropriate. More attention to patient education, or more structured training courses is likely to help to improve patient and carer confidence in the management of emergencies, as well as other aspects of the disease.

It is over 20 years since the diet adjusted for normal eating course was shown to improve both biochemical control of diabetes as well as the quality of life of participants (123). The vast majority of both patient and parent respondents to our questionnaire reported that they would attend a similar course structured towards CAH, with just under half reporting that they would like to know more about the medications they take for CAH, and frequent free text comments calling for further education. Confidence in managing the condition was high, although it should be noted that this decreases in acute illness and was lowest in emergencies. As such, further research should be targeted at creating a standardised programme with resources that can be shared nationwide to improve the education of the families of children and young people living with CAH.

Transition to adult services was not specifically questioned within the survey as this would only be applicable to a small section of the respondents. Nonetheless, five unprompted free text answers were focused on transition and a reminder that this is such a crucial stage in the management of chronic disease that requires appropriate support. Loss to follow up is most likely to occur when patients transition, and anxieties around transition have consistently been highlighted in patient and public involvement work carried out at our centre. The importance of the multidisciplinary team and the timing of further, age appropriate education at this crucial time cannot be underestimated (124). Further research into optimum transition care strategies is needed to reduce the high numbers of up to half of patients lost to follow up or with poor control that has been shown in the UK after transition (125).

This study has limitations related to the difficulty in assessing quality of care in a rare disease. Data were collected over 18 months, and collection began whilst services were still recovering from the difficulties of a global pandemic, the impact of which was significant (126), with difficulties in accessing services during the pandemic highlighted in one of the free text comments. There was a good geographical spread of responses from all of patients, clinicians, and carers, despite not all centres across the country taking part. Whilst the number of responses from each centre was variable, statistical weighting of the responses by the catchment area of the hospital and then number of clinicians from each centre that responded is a robust sensitivity analysis and showed no significant change in the proportion of responses to any of the questions.

Free text comments from families highlighted the benefits of conducting this service evaluation, and the importance of giving patients and families a voice to express their preferences. Audit of practice is highlighted as an important first step in the improvement of services, but should be followed by appropriate actions and further work to improve patient care (127). Results from this survey should be used to advocate for increased resources locally, as well as to direct future research into the best ways to support and educate patients and families with CAH.

3.07 Conclusion

Whilst overall satisfaction with paediatric CAH services in the UK is good, provision of education for those living with the disease is widely perceived as an unmet need. Improved access to allied health professionals and psychologists will help support families and improve patient outcomes. Further research to help standardise approaches to management and transition, and to develop effective educational resources is indicated.

3.08 Acknowledgements

This work would not be possible without the patients and the parents of the children with CAH, as well as the clinicians, who were kind enough to respond to this survey.

3.09 Author Contributions

NPK conceived the project and manuscript outline. NRL, IB and NPK conceived the questions, that were approved by the remaining authors within the BSPED Adrenal Special Interest Group after patient and public involvement involving SE and SB. NL performed data analysis with statistical support and help with data visualisation from GC, JD and ZL. SFA, SA, LB, JB, TC, LZ, JHD, MD, EG, RK, LP, AT, TR, FR, SE, SB facilitated data collection and provided support with data interpretation and production of the manuscript. All authors have contributed to improvements and approved the final version of the manuscript. NPK supervised the study.

3.10 Thesis contribution of chapter 3

A national service evaluation has been completed and published in *Hormone Research in Paediatrics* to achieve objective 2 (115). This evaluation showed care of children with CAH in the UK is considered favourable by families and clinicians, but improvements can be made in education about the disease. The evaluation provided objective evidence that whilst research is ongoing into novel biomarkers of disease in CAH, the majority of clinicians still focus on the measurement of 17-Hydroxyprogesterone and androstenedione when assessing disease control. As international guidance recommends (5), the measurement of blood pressure and growth is common to inform treatment strategies. This stakeholder engagement directed further analysis into these commonly used biomarkers, as well as blood pressure and growth in children with the disease in subsequent chapters (112-114), and informed the metrics to include within the prototype clinical decision support tool presented in chapter 7.

Chapter 4 – Modelling adrenal steroid profiles to inform monitoring guidance in congenital adrenal hyperplasia

4.01 Publication and contribution of others

This chapter was published in EBioMedicine and can be accessed at (https://doi.org/10.1016/j.ebiom.2025.105749) (114). Co-authors outside of the supervisory team contributed through coordinating the original phase 3 clinical trial, alongside review and feedback of the manuscript. The original draft of the manuscript and incorporation of feedback and improvements was done by Neil Lawrence. All authors agreed on the final version of the manuscript (full contributions in section 4.10).

Chapter 4 Co-authors:

Neil R. Lawrence[1], Jeremy Dawson[1], Zi-Qiang Lang[2], Alessandro Prete[3,4], Elizabeth S. Baranowski[3], Lina Schiffer[3,5], Angela E. Taylor[3], Aude Brac de la Perriere[6], Angelica Lindén Hirschberg[7], Anders Juul[8,9], Deborah P. Merke[10], John D. C. Newell-Price[1], D. Aled Rees[11], Nicole Reisch[12], Monica Stikkelbroeck[13], Philippe A. Touraine[14], Nils Krone[1], Brian Keevil[15], Gary S. Collins[16], Wiebke Arlt[5,17], Richard J. M. Ross[1]

Chapter 4 Author affiliations:

1: School of Medicine and Population Health, University of Sheffield, Sheffield, United Kingdom
2: School of Electrical and Electronic Engineering, University of Sheffield, Sheffield, United Kingdom
3: Department of Metabolism and Systems Science, School of Medical Sciences, College of Medicine and Health, University of Birmingham, United Kingdom
4: National Institute for Health and Care Research (NIHR) Birmingham Biomedical Research Centre, United Kingdom
5: MRC Laboratory of Medical Sciences, London, UK
6: Hopital Louis Pradel, Bron, France
7: Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden
8: Department of Growth and Reproduction, Copenhagen University Hospital - Rigshospitalet, Copenhagen, Denmark
9: Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark
10: National Institutes of Health Clinical Center and the Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD, USA
11: Cardiff University, Cardiff, United Kingdom
12: Medizinische Klinik und Poliklinik IV, Klinikum der Universität München, Munich, Germany
13: Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands
14: GH Pitie Salpetriere, Paris Cedex 13, France
15: Manchester University Hospital NHS Trust, Manchester, UK
16: Centre for Statistics in Medicine, University of Oxford
17: Imperial College London, London, UK

4.02 Abstract

Background

The recent development of modified-release hydrocortisone (Chronocort) gives the potential for disease control throughout 24 hours and the need to develop a rational regimen for monitoring patient's biochemical disease control.

Aim

To define the relationship between the commonly used markers of CAH biochemical control, 17OHP and androstenedione, in healthy controls and CAH patients.

Methods

Analysis of the 24hour 17OHP and androstenedione endocrine profiles in patients randomised to either standard glucocorticoid replacement therapy or Chronocort in a phase 3 clinical trial. Flexible splines were fit to carry out timeseries analysis of profiles and calculate area under the curve. Cross correlation and autocorrelation were used to assess rhythmicity and lag between markers. Bayesian multiple change point analysis was employed to assess the linearity of the relationship between 17OHP and Androstenedione.

Results

Healthy patients had a different relationship between androstenedione and 17OHP than those with CAH, having lower levels of 17OHP for levels of androstenedione. Area under curve of both 17OHP and androstenedione was lower in CAH patients treated with Chronocort, and lower in female patients with CAH than males. Cross-correlation of 17OHP and androstenedione showed that changes in both markers occurred at the same time with no time lag. The variation of 17OHP over time in individuals was greater than that of androstenedione. The relationship between 17OHP on androstenedione is log linear, but has a change point at 1.5nmol/L (95% confidence interval 1.4 to 1.6) of androstenedione and 4.5nmol/L (4.2 to 4.7) of 17OHP, after which higher levels of androstenedione are associated with a proportionally higher level of 17OHP on the natural log scale.

Conclusions

The relationship between androstenedione and 17OHP is not the same in CAH patients and healthy controls, with androstenedione levels either low or undetectable when 17OHP levels are at levels similar to healthy controls. This is likely because androstenedione is generated from 17OHP in uncontrolled CAH and from dehydroepiandrosterone in healthy subjects. Further investigation into the relationship between profile measurements at different times of day will inform the optimum time of day to assess point measurements of this biomarkers.

4.03 Research in context

Evidence before this study

We searched PubMed from inception to Mar 10, 2024, using the terms “congenital adrenal hyperplasia” AND “17-OH progesterone” OR “17OHP” OR “androstenedione” OR “11-ketotestosterone” OR “11-KT” OR “monitoring”. No language restrictions were applied. From the papers and reviews identified, there is no consensus on what hormones to measure and at what time of day to measure them when monitoring the biochemical control of CAH. The international guidelines simply state: “monitoring treatment through consistently timed hormone measurements relative to medication schedule and time of day.” There is therefore a need to define a rational monitoring strategy for patients with CAH.

Added value of this study

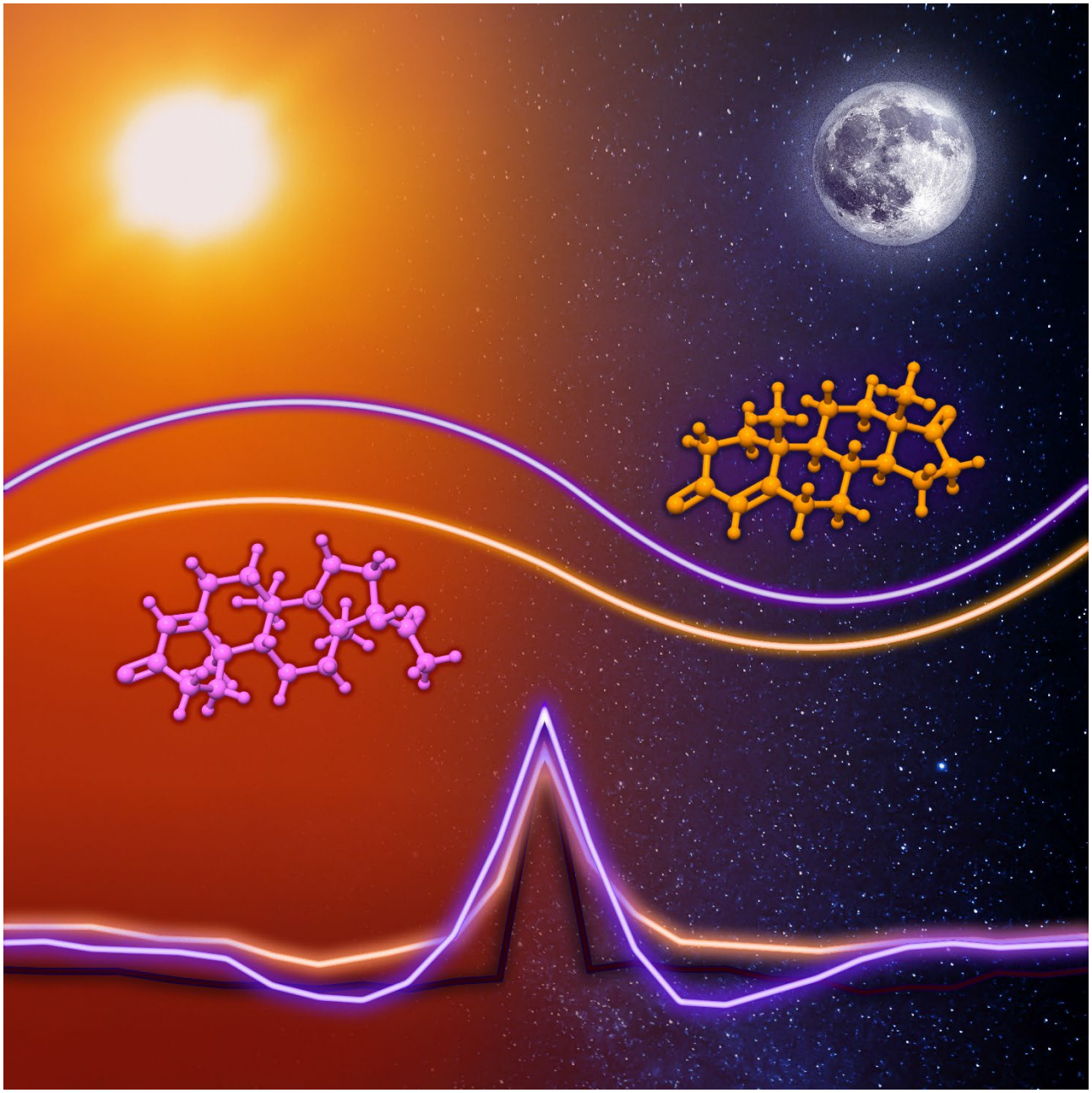
We have analysed and modelled the largest dataset of steroid profiles ever to be published with the aim of rationalising monitoring regimens in CAH. The 122-person multicentre phase 3 study of modified-release hydrocortisone (MRHC) versus standard treatment with repeated visits (4 x 24-hour profiles per participant) provides us with the largest dataset of CAH steroid profiles ever collected and is unlikely to ever be repeated. The pattern of steroid profiles within patients is strikingly similar, with cross correlation between them evidencing no time lag between the two most commonly used biomarkers of disease control; 17-Hydroxyprogesterone (17OHP) and androstenedione. The changepoint analysis we have conducted alongside the visualisation of profiles shows that androstenedione is more likely to be low, whilst 17OHP can still show elevated levels. The other androgens, including testosterone in women and 11-oxygenated androgens in both sexes, correlated with changes in 17OHP and androstenedione. The improved androgen control over 24 hours on MRHC shows that a single measurement during the day reflects disease control.

Implications of the available evidence

Our modelling analysis allows us to define monitoring regimens for patients with CAH. In CAH, androstenedione levels are proportionally lower for the same 17OHP level than those observed in healthy individuals. This means an androstenedione above the reference range demonstrates poor control, but a normal androstenedione does not mean tight control, and a low androstenedione does not imply over-treatment. A 17OHP <38nmol/l (1250 ng/dl) was associated with an androstenedione in the normal range <5nmol/l (143 ng/dl) in 95% of patients and in clinical trials has been used to reflect good control. On MRHC, which controls androgen levels over 24 hours, a single daytime sample of 17OHP and/or androstenedione is sufficient to monitor biochemical control.

4.04 Cover Art

Following submission to *EBioMedicine*, the work in chapter 4 was deemed appropriate to be included on the journal cover. The following cover was conceptualised and designed by Neil Lawrence initially rendered in Adobe Illustrator, with refinement in Adobe photoshop by Peter Drury from the media engagement team at the University of Sheffield. Unfortunately, the artwork was not selected for use as the final cover art on that month's issue of the journal.



4.05 Introduction

Congenital adrenal hyperplasia (CAH) is the commonest genetic endocrine disorder (128), and mutations in the *CYP21A2* gene encoding the enzyme 21-hydroxylase (21-OHD) account for approximately 95% of cases (129, 130). Deficiency in 21-hydroxylase blocks cortisol synthesis resulting in reduced cortisol feedback and consequently increased pituitary adrenocorticotropic hormone (ACTH) release, which in turn promotes the over-production of adrenal androgens. Patients with CAH therefore have two major problems: cortisol deficiency and androgen excess. In addition, many patients also have mineralocorticoid deficiency as 21-hydroxylase mediates a key step in aldosterone synthesis.

Androgens are generated through three pathways (131): the classic androgen pathway via dehydroepiandrosterone (DHEA) to testosterone (T) and dihydrotestosterone (DHT); the 11-oxygenated androgen pathway initiated by the conversion of androstenedione to 11-Hydroxyandrostenedione (11OHA4) and yielding the active 11-ketotestosterone (11KT); and the alternative pathway to DHT, that in health is active in the testis during male development in the fetus and the early neonatal period (132), but not active in childhood and adults. In CAH, androgen biosynthesis is upregulated due to the accumulation of 17-Hydroxyprogesterone (17OHP) prior to the 21-hydroxylase enzyme block. Excess 17OHP results in atypical, enhanced conversion of 17OHP to androstenedione by CYP17A1 17,20-lyase activity, which physiologically has a much higher preference for the conversion of 17OH-pregnenolone to DHEA. Accumulating 17OHP also drives increased androgen production via the alternative DHT pathway, and increased androstenedione feeds enhanced 11-oxygenated androgen pathway activity. In CAH patients on treatment, DHEA levels are usually low (133), as androstenedione is generated atypically from 17OHP rather than from DHEA; however, the precise regulatory mechanisms underlying this relative DHEA deficiency are not known.

The goals of treatment in CAH are to replace cortisol deficiency and control androgen excess. The major challenge to achieving these goals is the excessive rise in ACTH overnight as the lack of cortisol feedback is unopposed in patients receiving conventional, immediate-release glucocorticoids. As a consequence, patients often undulate between glucocorticoid overtreatment and androgen excess (7, 134). This poses a problem in deciding what to measure when monitoring glucocorticoid treatment in CAH patients. The latest guidelines recommend monitoring treatment through consistently timed hormone measurements relative to medication schedule and time of day (34). Traditionally 17OHP and androstenedione have been used as markers of adrenal androgen control, but it is now recognised that the 11-

oxygenated adrenal androgen pathway contributes to the adrenal androgenic activity (135); however, their measurement is not available in most clinical centres. Fundamentally, there is no consensus on what hormones to measure, when, how often, and how to interpret their results to optimize the biochemical control of patients with CAH.

Modified-release hydrocortisone hard capsules (MRHC), development name Chronocort (Efmody®, Diurnal, UK) is now available in Europe. Taken at bedtime, hydrocortisone is released during the early morning hours, resulting in an overnight and early morning rise in cortisol that resembles physiological profiles (36). As part of the phase 3 clinical trial (36), detailed 24-hour serum steroid profiles of 17OHP and androstenedione were collected on standard treatment and MRHC and in a sub-cohort of patients diurnal salivary samples were collected for steroid analysis by tandem mass spectrometry. This provides a unique dataset facilitating a detailed analysis of the steroid profiles in patients with CAH on standard treatment and MRHC, respectively. We compared these data to 24-hour steroid profiles measured in healthy participants to develop rational biochemical monitoring protocols in CAH.

4.06 Methods

Patients: Patients had classic 21-OHD CAH diagnosed in childhood, adequate mineralocorticoid replacement with renin less than 2 times the upper limit of normal, and were on stable glucocorticoid therapy over the preceding six months. Exclusion criteria included the use of medication interfering with glucocorticoid metabolism, bilateral adrenalectomy and night shift work. The study protocols for the phase III extension study were approved by local Ethics/Institutional Review Boards and the Medicines and Healthcare Products Regulatory Agency (NCT03062280, Eudract 2015-005448-32). The trials were performed in accordance with the principles of the Declaration of Helsinki. We screened 138 patients, 122 were randomised, and 117 completed the study, with 477 24-hour steroid profiles available for analysis (Table 4.1). Patients underwent blood sampling for 24-hour profiles at baseline, 4, 12, and 24 weeks; blood samples for measurement of 17OHP and androstenedione were taken every 2 hours from 15:00h to 15:00h the next day, with serum testosterone measured at 07:00h. A sub-cohort of 12 patients (9 females) from a single centre (Birmingham, United Kingdom) underwent 2-hourly saliva sampling by passive drooling between 07:00h and 23:00h. Patients, stratified by baseline glucocorticoid treatment, were randomized to receive MRHC or continue on standard glucocorticoid therapy. MRHC was prescribed as 5, 10 or 20 mg capsules and the initial dose was the hydrocortisone dose equivalent to their baseline therapy (5x for prednis(ol)one and 80x for dexamethasone) with approximately $^{1}/_{3}$ of the daily dose taken at 07:00h and $^{2}/_{3}$ of the daily dose taken at 23:00h. Standard treatment was a mixture of regimens that included

hydrocortisone, prednisolone, prednisone and dexamethasone. At baseline, 84% of patients were taking a dose of standard glucocorticoid after 18:00h, and 84% of patients were diagnosed as salt-wasting. At 4 and 12 weeks, dose titrations were made for both treatment groups, using identical rules, following centralized advice by two independent physicians blinded to all data except 24-hour steroid profiles and an investigator-completed adrenal insufficiency checklist. The blinded titrators considered morning and/or evening dose adjustments of either MRHC or standard glucocorticoid using 17OHP/ androstenedione measurements from the 24-hour profile and adrenal insufficiency symptom questionnaire.

Healthy Participants: The healthy control group comprised data from 20 healthy participants from a cohort previously reported (136) (Table 4.1). These 20 participants underwent 24-hour frequent serum sampling, with blood drawn in 20-minute intervals from 09:00h through to 09:00h the following day.

Steroid analysis: Serum steroids (17OHP, androstenedione, testosterone) in the serum samples from the phase 3 study were measured by liquid chromatography-tandem mass spectrometry (LC-MS/MS; Q2 Solutions, USA). Serum 17OHP, androstenedione, and testosterone were measured in the healthy controls by LC-MS/MS as previously described (137). Salivary steroids (17OHP, androstenedione testosterone, 11OHA4, 11KT) were measured by Liquid chromatography-tandem mass spectrometry (LC-MS/MS) as previously described (138).

Statistical Analysis: Data analysis was carried out in *R: A Language and environment for statistical computing* (R Foundation for Statistical Computing, Vienna, Austria), packages employed listed in Appendix Table S4.1. Missing profile measurements were interpolated between points within each profile or imputed with the adjacent value if the missing point was at the beginning or end of the profile. Profiles were modelled using a cubic smoothing spline function that calculated the area under the curve (AUC). Natural log (Ln) transformations were carried out to achieve normality of skewed variables. Linear mixed effects models with sine and cosine terms were employed to quantify the circadian rhythm of markers. Bayesian multiple change point analysis was employed on log-transformed values to assess for variation in the relationship between markers at different concentrations. Model fit was assessed by $\hat{R}$, a statistic indicating a suitable representation of the data at values between 1.0 and 1.1 (139).

Table 4.1 – Patient demographics and summary statistics of serum 17-Hydroxyprogesterone and androstenedione

	Healthy participants	All CAH Patients	CAH patients on standard glucocorticoid replacement*	CAH patients on Chronocort
Number of patients	20 Male = 13, Female = 7	122† Male = 44, Female = 78	61 Male = 25, Female = 36	61 Male = 19, Female = 42
Age: mean years (standard deviation)	28.4 (12.3)	36.3 (11.6)	37.5 (12.8)	35.2 (10.3)
Number of 24-hour profiles	20‡	477	301	176
Number of readings	1440§	6202	3914	2288
Serum 17OHP (nmol/l) from 2-hourly sampling over 24 hours				
Mean	3.7	33.7	43.1	17.5
Standard deviation	3.0	79.5	93.9	40.7
Median	2.7	5.1	6.7	3.5
Quartile 1 to quartile 3	1.5 to 4.9	1.8 to 24.3	2.2 to 33.6	1.2 to 11.9
Amplitude of cosinor model ♯	0.44 (95% CI: 0.33 to 0.54)	-	7.73 (95% CI: 7.67 to 7.79)	0.83 (95% CI: 0.77 to 0.90)
Serum androstenedione (nmol/l) from 2-hourly sampling over 24 hours				
Mean	4.0	3.6	4.4	2.4
Standard deviation	3.6	6.1	7.1	3.4
Median	3.0	1.5	1.7	1.4
Quartile 1 to quartile 3	2.1 to 4.6	0.7 to 3.5	0.7 to 4.4	0.6 to 2.7
Amplitude of cosinor model ♯	0.71 (95% CI: 0.64 to 0.78)	-	0.50 (95% CI: 0.47 to 0.54)	0.03 (95% CI: 0.01 to 0.05)

*Summary statistics for biomarker readings are calculated including readings from patients that were later randomised to take Chronocort, but at the time of their first profile, were still taking standard glucocorticoid replacement

† 117/122 patients completed all four study visits

‡ 1/20 patients only had profiling data available for Androstenedione, and not 17OHP

§ 1368 values available for 17OHP as one patient only had profiling of Androstenedione conducted

♯ full cosinor model parameters in supplementary table # linear mixed effects model. Confidence intervals estimated by calculating across 1000 bootstrap replications

17OHP=17-Hydroxyprogesterone

4.07 Results

Serum 17OHP and androstenedione: The 24-hour absolute levels of serum 17OHP in nmol/l overlaid on androstenedione 10*x*nmol/l (to allow comparison as in CAH absolute levels of androstenedione are ~10x lower than 17OHP) showed similar profiles, with changes in 17OHP and androstenedione paralleling each other in the CAH patients (Figure 4.1). The same was true for the profiles from healthy volunteers although the absolute levels of 17OHP and androstenedione were of the same magnitude and generally lower (Table 4.1). In healthy participants, there was a low amplitude of 17OHP (0.4 nmol/l) and Androstenedione (0.7 nmol/l) (Appendix Supplementary Analysis 4.1, Appendix Table S4.2). In contrast, patients with CAH had large variability in serum 17OHP levels exhibiting a diurnal rhythm with an amplitude of 7.7 nmol/l and peak at 07:54h in those taking standard glucocorticoid replacement. During the trial, control improved in both arms but more markedly in the MRHC arm and the amplitude of the 17OHP rhythm reduced to 0.8nmol/l in those on MRHC. Similar results with lower absolute levels were seen for androstenedione.

Relationship between serum 17OHP & androstenedione levels: In both healthy participants and patients with CAH, there was no lag on cross-correlation of 17OHP with androstenedione confirming that they change in parallel (Figure 4.2). In healthy participants, mean 24-hour 17OHP and androstenedione levels were very similar, whereas in CAH patients 17OHP levels were generally much greater than those of androstenedione (Table 4.1). In CAH patients, Bayesian multiple change point analysis assessing ln 17OHP on ln androstenedione converged on a changepoint of 17OHP at 4.5 nmol/l (149 ng/dl) (95%, CI:4.0 to 5.0 nmol/l, $\hat{R} < 1.007$) (Figure 4.3, Supplementary Analysis 4.2, Appendix Tables S4.3-S4.4). Below the changepoint, there was a greater decrease in androstenedione for a similar decrease in 17OHP. The changepoint did not differ between either male or female patients or whether on MRHC or conventional, immediate-release glucocorticoid. The relationship between serum 17OHP and androstenedione was different for healthy participants who, for similar levels of 17OHP, had higher levels of androstenedione than CAH patients. This means that in CAH patients, when the androstenedione was at the upper limit of normal, the 17OHP was elevated, and when the 17OHP was in the normal range, the androstenedione levels were suppressed: a 17OHP <38nmol/l (1250 ng/dl) was associated with an androstenedione in the normal range <5nmol/l (143ng/dl) in 95% of patients.

Figure 4.1 – Profiles of 17OHP and androstenedione in healthy participants and CAH patients
These profiles show the relationship between 17OHP and androstenedione within individuals. For CAH patients 17OHP in nmol/l is overlaid on androstenedione 10xnmol/l to allow comparison as in CAH absolute levels of androstenedione are ~10x lower than 17OHP whereas in healthy participants they have similar absolute levels of 17OHP and androstenedione. 17OHP=17OH-Progesterone

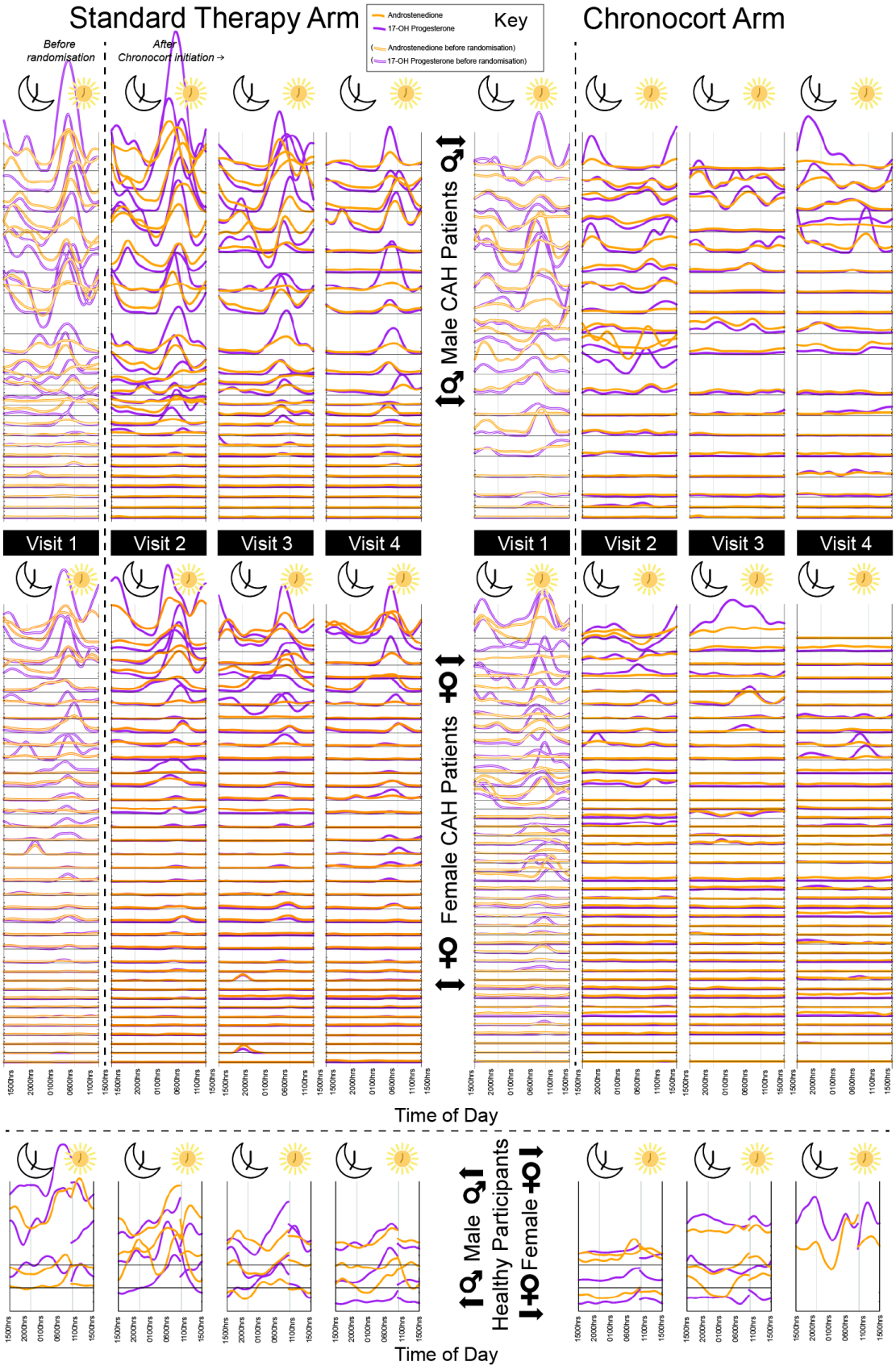


Figure 4.2 – Cross-correlation plot of 17OHP and androstenedione in healthy participants and CAH patients

This cross-correlation shows there is no lag between markers, with peak at zero demonstrating that the markers change in parallel over time. Points show mean of the estimate within groups, error bars show standard error of the mean, grey lines in background show individual patient cross-correlation plots. 17OHP=17OH-Progesterone

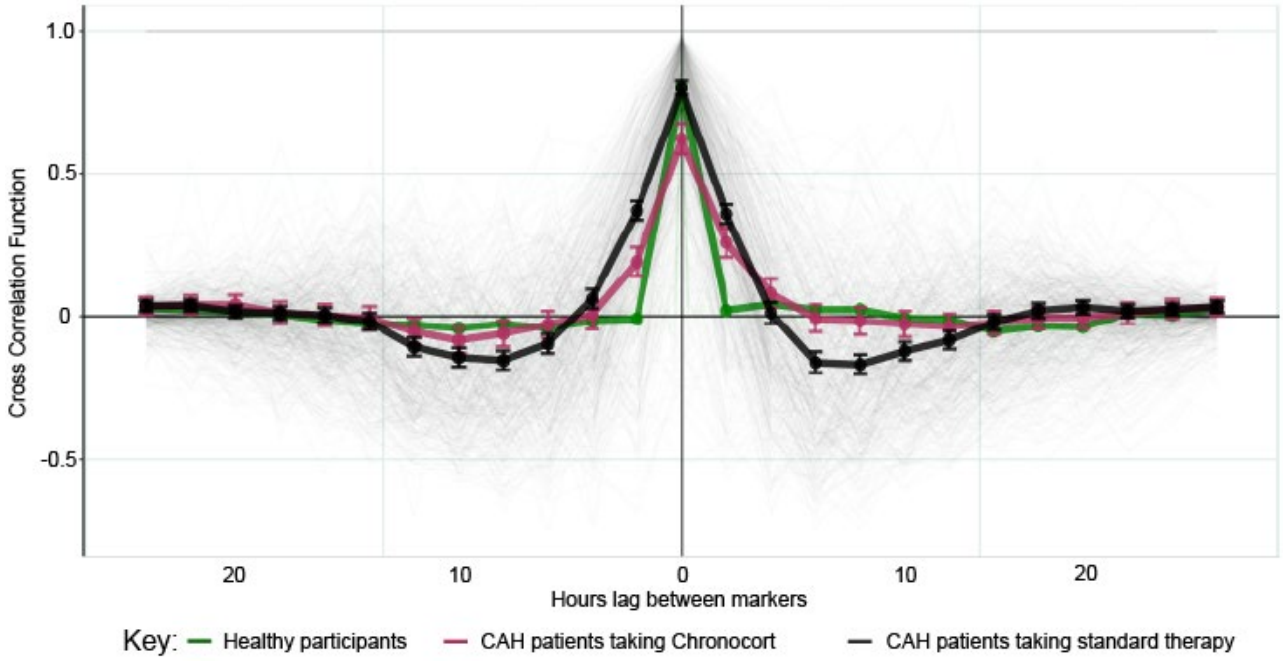
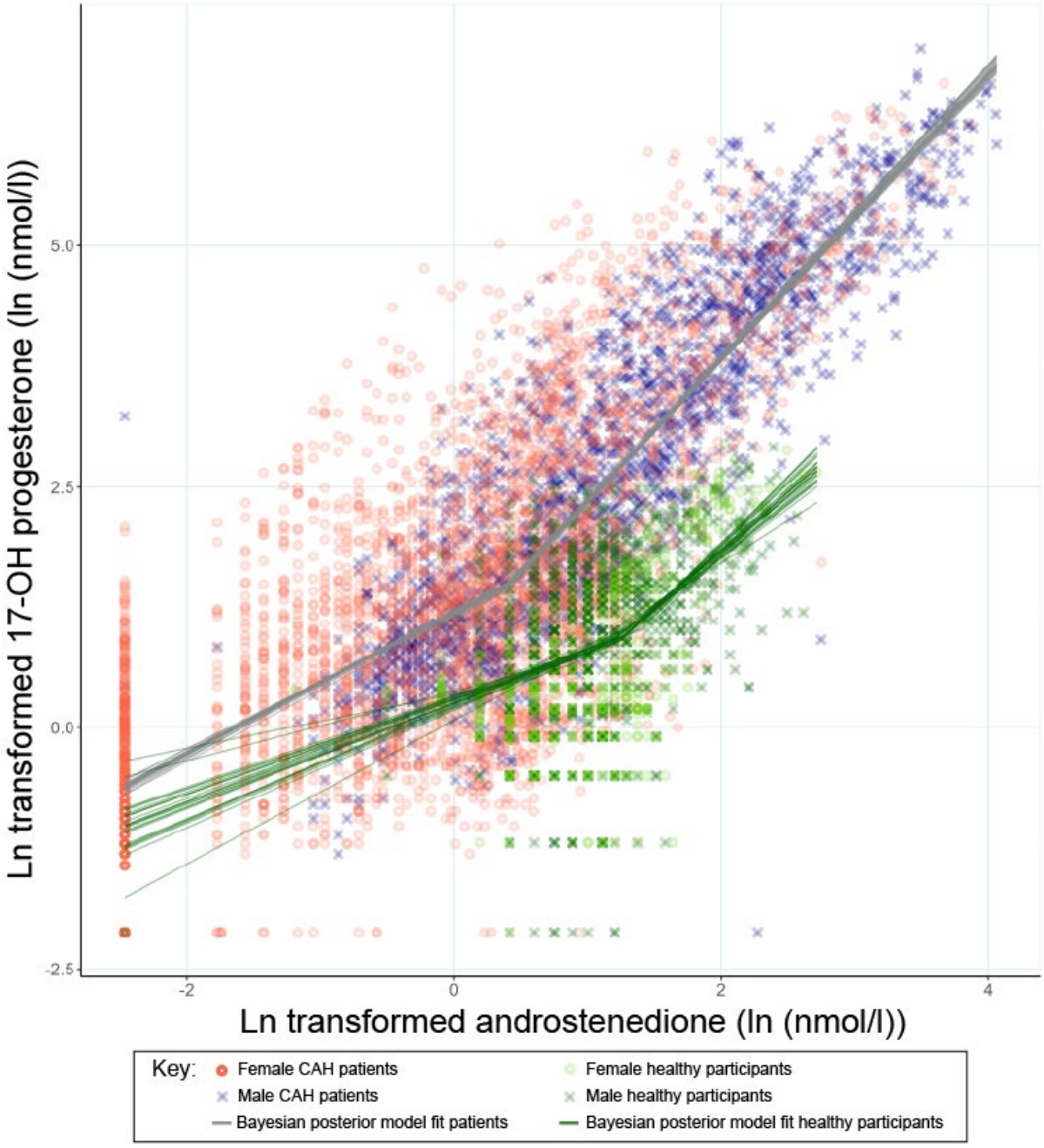


Figure 4.3 – Regression of Ln transformed AUC 17OHP on androstenedione in healthy participants and CAH patients.

In healthy participants there are higher level of androstenedione for same levels of 17OHP. Bayesian changepoint analysis of 17OHP on androstenedione converged on changepoint in CAH patients at 4.5 nmol/l 17OHP (149 ng/dl) ($\hat{R}$ < 1.008), compared to changepoint in healthy participants at 2.4nmol/l (80.0 ng/dl) ($\hat{R}$ < 1.020). Grey lines show 25 randomly selected Bayesian posterior model fits close together, thus showing stable convergence of the model.

17OHP=17OH-Progesterone



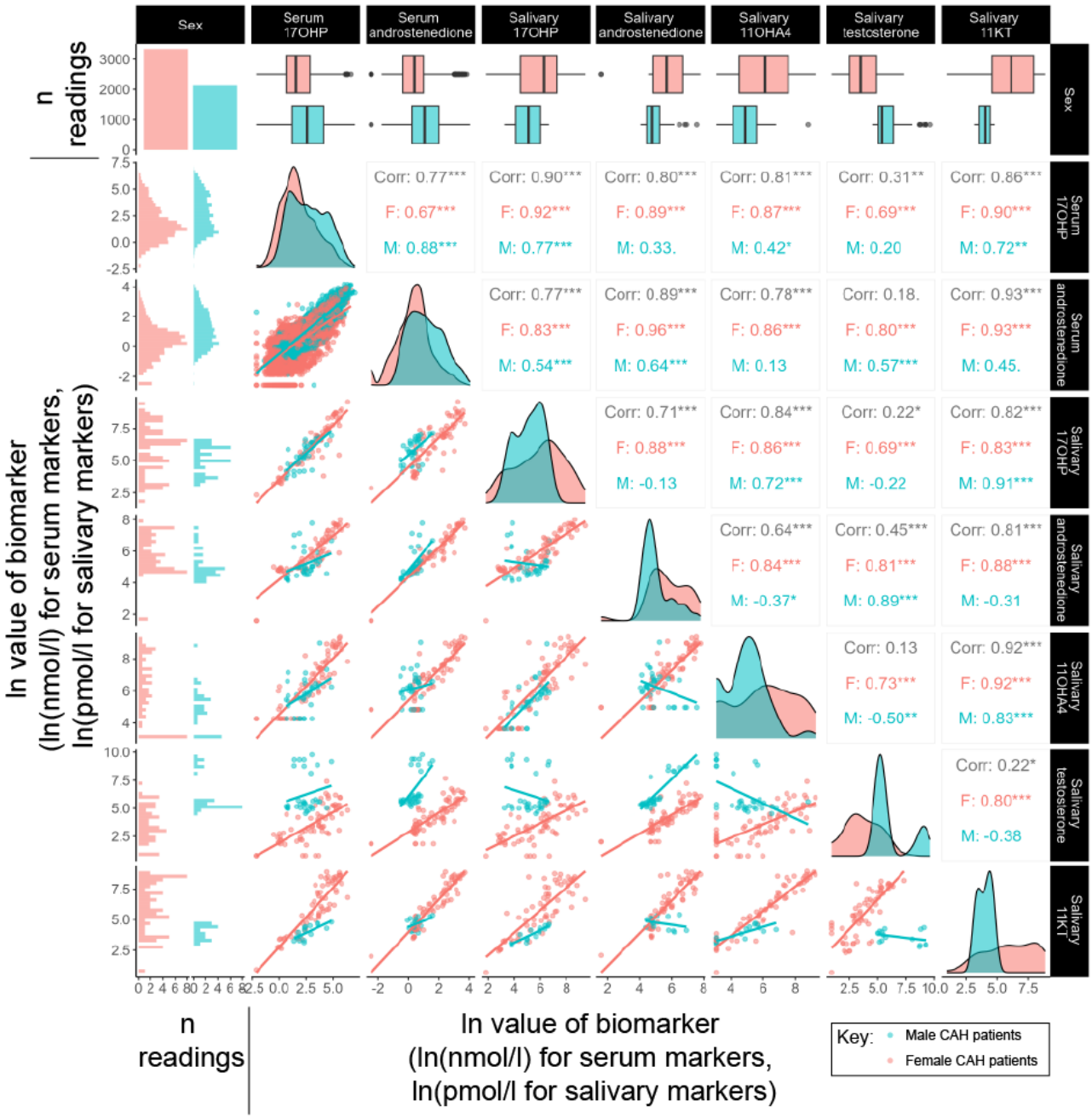
Relationship between serum 17OHP, androstenedione & testosterone: In healthy women, testosterone levels were higher for the same level of 17OHP than in women with CAH, whereas for androstenedione, the relationship between testosterone and androstenedione was similar for healthy women as in those with CAH (Appendix Supplementary Analysis 4.3, Appendix Table S4.5, Appendix Figure S4.1). In male patients, there was a very weak negative correlation between 17OHP, androstenedione and testosterone (17OHP: R^2_{adj}=0.016, p=0.10, androstenedione: R^2_{adj}=0.043, p=0.02).

Salivary profiles of 17OHP, androstenedione, 11OHA4, testosterone, & 11 ketotestosterone: Absolute profiles of all five salivary steroids measured overlaid on serum 17OHP and androstenedione profiles appropriately scaled showed similar patterns in CAH patients. Correlation of each salivary marker was strong within female CAH patients, ranging from r=0.7 to 0.9 for each of the five salivary steroids with both serum 17OHP and androstenedione. Correlations were weaker in male CAH patients (Figure 4.4, Appendix Table S4.6, Appendix Figure S4.2).

Dose of glucocorticoid replacement and disease control: There was a wide variation in total daily replacement glucocorticoid dose across the cohort from 10mg to 80mg hydrocortisone equivalent. Overall mean (SD) total daily hydrocortisone equivalent at time of profile measurement was 28.8mg (10.6) mg and was similar in those taking standard glucocorticoid therapy 28.9 (10.9) mg and those taking MRHC 28.8 (10.0) mg. Comparing AUC of serum 17OHP and androstenedione within patients on the same dose at two different visits, we saw similar control between the two visits (men: R^2_{adj}=0.79, 0.90, women: R^2_{adj}=0.44, 0.55). The distribution of serum 17OHP was similar across the dose range (slope = -0.007 standard error of the mean (SEM) = 0.008, p=0.34) (Supplementary Analysis 4.4, Appendix Tables S4.6-S4.7, Appendix Figures S4.3-S4.4), with 187/477 profiles measured within the range of serum 17OHP AUC seen in healthy controls. For serum 17OHP the distribution above and below the healthy controls was similar (140 above, 150 below), whereas for androstenedione AUC, the majority had an AUC below that of healthy controls (none above, 357 below).

Figure 4.4 – Matrix of Biomarker Correlations

Correlations between serum 17OHP, androstenedione and salivary 17OHP, androstenedione, 11OHA4, Testosterone and 11KT. Point measurements ln transformed measured at the same time of day. Saliva measured within 9 patients across 14 visits at one study centre



4.08 Discussion

We investigated the relationship between adrenal steroid hormones in healthy participants and patients with CAH to better understand the biochemical monitoring of CAH. Serum 17OHP and androstenedione showed similar changes over time in both healthy participants and CAH patients but absolute levels and the relationship between 17OHP and androstenedione differed between healthy participants and CAH patients with CAH patients having a proportionally lower androstenedione for the same level of 17OHP. Salivary 17OHP and androstenedione showed similar changes and correlated with salivary testosterone and 11-ketotestosterone in women. Measurements of 17OHP and androstenedione made on different days whilst receiving the same glucocorticoid dose were similar. Patients on MRHC had flatter profiles, similar to healthy participants, such that a single measurement of either 17OHP or androstenedione reflected the overall 24-hour levels. In contrast, patients on conventional, immediate-release glucocorticoid treatment generally had a diurnal rhythm with an increased amplitude of both 17OHP and androstenedione in the morning.

The 24-hour serum profiles of 17OHP and androstenedione showed parallel changes in both healthy participants and patients with CAH with no lag on cross-correlation. This means 17OHP and androstenedione have a similar serum half-life and provide similar information about androgen control over time. In healthy participants, profiles of serum 17OHP and androstenedione were relatively flat with low amplitude (<1nmol/l), suggesting no clinically significant circadian rhythm. At baseline, most CAH patients on conventional, immediate-release glucocorticoid treatment had a diurnal rhythm of both markers, with a rise over the early hours peaking in the morning consistent with the fact that standard glucocorticoid replacement cannot control the overnight rise in ACTH. This has led clinicians to measure steroid profiles on standard treatment and to debate whether to take measurements before or after the morning glucocorticoid dose, as a single measurement does not reflect the overall 24-hour biochemical control (133). In contrast, patients on MRHC had flatter profiles (amplitude <1nmol/l for 17OHP), similar to healthy participants. This demonstrates that MRHC treatment can be monitored in most patients by taking a sample at any time in the daytime and independent of the time of MRHC intake as the result will reflect the 24-hour levels. MRHC is a modified-release formulation of hydrocortisone with delayed release such that a nighttime dose provides a physiological cortisol rise overnight to peak in the morning with a second dose taken in the morning to fill in the diurnal rhythm. Thus, monitoring on MRHC doesn't need to be done before or after dosing but can be done any time in a morning or

afternoon clinic, and a single measurement will generally reflect control over the 24 hours, as 17OHP and androstenedione levels do not show a diurnal rhythm on MRHC.

The relationship between serum 17OHP and androstenedione differed between healthy participants and patients with CAH. In healthy participants, absolute levels of 17OHP and androstenedione were similar but CAH patients had proportionally much lower androstenedione levels for the same level of 17OHP. In addition, our results showed a change point in the relationship between 17OHP and androstenedione in CAH patients whereby when 17OHP is in the reference range androstenedione levels were proportionally even lower. This can be explained by the pathophysiology of CAH, where 17OHP accumulates prior to the 21-hydroxylase enzyme block. The excess levels of 17OHP result in the atypical conversion of 17OHP to androstenedione by CYP17A1 17,20-lyase activity, which physiologically has a much higher preference for the conversion of 17OH-pregnenolone to DHEA, which in CAH appears suppressed. The excess 17OHP drives increased androgen production through androstenedione, which also feeds the 11-oxygenated androgen pathway. Thus, when 17OHP levels are brought into the normal range, androstenedione is neither generated from 17OHP nor DHEA. The relevance of this to monitoring patients with CAH receiving MRHC treatment is that elevated androstenedione levels, even in the upper half of the reference range, reflect poor control of androgens and, conversely, low androstenedione does not necessarily reflect glucocorticoid overtreatment. This is evident from our study where a proportion of patients on total daily doses of MRHC 10-15 mg/day still had very low androstenedione levels in whom one would be hesitant to reduce such a low daily dose any further to avoid giving inadequate hydrocortisone replacement for adrenal insufficiency. It is unclear why androgen biosynthesis does not reset itself once better disease control is achieved. One hypothesis could be that there is an intra-adrenal feedback loop, with persistent androgen excess resulting in downregulation of CYP17A1 or specifically CYP17A1 17,20 lyase activity. Future research will need to investigate whether the relationship between 17OHP and androstenedione normalises as treatment improves.

The salivary steroid profiles of 17OHP and androstenedione correlated well with the serum profiles suggesting that where salivary measurements are available, they could be used instead of serum measurements. This is consistent with the previously published data of single measurements made in the day in a large cohort of children with CAH and healthy controls (140). This has the advantage that salivary steroids are stable at room temperature and can be collected at home and sent to the laboratory, avoiding the need for venipuncture. In women, the salivary testosterone, 11OHA4 and 11-ketotestosterone all strongly correlated with the serum and salivary 17OHP and

androstenedione, suggesting that controlling 17OHP and androstenedione also controls the other androgen pathways in the adrenal gland as has previously been shown with urine analysis (141).

The dose of glucocorticoid given either as standard treatment or MRHC did not correlate with control of 17OHP across patients, despite improved control in individual patients with increasing the glucocorticoid dose. All the patients studied had proven established classic CAH but there was a wide spectrum of total daily hydrocortisone equivalent dose from 5mg to 80mg, with the spectrum of control similar at each dose. Compliance was good in the study which was closely monitored. Half of the patients (50.3%) were taking either standard treatment or MRHC with hydrocortisone equivalent doses above that recommended from adult adrenal replacement (25mg/day). It is not clear what factors apart from glucocorticoid dose and circadian delivery of hydrocortisone determine the level of control, and this is something to be addressed by future research.

Current guidelines recommend that 17OHP should not be titrated into the reference range as this risks overtreatment with glucocorticoid; however, no recommendation on the absolute level is given (34). Cohort studies have suggested that “optimal control” would be a 17OHP <36nmol/l (1200 ng/dl), which is below 3-4x the upper limit of normal (ULN) (142, 143). Our analysis showed that a 17OHP <38nmol/l (1250 ng/dl) was associated with an androstenedione in the normal range <5nmol/l (143ng/dl) in 95% of patients. This level of 17OHP is similar to that used in the cohort studies to define optimal control and was used in the clinical trial. However, the absolute level of control of adrenal androgens in CAH depends on the clinical goals for the patients, often dependent on age and sex. During childhood, optimising growth and preserving fertility are important targets in the management of CAH. In young adults, fertility becomes increasingly important, and in older adults avoiding iatrogenic glucocorticoid excess becomes important. Among the patients on MRHC in the clinical trial, a number of female patients became pregnant, as well as the female partners of male patients, suggesting that controlling the adrenal androgens throughout the 24 hours is important for fertility (36). Our results show that if the 17OHP is controlled within 3-4xULN, then the androstenedione will be in the reference range or low. Based on the close relationship of androstenedione to testosterone and 11-ketotestosterone in women, elevated androstenedione levels are likely to be associated with virilisation and infertility. Therefore, in women prioritising fertility treatment, titrating androstenedione into the lower half of the reference range is likely to be more successful; in parallel to androstenedione, 17OHP and progesterone which typically accumulate in CAH will normalise, supporting improved implantation and early pregnancy outcomes. The absolute level of 17OHP and androstenedione used to titrate the hydrocortisone dose should depend upon the goals of treatment and the

patient's current dose and treatment regimen. Accepting that 17OHP is higher than androstenedione, both measurements give a similar reflection of control and a 17OHP <38nmol/l and an androstenedione of <5nmol/l reflect reasonable control for most patients.

The limitations of this study include the fact that not all steroid measurements were made in all participants and that serum steroid measurements were carried out by different assays in study participants and healthy controls, albeit both were established and validated LC-MS/MS assays. Relatively few patients had 11-oxygenated androgens measured but these were detailed profiles and this is the largest dataset available, and it is unlikely that an equivalent dataset will be created. The density of data with 2-hourly 24-hour profiles being repeated 4 times in a cohort of over 100 patients with CAH provides a reference set against which comparisons can be made.

In conclusion, measuring either 17OHP and or androstenedione gives similar information about the control of adrenal androgens; however, clinicians should be aware that in CAH patients androstenedione levels are proportionally lower than 17OHP, so where tight control is required, such as a woman wishing to be pregnant, androstenedione levels may be very low, and the hydrocortisone dose should not be reduced below a typical adrenal replacement dose. On MRHC, a single measurement of either 17OHP and/or androstenedione can be taken in the morning or early afternoon, independent of the time of drug dosing, and used to monitor treatment. Either serum or saliva measurements can be used as the levels parallel each other.

4.09 Disclosure summary

RJR is a Director of Diurnal Group Plc. AP receives unrelated research funds from Diurnal Limited and HRA Pharma. DPM received unrelated research funds from Diurnal Limited, Neurocrine Biosciences and Adrenas Therapeutics, all through the National Institutes of Health Cooperative Research and Development Agreements. GC is a NIHR Senior Investigator. AP delivered results for this study through the NIHR Birmingham Biomedical Research Centre. The views expressed in this article are those of the author(s) and not necessarily those of the NIHR, or the Department of Health and Social Care.

4.10 Author Contributions

RR, DM, WA and JN-P. conceptualised the original study and developed the methodology. RR and DPM acquired funding. AP, EB, LS, AT, AP, AH, AJ, DM, JN-P, AR, NR, MS, PT, BK, WA and RR provided resources and project administration to carry out the study. GSC, JD, ZL, NK and RR provided supervision and support with software, visualisation and formal analysis. NL and RR formally accessed, verified and curated the underlying data. NL wrote the original draft, with all co-authors supporting the writing with review and approval of the final manuscript.

4.11 Thesis contribution of chapter 4

A comprehensive analysis of the largest dataset of steroid profiles from patients with CAH has been completed and published in *eBiomedicine* (114), quantifying the similarity in profiles of 17-Hydroxyprogesterone and androstenedione and thus achieving objective 3. This research directly informed the assessment of these biomarkers in the subsequent chapters, when these biomarkers were analysed in the context of blood pressure and growth of children with CAH. Both biomarkers were assessed for statistical significance within models, but models were made more parsimonious by only including 17-Hydroxyprogesterone to reduce the degrees of freedom, knowing that due to the collinearity this was unlikely to cause significant detriment to the conclusions inferred (112, 113).

Chapter 5 – Blood pressure and its associations in 554 children and young people with CAH

5.01 Publication and contribution of others

This chapter was published in the European Journal of Endocrinology and can be accessed at (https://doi.org/10.1093/ejendo/lvaf060) (113). Co-authors outside of the supervisory team contributed through coordinating data collection for their centre, alongside review and feedback of the manuscript. The original draft of the manuscript and incorporation of feedback and improvements was done by Neil Lawrence. All authors agreed on the final version of the manuscript (full contributions in section 5.10).

Chapter 5 Co-authors

Neil R. Lawrence[1], Irina Bacila[1], Joseph Tonge[1], Jeremy Dawson[2,3], Gary S. Collins[4], Zi-Qiang Lang[5], Jillian Bryce[6], Malika Alimussina[6], Minglu Chen[6], Salma Rashid Ali[6,7], Safwaan Adam[8], Erica L.T. van den Akker[9], Tânia Aparecida Sartori Sanchez Bachega[10], Federico Baronio[11], Niels Holtum Birkebæk[12], Walter Bonfig[13,14], Hedi Claahsen - van der Grinten[15,16], Martine Cools[17,18], Eduardo Correa Costa[19], Miguel Debono[20], Liat de Vries[21,22], Christa E. Flück[23], Gabriella Gazdagh[24], Ayla Güven[25], Sabine E. Hannema[26], Violeta Iotova[27], Hetty J. van der Kamp[28], Ruth Krone[29,30], Sofia Leka-Emiri[31], María Clemente-León[32], Corina Raducanu Lichiardopol[33],

Renata L. Markosyan[34], Tatjana Milenkovic[35], Mirela Costa de Miranda[36], Uta Neumann[37],

John Newell-Price[1,20], Şükran Poyrazoğlu[38], Ursina Probst-Scheidegger[39], Gianni Russo[40],

Luisa De Sanctis[41,42], Sumudu Nimali Seneviratne[43], Marianna Rita Stancampiano[44], Rieko Tadokoro-Cuccaro[45], Ajay Thankamony[45], Ana Vieites[46], Malgorzata Wasniewska[47], Diego Yeste[48,49], Jeremy Tomlinson[50], S. Faisal Ahmed[6,7], and Nils Krone[1,51]

Author affiliations:

1: Division of Clinical Medicine, School of Medicine and Population Health, University of Sheffield, Sheffield S10 2TN, United Kingdom
2: Management School, University of Sheffield, Sheffield S10 2TN, United Kingdom
3: Division of Population Health, School of Medicine and Population Health, University of Sheffield, Sheffield S10 2TN, United Kingdom
4: Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford OX1 2JD, United Kingdom
5: Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield S10 2TN, United Kingdom
6: Office for Rare Conditions, Royal Hospital for Children & Queen Elizabeth University Hospital, Glasgow G51 4TF, United Kingdom
7: Developmental Endocrinology Research Group, University of Glasgow, Glasgow G12 8QQ, United Kingdom
8: Department of Endocrinology, The Christie Hospital, Manchester M20 4BX, United Kingdom
9: Department of Pediatric Endocrinology, Sophia Children's Hospital, Erasmus Medical Centre, PO Box 2040, 3000 CA, Rotterdam, Netherlands
10: Hormones and Molecular Genetics Laboratory LIM 42, Department of Internal Medicine, University of Sao Paulo, Sao Paulo 05508-090, Brazil
11: Department Hospital of Woman and Child, Pediatric Unit, Endo-ERN Center for Rare Endocrine Diseases, IRCCS AOUBO, Bologna 40138, Italy
12: Department of Pediatrics and Adolescent Medicine, Aarhus University Hospital, Aarhus 8200, Denmark
13: Department of Pediatrics, Technical University Munich, Munich 80333, Germany
14: Department of Pediatrics, Klinikum Wels-Grieskirchen, Wels 4600, Austria
15: Department of Pediatric Endocrinology, Radboud University Medical Centre, 6500 HB Nijmegen, Netherlands
16: Amalia Children's Hospital, Radboud University Medical Centre, 6500 HB Nijmegen, Netherlands

17: Department of Internal Medicine and Pediatrics, Ghent University, 9000 Ghent, Belgium
18: Department of Pediatric Endocrinology, Ghent University Hospital, 9000 Ghent, Belgium
19: Pediatric Surgery Service, Hospital de Clínicas de Porto Alegre, Porto Alegre, CEP 90410-000, Brazil
20: Endocrinology Department, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield S10 2JF, United Kingdom
21:Institute for Diabetes and Endocrinology, Schneider's Children Medical Center of Israel, Petah-Tikvah 4920235, Israel
22: Faculty of health and medical sciences, Tel-Aviv University, Tel-Aviv 69978, Israel
23: Paediatric Endocrinology, Diabetes and Metabolic medicine, Medizinische Universitätskinderklink, 3010 Bern, Switzerland
24: Wessex Clinical Genetics Service, University Hospital Southampton, Southampton SO16 6YD, United Kingdom
25: Pediatric Endocrinology, Baskent University Istanbul Hospital, Istanbul 06490, Turkey
26: Department of Paediatrics, Amsterdam UMC location Vrije Universiteit, 1081 HV Amsterdam, The Netherlands
27: Department of Paediatrics, Medical University of Varna, Varna 9002, Bulgaria
28: Pediatric Endocrinology Wilhelmina Children's Hospital, University Medical Centre Utrecht, 3584 CX Utrecht, The Netherlands
29: Centre for Endocrinology, Diabetes and Metabolism, University of Birmingham, Birmingham B15 2TT, United Kingdom
30: Birmingham Women's & Children's Hospital, Steelhouse Lane, Birmingham B4 6NH, United Kingdom
31: Department of Endocrinology-Growth and Development, "P.& A. KYRIAKOU" Children's Hospital, Athens 11527, Greece
32: Paediatric Endocrinology, University Hospital Vall d'Hebron. CIBER de Enfermedades Raras (CIBERER) ISCIII, Barcelona 8035, Spain
33: Department of Endocrinology, University of Medicine and Pharmacy Craiova, Craiova 200349, Romania
34: Department of Endocrinology, Yerevan State Medical University, Yerevan 0025, Armenia
35: Department of Endocrinology, Institute for Mother and Child Healthcare of Serbia "Dr Vukan Č upić", 11070 Belgrade, Serbia
36: Department of Internal Medicine, University of Sao Paulo, Sao Paulo 05508-070, Brazil
37: Clinic for Pediatric Endocrinology and Diabetology and Center for Chronically Sick Children, Charite—Universitätsmedizin, 10117 Berlin, Germany
38: Paediatric Endocrinology Unit, Istanbul University, Istanbul Faculty of Medicine, Istanbul 34093, Turkey
39: Pediatric Department, Kantonsspital Winterthur, Winterthur 8400, Switzerland
40: Department of Paediatrics, Endocrine Unit, Scientific Institute San Raffaele, Endo-ERN Center for Rare Endocrine Diseases, 20132 Milan, Italy
41: Paediatric Endocrinology, Regina Margherita Children's Hospital, 10126 Torino, Italy
42: Department of Public Sciences and Pediatrics, University of Torino, 10125 Torino, Italy
43: Faculty of Medicine, University of Colombo, Colombo 03, Sri Lanka
44: Department of Paediatrics, Endocrine Unit, Scientific Institute San Raffaele, 20132 Milan, Italy
45: Department of Paediatrics, Biomedical Campus, University of Cambridge, Cambridge CB2 1TN, United Kingdom
46: Centro de Investigaciones Endocrinológicas "Dr. César Bergadá" (CEDIE), CONICET—FEI—División de Endocrinología, Hospital de Niños R. Gutiérrez, C1425EFD Buenos Aires, Argentina
47: Department of Human Pathology in Adulthood and Childhood, University of Messina, 98122 Messina, Italy
48: Paediatric Endocrinology Service, Hospital Universitario Vall d'Hebron, 08035 Barcelona, Catalunya, Spain
49: CIBER de Enfermedades Raras (CIBERER) ISCIII, Barcelona 8035, Spain
50: Oxford Centre for Diabetes, Endocrinology and Metabolism, NIHR Oxford Biomedical Research Centre, University of Oxford, Churchill Hospital, Oxford OX1 2JD, United Kingdom
51: Department of Endocrinology, Sheffield Children's Hospital NHS Foundation Trust, Sheffield S10 2TH, United Kingdom

5.02 Significance Statement

We used advanced regression modelling to study a large group of children with CAH. We showed that increased levels of renin, androstenedione and 17OHP were associated with lower blood pressure (BP), although BP was higher than normative values, with a larger differential at younger ages. Although increased doses of mineralocorticoid resulted in a higher BP, the effect size was marginal at 1mmHg per 100 micrograms of fludrocortisone. Further long-term research into cardiovascular outcomes in patients with CAH will help us understand whether increased BP at young ages has any adverse clinical outcomes.

5.03 Abstract

Background

Congenital Adrenal Hyperplasia (CAH) due to 21-hydroxylase deficiency (21-OHD) affects approximately 1 in 15,000 individuals. We leveraged the power of multicentre registry data to assess the trend and predictors of blood pressure (BP) within children and young persons with 21-OHD to inform monitoring strategies.

Method

Data from the International CAH Registry in patients younger than 20 years was compared to normative values. Values of BP were modelled to create reference curves, multiple change point analysis applied to quantify the difference with normative data. Covariate adjustment was informed by a directed acyclic graph, prior to joint outcome regression modelling to accurately assess predictors of BP.

Results

A total of 6436 visits within 554 patients (52.5% females) showed BP-Standard deviation scores (SDS) were higher at younger ages. Patients under five years had systolic BP-SDS of 1.6 (Q1:0.6-Q3:2.7) decreasing to 1.0 (Q1:0.2-Q3:1.8) over five years, equating to 31.0% over the 95th centile decreasing to 15.0%. Higher doses of fludrocortisone were associated with a small increase in systolic BP equivalent to 1.2mmHg with every 100 micrograms extra fludrocortisone. Renin of 100µU/ml was associated with 4.6mmHg lower systolic BP than a renin of 1µU/ml, higher 17OH-progesterone and androstenedione also predicted lower systolic and diastolic BP ($p<0.05$).

Conclusion

Higher BP in children with 21-OHD is common and particularly pronounced at a younger age, but may not be attributable to excessive mineralocorticoid replacement. There is a need to improve our understanding of the determinants of this raised BP as well as its long-term effects.

5.04 Background

Congenital Adrenal Hyperplasia (CAH) is the most common form of inherited adrenal insufficiency, affecting between 1 in 10,000 to 1 in 15,000 people. It is caused in over 90% of cases by deficiency of the enzyme 21-hydroxylase which converts 17OH-progesterone (17OHP) to 11-deoxycortisol, the main substrate used in the production of cortisol. Patients with classic 21-hydroxylase deficiency (21-OHD) need lifelong treatment with glucocorticoid replacement, most commonly hydrocortisone in childhood (2, 5). Most patients are also at risk of salt wasting due to aldosterone deficiency caused by the lack of conversion of progesterone to deoxycorticosterone. This mineralocorticoid deficiency is treated with fludrocortisone. Salt replacement is also recommended at young ages, although use is variable (5, 45, 144).

Blood pressure (BP) in children with 21-OHD is contentious (145), small studies in under 24 patients reporting normal BP (146, 147), but studies investigating 24 hour ambulatory reading in 38 or fewer patients finding elevated BP (148-150). Others, including analysis of registry data from 716 children in two countries report that high BP readings are a transient problem in early childhood that resolves (45, 151). This previous registry study was unable to assess biomarkers of disease control in relation to BP (45).

High doses of fludrocortisone in 21-OHD to prevent salt loss can cause hypertension, as can high doses of glucocorticoids (22, 45, 152). However, the optimum dose of fludrocortisone has not been studied, and the extent to which hypertension in children with 21-OHD correlates with long term adverse outcomes or co-morbidities is unknown (5, 21, 145). As cardiovascular diseases remain the leading cause of global mortality (153) there is a pressing need to understand how hormone replacement and biomarkers of disease control impacts on BP in patients with 21-OHD, to inform appropriate prevention and monitoring strategies.

The International Congenital Adrenal Hyperplasia (CAH) Registry provides rich longitudinal data from CAH patients with 21-OHD (154). We set out to use advanced statistical modelling to assess the trend in BP throughout childhood and compare this to normative data from the National Heart, Lung, and Blood Institute (NHLBI) (155), and assess the impact of different aspects of patient treatment on BP in children with 21-OHD.

5.05 Methods

This retrospective cohort study included children with a diagnosis of 21-OHD with consent for data sharing with the I-CAH Registry. No patients were excluded. Data were extracted on 21/12/2021 and analysis was restricted to visits of patients under the age of 20 years. We carried out data pre-processing and clarification by longitudinal visualisation of variables with liaison with contributing centres to correct data entry errors.

Missing data

Missing data was assessed using a hierarchical hybrid approach of spline interpolation between longitudinal points for height and weight, last observation carried forward or next observation carried backward for dosing, and joint modelling multilevel multiple imputation for biomarkers and BP values. Analysis was conducted in ten imputation sets and estimates combined with Rubin's rules (156). A sensitivity analysis assessing the impact of missing data was performed by repeating all analyses with cases with complete data only (Appendix Supplementary Methods 5.1).

Statistical Analysis

Summary and reference values

Statistical Analysis was carried out in *R, a language and environment for statistical computing* (R Foundation for Statistical Computing, Vienna, Austria; packages in Appendix Table S5.1). Summary statistics were calculated using the median and interquartile range of continuous variables. Standard deviation scores (SDS) were derived by comparing to World Health Organisation (WHO) reference standards for growth (157), and NHLBI normative data over 1 year of age for BP (155, 158). Absolute BP values were modelled with a Lambda-Mu-Sigma (LMS) approach to create smoothing reference curves. We subtracted the NHLBI median BP for age from the I-CAH registry median BP for age, and conducted multiple change point analysis to assess the age at which BP in those with CAH plateaued above normative values.

Dosing and biomarkers

Glucocorticoid dose, fludrocortisone dose and salt replacement were summed as total daily doses. Glucocorticoids were converted to hydrocortisone equivalent by using British National formulary specified conversion ratios (Appendix Table S5.2, hydrocortisone(mg)=prednisolone(mg)x4) (159). Biomarkers assessed included 17OHP, androstenedione and renin. Lower and upper limits of detection and units of biomarkers were standardised across centres, and plasma

renin activity (PRA) converted to renin (Appendix Table S5.3, PRA(nmol/l/hr)x0.158=renin(μIU/ml)) (160). Biomarkers were ln transformed prior to multivariable modelling to better approximate normality.

Covariate adjustment

A directed acyclic graph was developed with domain experts (Appendix Figure S5.1) to ensure appropriate covariate adjustment sets to estimate the effect of each variable of interest on BP. The aim was to adjust for confounders that affect both the exposure and outcome of interest. If the effect of an exposure is mediated through a variable, that variable should not be adjusted for, to avoid collider bias (161).

Regression modelling

To estimate predictors of BP, we applied multilevel joint modelling regression, simultaneously assessing both systolic and diastolic BP as our outcome variables. We used the directed acyclic graph to select appropriate covariate adjustment sets for each target of estimation. We used both a treatment centre level and patient level random intercept with a random slope applied for age to account for multiple measures and varying trajectories within patients.

5.06 Results

Patient biometrics

This retrospective observational study included 554 patients (52.5% female, 46.4% male, 1.1% not assigned) from 35 centres across 18 countries (Table 5.1, imputed statistics Appendix Table S5.4). There was a total of 6436 visits with a median of 9 visits per patient (Quartile 1 (Q1):6 to Quartile 3 (Q3):16), visits spanning a median of 3.2 years (Q1:2.5 to Q3:7.3) within patients. Median age at visit was 3.0 years (Q1:1.0 to Q3:7.7),a greater proportion of visits at younger ages reflecting more frequent assessment of younger patients and attrition from registry data entry (Appendix Figure S5.2).

1871 # Table 5.1 – Summary statistics

Sex assigned at birth:	Male	Female	Not assigned	Total sample
Number of countries	17	17	4	18
Number of centres	31	31	4	35
Number of patients	257	291	6	554
Number of visits	3018	3361	57	6436
Number of visits per patient **Median (Q1 to Q3)**	9 (7 to 16)	9 (5 to 16)	8 (2 to 11)	9 (6 to 16)
Number of years visits spanned within patients **Median (Q1 to Q3)**	3.2 (2.7 to 6.9)	3.2 (2.1 to 8.0)	2.1 (0.2 to 3.1)	3.2 (2.5 to 7.3)
Age of patients at youngest visit (years) **Median (Q1 to Q3)**	0.13 (0.04 to 0.99)	0.21 (0.04 to 3.03)	0.08 (0.03 to 0.26)	0.16 (0.04 to 2.17)
Age of patients at most recent visit (years) **Median (Q1 to Q3)**	5.70 (3.09 to 11.47)	6.01 (3.07 to 14.35)	2.25 (0.46 to 3.17)	5.81 (3.07 to 12.87)
Systolic BP at visit (mmHg) **Median, (n) [Q1 to Q3]‡**	107 (n=1556) [97 to 118]	105 (n=1652) [96 to 116]	99 (n=21) [85 to 107]	106 (n=3229) [97 to 117]
Systolic BP SDS at visit‡ **Median (n) [Q1 to Q3]**	1.5 (n=1492) [0.5 to 2.4]	1.1 (n=1564) [0.3 to 2.1]	-	1.3 (n=3056) [0.4 to 2.2]
Diastolic BP at visit (mmHg) **Median, (n) [Q1 to Q3]‡**	64 (n=1542) [57 to 71]	64 (n=1645) [57 to 70]	60 (n=20) [50 to 71.25]	64 (n=3207) [57 to 70]
Diastolic BP SDS at visit‡ **Median (n) [Q1 to Q3]**	1.1 (n=1478) [0.4 to 2]	0.9 (n=1559) [0.3 to 1.6]	-	1.0 (n=3037) [0.3 to 1.8]
Visits prescribed hydrocortisone* **n (%) [missing n, % missing]‡**	2200 (72.9%) [610, 20.2%]	2214 (65.9%) [648, 19.3%]	35 (61.4%) [22, 38.6%]	4449 (69.1%) [1280, 19.9%]
Total Hydrocortisone equivalent at visit per BSA (mg/m^2)†‡ **Median (n) [Q1 to Q3]**	14.1 (n= 2237) [9.8 to 15.5]	14.5 (n= 2438) [9.9 to 15.7]	18.4 (n= 25) [13.5 to 20.3]	14.3 (n= 4700) [9.9 to 15.6]
Visits prescribed fludrocortisone **n (%) [missing n, % missing] ‡**	2577 (85.4%) [180, 6.0%]	2754 (81.9%) [140, 4.2%]	51 (89.5%) [6, 10.5%]	5382 (83.6%) [326, 5.1%]
Total Fludrocortisone at visit per body surface area (when prescribed) ($\mu g/m^2$) ‡ **Median (n) [Q1 to Q3]**	318 (n= 2442) [103 to 396]	292 (n= 2527) [99 to 321]	555 (n= 38) [207 to 484]	307 (n= 5007) [102 to 356]
Visits prescribed salt **n (%), [missing n, % missing] ‡**	400 (13.3%) [290, 9.6%]	486 (14.5%) [204, 6.1%]	35 (61.4%) [5, 8.8%]	921 (14.3%) [499, 7.8%]
Renin (µIU/ml) **Median (n) [Q1 to Q3] ‡**	5.0 (n=1059) [0.3 to 69.1]	4.0 (n=1041) [0.4 to 54.0]	0.4 (n=16) [0.1 to 3.2]	4.5 (n=2116) [0.4 to 61.8]
17OH-Progesterone (nmol/l) **Median (n) [Q1 to Q3]**	18.2 (n=1380) [2.0 to 100.0]	26.9 (n=1325) [3.0 to 140.0]	12.1 (n=31) [4.0 to 62.8]	21.18 (n=2736) [2.72 to 115.0]
Androstenedione (nmol/l) **Median (n) [Q1 to Q3]**	0.1 (n=1285) [0.1 to 3.0]	1.0 (n=1211) [0.1 to 6.0]	0.1 (n=29) [0.1 to 1.0]	0.7 (n=2525) [0.1 to 3.5]

1872 n=Number; BSA=Body surface area; Q1=Quartile 1; Q3=Quartile 3; SDS=Standard deviation score
1873 ‡Summary statistics of imputed values in Appendix Table S5.4
1874 *Remaining visits patients prescribed either cortisone acetate (n=352), dexamethasone (134), prednisone (37), prednisolone (105),
1875 methylprednisolone (1), mixed dosing (59) or no glucocorticoid (86) (Appendix Table S5.5)
1876 †Hydrocortisone equivalent calculated by multiplying preparations by the following factors: prednisolone/prednisone x4; dexamethasone x80;
1877 cortisone acetate x0.8; methylprednisolone x5 (Appendix Table S5.2 for full frequency tables of preparations)

Biomarkers

Renin was measured in 32.9% of visits with median value of 4.5μIU/ml (Q1:0.4 to Q3:61.8), androstenedione in 39.2% of visits with median of 0.7nmol/l (Q1:0.05 to Q3:3.5) and 17OHP in 42.5% of visits with median of 21.2nmol/l (Q1:2.7 to Q3:115.0). Glucocorticoid treatment consisted of hydrocortisone in 69% cases (alternative preparations Appendix Table S5.5), at a median dose of 14.3mg/m^2 (Q1:9.9 to Q3:15.6) hydrocortisone equivalent; fludrocortisone and salt supplements were prescribed in 84% and 14% cases respectively. Median height SDS at visit was -0.3 (Q1:-1.3 to Q3:0.6), with median BMI SDS of 0.5 (Q1:-0.3 to Q3:1.3) (Figure 5.1).

Blood pressure

The BP-SDS was higher at younger ages, patients under five having median systolic SDS 1.6 (Q1:0.6 to Q3:2.7), those over five having median systolic SDS of 1.0 (Q1:0.2 to Q3:1.8). This equated to 31.0% over the 95th centile for age and sex under five years, decreasing to 15.0% in those over five. For diastolic, median BP-SDS decreased from 1.6 (Q1:0.8 to Q3:2.5) to 0.6 (Q1:0.1 to Q3:1.2), proportions over 95th centile decreasing from 27.5% to 3.3%.
In absolute terms, modelled median systolic BP was 23mmHg higher at age 1 decreasing to 7mmHg higher at age 10 in males. Equivalent readings in females were 18mmHg higher decreasing to 9mmHg higher, diastolic BP showing a similar trajectory (Figure 5.2, Table 5.2).

Blood pressure changepoint analysis in comparison to normative values

Multiple change point analysis estimated the difference of median BP in patients above normative data stopped decreasing in males at age 11.5 years for systolic and 5.9 years for diastolic. In females, this occurred later at age 13.1 years for systolic and 7.0 years for diastolic. Following the change points, the median BP in male patients was 9.2mmHg above normative for systolic, 7.3mmHg above for diastolic and in females 6.4mmHg above for systolic and 5.6mmHg above for diastolic (Appendix Table S5.6, Appendix Figure S5.3).

Predictors of blood pressure in boys and girls

The directed acyclic graph (Appendix Figure S5.1) highlighted renin, androstenedione and 17OHP as mediators of the effect of drug doses on BP. To estimate the total effect of medications on BP, the independent variables were restricted to the covariates of age, sex, height, weight (Figure 5.3) and other drug doses. To estimate the extent to which each biomarker predicted BP, ancestor variables of drug doses were avoided, and covariates age, sex, height and weight controlled for (Table 5.3, full models Appendix Table S5.7).

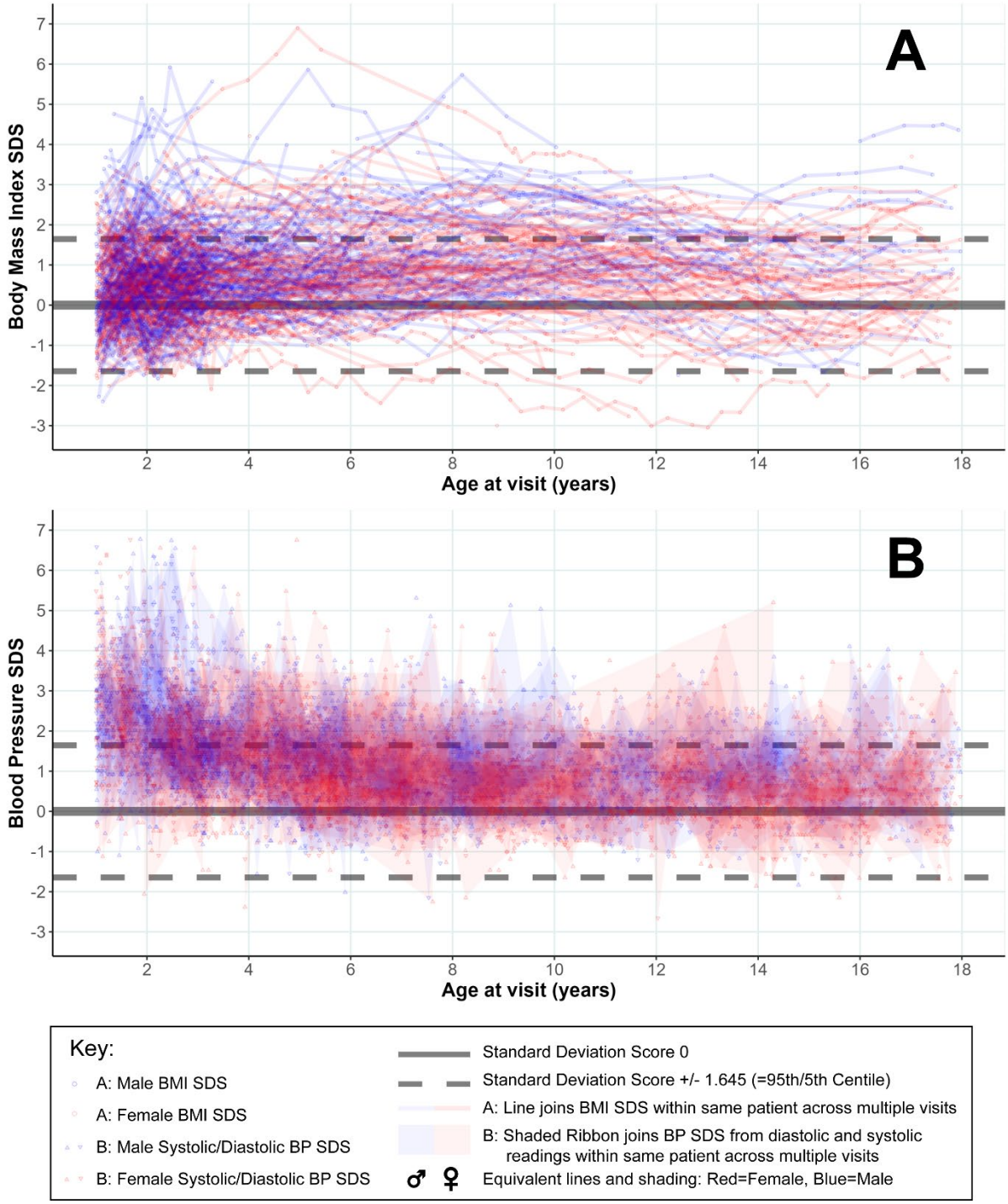
Higher renin, higher 17OHP and higher androstenedione all predicted lower BP. This translated to patients with a renin of 100µU/ml having systolic BP 4.6 mmHg lower and diastolic BP 2.3 mmHg lower than patients with a renin of 1µU/ml. Patients with a 17OHP of 100nmol/l had systolic BP 2.9mmHg lower and diastolic BP 2.3mmHg lower than patients with a 17OHP of 1nmol/l. Patients with androstenedione of 10nmol/l had systolic BP 1.7mmHg lower and diastolic BP 1.4mmHg lower than patients with an androstenedione of 1nmol/l.

Higher doses of fludrocortisone were associated with higher systolic and diastolic BP, but dose of glucocorticoid and salt did not have any consistent significant effect. However, whilst statistically significant, the effect of fludrocortisone on BP was clinically small, with the equivalent of 100 micrograms of extra fludrocortisone being associated with an increase in systolic BP of 1.2mmHg and diastolic BP of 0.8mmHg.

Sensitivity analyses

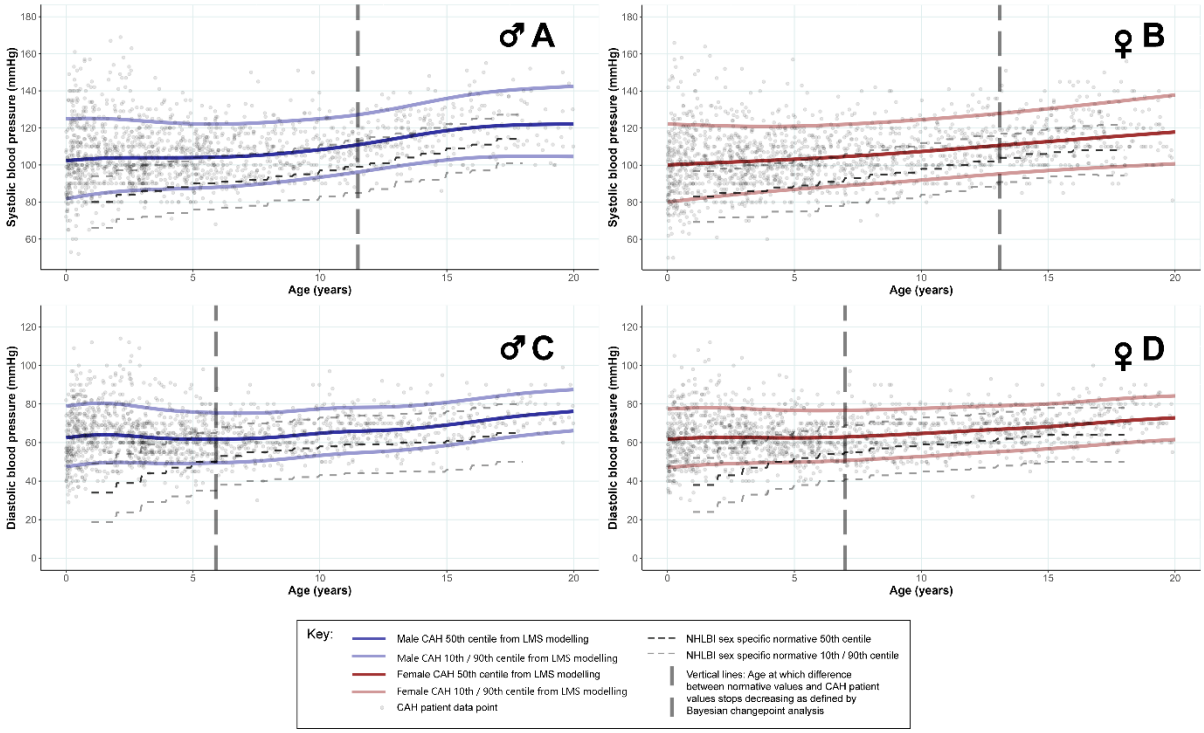
Bayesian joint models run without imputed data did not show any significant difference in the size or direction of the estimates of interest. Models estimated using SDS for biometrics and BP as well as doses per body surface area also showed no significant difference (Appendix Table S5.7).

Figure 5.1 – Variation of BMI-SDS and BP-SDS on age



SDS: Standard deviation score

Figure 5.2 – LMS modelling of blood pressure



A = Male systolic BP modelling; B = Female systolic BP modelling; C = Male diastolic BP modelling; D = Female diastolic BP modelling.

BP = Blood Pressure; LMS = Lambda, Mu, Sigma; CAH = Congenital adrenal hyperplasia; NHLBI = National Heart, Lung and Blood Institute (162).

CAH LMS patient centiles fit individually to each of 10 imputed datasets, with centile estimates combined using Rubin's rules

Table 5.2 – Blood pressure by sex and age

Age (Years)	Male systolic	Male diastolic	Female systolic	Female Diastolic
	Median (mmHg) (10^{th}-90^{th} centile) [Difference of median of CAH patients above normative values (mmHg)]			
1	103.3 (84.0 to 125.1) [23.3]	63.8 (49.1 to 80.3) [29.8]	100.8 (81.7 to 121.7) [17.8]	62.4 (48.1 to 77.9) [24.4]
2	103.8 (85.6 to 124.7) [19.8]	63.9 (49.7 to 80.0) [24.9]	101.4 (83.2 to 121.3) [16.4]	62.7 (48.8 to 78.0) [19.7]
3	103.9 (86.5 to 123.8) [17.9]	63.0 (49.4 to 78.4) [19.0]	101.9 (84.6 to 121.0) [15.9]	62.6 (49.2 to 77.6) [15.6]
4	103.9 (87.0 to 122.8) [15.9]	62.1 (49.1 to 76.8) [15.1]	102.5 (85.8 to 120.9) [14.5]	62.5 (49.5 to 77.1) [12.5]
5	104.0 (87.4 to 122.2) [14.0]	61.8 (49.2 to 75.9) [11.8]	103.2 (86.9 to 121.1) [14.2]	62.4 (49.8 to 76.7) [10.4]
6	104.2 (88.0 to 122.1) [13.2]	61.8 (49.5 to 75.4) [8.8]	103.9 (87.9 to 121.5) [12.9]	62.6 (50.2 to 76.6) [8.6]
		Plateau in difference after 6 years		
7	104.7 (88.9 to 122.3) [12.7]	62.0 (50.0 to 75.2) [7.0]	104.6 (88.9 to 122.0) [11.6]	62.9 (50.7 to 76.7) [7.9]
				Plateau in difference after 7 years
8	105.6 (90.2 to 123.0) [11.6]	62.6 (50.8 to 75.5) [6.6]	105.5 (89.9 to 122.8) [10.5]	63.5 (51.3 to 77.0) [6.5]
9	106.8 (91.7 to 124.0) [11.8]	63.6 (52.0 to 76.4) [6.6]	106.4 (90.9 to 123.6) [10.4]	64.1 (52.0 to 77.3) [6.1]
10	105.6 (90.2 to 123.0) [11.6]	64.8 (53.3 to 77.5) [6.8]	105.5 (89.9 to 122.8) [10.5]	63.5 (51.3 to 77.0) [6.5]
	Plateau in difference after 11 years			
12	112.0 (97.2 to 128.1) [11.0]	66.1 (55.1 to 78.3) [7.1]	109.5 (94.1 to 126.6) [7.5]	66.2 (54.4 to 78.7) [5.2]
			Plateau in difference after 13 years	
14	116.5 (101.1 to 133.2) [10.5]	67.7 (57.0 to 79.5) [7.7]	111.7 (96.2 to 129.1) [5.7]	67.5 (55.9 to 79.5) [4.5]
16	120.1 (103.9 to 138.1) [9.1]	70.7 (60.2 to 82.4) [7.7]	113.8 (97.9 to 131.9) [5.8]	69.2 (57.8 to 81.0) [5.2]
18	121.7 (104.7 to 140.9) [7.7]	74.0 (63.6 to 85.6) [9.0]	115.9 (99.4 to 134.8) [7.9]	71.3 (59.9 to 82.9) [7.3]

Blood pressure median and centiles derived from Lambda, Mu, Sigma modelling across all data within sex. Normative BP data was derived from the National Heart, Lung, and Blood Institute guidelines (155). Plateau in difference calculated by Bayesian multiple change point analysis (Appendix Table S5.6)

Table 5.3 – Predictors of blood pressure

Target of estimation	Appropriate covariate adjustment set*	Estimate of effect on systolic BP (95% CI)	Estimate of effect on diastolic BP (95% CI)	R^2 systolic model (95% CI)	R^2 diastolic model (95% CI)
Effect of daily hydrocortisone equivalent dose (mg) on BP	*Age, Sex, Height, Weight, Other medication dosing*	0.052 (-0.057 to 0.162)	0.031 (-0.045 to 0.107)	0.27 (0.14 to 0.40)	0.17 (0.06 to 0.28)
Effect of daily fludrocortisone dose (µg) on BP		0.012** (0.005 to 0.020)	0.008** (0.003 to 0.014)		
Effect of Daily salt dose (g) on BP		-0.69 (-2.14 to 0.77)	-0.56 (-1.54 to 0.42)		
Extent ln Renin (ln (µU/ml)) predicts BP	*Age, Sex, Height, Weight*	-1.00** (-1.47 to -0.53)	-0.71** (-1.13 to -0.29)	0.32 (0.19 to 0.45)	0.21 (0.09 to 0.32)
Extent ln 17OHP (ln (nmol/l)) predicts BP		-0.64** (-1.00 to -0.27)	-0.50** (-0.78 to -0.22)	0.31 (0.17 to 0.45)	0.20 (0.07 to 0.32)
Extent ln androstenedione (ln (nmol/l)) predicts BP		-0.73** (-1.18 to -0.28)	-0.61** (-0.96 to -0.25)	0.31 (0.17 to 0.45)	0.20 (0.08 to 0.32)

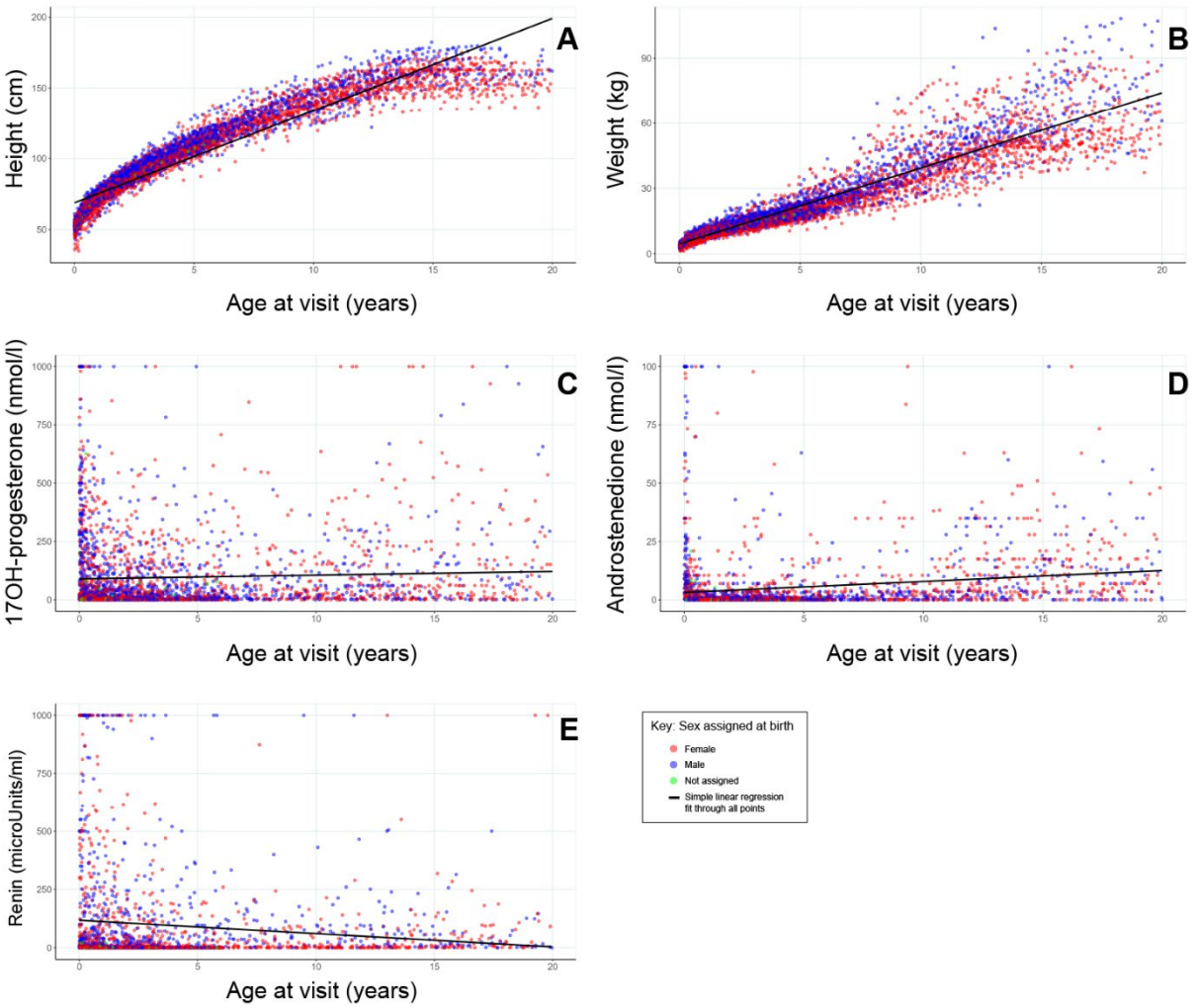
Interpretation of statistically significant coefficients at clinically meaningful values (reverse ln transformed where appropriate):
100 micrograms of extra fludrocortisone was associated with an increase in systolic BP of 1.2 mmHg and diastolic BP of 0.8mmHg
Patients with a renin of 100µU/ml having systolic BP 4.6mmHg lower and diastolic BP 3.3mmHg lower than patients with a renin of 1µU/ml
Patients with a 17OHP of 100nmol/l had systolic BP 2.9mmHg lower and diastolic BP 2.3mmHg lower than patients with a 17OHP of 1nmol/l
Patients with an androstenedione of 10nmol/l had systolic BP 1.7mmHg lower and diastolic BP 1.4mmHg lower than patients with an androstenedione of 1nmol/l

* Appropriate adjustment sets were applied informed by the directed acyclic graph to avoid bias introduced by conditioning on mediating variables or ancestors of the variable of interest (Appendix Figure S5.1)

** Statistically significant estimates as estimated by Bayesian joint modelling across 400 iterations with 10 bootstrap replications of each of 10 imputed datasets and combining estimates using Rubin's rules

BP = Blood pressure; 17OHP = 17-Hydroxyprogesterone; CI = Confidence interval. Full model estimates and sensitivity analyses in Appendix Table S5.7

1944 Figure 5.3 – Variation of height, weight and biomarkers on age



1945

1946 A = Height on age; B = Weight on age; C = 17-Hydroxyprogesterone on age; D = Androstenedione on age; E = Renin on age

1947 Simple linear regression line plotted through all points to show trend of data with increasing age

5.07 Discussion

We reviewed data from over 6000 clinic visits of over 550 patients with CAH under 20 years of age and found that the BP was higher than normative values. This increase was greater at younger ages, and similar in boys and girls. Joint outcome regression modelling showed only a small average increase in BP due to mineralocorticoid replacement, and no significant effect on BP from glucocorticoid or salt replacement. Higher levels of renin, 17OHP and androstenedione all predicted lower BP when controlling for age, sex, height and weight.

Higher BP in children with CAH has previously been attributed to inappropriately high mineralocorticoid replacement. We saw a low regression coefficient when modelling predictors of BP consistent with an extra 100 micrograms of fludrocortisone causing approximately 1mmHg increase in systolic and diastolic BP. This corroborates the findings of other studies showing higher BP at larger doses of fludrocortisone in childhood (42, 151). However, our much larger cohort in combination with robust modelling appropriately adjusting for known confounders allowed us to estimate a reliable effect size. This effect size was low, and can reassure clinicians that appropriate mineralocorticoid replacement is not likely to drive a patient into clinically significant hypertension. This low effect size also explains how smaller studies have failed to show any significant difference in BP with fludrocortisone dose (163, 164).

The lack of effect of daily hydrocortisone equivalent dose on BP in this study is further evidence that there is a variable dose requirement between patients with CAH, even when accounting appropriately for their age, sex, height and weight. The difference in regression coefficients between glucocorticoid and mineralocorticoid doses highlights that these doses should not be combined during analysis, but considered individually as they have separate pharmacological affects (165).

The negative correlation between renin and BP that we have shown highlights that this marker does have value when monitoring patients with CAH (2, 5, 151). Our large sample, careful transformation and handling of covariates has likely contributed to this finding where previous work has shown no association (166). Nonetheless, a model explaining only 1/3 of the variability in BP indicates why this marker is sometimes challenging to interpret in isolation within an individual patient, in part due to its variation with postural position (167, 168). Renin measurement should thus be standardised, with results interpreted alongside clinical measurements of BP and electrolytes and considered

against potential novel biomarkers within future studies to understand optimum mineralocorticoid replacement in CAH.

Higher 17OHP and androstenedione have been shown to be associated with lower BP in CAH in other studies (44, 147). In vitro, 17OHP has been shown to bind to the mineralocorticoid receptor and antagonise the effect of aldosterone (169), consistent with our results that higher levels of 17OHP are associated with lower BP. The large regression coefficients of patient weight within our model, consistent with the undisputed knowledge that heavier children are more likely to have higher BP (170), show the importance of adjusting appropriately for covariates and the value of guiding this analysis by the use of domain expertise mapped within a directed acyclic graph. However, the association of higher 17OHP and androstenedione levels with lower BP might also reflect reduced adherence to replacement therapy in respective individuals.

Salt was prescribed in less than 15% of visits and largely concentrated in those under 5 years, which may have contributed to its lack of a statistically significant effect on BP. However, our sensitivity analysis restricting analysis to patients under five showed similar results, and is consistent with another I-CAH study that showed no difference in BP between CAH patients on salt and those without (42), although the patients in that study not on salt were taking higher doses of mineralocorticoid replacement. Overall, appropriate salt replacement in CAH should not be considered to have the same effects of excess dietary salt in adulthood.

The limitations of this study are highlighted by the large proportion of missing data for BP and biomarkers that is typical for large, real-world data sets. Robust multiple imputation techniques that have shown results similar to those produced by complete case analysis is reassuring. However, measurements of both BP and biomarkers vary significantly between centres, and therefore put the findings at risk of regression dilution bias. This centre effect has been controlled for appropriately as a centre level random effect in the modelling, but may still have some residual effect on the estimates presented. Our modelling assumes linearity in relationships that are likely non-linear, and the effect sizes we report should not be extrapolated beyond the characteristics of the stable population from which we have estimated them.

Accurate measurement of BP in children in an outpatient setting is challenging, with this study limited by a lack of standardisation in measurement between centres. Patients may be upset undergoing blood tests at the same visit and

thus have artificially raised BP due to duress. However, such duress also risks artificially raising steroid biomarkers, yet we found higher steroid biomarkers associated with lower BP. Unfortunately, we have been unable to assess compliance in this study, an unmeasured covariate that may be associated with other known covariates such as weight. However, CAH is a rare disease, and this data from over 6000 clinic visits is only possible thanks to collaboration across continents facilitated by an international registry. With variable numbers of visits within patients our multilevel joint modelling design with careful covariate adjustment and appropriate support from professional statisticians has facilitated valuable insights from a complex dataset.

The higher BP-SDS shown in CAH patients in this study, and the relatively higher BP at younger ages warrants further investigation. Studies have shown vascular remodelling related to higher BP in the aorta and carotids in young patients, and an increase in prevalence of left ventricular diastolic dysfunction evident in even small cohorts of young patients (145, 149, 171). However, the impact of different treatment strategies on these outcomes is unknown. Further engagement with long term disease registries and linking of datasets between children and adult services will help establish how best to respond to raised BP in patients with CAH. Independent patient data meta-analysis would help to assess whether other metrics of cardiovascular health measured in smaller studies should be regularly monitored to improve patient outcomes and quality of life.

5.08 Conclusion

Higher BP in children with CAH is commonly observed and is particularly pronounced at a younger age. These higher readings are not explained by excessive mineralocorticoid or salt replacement alone, nor are they associated with poor disease control, higher levels of 17OHP and androstenedione being associated with lower BP. There is currently no evidence that BP is a significant problem in children with 21-OHD CAH, although there is a need to further our understanding of the determinants of the raised BP in younger children with CAH, and whether this has any long-term consequences. Future research assessing the impact of different dosing regimens on cardiac function would further our understanding of the underlying pathophysiological processes.

5.09 Acknowledgments and Funding

Neil Lawrence is supported by the National Institute for Health and care Research (NIHR) by a doctoral research fellowship (NIHR302559). Gary Collins is supported by a Cancer Research UK programme grant (C49297/A27294).

Professor Gary Collins is a NIHR Senior Investigator. The views expressed in this article are those of the author(s) and not necessarily those of the NIHR, or the Department of Health and Social Care. We would like to thank all the children and families who have consented to share their data with the I-CAH Registry, without whom this research would not be possible.

5.10 Authors' contributions

Neil Lawrence (Data curation [supporting], Formal analysis [lead], Funding acquisition [equal], Investigation [equal], Methodology [lead], Project administration [equal], Software [lead], Visualization [lead], Writing—original draft [lead], Writing—review & editing [equal]), Irina Bacila (Data curation [supporting], Writing—review & editing [supporting]), Joseph Tonge (Data curation [supporting], Writing—review & editing [supporting]), Jeremy Dawson (Formal analysis [supporting], Methodology [supporting], Software [supporting], Visualization [supporting], Writing— review & editing [supporting]), Gary Collins (Formal analysis [supporting], Methodology [supporting], Software [supporting], Supervision [supporting], Visualization [supporting], Writing—review & editing [supporting]), Zi-Qiang Lang (Formal analysis [supporting], Writing—review & editing [supporting]), Jillian Bryce (Data curation [supporting], Project administration [lead], Writing—review & editing [supporting]), Malika Alimussina (Data curation [supporting], Project administration [supporting], Writing—review & editing [supporting]), Minglu Chen (Project administration [supporting], Writing—review & editing [supporting]), Salma Ali (Project administration [supporting], Writing— review & editing [supporting]), Safwaan Adam (Data curation [supporting], Writing—review & editing [supporting]), Erica van den Akker (Data curation [supporting], Writing—review & editing [supporting]), Tania Bachega (Data curation [supporting], Writing—review & editing [supporting]), Federico Baronio (Data curation [supporting], Writing—review & editing [supporting]), Niels Birkebæk (Data curation [supporting], Writing—review & editing [supporting]), Walter Bonfig (Data curation [supporting], Writing—review & editing [supporting]), Hedi L. Claahsen—van der Grinten (Data curation [supporting], Writing—review & editing [supporting]), Martine Cools (Data curation [supporting], Writing—review & editing [supporting]), Eduardo Costa (Data curation [supporting], Writing—review & editing [supporting]), Miguel Debono (Data curation [supporting], Writing—review & editing [supporting]), Liat de Vries (Data curation [supporting], Writing—review & editing [supporting]), Christa Flueck (Data curation [supporting], Writing—review & editing [supporting]), Gabriella Gazdagh (Data curation [supporting], Writing—review & editing [supporting]), Ayla Güven (Conceptualization [supporting], Writing—review & editing [supporting]), Sabine E. Hannema (Data curation [supporting], Writing—review & editing [supporting]), Violeta Iotova (Data curation [supporting], Writing—review &

editing [supporting]), Hetty van der Kamp (Data curation [supporting], Writing—review & editing [supporting]), Ruth Krone (Data curation [supporting], Writing—review & editing [supporting]), Sofia Leka-Emiri (Data curation [equal], Writing—review & editing [equal]), Maria Clemente (Data curation [supporting], Writing—review & editing [supporting]), Corina Lichiardopol (Data curation [supporting], Writing—review & editing [supporting]), Renata Markosyan (Conceptualization [supporting], Writing—review & editing [supporting]), Tatjana Milenkovic (Data curation [equal], Writing—review & editing [equal]), Mirela Miranda (Data curation [supporting], Writing—review & editing [supporting]), Uta Neumann (Data curation [equal], Writing—review & editing [equal]), John Newell-Price (Data curation [supporting], Writing— review & editing [supporting]), Sukran Poyrazoglu (Data curation [supporting], Writing—review & editing [supporting]), Ursina Probst-Scheidegger (Data curation [supporting], Writing—review & editing [supporting]), Gianni Russo (Data curation [supporting], Writing—review & editing [supporting]), Luisa de Sanctis (Data curation [supporting], Writing— review & editing [supporting]), Sumudu Nimali Seneviratne (Data curation [supporting], Writing—review & editing [supporting]), Marianna Rita Stancampiano (Data curation [supporting], Writing—review & editing [supporting]), Rieko Tadokoro-Cuccaro (Data curation [supporting], Writing—review & editing [supporting]), Ajay Thankamony (Data curation [supporting], Writing—review & editing [supporting]), Ana Vieites (Data curation [supporting], Writing—review & editing [supporting]), Malgorzata Wasniewska (Data curation [supporting], Writing—review & editing [supporting]), Diego Yeste (Data curation [supporting], Writing—review & editing [supporting]), Jeremy Tomlinson (Data curation [supporting], Project administration [supporting], Supervision [supporting], Writing—review & editing [supporting]), S Ahmed (Data curation [supporting], Funding acquisition [lead], Project administration [lead], Resources [supporting], Writing—review & editing [supporting]), and Nils Krone (Conceptualization [lead], Data curation [supporting], Formal analysis [supporting], Methodology [supporting], Supervision [lead], Visualization [supporting], Writing—original draft [supporting], Writing— review & editing [lead]).

5.11 Availability of data and materials

All code associated with this analysis can be found at:

https://github.com/neilxlawrence/I-CAH_Blood_Pressure

Requests for access to data must be sought through SDM Registries:

https://sdmregistries.org/data-access/

5.12 Thesis contribution of chapter 5

Longitudinal analysis of from 554 patients with CAH has been completed and published in the *European Journal of Endocrinology* to achieve objective 4 (113). The lambda, mu, sigma modelling of systolic and diastolic blood pressure within this research is then used directly to display disease specific centiles of blood pressure within the clinical decision support tool presented within chapter 7 of this thesis.

Chapter 6 - Adiposity rebound and height velocity in patients with Congenital Adrenal Hyperplasia

6.01 Publication and contribution of others

This chapter has been accepted for publication in the *European Journal of Endocrinology* in March 2026. Co-authors outside of the supervisory team contributed through coordinating data collection for their centre, alongside review and feedback of the manuscript. The original draft of the manuscript and incorporation of feedback and improvements was done by Neil Lawrence. All authors agreed on the final version of the manuscript.

Co-authors

Neil R. Lawrence, Irina Bacila, Joseph Tonge, Chamila Balagamage, Jeremy Dawson, Gary S. Collins, Zi-Qiang Lang, Jillian Bryce, Malika Alimussina, Minglu Chen, Salma R. Ali, Nadia Amin, Nermine H Amr, Fathima A. Anverdeen, Tânia Bachega, Magdalena Banaszak-Ziemska, Federico Baronio, Niels Holtum Birkebæk, Walter Bonfig, María Clemente-León, Martine Cools, Justin H Davies, Liat de Vries, Christiaan de Bruin, Heba Elsedfy, Christa E. Flück, Antony Fu, Gabriella Gazdagh, Alina German, Evelien Gevers, Evgenia Globa, Ayla Güven, Sabine E. Hannema, Violeta Iotova, Dominika Janus, H. El Kaddouri, Hetty J van der Kamp, Ruth Krone, Nina Lenherr-Taube, Otilia Marginean, Renata Markosyan, Inas Mazen, Harriet Miles, Mirela Costa de Miranda, Klaus L. Mohnike, Cheryl Morris, Anuja Natarajan, Uta Neumann, Marek Niedziela, Rita Ortolano, Alegria Ferri Perez, Şükran Poyrazoğlu, Ursina Probst-Scheidegger, Tabitha Randell, D. Aled Rees, Gianni Russo, Mariacarolina Salerno, Luisa De Sanctis, Valérie Schwitzgebel Luscher, Sumudu Nimali Seneviratne, Savitha Shenoy, Margarett Shnorhavorian, Marianna Rita Stancampiano, Rieko Tadokoro-Cuccaro, Ajay Thankamony, Agustini Utari, Ana Vieites, Malgorzata Wasniewska, Diego Yeste, Jeremy W. Tomlinson, S. Faisal Ahmed, Nils Krone

6.02 Significance Statement

We used non-linear regression modelling to quantify growth in a large group of children with congenital adrenal hyperplasia. We found large variability in their age at adiposity rebound, occurring early before age 5 years in over 80% of the cohort. The latest peak height velocity was significantly blunted and earlier compared to healthy patient studies, at mean age 8.4 years (SD 3.0) in boys and 9.0 years (SD 1.6) in girls. Multivariate modelling showed girls prescribed higher doses of glucocorticoids associated with significantly higher BMI in adolescence and earlier adiposity rebound. Adiposity rebound can be considered as an early surrogate outcome measure for future studies into early-life treatment strategies.

6.03 Abstract

Objective

Adiposity rebound is the first rise in BMI that occurs after the initial decrease during infancy. Early adiposity rebound, before age 5, is a risk factor for later obesity and metabolic problems. We investigated adiposity rebound in children with Congenital Adrenal Hyperplasia due to 21-hydroxylase deficiency (CAH).

Design

Longitudinal observational registry study.

Methods

Height, weight and BMI from patients younger than 20 years in the I-CAH Registry was described by non-linear mixed-effects models. Covariates of glucocorticoid dose, mineralocorticoid dose, 17-Hydroxyprogesterone were assessed on growth and bone age.

Results

A total of 10,261 visits within 573 patients (43.6% male) showed significant variation in age at latest peak height velocity (8.4 years (SD=3.0) in boys; 9.0 years (SD=1.6) in girls). Peak height velocity was more blunted in boys (7.7 cm/year (SD=1.4)) than girls (7.4 cm/year (SD=1.3)) in comparison to normative values. Adiposity rebound occurred earlier than age 5 years in 82% of the cohort, mean age 3.7 years (SD=1.3) in boys and 3.9 years (SD=0.9) in girls. Girls prescribed higher doses of glucocorticoid were associated with heavier weight in adolescence and earlier adiposity rebound. Bone age was increasingly advanced in those prescribed higher doses in both sexes.

Conclusions

There is a large variation in the timing of adiposity rebound and SITAR-estimated latest peak height velocity in children with CAH. In addition to identifying individuals with CAH who may be at risk of adverse cardiometabolic outcomes these metrics may serve as early surrogate outcomes in future research investigating early-life treatment strategies.

6.04 Background

Congenital Adrenal Hyperplasia (CAH) is the most common form of inherited adrenal insufficiency, affecting approximately 1 in 15,000 individuals. In over 90% of cases, it is due to 21-hydroxylase deficiency (21-OHD) causing impaired conversion of 17OH-progesterone (17OHP) to 11-deoxycortisol, resulting in a deficiency of cortisol, and sex hormone excess. Patients require lifelong glucocorticoid replacement and, in most cases, mineralocorticoid replacement. International consensus guidance recommends a hydrocortisone daily equivalent dose of 10-15mg/m^2 per day (2, 5).

In healthy children, body mass index (BMI) increases after birth before falling after approximately one year, then increasing in early childhood after approximately age six. Adiposity rebound is the last nadir in BMI in early childhood, that corresponds to a decline in fat mass in relation to height. This follows a different trajectory between sexes and is frequently reported at an earlier age in girls than boys, although some studies find no difference in timing with sex. Earlier adiposity rebound in otherwise healthy children is associated with obesity in later life and cardiometabolic diseases in adulthood (172, 173).

Children with CAH have an earlier adiposity peak and rebound compared to healthy children. Precise estimates differ between studies, from as early as 1.7 years in a UK study to 3.8 years in a study of 194 children from the USA (47, 48, 174). The reasons for the early timing of adiposity rebound in CAH are poorly understood and the factors influencing this difference in body composition remain elusive. Multicentre studies have also shown increased prevalence of adverse cardiometabolic markers in adolescents with CAH (15, 175). A better understanding of the factors that contribute to the early adiposity rebound observed in children with CAH and whether this increases the likelihood of later obesity will help to improve monitoring and treatment strategies, ultimately leading to better health outcomes.

Latest peak height velocity has been shown to be blunted in 21OHD, although there are differing estimates about magnitude and timing (11, 12, 176), and studies are yet to quantify this using SuperImposition by Translation And Rotation modelling (SITAR), an effective statistical strategy to model growth in childhood (177, 178). Bone age is known to be inappropriately advanced in this condition, although findings as to how this varies with dose are conflicting (5, 11, 179).

The International Congenital Adrenal Hyperplasia (CAH) Registry provides rich longitudinal data in children with CAH due to 21OHD (154). We set out to use advanced statistical modelling to quantify growth in children with CAH, and assess associations between growth, doses of glucocorticoid replacement and levels of disease control. (11, 12, 176)(177, 178)(5, 11, 179)

6.05 Methods

This retrospective cohort study included children and young people under 20 years of age diagnosed with 21-OHD and consented for data sharing with the I-CAH Registry. Ethical approval for the registry has been approved by the West of Scotland research ethics committee (24/WS/0059). Data were extracted on 25/01/2024 and pre-processed, with subsequent clarification of outliers by liaison with contributing centres, to correct data entry errors.

Missing data

Missing data was accommodated using a combined approach of spline interpolation between longitudinal points for height, weight and bone age and last observation carried forward or next observation carried backward for dosing (Appendix supplementary methods 6.01 (180)).

Statistical Analysis

Summary and reference values

Statistical Analysis was performed in *R, a language and environment for statistical computing* (181). Continuous variables were summarised with median and interquartile range. Standard deviation scores (SDS) were derived by comparing to World Health Organisation (WHO) reference standards for growth (155, 157, 158).

Treatment dose and disease control

Glucocorticoid doses were converted to hydrocortisone equivalent (hydrocortisone(mg)=prednisolone(mg)x4, other ratios in Appendix Table S6.01 (180)) and summed to total daily dose. Biomarkers 17OHP and Androstenedione were measured in local laboratories with different assays, converted to standard units of nmol/l (Appendix Table S6.01 (180)) and natural log transformed to better approximate normality.

Longitudinal modelling of growth

Weight, height and BMI were modelled on the natural log of age in each sex separately using SITAR (50). This is a mixed effects model that fits an average b-spline curve through the data applying random effects that are allowed to vary in three dimensions for each individual participant: Size, a translation on the y axis; timing, a translation on the x axis; and intensity, horizontal stretching or compression of the curve equivalent to a rotation in the anteroposterior plane of the graph. These three adjustments allow for the estimation of a growth pattern that is a similar shape in all children, but comprises a sample of children who vary in stature, experience growth 'spurts' at variable times in childhood, and have a variable velocity of growth during such growth spurts. SITAR allows us to describe an overall 'average' pattern of growth along with a robust estimation of the variability of growth across the sample population. Alternatives, including the Lambda-Mu-Sigma model, cannot estimate the variability in the timing of growth between participants (178).

The flexibility of a SITAR curve increases with pre-defined degrees of freedom. Due to the complexity of the parameters of model estimation, curves were fit to children with adequate data points between specific age ranges, with optimum degrees of freedom defined by minimising the Bayesian information criterion (BIC). Age and BMI at adiposity rebound, and age and speed of latest peak height velocity were extracted from models transformed onto the absolute scale. Comparison of mean growth metrics was made to data from studies of healthy children using two-sided t-tests. Height and weight for each participant at age 18 was extracted from SITAR models and compared against adiposity rebound and height velocity metrics for each participant by linear regression.

Covariate modelling

To assess the impact of dose and disease control on growth in this cohort, the covariates of dose of hydrocortisone, dose of fludrocortisone, level of 17OHP and level of androstenedione were assessed. As SITAR cannot estimate covariates during model fitting, the modelling was simplified to a cubic spline with a two-level participant level random intercept nested within treatment centre. Each covariate was interacted with each element of the cubic spline basis function, optimum degrees of freedom defined by lowest BIC. This simplified modelling strategy did not require restriction on participant data by age or number of data points, but did require all covariates to be available at each clinic visit for data from that visit to be included. Both glucocorticoid and mineralocorticoid doses were included in the model, but models were tested with a combination of both 17OHP and androstenedione or each biomarker individually to define the most efficient model. Clinically meaningful values of covariates defined by expert consensus

were inserted into models to describe model trajectories, confidence intervals calculated by bootstrapping across 500 replications (Appendix Supplementary methods 6.01 (180)).

Bone age

Bone age advancement was calculated by bone age minus chronological age. Bone age was calculated locally and thus methodology differed between centres. Due to reduced frequency of this measurement, a basic cubic model was used to estimate the change in bone age advancement on age (26). The model was fit with similar covariates as growth models, and then reassessed using just dose of glucocorticoid replacement as a single interacted covariate to maximise the number of visits available for modelling. A sensitivity analysis was conducted restricting analysis to the most commonly used method of bone age estimation to assess the impact of different methods of bone age estimation.

6.06 Results

Biometrics and biomarkers

This observational registry study included 746 children with 21-OHD (43.6% male) from 22 countries treated in 47 different centres, 573 children having sufficient data for modelling of height and 457 for modelling of adiposity rebound (Appendix Figure S6.01 (180)). Overall,11,460 visits were distributed with median 11 visits (Q1:6 to Q3:21) within each participant, 10,261 visits contributing to height modelling and 6,901 to adiposity rebound modelling. Visits occurred between 23/1/2003 and 17/1/2024, 6388/11460 (56%) occurring in 2017 or beyond. There was a greater proportion of visits at younger ages, with median age at visit 5.1 years (Q1:1.9 to Q3:9.2) (Table 6.1). Attrition meant only 206/573 (36%) of participants had data available beyond the age of 12 years for height modelling.

Glucocorticoid treatment was most commonly with hydrocortisone (84.9% of visits, Appendix Table S6.02 (180)), 83.5% of visits with fludrocortisone prescribed. Of 1430/11460 (12.5%) visits with a recorded biomarker available, 17OHP recorded in 1888/11460 (16.5%) visits (median 29.8nmol/l Q1:4.2 to Q3:112.0), androstenedione recorded in 1786/11460 (15.6%) visits (median 1.7nmol/l Q1: 0.2 to Q3: 4.4), with 1430/11460 (12.5%) visits having both biomarkers available for analysis. Height SDS at visits was median -0.1 (Q1:-1.1 to Q3:0.9) and BMI SDS median 0.7 (Q1:-0.2 to Q3:1.5).

Growth modelling and adiposity rebound

There was significant variation in age at latest peak height velocity occurring at mean 8.4 years (between participant SD=3.0) in boys and 9.0 years (SD=1.6) in girls (Figure 6.1). Derived latest peak height velocity was 7.7 cm/year (SD=1.4) in boys and 7.4 cm/year in girls (SD=1.3), significantly earlier and lower than that derived from healthy children in the UK Avon Longitudinal Study of Parents and Children (10.6cm/year at 13.5 years in boys; 7.7cm/year at 11.7 years in girls (2-sided t-tests, $p<0.05$)) (177). Optimum SITAR fit was 5 and 6 degrees of freedom for height and weight respectively, $R^2_{marginal}$ 0.91-0.99 (Table 6.2, Appendix Table S6.03 (180)).

BMI modelled directly by SITAR estimated adiposity rebound in boys at mean of 3.7 years (SD:1.3) and 3.9 years in girls (SD:0.9), with 80.4% of males and 82.8% of females occurring early before age five. This was also earlier than the age calculated in a recent large healthy participant cohort of 5.1 years (SD:1.3) in both boys and girls (2-sided t-tests, $p<0.05$)) (182). A sensitivity analysis showed these metrics did not change significantly when interpolated data was not used in the calculations (Table S6.04, Figure S6.02). Residual plots showed individual patient estimates were not biased towards the amount or timing of data contributing to the SITAR models (Appendix Figure S6.03).
Earlier adiposity rebound and greater BMI at adiposity rebound was associated with heavier SITAR predicted weight at 18 years in both sexes ($p<0.05$). Age at adiposity rebound was not associated with predicted height at 18 years ($p>0.05$), but earlier age at peak height velocity was associated with shorter height at 18 years ($p<0.05$, Appendix Table S6.05, Appendix Figure S6.04 (180)).

Covariate modelling

Optimum model fit with covariates was obtained by using 17OHP alone, in comparison to androstenedione alone or both markers combined (lower BIC). There were reduced visits available with covariates for modelling (1785/11460 (15.6%) visits), leading to an optimum model fit achieved with lower degrees of freedom of 3, unable to model the latest peak height velocity within this smaller sub section of the population. Incorporating the covariates of hydrocortisone, fludrocortisone and 17OHP improved the proportion of variance explained by the models. The $R^2_{marginal}$ increased by 2.3% boys and 1.3% in girls for height, and by 6.7% and 4.8% in girls for weight (full model parameters in Appendix Tables S6.06-S6.07 (180)).

Regression modelling assesses the association of variability of continuous variables, rather than artificially dichotomising the data into groups (26). Non-linear interactions are illustrated by bootstrapping clinically meaningful

values decided by expert consensus into each model to derive representations of participants taking a dose of 8mg/m^2/day as low dose versus a high dose of 18mg/m^2/day, and a consistent level of 17OHP of 10nmol/l to demonstrate good control versus a consistent level of 17OHP of 38nmol/l to demonstrate poor control (Figure 6.2). Doses were selected to compare the lowest recommended physiological replacement dose in adrenal insufficiency (183) against a relatively high value, but one that was prescribed in a reasonable proportion of visits within the data set (80th centile dose was 17.5mg/m^2/day) (Appendix Figure S6.05). Alternative example values are demonstrated in Appendix Table S6.08, Appendix Figure S6.06 (180).

Larger doses of glucocorticoid replacement were associated with greater weight gain in girls. This achieved a statistically significant difference after the age of 10 where girls on higher doses with poor disease control had greater weight and BMI (Table 6.3, Figure 6.3). Adiposity rebound occurred 24.0 months earlier in girls (95% CI 3.6 to 55.9) requiring higher doses with worse disease control compared to those with lower doses and better disease control. However, this difference in adiposity rebound was not robust to all sensitivity analyses. There was no significant difference in timing with boys (7.2 months earlier, 95% CI -1.2 to 40.6).

Care must be taken when interpreting these covariate models. Alternative example values inserted are demonstrated in Appendix Table S6.08, Appendix Figure S6.06 (180). These show similar trajectories, as does a sensitivity analysis when modelling was carried out without interpolated or carried values (Appendix Table S6.09, Appendix Figure S6.07). However, sensitivity analysis did change whether metrics of growth had a statistically significant difference between example comparisons. The most robust finding was an increase in weight and BMI in girls on higher doses in early adolescence, consistent across all permutations of sensitivity analysis.

Bone age advancement

Across the cohort bone age advancement peaked at age 10.4 (95% CI 8.3 to 17.0, $n_{participants}$=204) years in boys and 8.7 (95% CI 7.4 to 10.9, $n_{participants}$=240) years in girls with mean advancement of 2.1 (95% CI 1.7 to 2.6) and 1.2 (0.9 to 1.6) years respectively (Appendix Table S6.8, Appendix Figure S6.5 (180)). Covariate modelling with all doses and levels of disease control did not show statistical significance, impacted by reduced visits ($n_{participants}$=144, n_{visits}=572) with those covariates available for modelling. However, the more parsimonious cubic model incorporating dose of hydrocortisone equivalent alone ($n_{participants}$=415, n_{visits}=3540) showed children prescribed higher doses within this data had greater bone age advancement between the ages of 2.4 to 9.8 years in boys and 3.0 to 10.8 years in girls. This

difference peaked at 0.9 (1.4 to 0.5) and 0.6 (1.5 to 0.5) years more advanced for those modelled on 18mg/m^2/day compared to 8mg/m^2/day in boys and girls respectively (Figure 6.4). Results were similar when restricted to models trained only on bone age estimated by the Greulich & Pyle method (81% of measurements) (Appendix Table S6.11, Appendix Figure S6.09).

6.07 Discussion

We have used advanced statistical modelling on multicentre registry data from children with CAH to quantify the blunted latest peak height velocity and early adiposity rebound seen in this condition, as well as significant variability in those metrics between individuals. We have shown associations with increased weight gain, detrimental BMI and advanced bone age in those prescribed higher doses of glucocorticoids.

The blunting of latest peak height velocity was more marked in boys than girls. Whilst one previous study of 598 individuals with CAH noted an earlier growth spurt but reported no detriment to height velocity (184), the majority of studies assessing height velocity in CAH have found blunting (11, 12, 176). None of these studies have employed SITAR, a method more appropriate to assess between participant variability (50). The large standard deviation we have found in magnitude-of and age-at latest peak height velocity highlights the importance of modelling repeated measures appropriately to reduce the risk of aggregation bias (178), and is likely to explain why we have estimated the mean peak height velocity of 7-8cm/year in CAH slightly higher than others. Other powerful techniques including the quadratic-exponent-pubertal-stop technique (QEPS) may provide more detailed insights into growth in this disease in the future (185). This method requires more data points per patient and was unable to converge in this data.

Mean age at adiposity rebound estimated by SITAR modelling was similar between sexes in this cohort, in over four-fifths of the cohort occurring before 5 years of age, a much higher proportion than the 40% of healthy children reported in a recent meta-analysis (186). Early adiposity rebound, has been associated with adverse metabolic profiles including higher glucose and insulin resistance in healthy children at age seven, as well as increased risk of overweight and obesity in adolescence and adulthood (46, 173, 182, 187). Our results highlight this association with later obesity is also the case in CAH. We must consider the possibility of a confounding obesogenic environment or metabolome causing the change in both metrics, but it may be possible that this early shift in adiposity reflects metabolic changes that lead to later obesity. Whether directly causal or simply an associated risk factor, early adiposity rebound can

highlight children that are at greater risk of obesity later in life, and can be used as an early outcome metric when researching different early treatment strategies.

Age at adiposity rebound in CAH has been reported even earlier than our analysis, including 1.7 years in 22 children (48), 3.0 years in 29 children (49), 3.8 years in 16 children (188) and 3.3 years in 42 children (47). The largest cohort to date included 515 children and an adiposity rebound of 3.9 years in boys and 3.3 years in girls, although unfortunately this was derived from LMS modelling that did not allow an estimate of between patient variability or quantification of height velocity. The non-linear SITAR modelling and large cohort is a significant strength of our work, allowing us to robustly estimate the variability in adiposity rebound. The standard deviation of 1.3 years in males shows the extent of this variability, highlighting how this metric can differentiate between individuals at an early age at greater risk of adverse metabolic outcomes. Our estimate of a later adiposity rebound than previous studies may be indicative of improving management of the condition, likely due to improving iterations of international guidance over the last two decades (5).

Our multivariable modelling of height and weight showed girls prescribed higher doses and with poorer disease control gained more weight in later childhood and adolescence, driving an increase in BMI. The higher dose and poorer disease control was also associated with earlier adiposity rebound. Trajectories of BMI were similar in boys but not statistically significant, due in part to slightly less data and larger variability. This differs from previous studies that report no association between dose, disease control and adiposity in CAH in much smaller cohorts (49, 188). The multivariate modelling must be interpreted with care, model estimates having wider confidence intervals at later ages, driven by less than a third of patients having any data beyond 12 years of age. However, the SITAR modelling done on a larger proportion of the cohort did show an earlier peak in height velocity was associated with shorter predicted height at age 18. Future work with data from the I-CAH registry will aim to assess a greater number of individuals followed to final height, highlighting the importance of ongoing data collection.

The associations we have shown between growth and glucocorticoid dose highlight the controversy in applying disease specific growth charts for this condition. Whilst patients with CAH have different growth trajectories to healthy patients, growth charts recently published developed from a heterogenous population of patients with CAH must be used with care, as comparisons may be made to data from patients that have had an unreported level of disease control (189). Bone age advancement was higher in individuals prescribed higher doses in this cohort.

Correlation does not necessarily mean causation, as it is undisputed that underdosing of glucocorticoid replacement in CAH is likely to cause bone age advancement. International guidance since 2002 has helped clinicians target appropriate dosing strategies for children with CAH, recommending regular review and dose adjusted dependent upon clinical need (5). Individuals deemed to have poor control, as indicated by bone age advancement or biomarkers of disease control, are therefore likely to be prescribed higher doses, indicative of the heterogeneity of treatment effect that we see from glucocorticoid replacement in CAH. Clinicians should aim to optimise growth in children with CAH towards normal growth.

A lack of association between bone age and glucocorticoid dose has often been shown in other studies (15, 190, 191). Karishma et al. found that those with higher doses had lower bone mineral density, although, those with low bone mineral density were significantly older within their cohort, and their comparison did not account for age (192). The non-linear relationship between bone age advancement and age we report here, similar to that shape reported by Bretones et al (11), highlights the importance of accounting for non-linearity in such analysis. Al-rayess et al found lower bone age SDS in children on lower doses, driven by the availability of glucocorticoids as suspension rather than tablets (179). That study design was able to take advantage of differences in available treatment preparations between their groups, inferring that those on tablet preparations were more likely to be receiving greater doses than they required, and thus a reason as to why their data showed the opposite association to that we report here. Our finding in a larger cohort of observational data highlights that bone age is a lagging indicator of previously inadequate disease control (5).

Limitations of this study include only 15% of visits having adequate data available for multivariate modelling, and the resultant simplification of multivariate models that had fewer degrees of freedom. Multivariate examples assume a consistent dose throughout childhood and adolescence, whereas patients may require higher doses during puberty due to the downregulation of 11β-Hydroxysteroid dehydrogenase type 1. SITAR modelling could only be carried out by restricting data to individuals with adequate data points for models to converge. We have not been able to assess timing of replacement doses, clinical pubertal staging, body fat percentage or other measures such as insulin resistance or bone density. However, the number of individuals modelled here provides estimates for adiposity rebound in CAH in the largest cohort to date. Age at diagnosis and ethnicity of patients as well as methods of measurement of biometrics, bone age and biomarkers vary between centres, and thus effect sizes must be interpreted with caution. This centre effect has been controlled for appropriately as a centre level random effect in

multivariate modelling but may still have some residual effect on the estimates presented, and caution must be applied before considering that associations will be consistent across different ethnic groups. Significant attrition with age means we must be cautious about the overall trajectories of growth presented. However, the longitudinal non-linear modelling complies with recent state-of-the-art guidance on such analysis (26). Age at adiposity rebound is highly variable in this cohort of children with CAH, early adiposity rebound and early increases in height velocity associated with detrimental growth trajectories that highlight patients at greater risk of adverse metabolic health outcomes in later life. Age at adiposity rebound thus has the potential to be used as a surrogate marker of long-term health outcomes when comparing treatment strategies for CAH in controlled studies in the future.

6.08 Conclusion

This study quantifies an early adiposity rebound in children with CAH and significantly blunted latest peak height velocity, with large variation between participants. Biomarkers of poor disease control and higher prescribed glucocorticoid doses were associated with earlier adiposity rebound and detrimental BMI in adolescence in girls. Participants of both sexes prescribed higher doses, indicative of more significant disease or possibly poor adherence, were associated with earlier bone age advancement. The timing of adiposity rebound and latest peak height velocity can allow the identification of individuals with CAH who may be at risk of adverse long-term cardiometabolic outcomes, and may be used as early surrogate outcomes in future research investigating early-life treatment strategies.

6.09 Data availability statement

All code associated with this analysis can be found at:

https://github.com/neilxlawrence/I-CAH_Adiposity_rebound

Requests for access to data must be sought through SDM Registries:

https://sdmregistries.org/data-access/

6.10 Acknowledgements and Funding

NRL is supported by the National Institute for Health and care Research (NIHR) by a doctoral research fellowship (NIHR302559). GSC is supported by a Cancer Research UK programme grant (C49297/A27294), and is a NIHR Senior

Investigator. The views expressed in this article are those of the author(s) and not necessarily those of the NIHR, or the Department of Health and Social Care.

We would like to thank all the children and families who have consented to share their data with the I-CAH Registry, without whom this research would not be possible.

6.11 Thesis contribution of chapter 6

Non-linear spline modelling on data from 573 patients with CAH has been completed and accepted for publication in *European Journal of Endocrinology* to achieve objective 5 (112). The differences in growth trajectory quantified within the patients of this cohort in comparison to healthy cohorts directly informed the decision to show disease specific centiles within the prototype clinical decision support tool described in the next chapter. The pre-processing of the data for objective 5 provided a suitable dataset for lambda-mu-sigma modelling to be carried out to produce the necessary reference centiles in the next chapter.

Table 6.1 – Summary statistics

	Total	Male	Female
Number of patients (n)	746	325	421
Number of visits (n)	11460	5690	5770
Number of countries (n)	22	20	22
Number of centres (n)	47	40	47
Number of visits per patient Median (Q1 to Q3)	11 (6 to 21)	14 (8 to 24)	10 (6 to 18)
Number of years visits spanned within patients Median (Q1 to Q3)	6.5 (2.9 to 10.7)	7.7 (3.6 to 11.3)	5.5 (2.4 to 10.0)
Age of patient at youngest visit (years) Median (Q1 to Q3)	0.3 (0.1 to 3.4)	0.2 (0.1 to 2.9)	0.3 (0.1 to 3.7)
Age of patient at most recent visit (years) Median (Q1 to Q3)	9.3 (5.6 to 13.6)	10.0 (6.4 to 13.6)	8.8 (5.0 to 13.5)
Number of visits with height available (% interpolated)‡	11251 (4.1)	5600 (3.9)	5651 (4.3)
Height SDS at visit (Median (Q1 to Q3)	-0.1 (-1.1 to 0.9)	0.0 (-1.0 to 1.0)	-0.2 (-1.1 to 0.7)
BMI SDS at visit (Median (Q1 to Q3)	0.7 (-0.2 to 1.5)	0.8 (-0.1 to 1.5)	0.6 (-0.3 to 1.5)
Number of visits with bone age available (% interpolated)‡	3803 (62.4)	2116 (64.5)	1687 (59.8)
Bone age advancement at visit (Median (Q1 to Q3)	0.9 (-0.3 to 2.5)	0.9 (-0.4 to 2.8)	0.9 (-0.2 to 2.2)
Visits with hydrocortisone equivalent dose available (n) (% carried to interpolate)	n=10088 (46.0%)	n=5129 (47.7%)	n=4959 (44.2%)
Dose of hydrocortisone equivalent at visit per BSA (mg/m^2/day) (Median (Q1 to Q3)†	13.8 (9.7 to 16.1)	14.1 (10 to 16.4)	13.5 (9.5 to 15.7)
Visits with fludrocortisone dose available (n) (% carried to interpolate)	n=10804 (1.7%)	n=5381 (1.3%)	n=5423 (2.2%)
Dose of fludrocortisone per day (micrograms) (Median (Q1 to Q3)	123.5 (50 to 200)	128.8 (50 to 200)	118.3 (50 to 150)
17-Hydroxyprogesterone (nmol/l) (Median [n] (Q1 to Q3)	29.8 [n=1888] (4.2 to 112.0)	29.0 [n=1041] (4.5 to 99.9)	30.3 [n=847] (3.9 to 128.3)
Androstenedione (nmol/l) (Median [n] (Q1 to Q3)	1.7 [n=1786] (0.2 to 4.4)	1.4 [n=987] (0.0 to 3.5)	2.0 [n=799] (0.3 to 5.6)

n=Number; Q1=Quartile 1; Q3=Quartile 3; SDS=Standard deviation score; BSA=Body surface area

†Hydrocortisone equivalent calculated by multiplying preparations by the following factors: prednisolone/prednisone x4; dexamethasone x27; cortisone acetate x0.8; methylprednisolone x5 (Appendix Table S6.01 for all conversion factors (180)). 84.9% of visits prescribed hydrocortisone (Appendix Table S6.02 for proportions of preparations prescribed (180)). Missing data for dose was carried forward or backward within clinically plausible limits set by age (Appendix Supplementary Methods S6.01 (180))

‡ Missing data for growth and bone age was interpolated between visits using splines, but not extrapolated beyond first or last available data point for each individual patient.

2464 Table 6.2 – SITAR Model derived metrics of growth

	Male patients			Female patients		
Age at latest peak height velocity (years) Mean (SD)	8.4 (3.0)			9.0 (1.6)		
Latest peak height velocity (cm/year) Mean (SD)	7.7 (1.4)			7.4 (1.3)		
Age at adiposity rebound Mean (SD)	3.7 (1.3)			3.9 (0.9)		
BMI at adiposity rebound Mean (SD)	16.2 (1.5)			16.0 (1.5)		
SITAR outcome variable	**Height (cm)**	**Weight (kg)**	**BMI (kg/m^2)**	**Height (cm)**	**Weight (kg)**	**BMI (kg/m^2)**
Number of patients used to fit SITAR model	266	267	219	307	309	238
Number of data points used to fit SITAR model	5196	5224	3030	5065	5101	2871
SITAR predicted value at age 18 Mean (SD)	166.9 (14.2)	74.8 (17.0)	27.1 (6.0)	157.2 (7.2)	68.9 (9.7)	28.0 (4.5)
WHO z-score of SITAR predicted value at age 18 Mean (SD)*	-1.2 (1.9)	-	1.3 (1.3)	-0.9 (1.1)	-	1.6 (1.0)
SITAR degrees of freedom	5	6	3	5	6	3
Restricted age range of patients to fit SITAR (years)	0.2 to 20	0.2 to 20	0.2 to 7	0.2 to 20	0.2 to 20	0.2 to 7
Root Mean Square Error	1.50	1.50	0.77	2.60	2.1	0.91
Marginal R^2	0.74	0.92	0.12	0.91	0.90	0.07
Conditional R^2	0.99	0.99	0.84	0.99	0.91	0.80

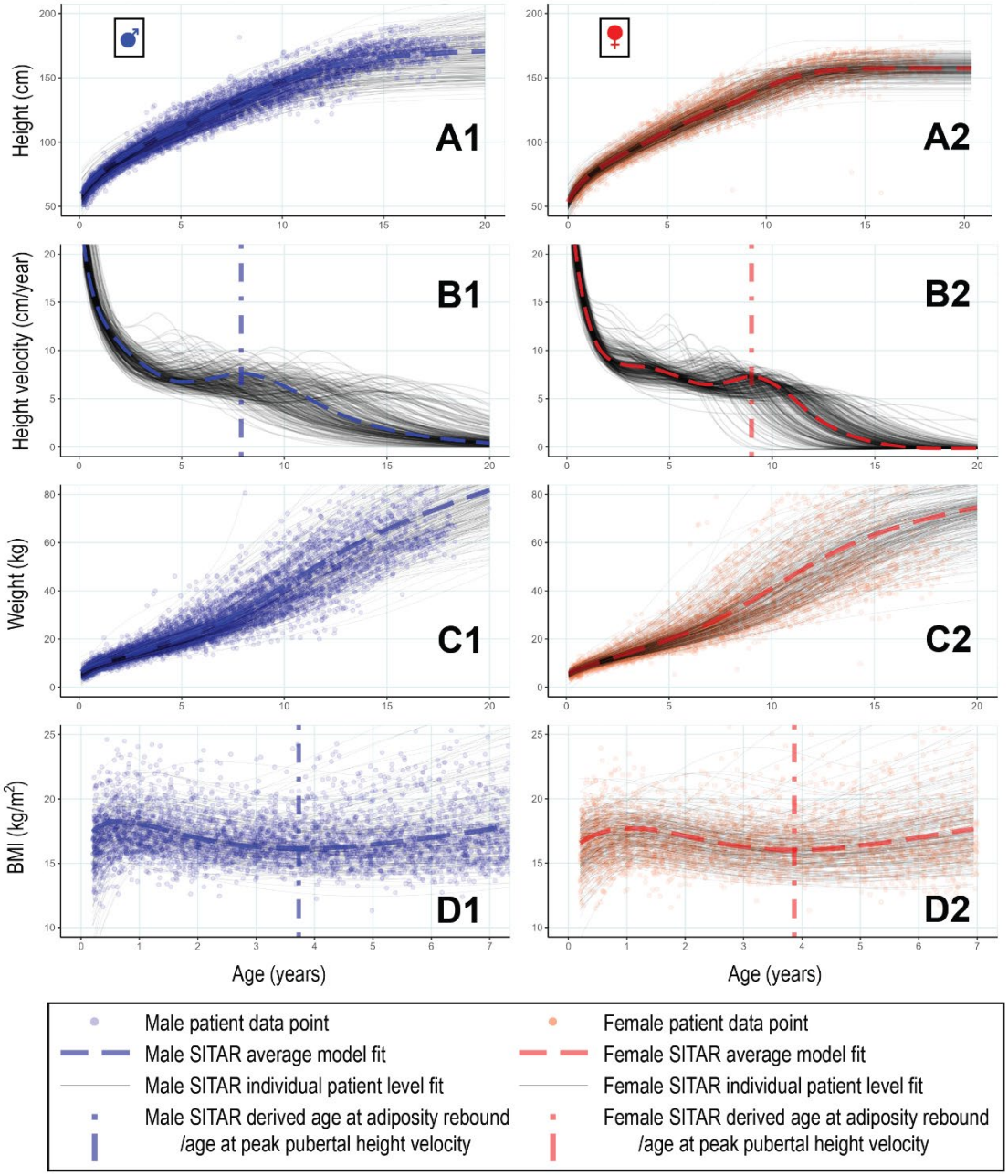
2465 SD=Standard deviation, variability defined by between patient random effects. SITAR: SuperImposition by Translation And Rotation. Optimum degrees of freedom defined by lowest Bayesian Information Criterion. Patients'
2466 included in models if they had six or more measures of the necessary biometric.
2467 *WHO do not provide reference standards for weight for age beyond 10 years

Table 6.3 – Covariate adjusted multivariable model metrics assessing disease associations with growth

Sex of patients in underlying model	Male					Female			
Data points with covariates available to train model	**982 Visits | 139 patients**					**803 Visits | 143 patients**			
Representation of model	**Low dose good control**	**Low dose poor control**	**High dose good control**	**High dose poor control**		**Low dose good control**	**Low dose poor control**	**High dose good control**	**High dose poor control**
Metrics inserted to demonstrate covariate model fit									
Hydrocortisone equivalent (mg/m^2/day)	**8**	**8**	**18**	**18**		**8**	**8**	**18**	**18**
Fludrocortisone dose (micrograms/day)	**100**	**100**	**100**	**100**		**150**	**150**	**150**	**150**
17-Hydroxyprogesterone (nmol/l)	**10**	**38**	**10**	**38**		**10**	**38**	**10**	**38**
Model derived estimates: (bold type indicates statistically significant difference between those on low dose with good control versus those with high dose and poor control)									
Height at age 8 (cm) (95% CI)	131.0 (128.9 to 132.9)	131.5 (129.3 to 133.3)	133.1 (131.2 to 134.9)	133.58 (131.47 to 135.62)		129.4 (127.8 to 131.0)	129.6 (127.9 to 131.2)	129.8 (127.1 to 132.3)	130.0 (127.0 to 132.7)
Height SDS at age 8 (95% CI)	0.7 (0.3 to 1.0)	0.7 (0.4 to 1.1)	1.0 (0.7 to 1.4)	1.12 (0.74 to 1.48)		0.5 (0.2 to 0.8)	0.5 (0.2 to 0.8)	0.6 (0.1 to 1.0)	0.6 (0.1 to 1.1)
Weight at age 8 (kg) (95% CI)	32.0 (29.3 to 34.1)	32.3 (29.7 to 34.7)	35.0 (32.2 to 38.4)	35.33 (32.28 to 38.89)		30.9 (28.9 to 33.1)	30.9 (28.8 to 33.0)	33.9 (28.8 to 38.9)	34.0 (28.8 to 39.0)
Weight SDS at age 8 (95% CI)	1.5 (1.0 to 1.9)	1.6 (1.0 to 2.0)	2.0 (1.5 to 2.6)	2.1 (1.56 to 2.65)		1.2 (0.9 to 1.6)	1.2 (0.8 to 1.6)	1.7 (0.8 to 2.4)	1.7 (0.8 to 2.4)
BMI at age 8 (kg/m^2) (95% CI)	18.6 (17.5 to 19.6)	18.7 (17.6 to 19.8)	19.7 (18.4 to 21.3)	19.75 (18.46 to 21.32)		18.5 (17.3 to 19.6)	18.4 (17.3 to 19.6)	20.1 (17.5 to 22.8)	20.0 (17.5 to 22.6)
BMI SDS at age 8 (95% CI)	1.6 (1.0 to 2.0)	1.6 (1.1 to 2.0)	2.0 (1.5 to 2.6)	2.03 (1.5 to 2.58)		1.3 (0.8 to 1.7)	1.3 (0.8 to 1.7)	1.9 (0.9 to 2.6)	1.8 (0.9 to 2.5)
Height at age 12 (cm) (95% CI)	152.3 (149.2 to 155.1)	152.5 (149.6 to 155.1)	155.2 (152.5 to 158.2)	155.4 (152.92 to 158.58)		148.5 (146.6 to 150.5)	148.9 (147.1 to 150.8)	149.4 (146.6 to 152.2)	149.8 (147.2 to 152.6)
Height SDS at age 12 (95% CI)	0.5 (0.0 to 0.9)	0.5 (0.1 to 0.9)	0.9 (0.5 to 1.3)	0.89 (0.54 to 1.34)		-0.4 (-0.7 to -0.1)	-0.3 (-0.6 to -0.1)	-0.3 (-0.7 to 0.2)	-0.2 (-0.6 to 0.2)
Weight at age 12 (kg) (95% CI)	47.1 (41.6 to 52.6)	47.8 (42.1 to 53.3)	50.3 (46.9 to 53.7)	50.88 (47.16 to 54.76)		43.6 (40.7 to 46.7)	44.3 (41.1 to 47.7)	48.2 (42.9 to 52.8)	48.9 (44.2 to 53.5)
BMI at age 12 (kg/m^2) (95% CI)	20.3 (18.3 to 22.1)	20.6 (18.5 to 22.3)	20.9 (19.7 to 22.0)	21.1 (20.0 to 22.1)		19.8 (18.7 to 21.1)	20.0 (18.8 to 21.4)	21.6 (19.7 to 23.2)	21.7 (20.1 to 23.3)
BMI SDS at age 12 (95% CI)	1.1 (0.4 to 1.7)	1.2 (0.4 to 1.7)	1.3 (0.9 to 1.6)	1.4 (1.0 to 1.6)		0.7 (0.3 to 1.1)	0.8 (0.3 to 1.2)	1.2 (0.7 to 1.6)	1.3 (0.8 to 1.7)
Age at adiposity rebound (95% CI)	3.6 (2.5 to 4.8)	3.5 (2.4 to 4.5)	2.9 (1.1 to 3.8)	2.9 (0.7 to 3.7)		4.8 (2.6 to 7.3)	4.8 (2.6 to 6.6)	2.5 (1.9 to 4.0)	2.6 (1.9 to 4.1)
BMI at adiposity rebound (kg/m^2) (95% CI)	16.5 (15.5 to 17.4)	16.2 (15.3 to 17.1)	16.6 (15.3 to 17.7)	16.3 (14.9 to 17.5)		17.5 (16.1 to 18.7)	17.3 (16.0 to 18.5)	16.3 (14.3 to 19.0)	16.2 (14.0 to 19.0)
BMI z-score at adiposity rebound (95% CI)	0.8 (-0.1 to 1.5)	0.6 (-0.3 to 1.3)	0.7 (-0.4 to 1.5)	0.5 (-0.8 to 1.4)		1.3 (0.6 to 1.9)	1.2 (0.5 to 1.8)	0.6 (-1.0 to 2.3)	0.6 (-1.1 to 2.2)

CI = Confidence interval. SDS=Standard deviation score (calculated from World Health Organisation normative standards, not available for weight above age 12). Doses and level of 17-Hydroxyprogesterone to define disease control defined by expert consensus, underlying model trained from visits extracted from the International Congenital Adrenal Hyperplasia Registry with all covariates available. Full multilevel spline model estimates reported in Appendix tables S6.06-6.07 (180). Metrics calculated for alternative doses and levels of disease control applied to the same models are in Appendix Table S6.08 (180). Comparisons between low dose good control and high dose poor control made across 500 bootstrapped replications to define statistical significance of the difference between metrics indicated by bold type

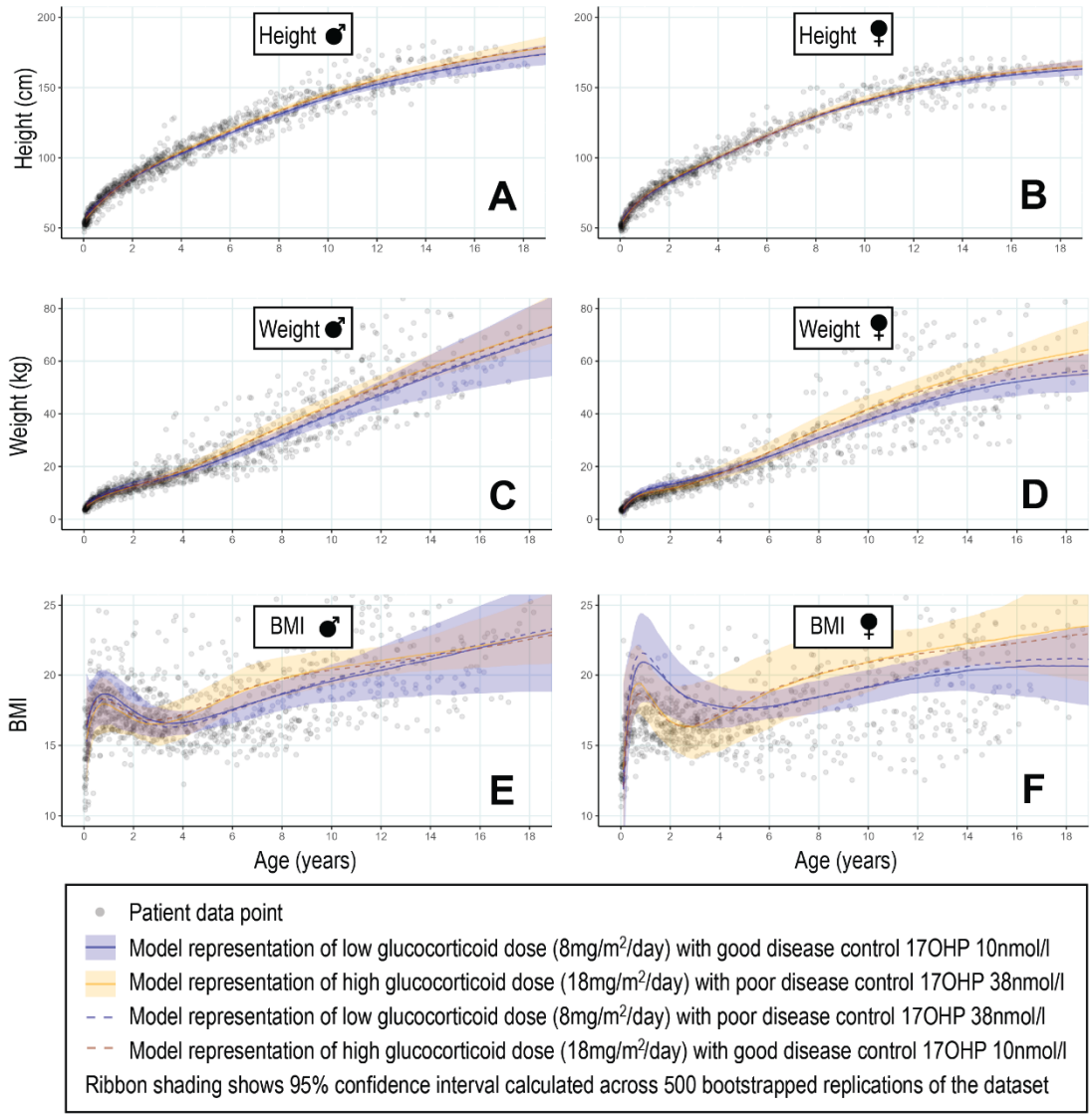
2472 Figure 6.1 – SITAR models to quantify height, weight and BMI



2473

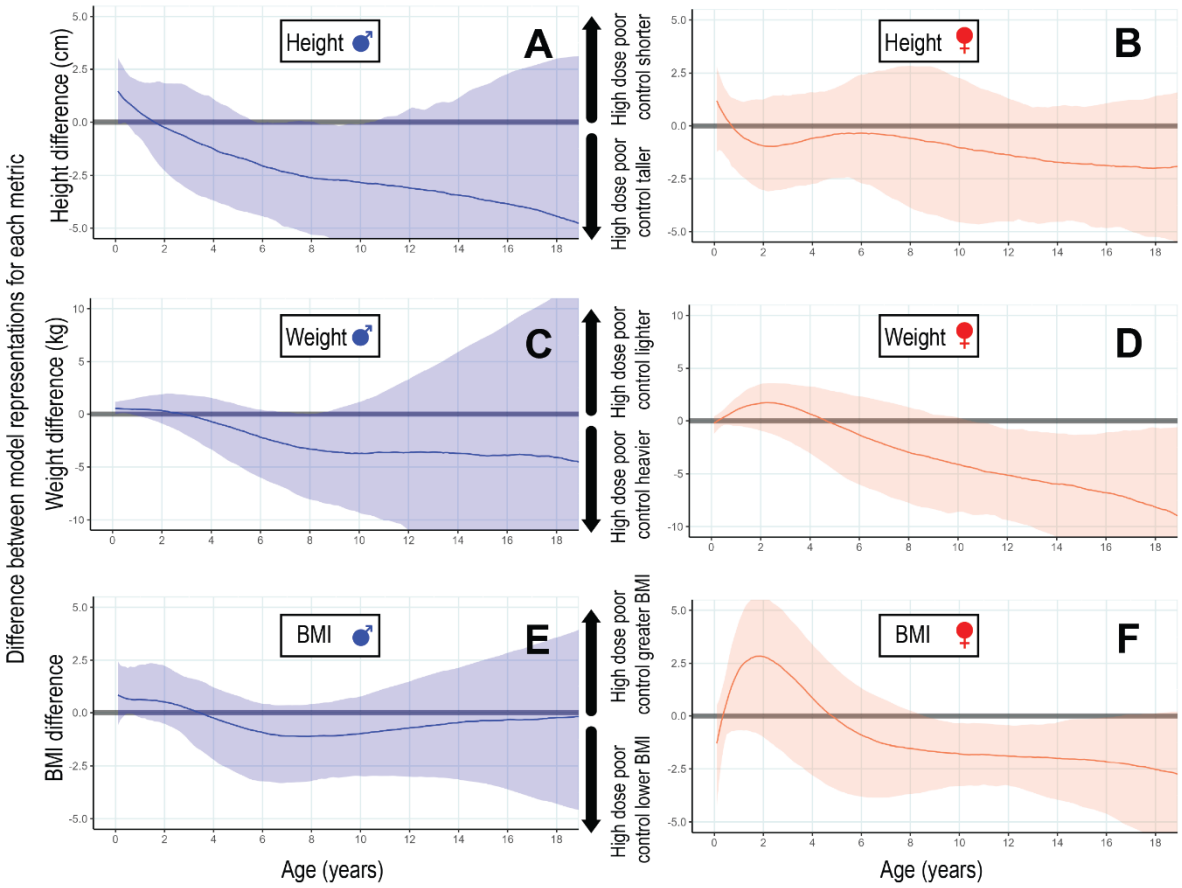
2474 Prefix A = Height; B = Height Velocity; C = Weight; D = BMI. Suffix 1=Male; 2=Female. SITAR: SuperImposition by Translation And Rotation

Figure 6.2 – Covariate models representations of different glucocorticoid doses and levels of 17-Hydroxyprogesterone



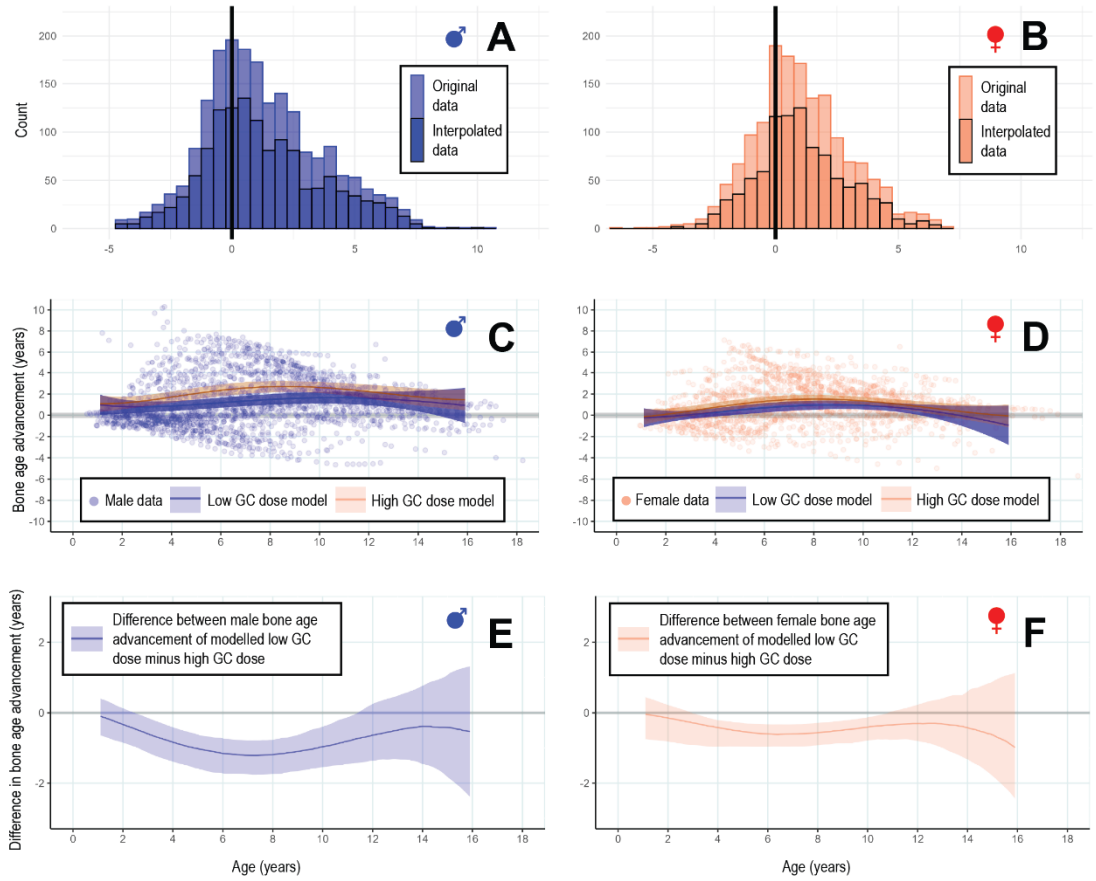
A = Male Height; B = Female Height; C = Male weight; D = Female weight. E = Male BMI; F = Female BMI. Figures show trajectories of growth for modelled values contained within Table 6.3. Alternative example values applied to models is found in Appendix Figure S6.05 (180).

Figure 6.3 – Bootstrapped differences between model representation of patient with low glucocorticoid dose and consistently good disease control versus patient with high glucocorticoid dose and consistently poor disease control



Model trajectories calculated from multivariate models with the following clinically meaningful values: Low glucocorticoid replacement dose defined as 8mg/m^2/day hydrocortisone equivalent, high defined as 18mg/m^2/day; Good disease control defined as 17-Hydroxyprogesterone consistently 10nmol/l, poor disease control defined as 38nmol/l. Differences calculated as a model representation of patients with low dose good control minus a model representation of patients with high dose and poor control. 95% Confidence interval shading produced by calculating difference in metrics across 500 bootstrap replications of the data. A: Male height difference, B: female height difference, C: male weight difference D: female weight difference, E: male BMI difference, F: female BMI difference

Figure 6.4 – Bone age advancement modelling



Data available for bone age modelling, displaying parsimonious bone age models with only dose of hydrocortisone equivalent added as a covariate. A&B: bone age data available, showing proportion of values interpolated between measurements, C&D: Bone age advancement (bone age – chronological age) in individuals overlaid with model representations of those prescribed low dose hydrocortisone equivalent (8mg/m^2/day) versus those prescribed high dose hydrocortisone equivalent (18mg/m^2/day), E&F: Difference between model representations of bone age advancement in those on high dose subtracted from those on low dose. A,C,E male data and models, B, D, F female data and models

Chapter 7 – Prototype clinical decision support tool for managing patients with congenital adrenal hyperplasia

This chapter has been written by Neil Lawrence, with advice from the supervisory team Jeremy Dawson, Zi-Qiang Lang, Gary Collins, and Nils Krone.

7.01 Background

The International Congenital Adrenal Hyperplasia Registry (I-CAH Registry) is a multicentre collaboration that collates real-world data from children with Congenital Adrenal Hyperplasia (CAH) for purposes of research. Now part of the wider International Registries for Rare Conditions Affecting Sex Development and Maturation (SDM registries), the I-CAH Registry evolved from part of the International Disorders of Sexual Development Registry, developed as part of the EuroDSD project in 2008 (154, 193). Engagement with SDM registries has increased year on year, the entire registry platform containing 10,179 patients from 167 centres across 46 countries as of June 2025 (194).

Longitudinal data analysis from the I-CAH Registry has been presented in this thesis to describe trends in blood pressure (Chapter 5) and growth (Chapter 6) in children with CAH. These have shown significant differences to normative data, with data on blood pressure showing a tendency towards higher readings at younger ages. Modelling of height, weight and body mass index (BMI) has shown an earlier adiposity rebound and significantly blunted pubertal growth spurt. Comparing measurements in patients with CAH to standard normative blood pressure and growth charts is therefore likely to show a disproportionate amount of measurements outside of what is considered the normal range.

CAH is a rare condition, affecting approximately 1 in 10,000 to 15,000 people (5). The I-CAH Registry has facilitated the analysis of one of the largest datasets of patients with this condition to date. However, the real-world nature of the data that is available does present challenges. Alongside a lack of standardisation of measurement protocols and laboratory analysis, there exists the simple but common limitation of mistakes in data entry. Data analysed within this thesis has been pre-processed by liaison with centres about clear data entry errors, a laborious process and one subject to limitations. Biologically implausible entries are often easy to identify, but more subtle entries are likely to

go undetected. The process of liaison with participating centres has also highlighted how research contributors are struggling to find time around clinical commitments to enter data into registries such as the I-CAH Registry. There is a significant lag between contribution of data to registry research, and presentation of results at academic meetings or in journal articles (9, 38). The SDM Registry team report a median of 3 years from project approval to publication. Consent of patients and data entry frequently occurs prior to research study development and approval. Establishing ways in which research centres can benefit from data entry earlier on in this process is likely to increase engagement with data entry, and reduce attrition, leading to richer, more powerful data available for longitudinal analysis. We set out to design a prototype clinical decision support tool for clinicians tailored towards patients with CAH to help visualise longitudinal measurements and contextualise them against modelling from previous I-CAH Registry data analysis.

7.02 Methods

Data from the I-CAH Registry was modelled in each sex separately using Lambda-Mu-Sigma for blood pressure as described and published previously (113). The same modelling technique was applied to height, weight and BMI from data extracted from the same registry in 2024 as part of the research reported in Chapter 6. Data was fit to a Box-Cox t-distribution, that estimates the mean (μ) and standard deviation (σ) as well as the amount of skew (ν) and size of the 'tail' (τ) of the data within the distribution. The model fit is estimated around a penalised B-spline that estimates the median of the data on the dependent variable, age. The model around the median is estimated such that the shape of the distribution would reduce to the more commonly known normal distribution when ν approaches zero and τ approaches infinity. Statistical analysis and application development was conducted in R 4.4.1, *a language and environment for statistical computing* (56).

The *Shiny* R library was employed to create a web application with a user interface designed to allow user entry of longitudinal data from patients with CAH. This was uploaded to the cloud hosting platform shinyapps.io to allow visualisation of this prototype within a web browser interface. A disclaimer was added to the application to ensure anyone viewing the application understands that it is a prototype, and should not be used in clinical practice.

User Interface

A user interface was designed within the Shiny application that separated patient level data from visit level data into a 'sidebar' and 'main panel' respectively. Within the main panel, the command 'tabPanel' was employed to subdivide the interface into different sections. The dynamic command 'observeEvent' was employed to allow the user to add data from multiple clinical assessment visits. The opportunity to download data as a comma separated values (CSV) file was incorporated alongside an option to upload previous data from such a file, to ensure users wouldn't have to start entering data from point zero every time the application was used. A further option to download a portable document format (PDF) report was incorporated to demonstrate how clinical decision support could be logged and saved in other electronic or paper records.

Visualisations

Longitudinal plots were created using the library 'ggplot2', overlaying data inserted by the user for each visit onto previously loaded centile lines developed from LMS modelling. The shiny command 'observe' was employed to ensure plots of growth and blood pressure rendered automatically when new data was inserted by the user into the application. Combined visualisations showing all longitudinal data entered into the application were shown in a separate tab named 'All Visits', alongside an extra tab that included instructions for application use. Centile lines were calculated and plotted at 2^{nd}, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th} and 98^{th} centiles.

Medication calculations

A tab within each visit was created to allow users to enter up to four different doses of glucocorticoid replacement. This tab allowed the user to declare the type of glucocorticoid replacement, and employed conversion ratios from the British National Formulary to equate these to a total daily hydrocortisone equivalent (159). Data from the growth tab was then used to calculate body surface area by the Mosteller formula (195), to give users a total daily hydrocortisone equivalent per body surface area to compare directly against the 10-15mg/m^2/day recommended in children by international guidelines (5).

7.03 Results

Underlying data

Data used to create condition specific LMS models of blood pressure, height, weight and BMI has been described within Chapter 5 and Chapter 6. The LMS model parameters are shown in Tables 7.1-7.5. The LMS models use the data to estimate centiles assuming a Box-Cox t-distribution. Due to the complexity of this fit, the parameters themselves are of limited use, but allow for the mathematical calculation of sample population centiles that aid interpretation when visualised (Figures 7.1-7.2).

Shiny application user interface

The web application can be accessed via the shiny servers at the following URL: https://endocrinology.shinyapps.io/cah_data_visualiser/ . The Shiny web interface employs a mixture of web technologies to render a user-friendly output. This interface automatically accounts for the size of the screen on which the user is loading the application. The screen loaded on a desktop computer is displayed with annotations in Figure 7.1. Excerpts from an example downloaded report are shown in Figure 7.2. The code underlying this application can be found at https://github.com/neilxlawrence/cah_data_visualiser .

7.04 Discussion

This chapter presents a user-friendly interface that offers instant visualisation of individual patient data designed to be applied to children living with CAH, displayed against centiles created from data extracted from the International CAH Registry. This interactive application shows a prototype of software that could incentivise clinicians to enter data to the I-CAH Registry, as it allows instant feedback that could help support clinical decisions and visualise results for families to improve shared decision making.

The I-CAH Registry accepts data entry through a web browser interface. The registry has gone through several iterations since its inception, both to the user interface and to the underlying database and non-user facing underlying database. However, no iteration has provided users with any facility to visualise data from patients within their centre, or to summarise data from patients within the centre to facilitate local audit or quality improvement. The application presented here shows a more user focussed design of a data entry system tailored towards automated data visualisation to promote clinical utility alongside research data collection.

The UK Medicines and Healthcare Regulatory Agency (MHRA) provide guidance on the development of software as a medical device, as do the United States Food and Drug Administration (FDA) (196, 197). The British standard BS EN 62366 outlines technical considerations for state of the art practice in considering ergonomics within the development of medical devices (198). However, the increasing use of artificial intelligence is providing new challenges that regulatory agencies are working hard to keep pace with (199). The prototype application presented here would not be classified as a medical device in its current form as it is displaying data against previously published research (200). However, development and incorporation of algorithmic dosing predictions that advise clinicians about management would cross the threshold into the development of a class IIa medical device (53). As CAH is a rare disease effecting between 1 in 10,000 to 1 in 15,000 people (5), developing software as a medical device to the state of the art would be unlikely to be economically viable (201). This prototype would be better employed to advocate for incremental improvements to the existing I-CAH Registry to improve the user experience of clinicians engaged in data entry and research.

The presentation of research in the scientific literature has evolved, with mainstream journals highlighting important aspects of research with panels such as 'What this study adds', and mandatory reporting of progressive research aspects such as the involvement of patients and the public in research (202). Graphical and multimedia abstracts are undoubtedly adding to the diversity of tools available for research communication (203). However, small studies have shown conflicting findings about how the inclusion of such novel abstracts contribute to citation or Altmetric scores (204, 205). The incorporation of other outputs including software and datasets within the contributions to knowledge and understanding of the UK research excellence framework means that there are increasing incentives to consider how to improve the impact of research beyond a traditional publication (206). However, some report that such incremental changes do not go far enough to shift researchers to focus on increasing the impact of their research, rather than just its rate of publication and citation (207).

Machine learning and artificial intelligence methods applied within data analysis and risk prediction are becoming increasingly sophisticated. Transparent reporting guidance recommends extending articles into supplementary material to detail the necessary information that would allow reproduction of the research (208). However, this is an imperfect solution, with others advocating for the removal of arguably arbitrary journal article word limits to prevent important information being inadequately reviewed when relegated to supplementary material (209). The significant

lack of reporting of final prediction models in any form shown within the systematic review in Chapter 2 highlights the importance of the reporting of data in relation to such studies that does not fit into the traditional model of journal article prose, tables and figures. The current practice of peer review could be advanced to encourage reviewers to consider whether alternative media should be requested to supplement research. Advocating for more basic web applications to display data science could encourage readers to interact with research articles in a more meaningful way.

The presented analysis and application have limitations. The growth centiles are based upon real-world registry data including a heterogenous collection of individuals with CAH from different countries with varying levels of underlying disease. As CAH is known to effect growth and blood pressure, visualisation in comparison to centiles created from individuals, some of whom are likely to have been living with poor disease control, may provide a suboptimal comparison. This should be weighed against comparing patients with CAH to healthy participant centiles, that would itself be limited as we have shown differences in blood pressure and growth in patients with CAH within this cohort. The application is only in prototype form, and has not undergone user testing. There is an absence of any incorporated safety mechanisms, such as implausible data input detection, which risks data entry error. However, the incorporation of instantly updating data visualisation does provide the user with visual feedback, that increases the likelihood of detection of clear outlying data by the user. The current I-CAH Registry platform does not provide this, and thus this prototype is suitable to present to the SDM Registries project management group and data access committee to advocate for the incorporation of active data visualisation within the next iteration of I-CAH Registry user interface.

7.05 Conclusion

Real-world registry data and LMS statistical modelling have been used to develop centile lines for blood pressure and growth in children with CAH. These centile lines have been incorporated into a user-friendly web application that allows the entry and download of individual patient data, with automated visualisation of measurements against derived centiles. The application also provides clinically relevant calculations of glucocorticoid replacement conversion between preparations, and calculation of total daily dose per body surface area. This application is not recommended for clinical use, but can be used to advocate for improved feedback and data visualisation for clinicians entering data into the International CAH Registry.

7.06 Thesis contribution of chapter 7

A prototype clinical decision support tool has been developed and presented in chapter 7 to achieve objective 6. This has been developed using insights from previous objectives. Figure 1.1 describing the chronology and contribution of each chapter and objective is reproduced here to contextualise the steps in the research:

7.07 Graphical thesis flow chart

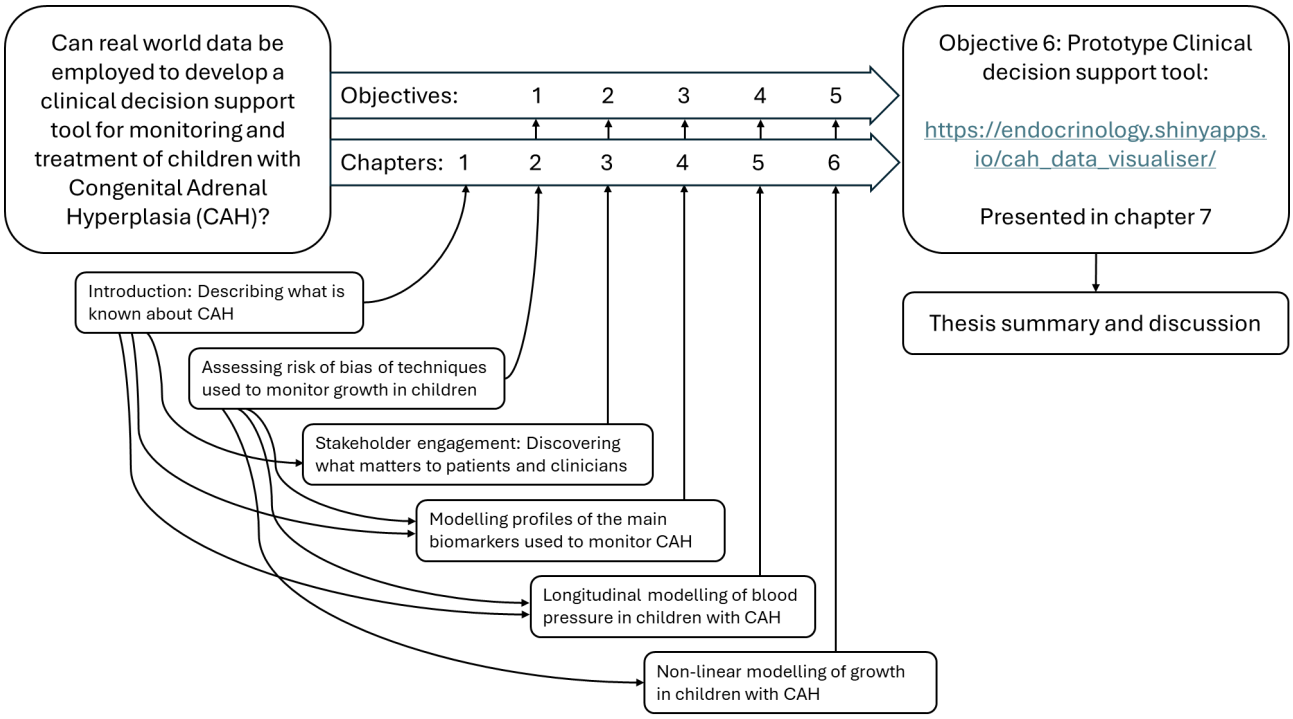


Table 7.1 - LMS model parameters Height

Dependent variable: Height (cm)	**Male LMS model parameters**			**Female LMS model parameters**		
Parameter	**Estimate**	**Standard error**	***p-value***	**Estimate**	**Standard error**	***p-value***
Intercept (μ)	4.284	0.001	**<0.001**	4.277	0.001	**<0.001**
μ	0.064	0.000	**<0.001**	0.063	0.000	**<0.001**
Intercept (σ)	-2.946	0.016	**<0.001**	-3.013	0.019	**<0.001**
σ	-0.002	0.002	0.303	-0.001	0.002	0.531
Intercept (ν)	0.303	0.326	0.352	0.967	0.389	**0.013**
ν	0.072	0.044	0.101	0.092	0.052	0.079
Intercept (τ)	2.965	0.240	**<0.001**	2.193	0.191	**<0.001**
τ	-0.005	0.031	0.867	0.134	0.050	0.007
Number of visits in model	5863			4574		
Number of patients in model	350			351		

LMS model fit in each sex with dependent variable of height. Mu is the generalised additive model estimate of the median of the dependent variable. Sigma estimates the scale or spread of the data around the median. Nu estimates the amount of skew of the spread of the data around the median. Tau estimates the kurtosis of the tail of the skew around the median. These four parameters allow representation of the Box-cox t-distribution, that allows for a more flexible deviation in ‘skewness’ from the normal distribution.

2678 Table 7.2 – LMS model parameters Weight

Dependent variable: Weight (kg)	**Male LMS model parameters**			**Female LMS model parameters**		
Parameter	**Estimate**	**Standard error**	***p-value***	**Estimate**	**Standard error**	***p-value***
Intercept (μ)	2.104	0.003	**<0.001**	2.067	0.004	**<0.001**
μ	0.149	0.001	**<0.001**	0.146	0.001	**<0.001**
Intercept (σ)	-1.890	0.017	**<0.001**	-2.015	0.019	**<0.001**
σ	0.017	0.002	**<0.001**	0.046	0.002	**<0.001**
Intercept (ν)	-0.194	0.106	0.067	0.206	0.119	0.084
ν	-0.030	0.013	**0.012**	-0.060	0.013	**<0.001**
Intercept (τ)	2.240	0.124	**<0.001**	2.093	0.144	**<0.001**
τ	0.035	0.019	0.068	0.052	0.025	**0.036**
Number of visits in model	5913			4631		
Number of patients in model	351			351		

2679

2680 Table 7.3 – LMS model parameters BMI

Dependent variable: BMI (kg/m^2)	**Male LMS model parameters**			**Female LMS model parameters**		
Parameter	**Estimate**	**Standard error**	***p-value***	**Estimate**	**Standard error**	***p-value***
Intercept (μ)	2.755	0.002	**<0.001**	2.735	0.003	**<0.001**
μ	0.019	0.000	**<0.001**	0.020	0.000	**<0.001**
Intercept (σ)	-2.391	0.017	**<0.001**	-2.488	0.020	**<0.001**
σ	0.036	0.002	**<0.001**	0.068	0.003	**<0.001**
Intercept (ν)	-1.269	0.157	**<0.001**	-0.702	0.173	**<0.001**
ν	-0.012	0.017	0.486	-0.028	0.017	0.100
Intercept (τ)	2.291	0.162	**<0.001**	1.581	0.117	**<0.001**
τ	0.074	0.031	**0.018**	0.142	0.030	**<0.001**
Number of visits in model	5861			4574		
Number of patients in model	350			351		

2681

2682 Table 7.4 – LMS model parameters Systolic Blood Pressure

2683

Dependent variable: Systolic Blood Pressure (mmHg)	**Male LMS model parameters**			**Female LMS model parameters**		
Parameter	**Estimate**	**Standard error**	***p-value***	**Estimate**	**Standard error**	***p-value***
Intercept (μ)	4.620	0.006	<0.001	4.591	0.006	<0.001
μ	0.008	0.001	<0.001	0.009	0.001	<0.001
Intercept (σ)	-1.797	0.027	<0.001	-1.876	0.031	<0.001
σ	-0.039	0.004	<0.001	-0.029	0.003	<0.001
Intercept (ν)	0.010	0.183	0.957	0.011	0.198	0.956
ν	-0.023	0.031	0.470	-0.022	0.027	0.407
Intercept (τ)	4.620	0.006	<0.001	2.624	0.340	<0.001
τ	0.008	0.001	<0.001	0.008	0.040	0.834
Number of visits in model	1556			1652		
Number of patients in model	215			244		

2684

2685 Table 7.5 – LMS model parameters Diastolic Blood Pressure

2686

Dependent variable: Diastolic Blood Pressure (mmHg)	Male LMS model parameters			Female LMS model parameters		
Parameter	**Estimate**	**Standard error**	**Parameter**	**Estimate**	**Standard error**	**Parameter**
Intercept (μ)	4.123	0.007	<0.001	4.099	0.007	<0.001
μ	0.006	0.001	<0.001	0.008	0.001	<0.001
Intercept (σ)	-1.513	0.030	<0.001	-1.656	0.031	<0.001
σ	-0.043	0.004	<0.001	-0.032	0.003	<0.001
Intercept (ν)	0.302	0.157	0.055	0.152	0.169	0.371
ν	-0.016	0.028	0.573	0.019	0.024	0.425
Intercept (τ)	0.506	0.083	<0.001	2.542	0.349	<0.001
τ	0.001	0.011	0.915	0.040	0.050	0.424
Number of visits in model	1542			1645		
Number of patients in model	215			244		

2687

Figure 7.1 – User interface

User interface of application with annotations describing features. Available at https://endocrinology.shinyapps.io/cah_data_visualiser/

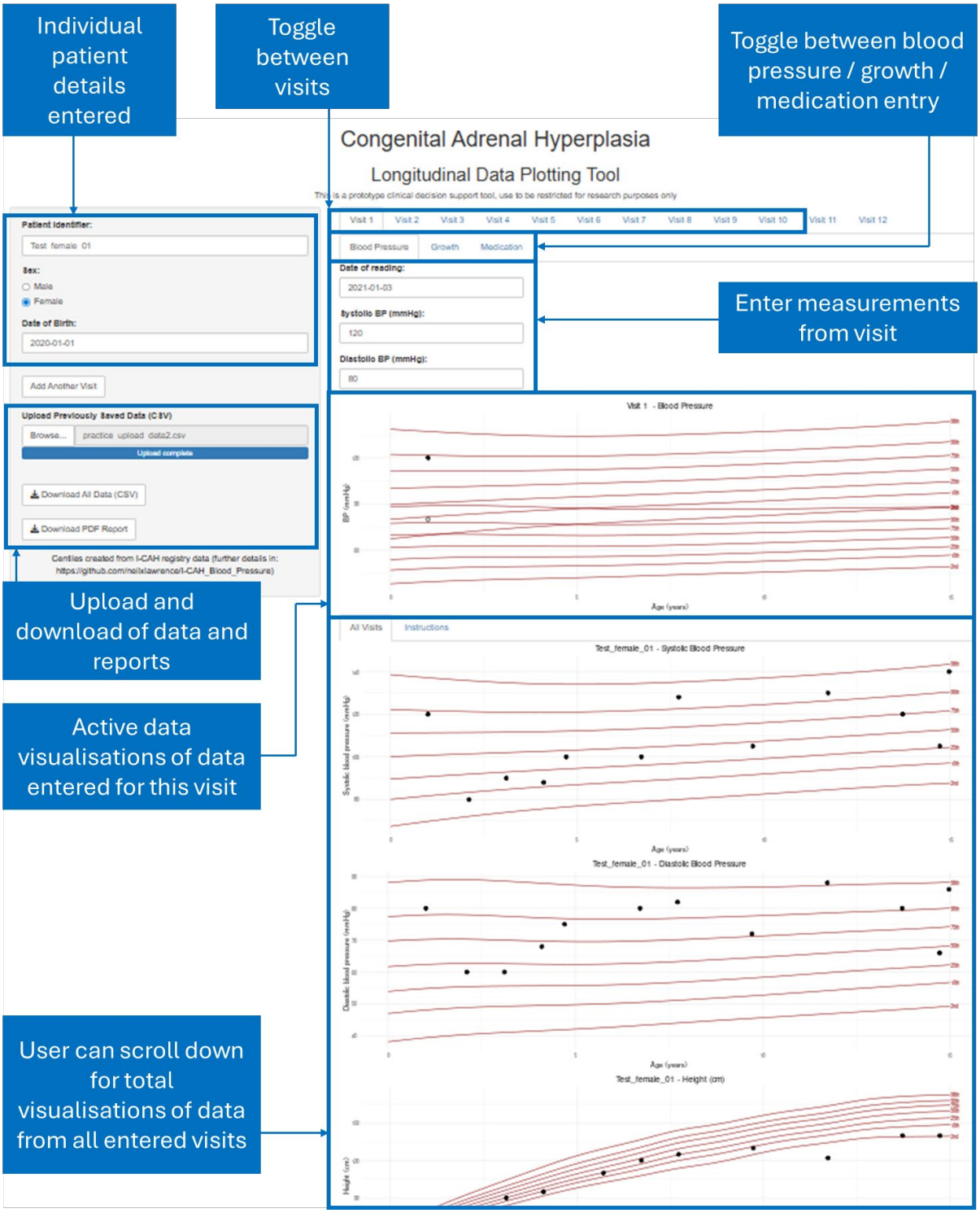
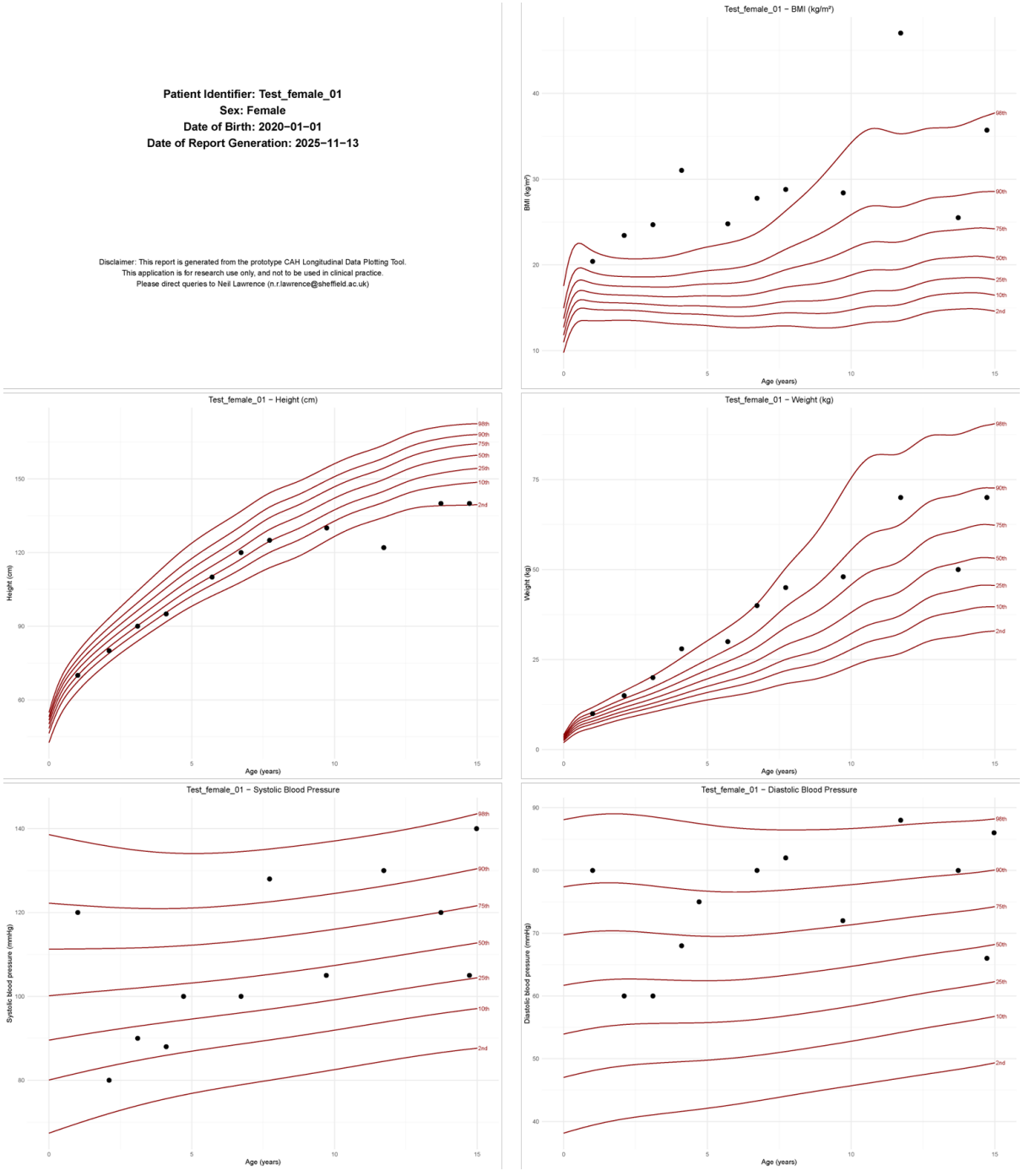


Figure 7.2 – Downloadable Report

Each page of an example downloaded pdf report, available from the web application shown in Figure 7.1.



Chapter 8 – Thesis Summary and Discussion

This final discussion has been written by Neil Lawrence, with advice from the supervisory team Jeremy Dawson, Zi-Qiang Lang, Gary Collins, and Nils Krone

8.01 Thesis Discussion

This thesis reports the development of a prototype clinical decision support tool for Congenital Adrenal Hyperplasia (CAH) that highlights the benefits of real time data visualisation and calculation of relative drug doses relevant to this disease. Incorporation of this tool into the online I-CAH Registry data entry platform would provide those entering data into the registry with quick feedback that would allow intuitive recognition of data entry errors as well as longitudinal visualisation of growth and blood pressure trajectories that clinicians use to inform their clinical decisions. This prototype clinical decision support tool can be found at (https://endocrinology.shinyapps.io/cah_data_visualiser/).

The thesis began with a systematic review into prediction models relevant to growth in paediatrics or fetal medicine, to inform the current standard of practice in this domain, hoping to learn from best practice methods. This review was confined to models published in 2022 to ensure a contemporary representation of current practices. Unfortunately, this review found a high risk of bias in all of the 241 models assessed (210), highlighting why so few of such published models translate into clinical practice (81). None of these models incorporated repeated measures assessments to inform predictions, whereas humans intuitively assess trends when forecasting (211), and clinicians are advised to do so by national guidance when assessing child growth (212). The insights from this review helped guide the analysis of data from the I-CAH Registry from individual patient prediction of disease outcomes, towards using more appropriate multilevel modelling techniques to provide statistical inferences about the course of CAH throughout childhood.

Chapter 3 reported a national service evaluation investigating management of CAH that helped frame analysis of data from patients with the condition within the context of the current provision of care across the UK. Data from 33 hospitals showed reassuringly good satisfaction across patients and families, although the provision of

education about the disease is highlighted as an area for improvement. This work led to investigation into the possibility of funding the development of a CAH education course, modelled around the success of the national dose adjustment for normal eating (DAFNE) course provided with great success to children and young people with type 1 diabetes (123). However, whilst CAH affects between 1 in 10,000 to 1 in 15,000 individuals, type 1 diabetes has over ten times the prevalence affecting over 1 in 1,000 (213). The development of a structured and standardised training course requires significant expertise and investment, and after discussion with funding bodies and those with expertise in funding calls this proposal was deemed unlikely to be successful. Instead, this work highlights how much disease education is important to children with chronic disease, and that health service providers should consider the adequate provision of time with clinicians that can provide bespoke advice to those with rare chronic disease that is suitable for their age, family circumstances and disease severity (214). Whilst a powerful sample size containing basic demographic data from respondents, we must acknowledge the limitations of this service evaluation that failed to achieve any responses from Wales or Northern Ireland, and only achieved responses as far north as Edinburgh and Glasgow in Scotland. There was also a lack of data about primary language of respondent, and insufficient support for non-English speakers to complete the questionnaire. Nonetheless, this survey provided important contemporaneous context to the analyses in the chapters that followed.

The management of CAH is informed by the measurement of blood biomarkers, most commonly 17-Hydroxyprogesterone (17OHP) and androstenedione (5). The fourth chapter of this thesis outlines an in-depth analysis of 24-hour profiles of these hormones in adult patients with CAH. As a phase 3 randomised controlled trial, with standardisation of measurement protocols and laboratory techniques, this data represents the highest standard of in vivo measurement one can expect. These results highlight that despite such standardisation, there is a significant amount of inter- and intra- patient variability in the trajectory of these hormones (Figure 3.1). The level of 17OHP and androstenedione measured at the same time correlate strongly, but the level of these hormones measured within the same patient on the same medication on different occasions can vary significantly, more so in women than in men (Appendix Figure S4.3). There is a lack of any correlation between the total dose of glucocorticoid replacement patients are taking and their levels of disease control, showing significant heterogeneity of disease and heterogeneity of treatment effect between individuals (Appendix Figure S4.4) (114). We know that in children, this variability also depends on

age (9, 135). This extensive modelling of the biomarkers within this disease is able to inform clinical interpretation, but highlights important limitations within the analysis of registry data, where biomarkers are often measured at different times under different conditions and in different laboratories. This informed the importance of applying longitudinal multilevel modelling to registry data, that allows inference from the trends within patients within hospitals, where group level differences can be estimated within the 'random effects' of such models (215). However, analysis of this adult data did show increased variability between female patients in comparison to male patients (Figure S4.3). One likely cause of this is the fluctuation in steroid levels at different times in the menstrual cycle, data for which was not collected in this study, and was not analysed in relation to data from female adolescents in the I-CAH Registry. Clinicians often take this into account when assessing female patients, thus highlighting the importance of future studies collecting data about the point in the menstrual cycle when steroid profiles are measured.

Research has previously described high blood pressure in patients with CAH, but not assessed the longitudinal changes within patients over time (19, 42-45). Chapter five of this thesis describes the analysis of data extracted from the I-CAH Registry for the purposes of assessing blood pressure within patients with CAH. The findings of a greater proportion of blood pressure measurements recorded at younger ages in children with this condition was further quantified by applying Bayesian multiple change point analysis. The application of LMS modelling across data from the whole cohort was used to describe changes in blood pressure on age, with 10^{th}, 50^{th} and 90^{th} centile equivalents extracted in tabular form to provide a disease specific reference for clinicians (113). This LMS modelling laid the basis for some of the visualisation tools employed within the final prototype clinical decision support tool. Whilst this analysis did show both high inter- and intra- individual variability in blood pressure, multivariable analysis showed a statistically significant association of the dose of fludrocortisone in patients on their blood pressure, when controlling for age, sex, height, weight and dose of hydrocortisone and salt replacement. This was in contrary to previous work with data from the I-CAH Registry that concluded that plasma renin measurements are unrelated to fludrocortisone replacement, despite international guidance recommending the measurement to help inform such mineralocorticoid replacement (5, 166). This previous work relied upon simple multiple regression not accounting for repeated measures within patients, leading to over emphasis of a null finding, and highlighting the importance of appropriate multilevel repeated measures analysis when dealing with such data (26, 166).

Analysis of a further extraction of data from the I-CAH Registry is described in Chapter 6, focusing on the growth of children with the condition. This analysis applies the technique of SuperImposition by Translation and Rotation (SITAR), a sophisticated type of non-linear multilevel modelling developed by Professor Tim Cole CBE, a supportive and engaged member of the PhD advisory panel (50). This analysis highlighted the benefit of applying SITAR to repeated measures data that follows a non-linear trend, allowing both an average description of adiposity rebound and pubertal growth spurts across the sample population, but also an estimate of the between patient variability in such metrics. Despite published advice to the contrary (178), researchers are still relying upon the LMS method to quantify these inflections in the trajectory of human growth, including recent work within CAH itself by a group in the United States of America who have published CAH disease specific growth charts in 2025 (189). The utility of this work is limited by the equations or code for the published trajectories not being made available, nor being incorporated into an online visualisation tool. There is also a lack of information within this publication about disease control or level of glucocorticoid replacement within the patients that contributed data to develop these trajectories (189). These aspects are important in relation to a disease where we know both disease control and level of glucocorticoid replacement will affect growth (34), and thus Chapter 6 of this thesis carefully and thoroughly describes these variables in the sample, to contextualise the growth modelling that has been applied. The application of multivariable non-linear modelling within this chapter confirms that these metrics do have a statistically significant association with different trajectories of growth. The code used to conduct the analysis within this thesis is available at https://github.com/neilxlawrence to allow thorough critique and promote the use of appropriate longitudinal modelling techniques to other data sets.

An important contrast must be noted in the factor employed to convert dosing of prednisolone to hydrocortisone throughout this thesis. The analysis in Chapter 4 reviews data from the phase 3 randomised controlled trial investigating modified-release hydrocortisone, a study that started in 2016 (36). This study employed a conversion factor of 5x hydrocortisone dose in mg to be equivalent to the dose of prednisolone in mg. There is some contention about the optimum conversion rate to employ when switching patients between hydrocortisone dosing and prednisolone dosing, as the drugs have different half-lives, complicating direct comparison. Some guidance recommends a conversion rate of 5x or greater for prednisolone to

hydrocortisone daily dosing in adrenal insufficiency (216). To maintain consistency within the modified-release hydrocortisone analysis, the analysis in chapter 4 employed the conversion factor of 5x. However, a conversion factor of 4x was employed when assessing data from the I-CAH Registry in chapter 5 and chapter 6 as this is consistent with the stronger evidence base cited and recommended within the British National Formulary and Endocrine Society guidelines (159, 217). Whilst the contention between conversion factors is a limitation of this research, most children within the I-CAH Registry data set were prescribed hydrocortisone only (73% of visits, Table 5.1). The longitudinal approach to modelling data also protects against bias, assessing differences in the progression of metrics within patients, and restricting the effect of this limitation to patients that switched between preparations of steroids, rather than those on a consistent preparation at varying doses across their visits. Nonetheless, this highlights the ongoing challenge that clinician's face when converting patients between different preparations of glucocorticoids, and the value that further research can have to clarify this contentious topic.

Chapter 7 described the development of a prototype clinical decision support tool, that is used to advocate for better active data analysis and visualisation within data registry platforms. The coordination of collaborators within studies analysing data from the I-CAH Registry (9, 218), as well as the national service evaluation (115) and previous contributions to the CAH-UK study (15, 116) has consistently highlighted the increased workload and stresses that both clinicians and researchers are experiencing. Professionals are increasingly entering data into registries in time that is not appropriately compensated, as doing it during busy clinics in between seeing patients is of no benefit. With active medication calculations and rapid data visualisation that the prototype in chapter 7 provides, clinicians can be provided with helpful insight into data that has the potential to make their clinical practice easier. The incorporation of such a tool into future iterations of the registry platform for SDM registries has the potential to make data entry a useful endeavour, rather than simply a demand on clinical researchers' time that takes many years to come to fruition as an objective research output. Future work would require investment in the assessment of the ergonomics of such a tool, as well as robust data governance procedures to be followed (52, 59). The prototype shown here should not be used in clinical practice, but can be used to advocate for improvement in data entry platforms to encourage engagement that can lead to larger datasets that can facilitate increasingly sophisticated multivariable analysis in the future.

Categorisation of classic CAH into sub types of 'salt wasting' and 'simple virilising' has traditionally been used to help distinguish between phenotypes of disease, and is discussed in international clinical guidelines (5). However, there is only a weak genotype-phenotype correlation with these categorisations (219). This leads to some researchers defining patients as salt wasting if they are ever treated with either salt replacement or mineralocorticoid replacement (42). Such categorisations can result in circular logic: Patients with salt wasting CAH require salt replacement, and patients are defined as salt wasting CAH if they require salt replacement. This has led many experts in CAH to advocate for an appreciation of a spectrum of salt wasting within the disease, rather than a strict dichotomisation (134). The analysis and modelling presented in this thesis has assessed the associations of dose of glucocorticoid, mineralocorticoid and salt replacement along a continuum, employing appropriate repeated measures techniques and covariable regression modelling to account for a spectrum of salt wasting across a heterogenous population. Nonetheless, the lack of data on underlying genetic mutations within the data presented in this thesis is certainly a limitation of the work. This highlights that genetic mutation data in CAH warrants further investigation in the context of clinical outcomes, and increasing the availability of these data within the I-CAH Registry will facilitate more detailed insights in the future (220).

The analysis within this thesis has provided novel insights into CAH that can inform clinicians managing patients with the condition, and direct future research to help improve management in the future. The national service evaluation provides a large contemporary snapshot of patient and clinician opinions that can be used to advocate for improvements in local practice to ensure patients are appropriately educated about a disease that, in cases of adrenal crisis, requires emergency management (115). The variability in biomarker measurements within patients on the same dose and preparation of replacement at different controlled visits highlights the limitations that clinicians must appreciate when interpreting these biomarkers. The strong correlation with salivary measures adds weight to the appropriate application of saliva monitoring instead of serum monitoring of biomarkers in children, a preference highlighted strongly within the patient and public involvement carried out to inform this research (Chapter 1) (114). Blood pressure measurements provide inherently noisy data due to large variability within patients, but the large cohort analysed in Chapter 4 and appropriate handling of covariates shows that dose of fludrocortisone does affect blood pressure, and the measurement of renin can predict blood pressure in such data sets. This contributes towards appreciating that

these metrics, whilst imperfect, do have value when measured in clinical practice, provided clinicians understand their limitations (113). The application of appropriate non-linear analysis in Chapter 6 has shown a large interindividual variability in the pubertal growth spurt in CAH and the timing of adiposity rebound, and that these measurements vary between patients dependent upon dose and disease control. This highlights that future research into growth trajectories in this disease is warranted, to understand optimum early years management strategies that are imprecisely defined within international guidance (5). Future research should utilise recommended repeated measurements and non-linear modelling with covariate adjustment to maximise insights (26).

8.02 Extent to which thesis meets aims and objectives

In summary, the thesis has addressed all the aims and objectives described in chapter 1 in the following ways:

8.02.01 Objective 1: Systematic Review into models predicting height and weight in fetus' and children

Thesis contribution: A systematic review has been completed and published in Diagnostic and Prognostic Research, showing the high risk of bias in models recently developed and validated to predict height and weight.

8.02.02 Objective 2: Patient and Public Involvement: Service evaluation into the provision of care of children with CAH in the UK

Thesis contribution: A national service evaluation has been completed and published in *Hormone Research in Paediatrics*, showing care of children with CAH in the UK is considered favourable by families and clinicians, but improvements can be made in education about the disease.

8.02.03 Objective 3: Analysis of data from a phase 3 clinical trial into the use of modified-release hydrocortisone in adults with CAH

Thesis contribution: A comprehensive analysis of the largest dataset of steroid profiles from patients with CAH has been completed and published in *eBiomedicine*, quantifying the similarity in profiles of 17-Hydroxyprogesterone and androstenedione.

8.02.04 Objective 4: Analysis of blood pressure data from the I-CAH Registry

Thesis contribution: Longitudinal analysis of from 554 patients with CAH has been completed and published in *European Journal of Endocrinology*, showing only a small effect size of dose of replacement fludrocortisone on blood pressure.

8.02.05 Objective 5: Analysis of growth data from the I-CAH Registry

Thesis contribution: Non-linear spline modelling on data from 573 patients with CAH has been completed and accepted for publication in *European Journal of Endocrinology (112)*. This has quantified an early average and large between patient variability in both adiposity rebound and peak height velocity.

8.02.06 Objective 6: Development of a prototype clinical decision support tool to increase utility of data entry into the I-CAH Registry

Thesis contribution: The prototype clinical decision support tool can be accessed https://endocrinology.shinyapps.io/cah_data_visualiser/ . This clinical decision support tool is an illustrative prototype that should not be used in clinical practice, but can be used to advocate for informative data visualisation to be incorporated into the data entry platform of the I-CAH Registry.

8.03 Practical implications of this research

This thesis has used statistical modelling to describe longitudinal variations and relationships between variables frequently used to inform clinical decisions about children with CAH. Equipping clinicians with more information about the magnitude and variability of the key metrics of this disease can help them contextualise clinical findings amongst those with a similar condition. However, one must be careful in-overinterpreting relationships within observational data, and should use insights to inform the design of robust clinical trials randomised to reduce the risk of selection bias.

The most relevant practical implications for clinicians and researchers derived from each objective are highlighted below:

8.03.01 Objective 1: Systematic Review into models predicting height and weight in fetus' and children

Practical implication: Prediction models require thorough assessment for bias prior to implementation in clinical practice. Clinicians and researchers should assess prediction models using systematic methods of bias

assessment such as PROBAST-AI. Prediction of growth in clinical practice is best carried out using traditional growth charts and simple formulae recommended by national and international bodies such as the Royal College of Paediatrics and Child Health and the Royal College of Obstetrics and Gynaecologists (70, 85).

8.03.02 Objective 2: Patient and Public Involvement: Service evaluation into the provision of care of children with CAH in the UK

Practical implication: Patients and families living with CAH would appreciate more education about their disease. Whilst challenging for a rare disease within a health system that is increasingly overburdened, further investment in research or novel methods of service delivery for education is worthy of consideration.

8.03.03 Objective 3: Analysis of data from a phase 3 clinical trial into the use of modified-release hydrocortisone in adults with CAH

Practical implication: Clinicians can use this analysis to help appreciate the difference in proportionality between 17-Hydroxyprogesterone and androstenedione in CAH to help their interpretation of these biomarkers, guided by thresholds outlined in this publication. This work also provides extra evidence to help advocate for the funding of modified release hydrocortisone to treat patients with CAH.

8.03.04 Objective 4: Analysis of blood pressure data from the I-CAH Registry

Practical implication: This longitudinal assessment helps focus future research into the assessment of blood pressure at early ages in CAH, and the importance of developing robust studies with hard outcome measures to help understand whether these higher readings seen at younger ages are having any detrimental health effects. In the meantime, this research should reassure clinicians that appropriate mineralocorticoid replacement in this condition will not necessarily put patients at risk of hypertension, and that renin does have value in monitoring the response to this treatment.

8.03.05 Objective 5: Analysis of growth data from the I-CAH Registry

Practical implication: This non-linear modelling has confirmed an early adiposity rebound and blunted pubertal growth spurt in CAH, but also quantified large variability around these metrics. This highlights the importance of assessing these metrics in relation to future clinical studies comparing early years treatment strategies, and highlights to clinicians the worth of monitoring these metrics within patients to highlight those where treatment or nutrition may be better optimised.

8.03.06 Objective 6: Development of a prototype clinical decision support tool to increase utility of data entry into the I-CAH Registry

Practical implication: The I-CAH and wider SDM registries have contributed widely to insights into rare conditions through a wide range of scientific publications (194). This clinical decision support tool helps to highlight how the significant time gap between registry data entry and output from scientific publications could be bridged by incorporating an active visualisation clinical decision support tool into the registry platform.

8.04 Thesis Conclusion

Longitudinal multilevel modelling has been applied throughout this thesis to describe and generate robust inferences from repeated measures data of patients with a rare disease. There is an increasing trend for researchers to favour complex machine learning algorithms, but challenges in their explainability and assessment persist, with clinicians becoming increasingly detached from methodology (210). This thesis describes complicated methods, but techniques that are more easily understood by clinicians and patients than their 'black-box' machine learning alternatives. Longitudinal multilevel modelling harvests the maximum power from data collected from patients at different time points, allowing for appropriate description that can help inform clinical practice. The prototype clinical decision support tool shows how data visualisation and automated calculations of clinical relevance could be incorporated into registry data entry platforms, to make the process of contribution more attractive and useful for frontline clinicians. Future work with the I-CAH Registry to integrate this approach into the SDM Registries web interface has the potential to increase engagement with registry research in the future.

8.05 Journal Articles published by the author within the PhD funding period:

Lawrence, N. R., Arshad, M. F., Pofi, R., Ashby, S., Dawson, J., Tomlinson, J. W., ... & Debono, M. (2023). Multivariable model to predict an ACTH stimulation test to diagnose adrenal insufficiency using previous test results. *Journal of the Endocrine Society*, *7*(12), bvad127.

Tonge, J. J., Stevens, O., Dawson, J., Hawley, D., Kerrison, C., Krone, N., ... & Lawrence, N. (2023). Assessing the Response of Biomarkers to Anti-Inflammatory Medications in PIMS-TS by Longitudinal Multilevel Modeling: Real-World Data from a UK Tertiary Center. *Pediatric Allergy, Immunology, and Pulmonology*, *36*(3), 94-103.

Ross, J., Fridman, M., Kelepouris, N., Murray, K., Krone, N., Polak, M., ... & Backeljauw, P. (2023). Factors associated with response to growth hormone in pediatric growth disorders: results of a 5-year registry analysis. *Journal of the Endocrine Society*, *7*(5), bvad026.

Balagamage, C., Lawrence, N. R., Krone, R., Bacila, I. A., & Krone, N. P. (2024). Blood pressure in children with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Hormone Research in Paediatrics*, *97*(4), 315-325.

Lawrence, N. R., Bacila, I., Dawson, J., Mahdi, S., Alvi, S., Cheetham, T. D., ... & Krone, N. P. (2024). Quality of Life in Children and Young People With Congenital Adrenal Hyperplasia—UK Nationwide Multicenter Assessment. *The Journal of Clinical Endocrinology & Metabolism*, *109*(1), e336-e346.

Bacila, I. A., Lawrence, N. R., Badrinath, S. G., Balagamage, C., & Krone, N. P. (2024). Biomarkers in congenital adrenal hyperplasia. *Clinical Endocrinology*, *101*(4), 300-310.

Doyle, L. M., Ahmed, S. F., Davis, J., Elford, S., Elhassan, Y. S., James, L., ... & Krone, N. P. (2024). Service evaluation suggests variation in clinical care provision in adults with congenital adrenal hyperplasia in the UK and Ireland. *Clinical Endocrinology*, *101*(4), 386-396.

Lawrence, N. R., Bacila, I. A., Collins, G. S., Dawson, J., Lang, Z. Q., Ji, X., ... & Krone, N. P. (2025). National Service Evaluation of the Quality of Care for Children and Young People with Congenital Adrenal Hyperplasia in the UK: Survey Responses from Patients and Clinicians. *Hormone Research in Paediatrics*, *98*(3), 296-306.

Lawrence, N. R., Bacila, I., Tonge, J., Dawson, J., Collins, G. S., Lang, Z. Q., ... & Krone, N. (2025). Blood pressure and its associations in 554 children and young people with congenital adrenal hyperplasia. *European journal of endocrinology*, *192*(5), 529-539.

Lawrence, N. R., Dawson, J., Lang, Z. Q., Prete, A., Baranowski, E. S., Schiffer, L., ... & Ross, R. J. (2025). Modelling adrenal steroid profiles to inform monitoring guidance in congenital adrenal hyperplasia. *EBioMedicine*, *116*.

Bacila, I. A., Lawrence, N. R., Alvi, S., Cheetham, T. D., Crowne, E., Das, U., ... & Krone, N. P. (2025). Leptin and adiponectin in children and young persons with congenital adrenal hyperplasia. *European journal of endocrinology*, *193*(3), 329-339.

Lawrence, N. R., Panchigar, K., Clark, S. J., Cole, T. J., Collins, G. S., Dawson, J. F., ... & Wright, N. (2025). Longitudinal modelling of growth in neonates exposed to antenatal steroids to quantify associations with final height: a cohort study. *Archives of Disease in Childhood*. Dec 15;111(1):56-61.

Lawrence NR, Bacila I, Tonge J, Tucker A, Dawson J, Lang ZQ, Krone NP, Dhiman P, Collins GS (2025). Risk of bias in machine learning and statistical models to predict height or weight: a systematic review in fetal and paediatric medicine. Diagn Progn Res. Dec 15;9(1):32.

Lawrence, N. R., Bacila, I., Tonge, J., Balagamage, C., Dawson, J., Collins, G. S., ... & Krone, N. (2026). Adiposity rebound and height velocity in patients with Congenital Adrenal Hyperplasia. European journal of endocrinology, lvag050 *[in press as of 13/03/2026]*.

9.01 List of abbreviations

11KT	11-ketotestosterone
11OHA4	11-Hydroxyandrostenedione
17OHP	17-Hydroxyprogesterone
21-OHD	21-hydroxylase deficiency
ACTH	Adrenocorticotropic hormone
AUC	Area under the curve
AI	Artificial Intelligence
BIC	Bayesian information criterion
BP	Blood pressure
CAH	Congenital Adrenal Hyperplasia
CSV	Comma separated values
CYP	Children and young people
DHEA	Dehydroepiandrosterone
DHT	Dihydrotestosterone
DAFNE	Dose adjustment for normal eating
FDA	Food and Drug Administration
I-CAH Registry	International Congenital Adrenal Hyperplasia Registry
LMS	Lambda-Mu-Sigma
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
MRHC	Modified-release hydrocortisone
MHRA	Medicines and Healthcare Regulatory Agency
NHLBI	National Heart, Lung, and Blood Institute
NIHR	National Institute of Health and Care Research
PDF	Portable document format
PRA	Plasma renin activity
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
PROBAST	Prediction model Risk Of Bias ASsessment Tool
Q1	Quartile 1
Q3	Quartile 3
SDS	Standard deviation score
SDM registries	International Registries for Rare Conditions Affecting Sex Development and Maturation
SITAR	SuperImposition by Translation And Rotation
TRIPOD-SRMA	Transparent reporting of multivariable prediction models for individual prognosis or diagnosis: checklist for systematic reviews and meta-analyses
UK	United Kingdom
USA	United States of America
WHO	World Health Organisation

10.01 References

1.Research Partnerships & Innovation. Guide to the thesis examination process for candidates of research degree programmes. 2025.

2.Claahsen–van der Grinten HL, Speiser PW, Ahmed SF, Arlt W, Auchus RJ, Falhammar H, et al. Congenital adrenal hyperplasia – current insights in pathophysiology, diagnostics and management. Endocrine Reviews. 2022;43(1):91-159.

3.Clayton PE, Miller WL, Oberfield SE, Ritzen EM, Sippell WG, Speiser PW. Consensus statement on 21-hydroxylase deficiency from the Lawson Wilkins Pediatric Endocrine Society and the European Society for Paediatric Endocrinology. The Journal of clinical endocrinology and metabolism. 2002;87(9):4048-53.

4.Speiser PW, Azziz R, Baskin LS, Ghizzoni L, Hensle TW, Merke DP, et al. Congenital adrenal hyperplasia due to steroid 21-hydroxylase deficiency: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab. 2010;95(9):4133-60.

5.Speiser PW, Arlt W, Auchus RJ, Baskin LS, Conway GS, Merke DP, et al. Congenital Adrenal Hyperplasia Due to Steroid 21-Hydroxylase Deficiency: An Endocrine Society Clinical Practice Guideline. The Journal of clinical endocrinology and metabolism. 2018;103(11):4043-88.

6.Webb E, Krone N. Current and novel approaches to children and young people with congenital adrenal hyperplasia and adrenal insufficiency. Best practice & research Clinical endocrinology & metabolism. 2015;29(3).

7.Han TS, Walker BR, Arlt W, Ross RJ. Treatment and health outcomes in adults with congenital adrenal hyperplasia. Nat Rev Endocrinol. 2014;10(2):115-24.

8.Sarafoglou K, Addo O, Turcotte L, Otten N, Wickremasinghe A, Pittock S, et al. Impact of hydrocortisone on adult height in congenital adrenal hyperplasia-the Minnesota cohort. The Journal of pediatrics. 2014;164(5):1141-6.e1.

9.Lawrence N, Bacila I, Dawson J, Bryce J, Ali S, van den Akker E, et al. Analysis of therapy monitoring in the International Congenital Adrenal Hyperplasia Registry. Clinical endocrinology. 2022.

10.Muthusamy K, Elamin M, Smushkin G, Murad M, Lampropulos J, Elamin K, et al. Clinical review: Adult height in patients with congenital adrenal hyperplasia: a systematic review and metaanalysis. The Journal of clinical endocrinology and metabolism. 2010;95(9):4161-72.

11.Bretones P, Riche B, Pichot E, David M, Roy P, Tardy V, et al. Growth curves for congenital adrenal hyperplasia from a national retrospective cohort. Journal of pediatric endocrinology & metabolism : JPEM. 2016;29(12).

12. Bonfig W, Bechtold S, Schmidt H, Knorr D, Schwarz H. Reduced final height outcome in congenital adrenal hyperplasia under prednisone treatment: deceleration of growth velocity during puberty. The Journal of clinical endocrinology and metabolism. 2007;92(5):1635-9.
13. Arlt W, Willis DS, Wild SH, Krone N, Doherty EJ, Hahner S, et al. Health status of adults with congenital adrenal hyperplasia: a cohort study of 203 patients. J Clin Endocrinol Metab. 2010;95(11):5110-21.
14. Falhammar H, Frisén L, Hirschberg A, Norrby C, Almqvist C, Nordenskjöld A, et al. Increased Cardiovascular and Metabolic Morbidity in Patients With 21-Hydroxylase Deficiency: A Swedish Population-Based National Cohort Study. The Journal of clinical endocrinology and metabolism. 2015;100(9).
15. Bacila I, Lawrence NR, Mahdi S, Alvi S, Cheetham TD, Crowne E, et al. Health status of children and young persons with congenital adrenal hyperplasia in the UK (CAH-UK): a cross-sectional multi-centre study. Eur J Endocrinol. 2022;187(4):543-53.
16. Finkielstain GP, Kim MS, Sinaii N, Nishitani M, Van Ryzin C, Hill SC, et al. Clinical characteristics of a cohort of 244 patients with congenital adrenal hyperplasia. J Clin Endocrinol Metab. 2012;97(12):4429-38.
17. Bachelot A, Golmard J, Dulon J, Dahmoune N, Leban M, Bouvattier C, et al. Determining clinical and biological indicators for health outcomes in adult patients with childhood onset of congenital adrenal hyperplasia. European journal of endocrinology. 2015;173(2).
18. Völkl TMK, Simm D, Beier C, Dörr HG. Obesity Among Children and Adolescents With Classic Congenital Adrenal Hyperplasia Due to 21-Hydroxylase Deficiency. Pediatrics. 2006;117(1):e98.
19. Subbarayan A, Dattani M, Peters C, Hindmarsh P. Cardiovascular risk factors in children and adolescents with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. Clinical endocrinology. 2014;80(4).
20. Mushannen T, Cortez P, Stanford FC, Singhal V. Obesity and Hypogonadism-A Narrative Review Highlighting the Need for High-Quality Data in Adolescents. Children (Basel). 2019;6(5).
21. Balagamage C, Lawrence NR, Krone R, Bacila IA, Krone NP. Blood Pressure in Children with Congenital Adrenal Hyperplasia Due To 21-Hydroxylase Deficiency. Hormone Research in Paediatrics. 2023.
22. Whitworth J, Brown M, Kelly J, Williamson P. Mechanisms of cortisol-induced hypertension in humans. Steroids. 1995;60(1).
23. Bacila I, Freeman N, Daniel E, Sandrk M, Bryce J, Ali SR, et al. International practice of corticosteroid replacement therapy in congenital adrenal hyperplasia: data from the I-CAH registry. Eur J Endocrinol. 2021;184(4):553-63.

24.Silva IN, Kater CE, Cunha CF, Viana MB. Randomised controlled trial of growth effect of hydrocortisone in congenital adrenal hyperplasia. Arch Dis Child. 1997;77(3):214-8.
25.Bonfig W, Pozza S, Schmidt H, Pagel P, Knorr D, Schwarz H. Hydrocortisone dosing during puberty in patients with classical congenital adrenal hyperplasia: an evidence-based recommendation. The Journal of clinical endocrinology and metabolism. 2009;94(10).
26.Lopez-Ayala P, Riley RD, Collins GS, Zimmermann T. Dealing with continuous variables and modelling non-linear associations in healthcare data: practical guide. BMJ. 2025;390:e082440.
27.Kristrom B, Jansson C, Rosberg S, AlbertssonWikland K. Growth response to growth hormone (GH) treatment relates to serum insulin-like growth factor I (IGF-I) and IGF-Binding protein-3 in short children with various GH secretion capacities. Journal of Clinical Endocrinology & Metabolism. 1997;82(9):2889-98.
28.Schonau E, Westermann F, Rauch F, Stabrey A, Wassmer G, Keller E, et al. A new and accurate prediction model for growth response to growth hormone treatment in children with growth hormone deficiency. European Journal of Endocrinology. 2001;144(1):13-20.
29.Ranke MB, Lindberg A, Cowell CT, Wikland KA, Reiter EO, Wilton P, et al. Prediction of response to growth hormone treatment in short children born small for gestational age: analysis of data from KIGS (Pharmacia International Growth Database). The Journal of Clinical Endocrinology & Metabolism. 2003;88(1):125-31.
30.Navarro CLA, Damen JA, Takada T, Nijman SW, Dhiman P, Ma J, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. bmj. 2021;375.
31.Bhui K, Mooney R, Stepney M, Mand K, Kirk M, Lawrence E, et al. Improving Research Inclusion: learning from NIHR and Research Council funded studies in England. NIHR Open Research. 2025;5:102.
32.Tonge J, Bacila I, Barker N, Aslam M, Salim E, Walsh J, et al., editors. Patient and public involvement: Techniques used to engage with children and young people about research in congenital adrenal hyperplasia. Endocrine Abstracts; 2023: Bioscientifica.
33.Merke D, Bornstein S. Congenital adrenal hyperplasia. Lancet (London, England). 2005;365(9477):2125-36.
34.Speiser PW, Arlt W, Auchus RJ, Baskin LS, Conway GS, Merke DP, et al. Congenital Adrenal Hyperplasia Due to Steroid 21-Hydroxylase Deficiency: An Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab. 2018;103(11):4043-88.
35.Debono M, Price JN, Ross RJ. Novel strategies for hydrocortisone replacement. Best Pract Res Clin Endocrinol Metab. 2009;23(2):221-32.

36. Merke DP, Mallappa A, Arlt W, Brac de la Perriere A, Linden Hirschberg A, Juul A, et al. Modified-Release Hydrocortisone in Congenital Adrenal Hyperplasia. J Clin Endocrinol Metab. 2021;106(5):e2063-e77.
37. European Medicines Agency. Efmody (hydrocortisone) - An overview of Efmody and why it is authorised in the EU. EMA/186267/2021; EMEA/H/C/005105 ed2021.
38. Bacila I, Freeman N, Daniel E, Sandrk M, Bryce J, Ali S, et al. International practice of corticosteroid replacement therapy in congenital adrenal hyperplasia: data from the I-CAH registry. European journal of endocrinology. 2021;184(4):553-63.
39. Miao H, Zhou Z, Yang S, Zhang Y. The association of triglyceride-glucose index and related parameters with hypertension and cardiovascular risk: a cross-sectional study. Hypertension Research. 2024;47(4):877-86.
40. Mooij CF, Kroese JM, Sweep FC, Hermus AR, Tack CJ. Adult patients with congenital adrenal hyperplasia have elevated blood pressure but otherwise a normal cardiovascular risk profile. PLoS One. 2011;6(9):e24204.
41. Falhammar H, Filipsson Nyström H, Wedell A, Thorén M. Cardiovascular risk, metabolic profile, and body composition in adult males with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. Eur J Endocrinol. 2011;164(2):285-93.
42. Neumann U, Van Der Linde A, Krone RE, Krone NP, Güven A, Güran T, et al. Treatment of congenital adrenal hyperplasia in children aged 0–3 years: a retrospective multicenter analysis of salt supplementation, glucocorticoid and mineralocorticoid medication, growth and blood pressure. European journal of endocrinology. 2022;186(5):587-96.
43. de Vries L, Lebenthal Y, Phillip M, Shalitin S, Tenenbaum A, Bello R. Obesity and Cardiometabolic Risk Factors in Children and Young Adults With Non-classical 21-Hydroxylase Deficiency. Front Endocrinol (Lausanne). 2019;10:698.
44. Maccabee-Ryaboy N, Thomas W, Kyllo J, Lteif A, Petryk A, Gonzalez-Bolanos MT, et al. Hypertension in children with congenital adrenal hyperplasia. Clinical Endocrinology. 2016;85(4):528-34.
45. Bonfig W, Roehl F, Riedl S, Dörr H, Bettendorf M, Brämswig J, et al. Blood Pressure in a Large Cohort of Children and Adolescents With Classic Adrenal Hyperplasia (CAH) Due to 21-Hydroxylase Deficiency. American journal of hypertension. 2016;29(2).
46. Cissé A, Lioret S, de Lauzon-Guillain B, Forhan A, Ong K, Charles M, et al. Association between perinatal factors, genetic susceptibility to obesity and age at adiposity rebound in children of the EDEN mother-child cohort. International journal of obesity (2005). 2021;45(8):1802-10.

47.Bhullar G, Tanawattanacharoen V, Yeh M, Kim W, Vidmar A, Geffner M, et al. Early Adiposity Rebound Predicts Obesity and Adiposity in Youth with Congenital Adrenal Hyperplasia. Hormone research in paediatrics. 2020;93(11-12):609–15.
48.Cornean RE, Hindmarsh PC, Brook CGD. Obesity in 21-hydroxylase deficient patients. Arch Dis Child. 1998;78:261-3.
49.Takishima S, Nakajima K, Nomura R, Tsuji-Hosokawa A, Matsuda N, Matsubara Y, et al. Lower body weight and BMI at birth were associated with early adiposity rebound in 21-hydroxylase deficiency patients. Endocrine journal. 2016;63(11):983-90.
50.Cole T, Donaldson M, Ben-Shlomo Y. SITAR--a useful instrument for growth curve analysis. International journal of epidemiology. 2010;39(6).
51.Lawrence NR, Panchigar K, Clark SJ, Cole TJ, Collins GS, Dawson JF, et al. Longitudinal modelling of growth in neonates exposed to antenatal steroids to quantify associations with final height: a cohort study. Archives of Disease in Childhood. 2025.
52.Ward ME, Sujan M, Pool R, Preston K, Huang H, Carrington A, et al. Integrating Human-Centred AI in Clinical Practice. Chartered Institute of Ergonomics and Human Factors. 2024.
53.*Cobbaert K*. Software as a Medical Device: Regulatory Affairs Professional Society,; 2021.
54.Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. Ann Intern Med. 2019;170(1):W1-W33.
55.Merke DP, Mallappa A, Arlt W, Brac de la Perriere A, Lindén Hirschberg A, Juul A, et al. Modified-release hydrocortisone in congenital adrenal hyperplasia. The Journal of Clinical Endocrinology & Metabolism. 2021;106(5):e2063-e77.
56.R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing , Vienna, Austria.; 2024.
57.Chang W, Cheng J, Allaire JJ, Sievert C, Schloerke B, Xie Y, et al. Shiny: web application framework for R. R package version 11119001. 2012.
58.Wickham H, Navarro D, Pedersen TL. ggplot2: Elegant Graphics for Data Analysis (3e)2025.
59.European Union. General Data Protection Regulation (GDPR). 2018.
60.Lawrence NR, Bacila I, Tonge J, Tucker A, Dawson J, Lang ZQ, et al. Risk of bias in machine learning and statistical models to predict height or weight: a systematic review in fetal and paediatric medicine. Diagn Progn Res. 2025;9(1):32.
61.Koletzko B, Chourdakis M, Grote V, Hellmuth C, Prell C, Rzehak P, et al. Regulation of early human growth: impact on long-term health. Ann Nutr Metab. 2014;65(2-3):101-9.

62.National Institute for Health and Care Excellence. NG201: Antenatal Care 2021 [Available from: www.nice.org.uk/guidance/ng201.
63.World Health Organisation & United Nations Children's Fund. Comprehensive Implementation Plan on Maternal, Infant and Young Child Nutrition. World Health Organisation; 2014. p. 1-22.
64.Lei X, Zhao D, Huang L, Luo Z, Zhang J, Yu X, et al. Childhood Health Outcomes in Term, Large-for-Gestational-Age Babies With Different Postnatal Growth Patterns. American Journal of Epidemiology. 2018;187(3):507-14.
65.Wright CM, Booth IW, Buckler JM, Cameron N, Cole TJ, Healy MJ, et al. Growth reference charts for use in the United Kingdom. Arch Dis Child. 2002;86(1):11-4.
66.Royal College of Paediatrics and Child Health. Early Years Neglect: Systematic Review. 2022.
67.Prado EL, Larson LM, Cox K, Bettencourt K, Kubes JN, Shankar AH. Do effects of early life interventions on linear growth correspond to effects on neurobehavioural development? A systematic review and meta-analysis. Lancet Glob Health. 2019;7(10):e1398-e413.
68.Campbell S. A short history of sonography in obstetrics and gynaecology. Facts Views Vis Obgyn. 2013;5(3):213-29.
69.Hadlock FP, Harrist RB, Sharman RS, Deter RL, Park SK. Estimation of fetal weight with the use of head, body, and femur measurements--a prospective study. Am J Obstet Gynecol. 1985;151(3):333-7.
70.Morris RK, Johnstone E, Lees C, Morton V, Smith G, Gynaecologists RCoOa. Investigation and Care of a Small-for-Gestational-Age Fetus and a Growth Restricted Fetus (Green-top Guideline No. 31). BJOG. 2024;131(9):e31-e80.
71.American Academy of Obstetrics and Gynaecology. Macrosomia: ACOG Practice Bulletin, Number 216. Obstet Gynecol. 2020;135(1):e18-e35.
72.Kingdom J, Ashwal E, Lausman A, Liauw J, Soliman N, Figueiro-Filho E, et al. Guideline No. 442: Fetal Growth Restriction: Screening, Diagnosis, and Management in Singleton Pregnancies. J Obstet Gynaecol Can. 2023;45(10):102154.
73.Bayley N. Tables for predicting adult height from skeletal age and present height. J Pediatr. 1946;28:49-64.
74.Tanner JM, Goldstein H, Whitehouse RH. Standards for children's height at ages 2-9 years allowing for heights of parents. Arch Dis Child. 1970;45(244):755-62.
75.Tanner JM, Whitehouse RH, Marshall WA, Carter BS. Prediction of adult height from height, bone age, and occurrence of menarche, at ages 4 to 16 with allowance for midparent height. Arch Dis Child. 1975;50(1):14-26.

76.Cole TJ, Wright CM. A chart to predict adult height from a child's current height. Annals of human biology. 2011;38(6):662-8.
77.Thodberg HH, Jenni OG, Caflisch J, Ranke MB, Martin DD. Prediction of adult height based on automated determination of bone age. J Clin Endocrinol Metab. 2009;94(12):4868-74.
78.Yang C, Kors JA, Ioannou S, John LH, Markus AF, Rekkas A, et al. Trends in the conduct and reporting of clinical prediction model development and validation: a systematic review. Journal of the American Medical Informatics Association. 2022;29(5):983-9.
79.Weissler EH, Naumann T, Andersson T, Ranganath R, Elemento O, Luo Y, et al. The role of machine learning in clinical research: transforming the future of evidence generation. Trials. 2021;22(1):1-15.
80.Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol. 2019;110:12-22.
81.Haug CJ, Drazen JM. Artificial Intelligence and Machine Learning in Clinical Medicine, 2023. N Engl J Med. 2023;388(13):1201-8.
82.Zou J, Schiebinger L. Ensuring that biomedical AI benefits diverse populations. EBioMedicine. 2021;67.
83.Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med. 2019;170(1):51-8.
84.Robson S, Martin W, Morris R. The investigation and management of the small-for-gestational-age fetus. Royal College of Obstetricians and Gynaecologists. 2013.
85.Royal College of Paediatrics and Child Health. Growth Charts [Available from: https://www.rcpch.ac.uk/ resources/growth-charts.
86.Shur N, Tigranyan A, Daymont C, Regier DS, Raturi S, Roshan Lal T, et al. The past, present, and future of child growth monitoring: A review and primer for clinical genetics. American Journal of Medical Genetics Part A. 2023;191(4):948-61.
87.Snell KIE, Levis B, Damen JAA, Dhiman P, Debray TPA, Hooft L, et al. Transparent reporting of multivariable prediction models for individual prognosis or diagnosis: checklist for systematic reviews and meta-analyses (TRIPOD-SRMA). BMJ. 2023;381:e073538.
88.Lawrence NL, Bacila I, Tonge J, Krone N, Tucker A, Dawson J, et al. Reporting of machine learning and statistical models to predict height or weight in fetal and paediatric medicine. PROSPERO2023.

89.Archer L, Snell KIE, Ensor J, Hudda MT, Collins GS, Riley RD. Minimum sample size for external validation of a clinical prediction model with a continuous outcome. Statistics in Medicine. 2021;40(1):133-46.
90.Christodoulou E, van Smeden M, Edlinger M, Timmerman D, Wanitschek M, Steyerberg EW, et al. Adaptive sample size determination for the development of clinical prediction models. Diagn Progn Res. 2021;5(1):6.
91.Riley R, Snell K, Ensor J, Burke D, Harrell F, Moons K, et al. Minimum sample size for developing a multivariable prediction model: Part I - Continuous outcomes. Statistics in medicine. 2019;38(7):1262-75.
92.Martin G, Riley R, Collins G, Sperrin M. Developing clinical prediction models when adhering to minimum sample size recommendations: The importance of quantifying bootstrap variability in tuning parameters and predictive performance. Statistical methods in medical research. 2021;30(12).
93.Riley R, Ensor J, Snell K, Harrell F, Martin G, Reitsma J, et al. Calculating the sample size required for developing a clinical prediction model. BMJ (Clinical research ed). 2020;368.
94.Ma J, Dhiman P, Qi C, Bullock G, van Smeden M, Riley RD, et al. Poor handling of continuous predictors in clinical prediction models using logistic regression: a systematic review. Journal of Clinical Epidemiology. 2023.
95.Riley RD, Cole TJ, Deeks J, Kirkham JJ, Morris J, Perera R, et al. On the 12th Day of Christmas, a Statistician Sent to Me. bmj. 2022;379.
96.Collins GS, Ogundimu EO, Cook JA, Manach YL, Altman DG. Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model. Statistics in medicine. 2016;35(23):4124-35.
97.Dhiman P, Ma J, Qi C, Bullock G, Sergeant JC, Riley RD, et al. Sample size requirements are not being considered in studies developing prediction models for binary outcomes: a systematic review. BMC Medical Research Methodology. 2023;23(1):1-11.
98.Pate A, Riley RD, Collins GS, Van Smeden M, Van Calster B, Ensor J, et al. Minimum sample size for developing a multivariable prediction model using multinomial logistic regression. Statistical Methods in Medical Research. 2023;32(3):555-71.
99.Collins GS, Dhiman P. Prediction models should contain predictors known at the moment of intended use. Aging Clinical and Experimental Research. 2023:1-2.
100.Austin PC, White IR, Lee DS, van Buuren S. Missing data in clinical research: a tutorial on multiple imputation. Canadian Journal of Cardiology. 2021;37(9):1322-31.

101. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Annals of internal medicine. 2015;162(1):W1-W73.
102. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. bmj. 2020;369.
103. Dhiman P, Ma J, Andaur Navarro CL, Speich B, Bullock G, Damen JAA, et al. Risk of bias of prognostic models developed using machine learning: a systematic review in oncology. Diagn Progn Res. 2022;6(1):13.
104. Haines K, McPeake J, Hibbert E, Mikkelsen M, Collins G, Hough C, et al. Prediction Models of Post-Intensive Care Syndrome: A Systematic Review and Risk of Bias Assessment. American Thoracic Society International Conference Abstracts. 2020(A41):A1606.
105. Dhiman P, Ma J, Gibbs VN, Rampotas A, Kamal H, Arshad SS, et al. Systematic review highlights high risk of bias of clinical prediction models for blood transfusion in patients undergoing elective surgery. Journal of Clinical Epidemiology. 2023.
106. Milner J, Arezina J. The accuracy of ultrasound estimation of fetal weight in comparison to birth weight: A systematic review. Ultrasound. 2018;26(1):32-41.
107. Vaughn-Valencia MM, Zhao YD, Edwards RK, Clifton S, Nadeau HC. Fetal weight extrapolation following a third-trimester ultrasound examination using the gestation-adjusted projection method: a systematic review and meta-analysis. American Journal of Perinatology. 2025.
108. Bone Xpert. Adult Height Predictor v3.1 2025 [Available from: https://bonexpert.com/ahp/.
109. Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ. 2025;388:e082505.
110. Rajula HSR, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. Medicina. 2020;56(9):455.
111. Pugliese R, Regondi S, Marini R. Machine learning-based approach: global trends, research directions, and regulatory standpoints. 2021.
112. Lawrence NR, Bacila I, Tonge J, Balagamage C, Dawson J, Collins GS, et al. Adiposity rebound and height velocity in patients with Congenital Adrenal Hyperplasia. European journal of endocrinology. 2026:lvag050.
113. Lawrence NR, Bacila I, Tonge J, Dawson J, Collins GS, Lang ZQ, et al. Blood pressure and its associations in 554 children and young people with congenital adrenal hyperplasia. Eur J Endocrinol. 2025;192(5):529-39.

114. Lawrence NR, Dawson J, Lang Z-Q, Prete A, Baranowski ES, Schiffer L, et al. Modelling adrenal steroid profiles to inform monitoring guidance in congenital adrenal hyperplasia. EBioMedicine. 2025;116.
115. Lawrence NR, Bacila IA, Collins GS, Dawson J, Lang Z-Q, Ji X, et al. National Service Evaluation of the Quality of Care for Children and Young People with Congenital Adrenal Hyperplasia in the UK: Survey Responses from Patients and Clinicians. Hormone Research in Paediatrics. 2025;98(3):296-306.
116. Lawrence NR, Bacila I, Dawson J, Mahdi S, Alvi S, Cheetham TD, et al. Quality of Life in Children and Young People With Congenital Adrenal Hyperplasia—UK Nationwide Multicenter Assessment. The Journal of Clinical Endocrinology & Metabolism. 2023:dgad405.
117. The United Kingdom Government. Results of the national conversation on rare diseases survey. https://www.gov.uk/government/publications/uk-rare-diseases-framework/the-uk-rare-diseases-framework#annex-a; Accessed 18/08/2023,2019.
118. McManus E, Meacock R, Parkinson B, Sutton M. Population level impact of the NHS Diabetes Prevention Programme on incidence of type 2 diabetes in England: An observational study. The Lancet Regional Health–Europe. 2022;19.
119. NHS Acute (Hospital) Trust Catchment Populations [Internet]. 2023. Available from: https://app.powerbi.com/view?r=eyJrIjoiODZmNGQ0YzItZDAwZi00MzFiLWE4NzAtMzVmNTUwMThmMTVlIiwidCI6ImVlNGUxNDk5LTRhMzUtNGIyZS1hZDQ3LTVmM2NmOWRlODY2NiIsImMiOjh9.
120. Lopez-Vargas P, Tong A, Crowe S, Alexander SI, Caldwell PHY, Campbell DE, et al. Research priorities for childhood chronic conditions: a workshop report. Archives of disease in childhood. 2019;104(3):237-45.
121. Dean AJ, Walters J, Hall A. A systematic review of interventions to enhance medication adherence in children and adolescents with chronic illness. Archives of disease in childhood. 2010;95(9):717-23.
122. Athanasiadis L, editor Psychological Evaluation of Patients with Congenital Adrenal Hyperplasia (CAH)2021.
123. Group DS. Training in flexible, intensive insulin management to enable dietary freedom in people with type 1 diabetes: dose adjustment for normal eating (DAFNE) randomised controlled trial. Bmj. 2002;325(7367):746.
124. Chawla R, Rutter M, Green J, Weidler EM, editors. Care of the adolescent patient with congenital adrenal hyperplasia: Special considerations, shared decision making, and transition. Seminars in pediatric surgery; 2019: Elsevier.

125. Gleeson H, Davis J, Jones J, O'Shea E, Clayton PE. The challenge of delivering endocrine care and successful transition to adult services in adolescents with congenital adrenal hyperplasia: experience in a single centre over 18 years. Clinical endocrinology. 2013;78(1):23-8.
126. Workforce Information Team RCoPaCH. Impact of COVID-19 on child health services between November 2020 and February 2021. 2021.
127. Paton JY, Ranmal R, Dudley J, Committee RCS. Clinical audit: still an important tool for improving healthcare. Archives of Disease in Childhood-Education and Practice. 2015;100(2):83-8.
128. Pang SY, Wallace MA, Hofman L, Thuline HC, Dorche C, Lyon IC, et al. Worldwide experience in newborn screening for classical congenital adrenal hyperplasia due to 21-hydroxylase deficiency. Pediatrics. 1988;81(6):866-74.
129. White PC, Speiser PW. Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. Endocr Rev. 2000;21(3):245-91.
130. Miller WL, Auchus RJ. The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. Endocrine Reviews. 2011;32(1):81-151.
131. Storbeck KH, Schiffer L, Baranowski ES, Chortis V, Prete A, Barnard L, et al. Steroid Metabolome Analysis in Disorders of Adrenal Steroid Biosynthesis and Metabolism. Endocr Rev. 2019;40(6):1605-25.
132. Reisch N, Taylor AE, Nogueira EF, Asby DJ, Dhir V, Berry A, et al. Alternative pathway androgen biosynthesis and human fetal female virilization. Proc Natl Acad Sci U S A. 2019;116(44):22294-9.
133. Debono M, Mallappa A, Gounden V, Nella AA, Harrison RF, Crutchfield CA, et al. Hormonal circadian rhythms in patients with congenital adrenal hyperplasia: identifying optimal monitoring times and novel disease biomarkers. Eur J Endocrinol. 2015;173(6):727-37.
134. Merke DP, Auchus RJ. Congenital Adrenal Hyperplasia Due to 21-Hydroxylase Deficiency. N Engl J Med. 2020;383(13):1248-61.
135. Bacila IA, Lawrence NR, Badrinath SG, Balagamage C, Krone NP. Biomarkers in congenital adrenal hyperplasia. Clin Endocrinol (Oxf). 2023.
136. Whitaker MJ, Debono M, Huatan H, Merke DP, Arlt W, Ross RJ. An oral multiparticulate, modified-release, hydrocortisone replacement therapy that provides physiological cortisol exposure. Clin Endocrinol (Oxf). 2014;80(4):554-61.
137. Schiffer L, Shaheen F, Gilligan LC, Storbeck KH, Hawley JM, Keevil BG, et al. Multi-steroid profiling by UHPLC-MS/MS with post-column infusion of ammonium fluoride. J Chromatogr B Analyt Technol Biomed Life Sci. 2022;1209:123413.

138. Schiffer L, Adaway JE, Arlt W, Keevil BG. A liquid chromatography-tandem mass spectrometry assay for the profiling of classical and 11-oxygenated androgens in saliva. Ann Clin Biochem. 2019;56(5):564-73.
139. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Statistical science. 1992;7(4):457-72.
140. Bacila I, Adaway J, Hawley J, Mahdi S, Krone R, Patel L, et al. Measurement of Salivary Adrenal-Specific Androgens as Biomarkers of Therapy Control in 21-Hydroxylase Deficiency. J Clin Endocrinol Metab. 2019;104(12):6417-29.
141. Jones CM, Mallappa A, Reisch N, Nikolaou N, Krone N, Hughes BA, et al. Modified-Release and Conventional Glucocorticoids and Diurnal Androgen Excretion in Congenital Adrenal Hyperplasia. J Clin Endocrinol Metab. 2017;102(6):1797-806.
142. Arlt W, Willis DS, Wild SH, Krone N, Doherty EJ, Hahner S, et al. Health Status of Adults with Congenital Adrenal Hyperplasia: A Cohort Study of 203 Patients. Journal of Clinical Endocrinology & Metabolism. 2010;95(11):5110-21.
143. Finkielstain GP, Kim MS, Sinaii N, Nishitani M, Van Ryzin C, Hill SC, et al. Clinical characteristics of a cohort of 244 patients with congenital adrenal hyperplasia. The Journal of clinical endocrinology and metabolism. 2012;97(12):4429-38.
144. Bonfig W, Roehl F, Riedl S, Brämswig J, Richter-Unruh A, Fricke-Otto S, et al. Sodium Chloride Supplementation Is Not Routinely Performed in the Majority of German and Austrian Infants with Classic Salt-Wasting Congenital Adrenal Hyperplasia and Has No Effect on Linear Growth and Hydrocortisone or Fludrocortisone Dose. Hormone research in paediatrics. 2018;89(1).
145. Krysiak R, Claahsen-van der Grinten HL, Reisch N, Touraine P, Falhammar H. Cardiometabolic Aspects of Congenital Adrenal Hyperplasia. Endocrine Reviews. 2024:bnae026.
146. Mooij C, Kapusta L, Otten B, Claahsen-van der Grinten H. Blood pressure in the first year of life in children with congenital adrenal hyperplasia due to 21-hydroxylase deficiency: a pilot study. Hormone research in paediatrics. 2010;74(5).
147. Ubertini G, Bizzarri C, Grossi A, Gimigliano F, Ravà L, Fintini D, et al. Blood Pressure and Left Ventricular Characteristics in Young Patients with Classical Congenital Adrenal Hyperplasia due to 21-Hydroxylase Deficiency. Int J Pediatr Endocrinol. 2009;2009:383610.
148. Liivak K, Tillmann V. 24-hour blood pressure profiles in children with congenital adrenal hyperplasia on two different hydrocortisone treatment regimens. J Pediatr Endocrinol Metab. 2009;22(6):511-7.

149. Akyürek N, Atabek ME, Eklioğlu BS, Alp H. Ambulatory blood pressure and subclinical cardiovascular disease in patients with congenital adrenal hyperplasia: a preliminary report. J Clin Res Pediatr Endocrinol. 2015;7(1):13-8.
150. Roche EF, Charmandari E, Dattani MT, Hindmarsh PC. Blood pressure in children and adolescents with congenital adrenal hyperplasia (21-hydroxylase deficiency): a preliminary report. Clin Endocrinol (Oxf). 2003;58(5):589-96.
151. Bonfig W, Schwarz HP. Blood pressure, fludrocortisone dose and plasma renin activity in children with classic congenital adrenal hyperplasia due to 21-hydroxylase deficiency followed from birth to 4 years of age. Clinical Endocrinology. 2014;81(6):871-5.
152. Growth of Patients with 21-Hydroxylase Deficiency: An Analysis of the Factors Influencing Adult Height | Pediatric Research 2021 [Available from: https://www.nature.com/articles/pr19975.
153. Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, et al. Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update From the GBD 2019 Study. J Am Coll Cardiol. 2020;76(25):2982-3021.
154. Tseretopoulou X, Bryce J, Chen M, McMillan M, Lucas-Herald AK, Ali SR, et al. The I-CAH Registry: A platform for international collaboration for improving knowledge and clinical care in congenital adrenal hyperplasia. Clin Endocrinol (Oxf). 2024;101(4):397-404.
155. National Heart Lung and Blood Institute. Expert panel on integrated guidelines for cardiovascular health and risk reduction in children and adolescents: summary report. Pediatrics. 2011;128 Suppl 5(Suppl 5):S213-56.
156. Carpenter JR, Bartlett JW, Morris TP, Wood AM, Quartagno M, Kenward MG. Multiple imputation and its application: John Wiley & Sons; 2023.
157. de Onis M, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. Development of a WHO growth reference for school-aged children and adolescents. Bull World Health Organ. 2007;85(9):660-7.
158. Gemelli M, Manganaro R, Mamì C, De Luca F. Longitudinal study of blood pressure during the 1st year of life. Eur J Pediatr. 1990;149(5):318-20.
159. British National Formulary. Equivalent anti-inflammatory doses of corticosteroids. 2023.
160. Trenkel S, Seifarth C, Schobel H, Hahn E, Hensen J. Ratio of serum aldosterone to plasma renin concentration in essential hypertension and primary aldosteronism. Experimental and clinical endocrinology & diabetes. 2002;110(02):80-5.
161. Tennant PW, Murray EJ, Arnold KF, Berrie L, Fox MP, Gadd SC, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. International journal of epidemiology. 2021;50(2):620-32.

162. Expert Panel on Integrated Guidelines for Cardiovascular Health and Risk Reduction in Children and Adolescents. Expert panel on integrated guidelines for cardiovascular health and risk reduction in children and adolescents: summary report. Pediatrics. 2011;128(Supplement_5):S213-S56.
163. Han TS, Conway GS, Willis DS, Krone N, Rees DA, Stimson RH, et al. Relationship between final height and health outcomes in adults with congenital adrenal hyperplasia: United Kingdom congenital adrenal hyperplasia adult study executive (CaHASE). J Clin Endocrinol Metab. 2014;99(8):E1547-55.
164. Mooij CF, van Herwaarden AE, Sweep F, Roeleveld N, de Korte CL, Kapusta L, et al. Cardiovascular and metabolic risk in pediatric patients with congenital adrenal hyperplasia due to 21 hydroxylase deficiency. J Pediatr Endocrinol Metab. 2017;30(9):957-66.
165. Gidlöf S, Hogling DE, Lönnberg H, Ritzén M, Lajic S, Nordenström A. Growth and Treatment in Congenital Adrenal Hyperplasia: An Observational Study from Diagnosis to Final Height. Hormone Research in Paediatrics. 2024;97(5):445-55.
166. Pofi R, Prete A, Thornton-Jones V, Bryce J, Ali SR, Faisal Ahmed S, et al. Plasma Renin Measurements are Unrelated to Mineralocorticoid Replacement Dose in Patients With Primary Adrenal Insufficiency. The Journal of Clinical Endocrinology & Metabolism. 2019;105(1):314-26.
167. Al-Zoubaidi D, Johannsen T, Jørgensen NR, Main KM. The effect of 30 min of supine rest on plasma renin concentrations in paediatric patients. Acta Paediatrica. 2024;113(11):2473-8.
168. Fukushige J, Shimomura K, Ueda K. Influence of upright activity on plasma renin activity and aldosterone concentration in children. European journal of pediatrics. 1994;153:284-6.
169. Quinkler M, Meyer B, Bumke-Vogt C, Grossmann C, Gruber U, Oelkers W, et al. Agonistic and antagonistic properties of progesterone metabolites at the human mineralocorticoid receptor. European journal of endocrinology. 2002;146(6):789-99.
170. Flynn JT, Kaelber DC, Baker-Smith CM, Blowey D, Carroll AE, Daniels SR, et al. Clinical Practice Guideline for Screening and Management of High Blood Pressure in Children and Adolescents. Pediatrics. 2017;140(3).
171. Wasniewska M, Balsamo A, Valenzise M, Manganaro A, Faggioli G, Bombaci S, et al. Increased large artery intima media thickness in adolescents with either classical or non-classical congenital adrenal hyperplasia. J Endocrinol Invest. 2013;36(1):12-5.
172. Pomi AL, Pepe G, Aversa T, Corica D, Valenzise M, Messina MF, et al. Early adiposity rebound: predictors and outcomes. Ital J Pediatr. 2024;50(1):98.
173. González L, Corvalán C, Pereira A, Kain J, Garmendia ML, Uauy R. Early adiposity rebound is associated with metabolic risk in 7-year-old children. Int J Obes (Lond). 2014;38(10):1299-304.

174. Sarafoglou K, Forlenza G, Yaw Addo O, Kyllo J, Lteif A, Hindmarsh P, et al. Obesity in children with congenital adrenal hyperplasia in the Minnesota cohort: importance of adjusting body mass index for height-age. Clinical endocrinology. 2017;86(5).
175. Panic Zaric S, Milenkovic T, Todorovic S, Mitrovic K, Cvetkovic D, Cehic M, et al. Metabolic Syndrome Spectrum in Children with Classic Congenital Adrenal Hyperplasia—A Comprehensive Review. Metabolites. 2025;15(2):89.
176. Maheshwari A, Khadilkar V, Gangodkar P, Khadilkar A. Long-term Growth in Congenital Adrenal Hyperplasia. Indian J Pediatr. 2019;86(2):154-8.
177. Cole TJ. Tanner's tempo of growth in adolescence: recent SITAR insights with the Harpenden Growth Study and ALSPAC. Ann Hum Biol. 2020;47(2):181-98.
178. Cole TJ. The LMS method should not be used to construct velocity centiles in puberty. Acta Paediatr. 2025;114(2):448-50.
179. Al-Rayess H, Addo OY, Palzer E, Jaber M, Fleissner K, Hodges J, et al. Bone Age Maturation and Growth Outcomes in Young Children with CAH Treated with Hydrocortisone Suspension. J Endocr Soc. 2022;6(2):bvab193.
180. Lawrence N, Bacila I, Tonge J, Balagamage C, Dawson J, Collins G, et al. Adiposity rebound and height velocity in patients with Congenital Adrenal Hyperplasia - Supplementary Material. 2025.
181. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing , Vienna, Austria.; 2023.
182. Fonseca MJ, Moreira C, Santos AC. Adiposity rebound and cardiometabolic health in childhood: results from the Generation XXI birth cohort. Int J Epidemiol. 2021;50(4):1260-71.
183. Mushtaq T, Wing B. Emergency and peri-operative management of adrenal insufficiency in children and young people: BSPED consensus guidance. 2023.
184. Hargitai G, Sólyom J, Battelino T, Lebl J, Pribilincová Z, Hauspie R, et al. Growth patterns and final height in congenital adrenal hyperplasia due to classical 21-hydroxylase deficiency. Results of a multicenter study. Horm Res. 2001;55(4):161-71.
185. Holmgren A, Niklasson A, Gelander L, Aronson AS, Nierop AF, Albertsson-Wikland K. Insight into human pubertal growth by applying the QEPS growth model. BMC pediatrics. 2017;17(1):107.
186. Zhou J, Zhang F, Qin X, Li P, Teng Y, Zhang S, et al. Age at adiposity rebound and the relevance for obesity: a systematic review and meta-analysis. Int J Obes (Lond). 2022;46(8):1413-24.
187. Hughes AR, Sherriff A, Ness AR, Reilly JJ. Timing of adiposity rebound and adiposity in adolescence. Pediatrics. 2014;134(5):e1354-61.

188. Matsubara Y, Ono M, Miyai K, Takizawa F, Takasawa K, Onishi T, et al. Longitudinal analysis of growth and body composition of Japanese 21-OHD patients in childhood. Endocr J. 2013;60(2):149-54.
189. Sarafoglou K, Mercado Munoz Y, Sukin C, Lteif A, Kyllo J, Miller BS, et al. The Importance of Disease Specific Growth Charts for Children with Congenital Adrenal Hyperplasia. The Journal of Clinical Endocrinology & Metabolism. 2025:dgaf554.
190. Girgis R, Winter JSD. The effects of glucocorticoid replacement therapy on growth, bone mineral density, and bone turnover markers in children with congenital adrenal hyperplasia. Journal of Clinical Endocrinology and Metabolism. 1997;82(12):3926-9.
191. Delvecchio M, Soldano L, Lonero A, Ventura A, Giordano P, Cavallo L, et al. Evaluation of impact of steroid replacement treatment on bone health in children with 21-hydroxylase deficiency. Endocrine. 2015;48(3):995-1000.
192. Karishma MI, Rai VR, Afridi SB, Riaz M, Laghari T. Bone Mineral Density in Children with Congenital Adrenal Hyperplasia. Bone. 2024;36(92.3):0.196.
193. Ahmed F. SDM registries website (sdmregistries.org) 2025 [Available from: sdmregistries.org.
194. Bryce J. SDM Registries Activity Report. June 2025.
195. Mosteller R. Simplified calculation of body surface area. New Engl J Med. 1987;317:1098.
196. Medicines and Healthcare products Regulatory Agency. Crafting an intended purpose in the context of software as a medical device (SaMD). United Kingdom Government; 2023.
197. United States Food and Drug Administration. Applying Human Factors and Usability Engineering to Medical Devices: guidance for industry and Food and Drug Administration staff. United States Food and Drug Administration; 2016.
198. International Electrotechnical Commission. IEC 62366-1; Application of usability engineering to medical devices. BSI Standards Limited 2020: The British Standards Institution; 2015.
199. Gerke S, Babic B, Evgeniou T, Cohen IG. The need for a system view to regulate artificial intelligence/machine learning-based software as medical device. NPJ digital medicine. 2020;3(1):53.
200. Medicines and Healthcare products Regulatory Agency. Medical device stand-alone software including apps (including IVDMDs). United Kingdom Government,2023.
201. Pearson I, Rothwell B, Olaye A, Knight C. Economic modeling considerations for rare diseases. Value in Health. 2018;21(5):515-24.
202. The British Medical Journal. BMJ Guidance for Authors. 2018.
203. Ramos E, Concepcion BP, editors. Visual abstracts: redesigning the landscape of research dissemination. Seminars in nephrology; 2020: Elsevier.

204. Kim Y, Lee J, Yoo J-J, Jung E-A, Kim SG, Kim YS. Seeing is believing: the effect of graphical abstracts on citations and social media exposure in gastroenterology & hepatology journals. Journal of Korean Medical Science. 2022;37(45).
205. Aggarwal V. Visual abstracts do not increase some impact scores more than conventional abstracts of clinical research: A retrospective cohort study. Health Information & Libraries Journal. 2021;38(4):259-67.
206. Research Excellence Framework. Contributions to Knowledge and Understanding (CKU) guidance. 2025.
207. Crompton JL. Beyond metrics: shifting measures of research impact from citations and journal rankings to meaningful impacts on society. International Journal of Contemporary Hospitality Management. 2025;37(5):1449-64.
208. Collins GS, Moons KG, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+ AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. bmj. 2024;385.
209. Pop M, Salzberg SL. Use and mis-use of supplementary material in science publications. BMC bioinformatics. 2015;16(1):237.
210. Lawrence NR, Bacila I, Tonge J, Tucker A, Dawson J, Lang Z-Q, et al. Risk of Bias in Machine Learning and Statistical Models to Predict Height or Weight: A Systematic Review in Fetal and Paediatric Medicine. Available at SSRN 4788669. 2024.
211. Kahneman D, Tversky A. Intuitive prediction: Biases and corrective procedures. 1977.
212. National Institute for Health and Care Excellence. Recognition and Management of Faltering Growth in Children. National Institute for Health and Care Excellence, London. 2017.
213. Vanderniet JA, Jenkins AJ, Donaghue KC. Epidemiology of type 1 diabetes. Current cardiology reports. 2022;24(10):1455-65.
214. Kneebone RL. Bespoke practice. The Lancet. 2017;389(10064):28-9.
215. Goldstein H, Browne W, Rasbash J. Multilevel modelling of medical data. Statistics in Medicine. 2002;21(21):3291-315.
216. Sharma A, Lazarus K, Papadopoulou D, Prabhudev H, Tan T, Meeran K, et al. Optimising prednisolone or prednisone replacement in adrenal insufficiency. Endocrine Connections. 2023;12(8).
217. Beuschlein F, Else T, Bancos I, Hahner S, Hamidi O, Van Hulsteijn L, et al. European Society of Endocrinology and Endocrine Society Joint Clinical Guideline: diagnosis and therapy of glucocorticoid-induced adrenal insufficiency. European journal of endocrinology. 2024;190(5):G25-G51.

218. Lawrence NR, Bacila I, Tonge J, Dawson J, Collins GS, Lang Z-Q, et al. Supplementary Materials: Blood pressure and its associations in 554 children and young people with CAH. 2024.
219. Krone N, Rose IT, Willis DS, Hodson J, Wild SH, Doherty EJ, et al. Genotype-Phenotype Correlation in 153 Adult Patients With Congenital Adrenal Hyperplasia due to 21-Hydroxylase Deficiency: Analysis of the United Kingdom Congenital Adrenal Hyperplasia Adult Study Executive (CaHASE) Cohort. Journal of Clinical Endocrinology & Metabolism. 2013;98(2):E346-E54.
220. Welch R, Balagamage C, Bryce J, Alimussina M, Chen M, Francisco AG, et al., editors. Establishing the current practice of the use of genetic testing for patients with congenital adrenal hyperplasia (CAH). Endocrine Abstracts; 2025: Bioscientifica.
221. Feng YD, Wang J, Tao ZB, Jiang HK. Development and validation of a nomogram to predict poor short-term response to recombinant human growth hormone treatment in children with growth disorders. J Endocrinol Invest. 2023;46(7):1343-59.
222. Liu Z, Han N, Su T, Ji Y, Bao H, Zhou S, et al. Interpretable machine learning to identify important predictors of birth weight: A prospective cohort study. Front Pediatr. 2022;10:899954.
223. van Dommelen P, Arnaud L, Koledova E. Curve matching to predict growth in patients receiving growth hormone therapy: An interpretable & explainable method. Front Endocrinol (Lausanne). 2022;13:999077.
224. Bekele WT. Machine learning algorithms for predicting low birth weight in Ethiopia. BMC Med Inform Decis Mak. 2022;22(1):232.
225. Chen L, Fan X, Mao K, Tolba A, Alqahtani F, Ahmed AM. Study of Multidimensional and High-Precision Height Model of Youth Based on Multilayer Perceptron. Comput Intell Neurosci. 2022;2022:7843455.
226. Iwama N, Obara T, Ishikuro M, Murakami K, Ueno F, Noda A, et al. Risk scores for predicting small for gestational age infants in Japan: The TMM birthree cohort study. Sci Rep. 2022;12(1):8921.
227. Feng Y, Zheng H, Fang D, Mei S, Zhong W, Zhang G. Prediction of late-onset fetal growth restriction using a combined first- and second-trimester screening model. J Gynecol Obstet Hum Reprod. 2022;51(2):102273.
228. Ikwuezunma I, Wang K, Raymond S, Badin D, Kreulen RT, Jain A, et al. Height Gain After Spinal Fusion for Idiopathic Scoliosis: Which Model Fits Best? J Pediatr Orthop. 2022;42(9):457-61.
229. Kim HS, Oh SY, Cho GJ, Choi SJ, Hong SC, Kwon JY, et al. A Predictive Model for Large-for-Gestational-Age Infants among Korean Women with Gestational Diabetes Mellitus Using Maternal Characteristics and Fetal Biometric Parameters. J Clin Med. 2022;11(17).

230. Jing G, Huwei S, Chao C, Lei C, Ping W, Zhongzhou X, et al. A predictive model of macrosomic birth based upon real-world clinical data from pregnant women. BMC Pregnancy Childbirth. 2022;22(1):651.
231. Deval R, Saxena P, Pradhan D, Mishra AK, Jain AK. A Machine Learning-Based Intrauterine Growth Restriction (IUGR) Prediction Model for Newborns. Indian J Pediatr. 2022;89(11):1140-3.
232. Lee HS, Kum CD, Rho JG, Hwang JS. Long-term effectiveness of growth hormone therapy in children born small for gestational age: An analysis of LG growth study data. PLoS One. 2022;17(4):e0266329.
233. Tang X, Zhao Y, Liu Q, Hu D, Li G, Sun J, et al. The Effect of Risk Accumulation on Childhood Stunting: A Matched Case-Control Study in China. Front Pediatr. 2022;10:816870.
234. Wu C, Sun J, Dong X, Cai L, Deng X, Zhang F, et al. Establishment of a New Equation for Ultrasonographic Estimated Foetal Weight in Chongqing: A Prospective Study. IMR Press. 2022;49(12):271.
235. Kim YR, Park G, Joo EH, Jang JH, Ahn EH, Jung SH, et al. First-trimester screening model for small-for-gestational-age using maternal clinical characteristics, serum screening markers, and placental volume: prospective cohort study. J Matern Fetal Neonatal Med. 2022;35(25):5149-54.
236. Blue NR, Allshouse AA, Heerboth S, Grobman W, Mercer B, Shanks A, et al. Derivation and assessment of a sex-specific fetal growth standard. J Matern Fetal Neonatal Med. 2022;35(25):9913-21.
237. Wu W, Zhang B, Li D, Yan M, Deng Q, Kang Y, et al. Development and validation of nomogram for prediction of low birth weight: A large-scale cross-sectional study in northwest China. The Journal of Maternal-Fetal & Neonatal Medicine. 2022;35(25):7562-70.
238. O'Dwyer V, Russell N, McDonnell B, Sharkey L, Mulcahy C, Higgins M. Antenatal prediction of fetal macrosomia in pregnancies affected by maternal pre-gestational diabetes. The Journal of Maternal-Fetal & Neonatal Medicine. 2022;35(25):7412-6.
239. Martín-Palumbo G, Duque Alcorta M, Atanasova VB, Rego Tejeda MT, Antolin Alvarado E, Bartha JL. Prenatal prediction of very late onset small-for-gestational age newborns in low-risk pregnancies. The Journal of Maternal-Fetal & Neonatal Medicine. 2022;35(25):9816-20.
240. Bergman M, Reichman O, Farkash R, Bin-Nun A, Samueloff A, Sapir AZ, et al. Sonographic growth curves versus neonatal birthweight growth curves for the identification of fetal growth restriction. The Journal of Maternal-Fetal & Neonatal Medicine. 2022;35(23):4558-65.
241. Anum I, Zahra SS, Kumari S, Wahid A, Tahir UA, Qureshi AI. Frequency of Correct Fetal Weight Estimation by Clinical and Ultrasound Methods in Pregnant Women. Pakistan Journal of Medical & Health Sciences. 2022;16(11):251-.

242. Islam Pollob SA, Abedin MM, Islam MT, Islam MM, Maniruzzaman M. Predicting risks of low birth weight in Bangladesh with machine learning. PloS one. 2022;17(5):e0267190.
243. Zhang J, Wu X, Song Q. Analytical Comparison of Risk Prediction Models for the Onset of Macrosomia Based on Three Statistical Methods. Disease Markers. 2022;2022.
244. Xu D, Shen X, Guan H, Zhu Y, Yan M, Wu X. Prediction of small-for-gestational-age neonates at 33–39 weeks' gestation in China: logistic regression modeling of the contributions of second-and third-trimester ultrasound data and maternal factors. BMC Pregnancy and Childbirth. 2022;22(1):661.
245. Chaliki AC, Nekkanti R, Chaliki SC, Chandra TJ. Johnsons Technique versus Hadlock – A Comparative Study to Estimate Foetus Weight. International Journal of Pharmaceutical and Clinical Research. 2022;14(10):499-503.
246. Duncan JR, Schenone C, Dorset KM, Goedecke PJ, Tobiasz AM, Meyer NL, et al. Estimated fetal weight accuracy in pregnancies with preterm prelabor rupture of membranes by the Hadlock method. The Journal of Maternal-Fetal & Neonatal Medicine. 2022;35(9):1754-8.
247. Dieste-Pérez P, Savirón-Cornudella R, Tajada-Duaso M, Pérez-López FR, Castán-Mateo S, Sanz G, et al. Personalized Model to predict small for gestational age at delivery using fetal biometrics, maternal characteristics, and pregnancy biomarkers: a retrospective cohort study of births assisted at a Spanish hospital. Journal of Personalized Medicine. 2022;12(5):762.
248. Yuan X, Han X, Jia C, Long W, Wang H, Yu B, et al. Investigation and application of risk factors of macrosomia based on 10,396 Chinese pregnant women. Frontiers in Endocrinology. 2022;13:837816.
249. Du J, Zhang X, Chai S, Zhao X, Sun J, Yuan N, et al. Nomogram-based risk prediction of macrosomia: a case-control study. BMC Pregnancy and Childbirth. 2022;22(1):1-8.
250. Blum WF, Ranke MB, Keller E, Keller A, Barth S, de Bruin C, et al. A Novel Method for Adult Height Prediction in Children With Idiopathic Short Stature Derived From a German-Dutch Cohort. Journal of the Endocrine Society. 2022;6(7):bvac074.
251. Mejorado-Molano FJ, Sanz-Calvo ML, Posada-Ayala A, Caballo-Roig N, Gavela-Pérez T, Mahillo-Fernández I, et al. Adult Height in Girls With Idiopathic Premature Adrenarche: A Cohort Study and Design of a Predictive Model. Frontiers in Endocrinology. 2022;13:852422.
252. Mlodawski J, Wolder D, Niziurski P, Adamczyk-Gruszka O, Głuszek S, Rokita W. Birth weight prediction by Lee formula based on fractional thigh volume in term pregnancies–is it helpful? Archives of Medical Science: AMS. 2022;18(1):79.

253. Lian C, Wang Y, Bao X, Yang L, Liu G, Hao D, et al. Dynamic prediction model of fetal growth restriction based on support vector machine and logistic regression algorithm. Frontiers in Surgery. 2022;9:951908.
254. Chilyabanyama ON, Chilengi R, Simuyandi M, Chisenga CC, Chirwa M, Hamusonde K, et al. Performance of machine learning classifiers in classifying stunting among under-five children in Zambia. Children. 2022;9(7):1082.
255. Khazri HB, Shimmi SC, Parash MTH. A multivariate analysis to propose linear models for the stature estimation in the Sabahan young adult population. Plos one. 2022;17(8):e0273840.
256. McKenna M, McKenna D, Zhou M, Sonek J, Wiegand S. Prediction of Neonatal Growth Restriction in Fetuses With Gastroschisis by Early Third Trimester Ultrasonography Utilizing Contemporary Birth Weight Percentiles. Journal of Ultrasound in Medicine. 2022;42(5):997-1005.
257. Lolli L, Johnson A, Monaco M, Di Salvo V, Atkinson G, Gregson W. The percentage of mature height as a morphometric index of somatic growth: a formal scrutiny of conventional simple ratio scaling assumptions. Pediatric Exercise Science. 2022;35(2):107-15.
258. Rauh M, Rasim K, Schmidt B, Schnabel A, Köninger A. Accuracy of the sonographic determination of estimated fetal weight in anhydramnios. Archives of Gynecology and Obstetrics. 2022:1-8.
259. Huang S, Chen Z, Chen R, Zhang Z, Sun J, Chen H. Analysis of risk factors and construction of a prediction model for short stature in children. Front Pediatr. 2022;10:1006011.
260. Cheng Y, Xu H, Rissel C, Phongsavan P, Buchanan L, Taki S, et al. Estimation and feasibility of correction modelling for mother-reported child height and weight at 2 years using data from the Australian CHAT trial. Scientific Reports. 2022;12(1):21353.
261. Kiefer M, Finneran M, Ware C, Foy P, Thung S, Gabbe S, et al. Prediction of large-for-gestational-age infant by fetal growth charts and hemoglobin A1c level in pregnancy complicated by pregestational diabetes. Ultrasound in Obstetrics & Gynecology. 2022;60(6):751-8.
262. Polfuss M, Liu T, Smith K, Murphy PS, Ward E, Thibadeau J, et al. Weight Status of Children Participating in the National Spina Bifida Patient Registry. Pediatrics. 2022;150(6):e2022057007.
263. Boute T, Rizzo G, Mappa I, Makatsariya A, Toneto BR, Moron AF, et al. Correlation between estimated fetal weight and weight at birth in infants with gastroschisis and omphalocele. The Journal of Maternal-Fetal & Neonatal Medicine. 2022;35(16):3070-5.
264. Coenen H, Braun J, Köster H, Möllers M, Schmitz R, Steinhard J, et al. Role of umbilicocerebral and cerebroplacental ratios in prediction of perinatal outcome in FGR pregnancies. Archives of Gynecology and Obstetrics. 2022:1-10.

265. Ozler S, Kozanhan B, Bardak O, Arıkan MN. Are body roundness index and a body shape index in the first trimester related to foetal macrosomia? Journal of Obstetrics and Gynaecology. 2022;42(3):396-402.
266. Roeckner JT, Odibo L, Odibo AO. The value of fetal growth biometry velocities to predict large for gestational age (LGA) infants. The Journal of Maternal-Fetal & Neonatal Medicine. 2022;35(11):2099-104.
267. Visentin S, Londero AP, Cataneo I, Bellussi F, Salsi G, Pilu G, et al. A prenatal standard for fetal weight improves the prenatal diagnosis of small for gestational age fetuses in pregnancies at increased risk. BMC Pregnancy and Childbirth. 2022;22(1):1-7.
268. Song J, Liu J, Liu L, Jiang Y, Zheng H, Ke H, et al. The birth weight of macrosomia influence the accuracy of ultrasound estimation of fetal weight at term. Journal of Clinical Ultrasound. 2022;50(7):967-73.
269. Ng HM, MacDonell S, Yap J, Peddie MC, Scott T, Haszard JJ. Predicting height from ulna length for the determination of weight status in New Zealand adolescents: A cross-sectional study. Journal of Human Nutrition and Dietetics. 2022;35(2):406-14.
270. Wahab RJ, Jaddoe VW, Voerman E, Ruijter GJ, Felix JF, Marchioro L, et al. Maternal body mass index, early-pregnancy metabolite profile, and birthweight. The Journal of Clinical Endocrinology & Metabolism. 2022;107(1):e315-e27.
271. Zhang Y, Wang Q, Xue M, Pang B, Yang M, Zhang Z, et al. Identifying factors associated with central obesity in school students using artificial intelligence techniques. Frontiers in Pediatrics. 2022;10:1060270.
272. Byeon H. Predicting South Korean adolescents vulnerable to obesity after the COVID-19 pandemic using categorical boosting and shapley additive explanation values: A population-based cross-sectional survey. Frontiers in Pediatrics. 2022;10:955339.
273. Carrillo-Balam G, Doi L, Marryat L, Williams AJ, Bradshaw P, Frank J. Validity of Scottish predictors of child obesity (age 12) for risk screening in mid-childhood: a secondary analysis of prospective cohort study data—with sensitivity analyses for settings without various routinely collected predictor variables. International Journal of Obesity. 2022;46(9):1624-32.
274. Wang Q, Yang M, Pang B, Xue M, Zhang Y, Zhang Z, et al. Predicting risk of overweight or obesity in Chinese preschool-aged children using artificial intelligence techniques. Endocrine. 2022;77(1):63-72.
275. Gomes D, Le L, Perschbacher S, Haas NA, Netz H, Hasbargen U, et al. Predicting the earliest deviation in weight gain in the course towards manifest overweight in offspring exposed to obesity in pregnancy: a longitudinal cohort study. BMC medicine. 2022;20(1):1-18.

276. He X, Jiang Z, Wu C, Zeng L, Qi M, Sun Y, et al. Development of a nutritional risk screening tool for preterm children in outpatient settings during a complementary feeding period: a pilot study. BMC pediatrics. 2022;22(1):1-12.
277. Huang K-H, Chen F-Y, Liu Z-Z, Luo J-Y, Xu R-L, Jiang L-L, et al. Prediction of pre-eclampsia complicated by fetal growth restriction and its perinatal outcome based on an artificial neural network model. Frontiers in Physiology. 2022:2272.
278. Lugo-Martinez J, Xu S, Levesque J, Gallagher D, Parker LA, Neu J, et al. Integrating longitudinal clinical and microbiome data to predict growth faltering in preterm infants. Journal of Biomedical Informatics. 2022;128:104031.
279. Price CR, Roeckner J, Odibo L, Odibo A. Comparing fetal biometric growth velocity versus estimated fetal weight for prediction of neonatal small for gestational age. The Journal of Maternal-Fetal & Neonatal Medicine. 2022;35(20):3931-6.
280. Han JH, Yoon SJ, Lee HS, Park G, Lim J, Shin JE, et al. Application of Machine Learning Approaches to Predict Postnatal Growth Failure in Very Low Birth Weight Infants. Yonsei Medical Journal. 2022;63(7):640.
281. Wang D, Schwinger C, Urassa W, Berhane Y, Strand TA, Fawzi WW. Comparing Attained Weight and Weight Velocity during the First 6 Months in Predicting Child Undernutrition and Mortality. The Journal of Nutrition. 2022;152(1):319-30.
282. Pradhan A, Mishra P, Tiwari S, Choure K, Gupta A. Prediction of Low Birth Weight by Quadruple Parameters in High-Risk Pregnancies. International Journal of Applied and Basic Medical Research. 2022;12(4):277.
283. Perichart-Perera O, Avila-Sosa V, Solis-Paredes JM, Montoya-Estrada A, Reyes-Muñoz E, Rodríguez-Cano AM, et al. Vitamin D Deficiency, Excessive Gestational Weight Gain, and Oxidative Stress Predict Small for Gestational Age Newborns Using an Artificial Neural Network Model. Antioxidants (Basel). 2022;11(3).
284. Chen Y, Cai C, Tan J, Lei X, Chen Q, Zhang J, et al. High-risk growth trajectory related to childhood overweight/obesity and its predictive model at birth. The Journal of Clinical Endocrinology & Metabolism. 2022;107(10):e4015-e26.

Appendix

A2 Supplementary material for Chapter 2

Table S2.1 – Search terms used

Searched via EMBASE and MEDLINE via OVID on 27/4/2023.

S2.1.1 Machine learning or modelling terms:

1	Machine Learning/
2	(machine adj1 (learn$ or model$)).ti,ab,kw.
3	Deep Learning/
4	(deep adj2 learn$).ti,ab,kw.
5	exp Supervised Machine Learning/
6	(supervised adj2 machine adj2 learn$).ti,ab,kw.
7	((support or relevance) adj2 vector adj2 (machine$ or classification$)).ti,ab,kw.
8	"Neural Networks (Computer)"/
9	(neural adj2 network$).ti,ab,kw.
10	((statistical or "statistical-learning") adj1 (learn$ or strateg$)).ti,ab,kw.
11	(multi adj2 layer adj1 perceptron$).ti,ab,kw.
12	(random adj2 forest$).ti,ab,kw.
13	"RF classifi$".ti,ab,kw.
14	(lasso or ridge or kernel or ensemble or bagging or bagged or bootstrap$ or boosting or boosted or fuzzy).ti,ab,kw.
15	((penali?ed or regulari?ed) adj2 ('likelihood' or 'regression' or 'logistic' or 'survival' or 'estimat$' or 'function$' or 'method$' or 'least' or 'ensemble')).ti,ab,kw.
16	((classification or regression or estimation or decision) adj2 tree$).ti,ab,kw.
17	(bayes$ adj1 network$).ti,ab,kw.
18	(nearest adj1 neighbo?r).ti,ab,kw.
19	(k-nearest adj1 neighbo?r).ti,ab,kw.
20	(elastic adj1 net).ti,ab,kw.
21	(naive adj1 bayes$).ti,ab,kw.
22	((nonparametric or "non-parametric") adj2 (model$ or analys$)).ti,ab,kw
23	(KNN or ANN or ANNs or RNN or RF or SVM or NB or CART or DT or MLP).ti,ab,kw.
24	Logistic Models/
25	(logistic adj4 (model$ or regression)).ti,ab,kw.
26	Linear Models/
27	(linear adj2 (model$ or regression)).ti,ab,kw.
28	(proportion$ adj2 odds adj2 regression).ti,ab,kw.
29	Least-Squares Analysis/
30	(least adj2 square$).ti,ab,kw.
31	Survival Analysis/
32	(survival adj1 (analys$ or model$)).ti,ab,kw.
33	Proportional Hazards Models/
34	(proportional adj1 hazard$).ti,ab,kw.
35	((cox or parametric) adj1 (regression or model$)).ti,ab,kw.
36	(semi adj2 parametric adj1 (regression or model$)).ti,ab,kw.
37	Disease-Free Survival/
38	Progression-Free Survival/
39	((disease or progression or event) adj2 free adj1 survival).ti,ab,kw.
40	(overall adj1 survival).ti,ab,kw.
41	formula$.ti,ab,kw.
42	equation$.ti,ab,kw.
43	"curve matching".ti,ab,kw.
44	"interpolation method$".ti,ab,kw.

S2.1.2 Performance terms:

45	Prognosis/
46	(prognos$ adj1 (modelling or modeling or model or models or predict$ or index or performance or nomogram or tools or ability or accuracy or probability or risk or factor$ or marker$ or biomarker$)).ti,ab,kw.
47	"risk model$".ti,ab,kw.
48	"predict$ the prognosis of".ti,ab,kw.
49	"predict$ the risk of".ti,ab,kw.
50	"predict$ the probability of".ti,ab,kw.
51	Probability/
52	(probability adj1 (modelling or modeling or model or models)).ti,ab,kw.
53	predict$ adj3 (modelling or modeling or model or models or nomogram or tools or performance or ability or index or accuracy or probability or risk or factor$ or marker$ or biomarker$).ti,ab,kw.
54	"candidate predictor$".ti,ab,kw.
55	"predictive clinical parameter$".ti,ab,kw.
56	((discrimination or discriminative or discriminatory) adj1 (accuracy or ability or performance or value or model or models or power or capacity or capabilit$ or efficiency)).ti,ab,kw.
57	(discriminability or c-index or c-statistic or concordance or DCA).ti,ab,kw.
58	"decision curve".ti,ab,kw.
59	(calibrat$ adj1 (plot$ or curve$ or slope$ or model or models)).ti,ab,kw.
60	(brier adj1 score$).ti,ab,kw.
61	(performance adj1 (classification or classifier or clinical or accuracy or validation or metrics or diagnostic or AUC)).ti,ab,kw.
62	(sensitivity or specificity or PPV or NPV).ti,ab,kw.
63	"correctly classified".ti,ab,kw.
64	"clinical accuracy".ti,ab,kw.
65	"positive predictive value$".ti,ab,kw.
66	"negative predictive value$".ti,ab,kw.
67	(classification or classifier).ti,ab,kw.
68	Area Under Curve/
69	"Area under the curve".ti,ab,kw.
70	"Area under the ROC curve".ti,ab,kw.
71	"Area under the ROC".ti,ab,kw.
72	"Area Under the Receiver Operat$ Characteristic$".ti,ab,kw.
73	ROC Curve/
74	"receiver operating characteristic$".ti,ab,kw.
75	(ROC or AUC or AUROC).ti,ab,kw.
76	"Hosmer-Lemeshow".ti,ab,kw.
77	"H-L test".ti,ab,kw.
78	"expected ratio".ti,ab,kw.
79	"observed ratio".ti,ab,kw.
80	"E:O ratio".ti,ab,kw.
81	r?squared.ti,ab,kw.
82	MSE$.ti,ab,kw.
83	Correlation.ti,ab,kw.
84	risk accumulation$.ti,ab,kw.
85	"root mean square$ error".ti,ab,kw.
86	"root mean square$ percentage error".ti,ab,kw.
87	RMSE$.ti,ab,kw.
88	explained varia$.ti,ab,kw.
89	error.ti,ab,kw.
90	("F?score" or "F score" or "F1 score" or "F2 score" or "F-1 score" or "F-2 score").ti,kw,ab.
91	("F?measure" or "f measure" or "f1 measure" or "f2 measure" or "f?1 measure" or "f?2 measure").ti,kw,ab.

92	r?2.ti,ab,kw.
93	recall.ti,ab,kw.
94	precision.ti,ab,kw.
95	accuracy.ti,ab,kw.
96	"detection rate".ti,ab,kw
97	"false positive rate".ti,ab,kw
98	"hamming loss".ti,ab,kw.
99	"jaccard score".ti,ab,kw.
100	"jaccard index".ti,ab,kw.
101	"youden score".ti,ab,kw.
102	"youden index".ti,ab,kw.

S2.1.3 Growth terms

103	growth$.ti,ab,kw.
104	height$.ti,ab,kw.
105	weight$.ti,ab,kw.
106	fat$.ti,ab,kw.
107	obes$.ti,ab,kw
108	macrosom$.ti,ab,kw.
109	stunt$.ti,ab,kw.
110	stature$.ti,ab,kw.
111	"failure to thrive".ti,ab,kw.

S2.1.4 Patient demographic terms

112	(adolescen$ or child$ or infan$ or neonat$ or neo-nat$ or born$ or paediatric$ or peadiatric$ or pediatric$ or perinat$ or fetus$ or foetus$ or fetal$ or foetal$).ti,ab,kw.
113	adolescent/ or child/ or infant/ or fetus/

S2.1.5 Combination of terms

114	Or/1-43	Machine learning or modelling terms
115	Or/44-92	Performance terms
116	Or/93-99	Growth terms
117	Or/100-101	Adolescent, Infant, Child or fetus terms
118	And/102-105	**Systematic Review Collection of terms**
119	limit 118 to yr="2022"	**Time restriction**
120	limit 119 to (english)	**Journal article and language restriction**
121	Deduplicate 120	**Final titles and abstracts to screen**

3821 Table S2.2 – PROBAST domains and signalling questions

3822 Screening questions used to inform domain ratings (54, 83)

PROBAST domain and signalling questions	Notes
Domain 1: PARTICIPANTS	
1.1 Were appropriate data sources used, e.g., cohort, randomized controlled trial, or nested case–control study data?	
1.2 Were all inclusions and exclusions of participants appropriate?	
Domain 2: PREDICTORS	
2.1 Were predictors defined and assessed in a similar way for all participants?	
2.2 Were predictor assessments made without knowledge of outcome data?	
2.3 Are all predictors available at the time the model is intended to be used?	
Domain 3: OUTCOMES	
3.1 Was the outcome determined appropriately?	
3.2 Was a prespecified or standard outcome definition used?	
3.3 Were predictors excluded from the outcome definition?	
3.4 Was the outcome defined and determined in a similar way for all participants?	
3.5 Was the outcome determined without knowledge of predictor information?	
3.6 Was the time interval between predictor assessment and outcome determination appropriate?	
Domain 4: ANALYSIS	
4.1 Were there a reasonable number of participants with the outcome?	
4.2 Were continuous and categorical predictors handled appropriately?	
4.3 Were all enrolled participants included in the analysis?	
4.4 Were participants with missing data handled appropriately?	
4.5 Was selection of predictors based on univariable analysis avoided?	
4.6 Were complexities in the data (e.g., censoring, competing risks, sampling of control participants) accounted for appropriately?	Not applicable to assessment of model validation
4.7 Were relevant model performance measures evaluated appropriately?	
4.8 Were model overfitting and optimism in model performance accounted for?	Not applicable to assessment of model validation
4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?	Not applicable to assessment of model validation

3823

3824 Table S2.3 – Study specific guidance for PROBAST ratings

Domain 1: Participants

1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study?

Factors decreasing risk of bias (Y/PY)

- prospective longitudinal cohorts (RCT or proper registry) with consistent methods for participant inclusion and exclusion – predefined predictors and outcome determination
- case control/cohort studies are low risk of bias if the original cohort/registry outcome frequency is adjusted for. Look for 'reweighting' or 'inverse sampling fraction' of the outcome.

Factors increasing risk of bias (N/PN)

- existing cohorts with potentially inconsistent participant inclusion/exclusion criteria – data collected for other purposes than developing and validating a prediction model (if a protocol is given this may reduce the risk of bias)
- RCTs have narrower eligibility for participants, they also need to ensure treatment is included as a predictor, if not this is high risk of bias
- ill-defined case-control/cohort studies are at high risk of bias
- non-nested case-control design

Notes:
Studies using questionnaire data are not necessarily at high risk of bias and are considered on a case by case basis. If the population is appropriately representative or weighted to make itself representative of the population in which it could be employed their use can be justified.

1.2 Were all inclusions and exclusions of participants appropriate?

Factors decreasing risk of bias (Y/PY)

- inclusion/exclusion appropriate to get representative sample of target population
- participants correspond to unselected participants of interest

Factors increasing risk of bias (N/PN)

- inappropriate inclusion/exclusion of participants e.g., subgroups of populations that are not representative of the target population
- includes participants who have already had the outcome e.g., including pre-operative transfusion patient when predicting intra- or post-operative transfusion or if the blood transfusion outcome is a self-reported outcome measure

Notes:
Inclusion and exclusion criteria should not include any attribute that is retrospectively applied, and cannot be known at the time of entry into the study (e.g. excluding babies born prematurely). Where studies exclude patients on the basis of missing data, this is considered within the analysis rating, not the domain of participants.

Domain 2: Predictors

2.1 Were predictors defined and assessed in a similar way for all participants?

Factors decreasing risk of bias (Y/PY)

- predictors defined and assessed in the same way
- definitions of predictors and their assessment were similar for all participants

Factors increasing risk of bias (N/PN)

- predictors not defined and assessed in the same way
- predictors involving subjective judgement/assessment or skilled training, relying on the ability of the assessor
- data from multiple sources likely to have used different definitions

Notes:
Where birth weight is defined without reference to precise equipment, this is not considered a high risk of bias due to the simplicity of the measurement, provided it was conducted in hospital

2.2 Were predictor assessments made without knowledge of outcome data?

Factors decreasing risk of bias (Y/PY)

- outcome information was stated as not used during predictor assessment or was clearly not (yet) available to those assessing predictors
- blinding of the outcome

Factors increasing risk of bias (N/PN)

- clear that outcome information was used when assessing predictors
- retrospectively recorded predictors

Notes:
Whilst blinding is to be considered, cases that do not specifically report blinding of the outcome can be considered on a chronological basis (e.g. ultrasonographic fetal measurements will be made at time of scan, and can be rated as PY, as birth weight cannot be known at the time)

2.3 Are all predictors available at the time the model is intended to be used?

Factors decreasing risk of bias (Y/PY)

- *for development studies,* the model can be used in the real world and predictors are available in clinical settings
- included predictors would be available at the time the model is intended to be used for prediction (e.g., if values from umbilical cord blood is used to predict the baby's weight, this would not be useful in clinical practice because umbilical cord blood is only taken after the baby is born)

Factors increasing risk of bias (N/PN)

- *for validation studies,* predictor data needed for the model is missing from the validation dataset

- predictors would not be available at the time the model is intended to be used for prediction

Domain 3: Outcome

3.1 Was the outcome determined appropriately?

Factors decreasing risk of bias (Y/PY)

- method of outcome determination has been used which is considered optimal or acceptable by guidelines or previous publications on the topic

Factors increasing risk of bias (N/PN)

- a clearly suboptimal method has been used that causes unacceptable error in determining outcome status in participants

Notes:
Where birth weight is defined without reference to precise equipment, this is not considered a high risk of bias due to the simplicity of the measurement, provided it was conducted in hospital

3.2 Was a prespecified or standard outcome definition used?

Factors decreasing risk of bias (Y/PY)

- pre-specified/standard objective outcome
- substantiated from clinical guidelines/previous studies/available protocol
- prespecified categories are used to group outcomes

Factors increasing risk of bias (N/PN)

- composite outcomes
- uses atypical thresholds, or creates multiple thresholds on continuous outcome
- outcomes created from the same dataset (e.g. via cluster analysis)

3.3 Were predictors excluded from the outcome definition?

Factors decreasing risk of bias (Y/PY)

- outcome determined without any predictor information

Factors increasing risk of bias (Y/PY)

- ≥1 of the predictors forms part of the outcome definition

Notes:
Using previous height or weight to predict future height or weight is sensible and not considered as the same predictor information. However, if current height or weight is used within the creation of another metric (e.g. height or weight trajectory) that is then used as a predictor, the model is at a high risk of bias

3.4 Was the outcome defined and determined in a similar way for all participants?

Factors decreasing risk of bias (Y/PY)

- outcomes were defined and determined in a similar way for all participants

Factors increasing risk of bias (N/PN)

- data collected for non-research purposes e.g., routinely collected data from registries

Notes:
Where birth weight is defined without reference to precise equipment, this is not considered a high risk of bias due to the simplicity of the measurement, provided it was conducted in hospital

3.5 Was the outcome determined without knowledge of predictor information?

Factors decreasing risk of bias (Y/PY)

- predictor information was not known when determining the outcome status
- outcome status determination is clearly reported as determined without knowledge of predictor information

Factors increasing risk of bias (N/PN)

- predictor information was used when determining the outcome status
- outcomes were clearly defined and determined in a different way for some participants
- outcomes requiring interpretation
- predictor information would be available at time of outcome determination (consider the potential consequences)

Notes:
Due to the objective nature of measuring height and weight, it is important to consider the context of the data collection prior to rating a model at high risk of bias due to a lack of information or failure to blind those measuring the outcome to predictor information

3.6 Was the time interval between predictor assessment and outcome determination appropriate?

Factors decreasing risk of bias (Y/PY)

- time interval between predictor assessment and outcome determination was appropriate to enable suitable assessment of the predictive accuracy

Factors increasing risk of bias (N/PN)

- time between predictor assessment and outcome determination is too long/too short to enable suitable assessment of the predictive accuracy

Notes:
Metrics reported from models developed or applied to changes in height or weight very soon after predictor assessment are likely to show inflated accuracy. This must be considered in the clinical context of when the prediction model is designed to be applied. For instance, a third trimester ultrasound to predict birthweight may be conducted late in

pregnancy close to when birthweight is attained. However, if the model is designed to be employed to inform about the decision to proceed to caesarean section, then this has valid clinical utility to be applied at that point in pregnancy, and is therefore not necessarily at risk of bias.

Domain 4: Analysis

4.1 Were there a reasonable number of participants with the outcome?

Factors decreasing risk of bias (Y/PY)

- *for model development studies,* if the outcome is categorical within the model, the number of participants with the outcome relative to the number of candidate predictor parameters is ≥20 (events per variable ≥20)*. If the outcome is continuous, consider a number of participants >100 as a minimum
- *for model validation studies,* if the number of participants with the outcome is ≥100, or participants>100 for continuous outcomes

Factors increasing risk of bias (N/PN)

- *for model development studies,* if the number of participants with the outcome relative to the number of candidate predictor parameters is <10)* , or number of participants <100 for a continuous outcome
- *for model validation studies,* if the number of participants with the outcome is <100, or participants<100 for continuous outcomes

Notes:

**For events per variable between 10 and 20, the item should be rated as either probably yes or probably no, depending on the outcome frequency, overall model performance, and distribution of the predictors in the model*

Consider the number of degrees of freedom used by categorical predictors. For instance, if a predictor is ethnicity that contains 3 categories of 'white', 'black' and 'other', this will use 2 degrees of freedom and require the estimation of 2 candidate predictor parameters. Extra parameters are estimated if there are extra terms added to assess non-linearity (e.g. quadratic terms used two degrees of freedom to estimate $\beta x + \beta x^2$*).*

Consider what parameter the model is estimating prior to considering events per variable. If the model itself is predicting a continuous outcome, but internal validation assesses those outcomes against a standardised dichotomous threshold, the model sample size should be considered against the continuous variable that it is assessing, not the number of categorical results above the threshold. However, if there are few results over or under a clinically significant threshold, this must be considered within domain 1 as a potential risk of bias due to a bias sample.

4.2 Were continuous and categorical handled appropriately?

Factors decreasing risk of bias (Y/PY)

- continuous kept as continuous predictors
- continuous predictors examined for nonlinearity – look for 'fractional polynomials' or 'restricted cubic splines'

Factors increasing risk of bias (N/PN)

- continuous predictors are dichotomised

- continuous predictors are categorised, especially using widely accepted clinical cut-offs**,** data driven cut-offs increase the risk of bias
- *for validation studies,* predictors are collected using different formats

4.3 Were enrolled participants included in the analysis?

Factors decreasing risk of bias (Y/PY)

- all participants enrolled in the study are included in the data analysis, or a low number are excluded

Factors increasing risk of bias (N/PN)

- some or a subgroup of participants are inappropriately excluded from the analysis, including participants with 'unclear' findings, missing data, or outliers

4.4 Were participants with missing data handled appropriately?

Factors decreasing risk of bias (Y/PY)

- no missing values of predictors or outcomes and the study explicitly reports that participants are not excluded based on missing data
- missing values are handled using multiple imputation
- comparing results with and without missing data

Factors increasing risk of bias (N/PN)

- missing data are omitted from the analysis
- method of handling missing data is clearly flawed, e.g., missing indicator method or inappropriate use of last value carried forward
- study had no explicit mention of methods to handle missing data

Notes:
If missing data information not reported then assume complete case analysis was conducted

4.5 Was selection of predictors based on univariable analysis avoided?

Factors decreasing risk of bias (Y/PY)

- predictors are not selected based on univariable analysis prior to multivariable modelling
- predictors selected on existing knowledge, they are reliable, consistent, applicable, available, and credible
- credible/a-priori predictors are forced into the model

Factors increasing risk of bias (N/PN)

- predictors are selected based on univariable analysis prior to multivariable modelling

Notes:
This question applies for development studies only

4.6 Were complexities in the data (e.g., censoring, competing risks, sampling of control participants) accounted appropriately?

Factors decreasing risk of bias (Y/PY)

- complexities in the data are accounted for appropriately
- clear that any potential data complexities have been identified appropriately as unimportant
- multilevel or random effects models for multiple outcome measures

Factors increasing risk of bias (N/PN)

- complexities in the data that could affect model performance are ignored

Notes:
For a complexity in the data to apply, the criticism must not have already been considered in any of the other screening questions. If a dataset has no complexities that aren't considered elsewhere, then this should be marked as 'Y' rather than 'NI'

4.7 Were relevant model performance measures evaluated appropriately?

Factors decreasing risk of bias (Y/PY)

- both calibration and discrimination* are evaluated appropriately (including relevant measures tailored for models predicting survival outcomes).

Factors increasing risk of bias (N/PN)

- both calibration and discrimination are not evaluated
- only goodness-of-fit tests, such as the Hosmer–Lemeshow test, are used to evaluate calibration
- if classification measures (like sensitivity, specificity, or predictive values) were presented using predicted probability thresholds derived from the data set at hand/non-clinical cut-offs

Notes:
If model is predicting a continuous outcome without a clinically significant threshold associated with it, it may be appropriate to only have assessed calibration. However, if predicting a continuous outcome in relation to a threshold that would change a clinical decision, discrimination around that threshold should be assessed

4.8 Was model overfitting, underfitting, and optimism in model performance accounted for?

Factors decreasing risk of bias (Y/PY)

- internal validation techniques, such as bootstrapping and cross-validation including all model development procedures, have been used to account for any optimism in model fitting, and subsequent adjustment of the model performance estimates have been applied

Factors increasing risk of bias (N/PN)

- no internal validation has been performed, or if internal validation consists only of a single random split-sample of participant data
- bootstrapping or cross-validation did not include all model development procedures including any variable selection

Notes:
This question applies for development studies only

4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?

Factors decreasing risk of bias (Y/PY)

- predictors and regression coefficients in the final model correspond to reported results from multivariable analysis

Factors increasing risk of bias (N/PN)

- predictors and regression coefficients in the final model do not correspond to reported results from multivariable analysis

Notes:
If final model is not presented, or model code is not available via supplementary material, then rate as NI. If model is presented once, then one can assume there is no conflicting model parameters. High risk of bias is considered when there is conflicting model information presented in different places within the paper or supplementary material.

Link to data collection proforma:

https://tinyurl.com/3aazxyen

3828 # Table S2.4 – List of all included studies

3829 Study code to correlate to Figure S2.4, model types rated, and total number of models assessed within each study

Title of study	Study code	Type of Study	No. of models rated
Development and validation of a nomogram to predict poor short term response to recombinant human growth hormone treatment in children with growth disorders (221)	A003	Development	2
Interpretable machine learning to identify important predictors of birth weight: A prospective cohort study (222)	A004	Development	8
Curve matching to predict growth in patients receiving growth hormone therapy: An interpretable & explainable method (223)	A005	Development	1
Machine learning algorithms for predicting low birth weight in Ethiopia (224)	A006	Development	8
Study of Multidimensional and High-Precision Height Model of Youth Based on Multilayer Perceptron (225)	A007	Development and external validation of the same model	1
Risk scores for predicting small for gestational age infants in Japan: The TMM BirThree cohort study (226)	A008	Development	3
Prediction of late-onset fetal growth restriction using a combined first- and second-trimester screening model (227)	A010	Development	3
Height Gain After Spinal Fusion for Idiopathic Scoliosis: Which Model Fits Best? (228)	A012	Validation	5
A Predictive Model for Large-for-Gestational-Age Infants among Korean Women with Gestational Diabetes Mellitus Using Maternal Characteristics and Fetal Biometric Parameters (229)	A013	Development of a model with validation of two already existing models	3
A predictive model of macrosomic birth based upon real world clinical data from pregnant women (230)	A014	Development	1
A Machine Learning Based Intrauterine Growth Restriction (IUGR) Prediction Model for Newborns (231)	A015	Development	10
Long-term effectiveness of growth hormone therapy in children born small for gestational age: An analysis of LG growth study data (232)	A016	Development	1
The Effect of Risk Accumulation on Childhood Stunting: A Matched Case-Control Study in China (233)	A020	Development	1
Establishment of a New Equation for Ultrasonographic Estimated Foetal Weight in Chongqing: A Prospective Study (234)	A022	Development and external validation of the same model with validation of 3 other already existing models	4
First-trimester screening model for small-for-gestational-age using maternal clinical characteristics, serum screening markers, and placental volume: prospective cohort study (235)	A023	Development	1

Title		Type of Study	No. of models rated
Derivation and assessment of a sex-specific fetal growth standard (236)	A025	Development	1
Development and validation of nomogram for prediction of low birth weight: a large-scale cross-sectional study in northwest China (237)	A026	Development	1
Antenatal prediction of fetal macrosomia in pregnancies affected by maternal pre-gestational diabetes (238)	A027	Development	1
Prenatal prediction of very late onset small-for-gestational age newborns in low-risk pregnancies (239)	A028	Development	3
Sonographic growth curves versus neonatal birthweight growth curves for the identification of fetal growth restriction (240)	A029	Validation	1
Frequency of Correct Fetal Weight Estimation by Clinical and Ultrasound Methods in Pregnant Women (241)	A031	Validation	2
Predicting risks of low birth weight in Bangladesh with machine learning (242)	A032	Development	2
Analytical Comparison of Risk Prediction Models for the Onset of Macrosomia Based on Three Statistical Methods (243)	A035	Development	3
Prediction of small-for-gestational-age neonates at 33-39 weeks' gestation in China: logistic regression modeling of the contributions of second- and third-trimester ultrasound data and maternal factors (244)	A036	Development of 4 models with validation of 2 already existing models	6
Johnsons Technique versus Hadlock - A Comparative Study to Estimate Foetus Weight (245)	A037	Validation	2
Estimated fetal weight accuracy in pregnancies with preterm prelabor rupture of membranes by the Hadlock method (246)	A038	Validation	1
Personalized Model to Predict Small for Gestational Age at Delivery Using Fetal Biometrics, Maternal Characteristics, and Pregnancy Biomarkers: A Retrospective Cohort Study of Births Assisted at a Spanish Hospital (247)	A039	Development of a model with validation of an already existing model	2
Investigation and Application of Risk Factors of Macrosomia Based on 10,396 Chinese Pregnant Women (248)	A040	Development	3
Nomogram-based risk prediction of macrosomia: a case-control study (249)	A041	Development	1
A Novel Method for Adult Height Prediction in Children With Idiopathic Short Stature Derived From a German-Dutch Cohort (250)	A043	Development of 10 models with validation of 3 already existing models	13
Adult Height in Girls With Idiopathic Premature Adrenarche: A Cohort Study and Design of a Predictive Model (251)	A045	Development	2
Birth weight prediction by Lee formula based on fractional thigh volume in term pregnancies: is it helpful? (252)	A046	Validation	1

Title		Type of Study	No. of models rated
Dynamic prediction model of fetal growth restriction based on support vector machine and logistic regression algorithm (253)	A049	Development	8
Performance of Machine Learning Classifiers in Classifying Stunting among Under-Five Children in Zambia (254)	A050	Development	5
A multivariate analysis to propose linear models for the stature estimation in the Sabahan young adult population (255)	A051	Development	1
Prediction of Neonatal Growth Restriction in Fetuses with Gastroschisis by Early Third Trimester Ultrasonography Utilizing Contemporary Birth Weight Percentiles (256)	A052	Validation	7
The Percentage of Mature Height as a Morphometric Index of Somatic Growth: A Formal Scrutiny of Conventional Simple Ratio Scaling Assumptions (257)	A053	Validation	1
Accuracy of the sonographic determination of estimated fetal weight in anhydramnios (258)	A054	Validation	1
Analysis of risk factors and construction of a prediction model for short stature in children (259)	A058	Development	1
Estimation and feasibility of correction modelling for mother-reported child height and weight at 2 years using data from the Australian CHAT trial (260)	A059	Development	6
Prediction of large-for-gestational-age infant by fetal growth charts and hemoglobin A1c level in pregnancy complicated by pregestational diabetes (261)	A063	Development of 3 models with validation of 1 already existing model	4
Weight Status of Children Participating in the National Spina Bifida Patient Registry (262)	A064	Development	1
Correlation between estimated fetal weight and weight at birth in infants with gastroschisis and omphalocele (263)	A070	Validation	3
Role of umbilicocerebral and cerebroplacental ratios in prediction of perinatal outcome in FGR pregnancies (264)	A078	Development of 2 models with validation of 2 already existing models	4
Are body roundness index and a body shape index in the first trimester related to foetal macrosomia? (265)	A081	Validation	3
The value of fetal growth biometry velocities to predict large for gestational age (LGA) infants (266)	A082	Development of 4 models with validation of 1 already existing model	5
A prenatal standard for fetal weight improves the prenatal diagnosis of small for gestational age fetuses in pregnancies at increased risk (267)	A086	Validation	3
The birth weight of macrosomia influence the accuracy of ultrasound estimation of fetal weight at term (268)	A091	Validation	1
Predicting height from ulna length for the determination of weight status in New Zealand adolescents: A cross-sectional study (269)	A099	Development of 1 model with validation of 2 already existing models	3

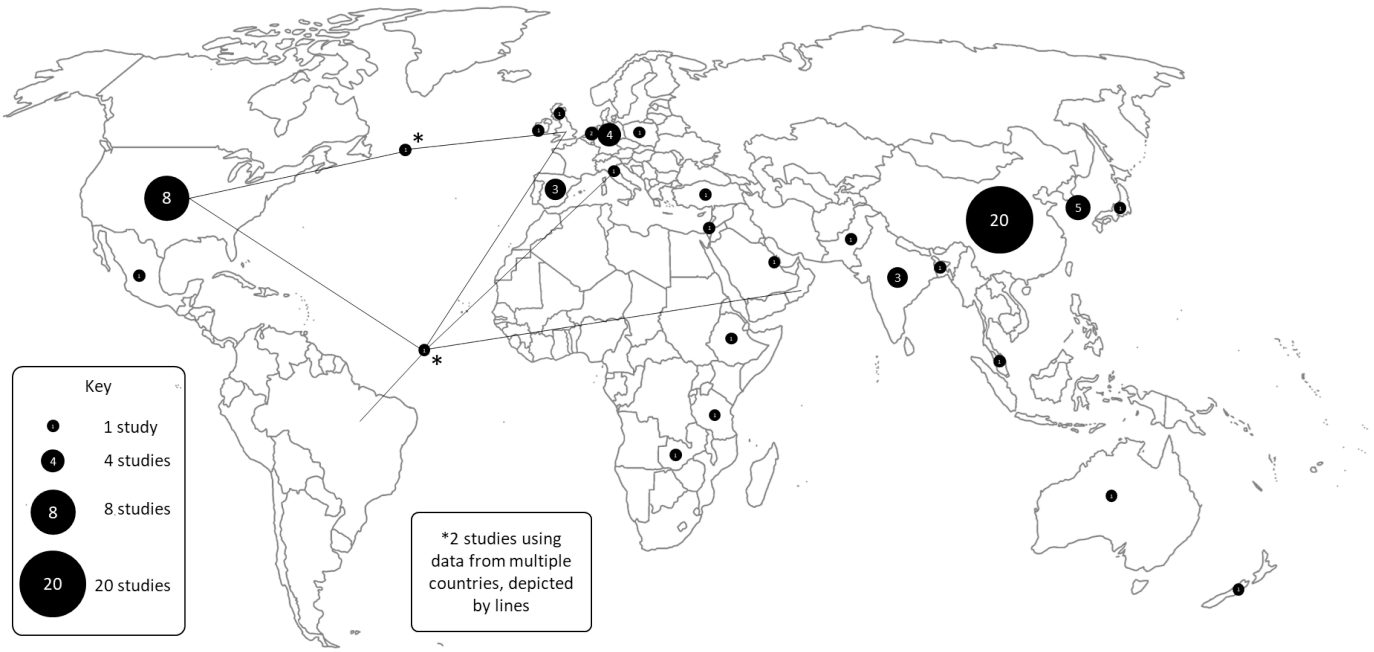
Title		Type of Study	No. of models rated
Maternal Body Mass Index, Early-Pregnancy Metabolite Profile, and Birthweight (270)	A108	Development	3
Identifying factors associated with central obesity in school students using artificial intelligence techniques (271)	A109	Development	11
Predicting South Korean adolescents vulnerable to obesity after the COVID-19 pandemic using categorical boosting and shapley additive explanation values: A population-based cross-sectional survey (272)	A117	Development	1
Validity of Scottish predictors of child obesity (age 12) for risk screening in mid-childhood: a secondary analysis of prospective cohort study data-with sensitivity analyses for settings without various routinely collected predictor variables (273)	A118	Development	2
Predicting risk of overweight or obesity in Chinese preschool-aged children using artificial intelligence techniques (274)	A120	Development	11
Predicting the earliest deviation in weight gain in the course towards manifest overweight in offspring exposed to obesity in pregnancy: a longitudinal cohort study (275)	A121	2 models with development and external validation of the same models, alongside 1 other model developed	3
Development of a nutritional risk screening tool for preterm children in outpatient settings during a complementary feeding period: a pilot study (276)	A133	3 models with development and external validation of the same models	3
Prediction of pre-eclampsia complicated by fetal growth restriction and its perinatal outcome based on an artificial neural network model (277)	A134	Development	1
Integrating longitudinal clinical and microbiome data to predict growth faltering in preterm infants (278)	A135	Development	6
Comparing fetal biometric growth velocity versus estimated fetal weight for prediction of neonatal small for gestational age (279)	A136	Development of 1 model with validation of 3 already existing models	4
Application of Machine Learning Approaches to Predict Postnatal Growth Failure in Very Low Birth Weight Infants (280)	A137	Development	25
Comparing Attained Weight and Weight Velocity during the First 6 Months in Predicting Child Undernutrition and Mortality (281)	A139	Validation	3
Prediction of Low Birth Weight by Quadruple Parameters in High-Risk Pregnancies (282)	A144	Development	1
Vitamin D Deficiency, Excessive Gestational Weight Gain, and Oxidative Stress Predict Small for Gestational Age Newborns Using an Artificial Neural Network Model (283)	A145	Development	1
High-risk Growth Trajectory Related to Childhood Overweight/Obesity and Its Predictive Model at Birth (284)	A146	Development	5

3830

Table S2.5 – Countries from which the participants used in articles were assessed

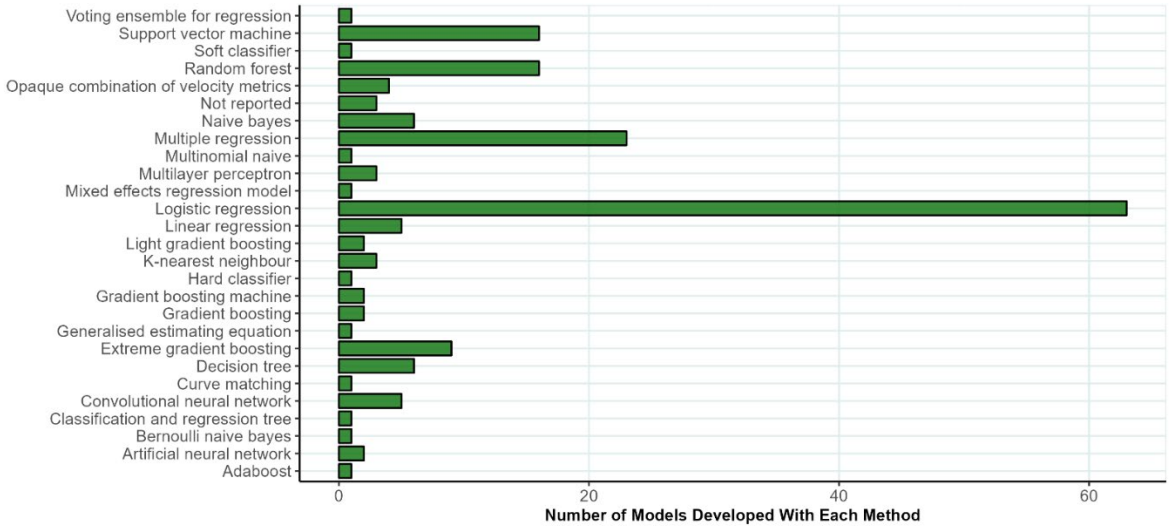
Country	n
Australia	1
Bangladesh	1
Brazil / Italy / Oman / UK / USA	1
China	20
Ethiopia	1
Germany	4
India	3
Ireland	1
Israel	1
Italy	1
Japan	1
Malaysia	1
Mexico	1
Netherlands	2
New Zealand	1
Pakistan	1
Poland	1
Qatar	1
Scotland	1
South Korea	5
Spain	3
Tanzania	1
Turkey	1
USA	8
USA & UK	1
Zambia	1

3834 Figure S2.1 – Map of countries from which participant data was used in studies



3835

3836 Figure S2.2 – Methods applied within prediction models assessed



3837

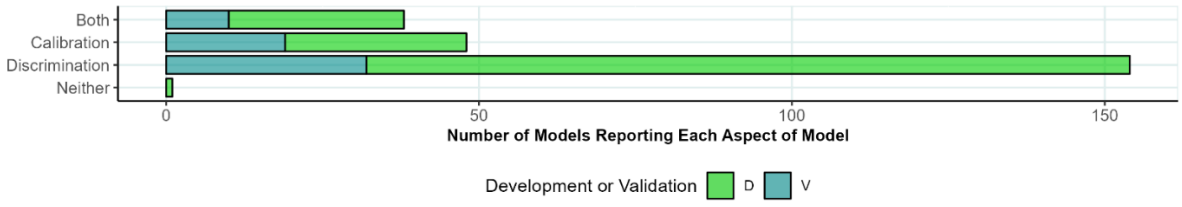
3838 Table S2.6 – Categories of methods employed for all models developed

Model Method	**n, %**
Regression Based Method (n=97)	
Logistic regression	63, 35.0%
Multiple linear regression	23, 12.8%
Linear regression	5, 2.7%
Opaque combination of velocity metrics (likely regression)	4, 2.2%
Mixed effects regression model	1, 0.6%
Generalised estimating equation	1, 0.6%
Flexible Machine Learning Method (n=45)	
Support vector machine	16, 8.9%
Naive bayes	6, 3.3%
Decision tree	6, 3.3%
Convolutional neural network	5, 2.7%
Multilayer perceptron	3, 1.6%
K-nearest neighbour	3, 1.6%
Artificial neural network	2, 1.1%
Multinomial naïve	1, 0.6%
Curve matching	1, 0.6%
Classification and regression tree	1, 0.6%
Bernoulli naive bayes	1, 0.6%
Ensemble Machine Learning Method (n=35)	
Random forest	16, 8.9%
Extreme gradient boosting	9, 5.0%
Gradient boosting	4, 2.2%
Light gradient boosting	2, 1.1%
Voting ensemble for regression	1, 0.6%
Soft classifier	1, 0.6%
Hard classifier	1, 0.6%
Adaboost	1, 0.6%
Model development method not reported	
Not reported	3, 1.6%

3839

3840

Figure S2.3 – Method of analysis of accuracy of predictions from models reviewed



3844 Table S2.7 – Individual question ratings for all developed and validated models

		Developed models (n=180) n, %					Validated models (n=61) n, %				
	Question	**Y**	**PY**	**PN**	**N**	**NI**	**Y**	**PY**	**PN**	**N**	**NI**
1.1	Were appropriate data sources used, e.g., cohort, randomized controlled trial, or nested case control study data?	n = 78, 43%	n = 5, 3%	n = 65, 36%	n = 14, 8%	n = 18, 10%	n = 40, 66%	n = 3, 5%	n = 3, 5%	n = 15, 24%	n = 0, 0%
1.2	Were all inclusions and exclusions of participants appropriate?	n = 30, 17%	n = 32, 18%	n = 28, 15%	n = 63, 35%	n = 27, 15%	n = 21, 35%	n = 10, 16%	n = 0, 0%	n = 27, 44%	n = 3, 5%
2.1	Were predictors defined and assessed in a similar way for all participants?	n = 39, 21%	n = 13, 7%	n = 55, 31%	n = 55, 31%	n = 18, 10%	n = 26, 42%	n = 11, 18%	n = 1, 2%	n = 20, 33%	n = 3, 5%
2.2	Were predictor assessments made without knowledge of outcome data?	n = 91, 51%	n = 32, 18%	n = 29, 16%	n = 13, 7%	n = 15, 8%	n = 52, 85%	n = 0, 0%	n = 0, 0%	n = 2, 3%	n = 7, 12%
2.3	Are all predictors available at the time the model is intended to be used?	n = 168, 93%	n = 0, 0%	n = 0, 0%	n = 10, 6%	n = 2, 1%	n = 61, 100%	n = 0, 0%	n = 0, 0%	n = 0, 0%	n = 0, 0%
3.1	Was the outcome determined appropriately?	n = 35, 19%	n = 57, 32%	n = 27, 15%	n = 50, 28%	n = 11, 6%	n = 24, 39%	n = 30, 49%	n = 3, 5%	n = 3, 5%	n = 1, 2%
3.2	Was a prespecified or standard outcome definition used?	n = 157, 87%	n = 12, 7%	n = 0, 0%	n = 11, 6%	n = 0, 0%	n = 54, 88%	n = 3, 5%	n = 0, 0%	n = 3, 5%	n = 1, 2%
3.3	Were predictors excluded from the outcome definition?	n = 168 93%	n = 0, 0%	n = 0, 0%	n = 12, 7%	n = 0, 0%	n = 55, 90%	n = 3, 5%	n = 0, 0%	n = 3, 5%	n = 0, 0%
3.4	Was the outcome defined and determined in a similar way for all participants?	n = 19, 10%	n = 92, 52%	n = 21, 12%	n = 29, 16%	n = 19, 10%	n = 19, 31%	n = 34, 56%	n = 1, 2%	n = 7, 11%	n = 0, 0%
3.5	Was the outcome determined without knowledge of predictor information?	n = 0, 0%	n = 76, 42%	n = 17, 10%	n = 31, 17%	n = 56, 31%	n = 0, 0%	n = 38, 63%	n = 2, 3%	n = 2, 3%	n = 19, 31%
3.6	Was the time interval between predictor assessment and outcome determination appropriate?	n = 88, 49%	n = 3, 2%	n = 19, 10%	n = 52, 29%	n = 18, 10%	n = 43, 71%	n = 6, 10%	n = 3, 5%	n = 7, 11%	n = 2, 3%

		Developed Models					Validated Models				
	Question	**Y**	**PY**	**PN**	**N**	**NI**	**Y**	**PY**	**PN**	**N**	**NI**
4.1	Were there a reasonable number of participants with the outcome?	n = 29, 16%	n = 58, 32%	n = 13, 8%	n = 49, 27%	n = 31, 17%	n = 21, 34%	n = 21, 34%	n = 3, 5%	n = 16, 27%	n = 0, 0%
4.2	Were continuous and categorical predictors handled appropriately?	n = 24, 13%	n = 14, 8%	n = 1, 1%	n = 121, 67%	n = 20, 11%	n = 53, 87%	n = 0, 0%	n = 0, 0%	n = 7, 11%	n = 1, 2%
4.3	Were all enrolled participants included in the analysis?	n = 0, 0%	n = 0, 0%	n = 1, 1%	n = 143, 79%	n = 36, 20%	n = 7, 11%	n = 0, 0%	n = 0, 0%	n = 37, 61%	n = 17, 28%
4.4	Were participants with missing data handled appropriately?	n = 11, 6%	n = 2, 1%	n = 19, 11%	n = 114, 63%	n = 34, 19%	n = 0, 0%	n = 2, 3%	n = 0, 0%	n = 38, 62%	n = 21, 35%
4.5	Was selection of predictors based on univariable analysis avoided?	n = 88, 49%	n = 3, 2%	n = 12, 6%	n = 50, 28%	n = 27, 15%					
4.6	Were complexities in the data (e.g. censoring, competing risks, sampling of control participants) accounted for appropriately?	n = 38, 21%	n = 0, 0%	n = 2, 1%	n = 115, 64%	n = 25, 14%	n = 8, 13%	n = 0, 0%	n = 2, 3%	n = 45, 74%	n = 6, 10%
4.7	Were relevant model performance measures evaluated appropriately?	n = 8, 4%	n = 13, 7%	n = 3, 2%	n = 156, 87%	n = 0, 0%	n = 4, 6%	n = 0, 0%	n = 3, 5%	n = 54, 89%	n = 0, 0%
4.8	Were model overfitting and optimism in model performance accounted for?	n = 16, 9%	n = 11, 6%	n = 11, 6%	n = 103, 57%	n = 39, 22%					
4.9	Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?	n = 29, 16%	n = 2, 1%	n = 3, 2%	n = 14, 8%	n = 132, 73%					

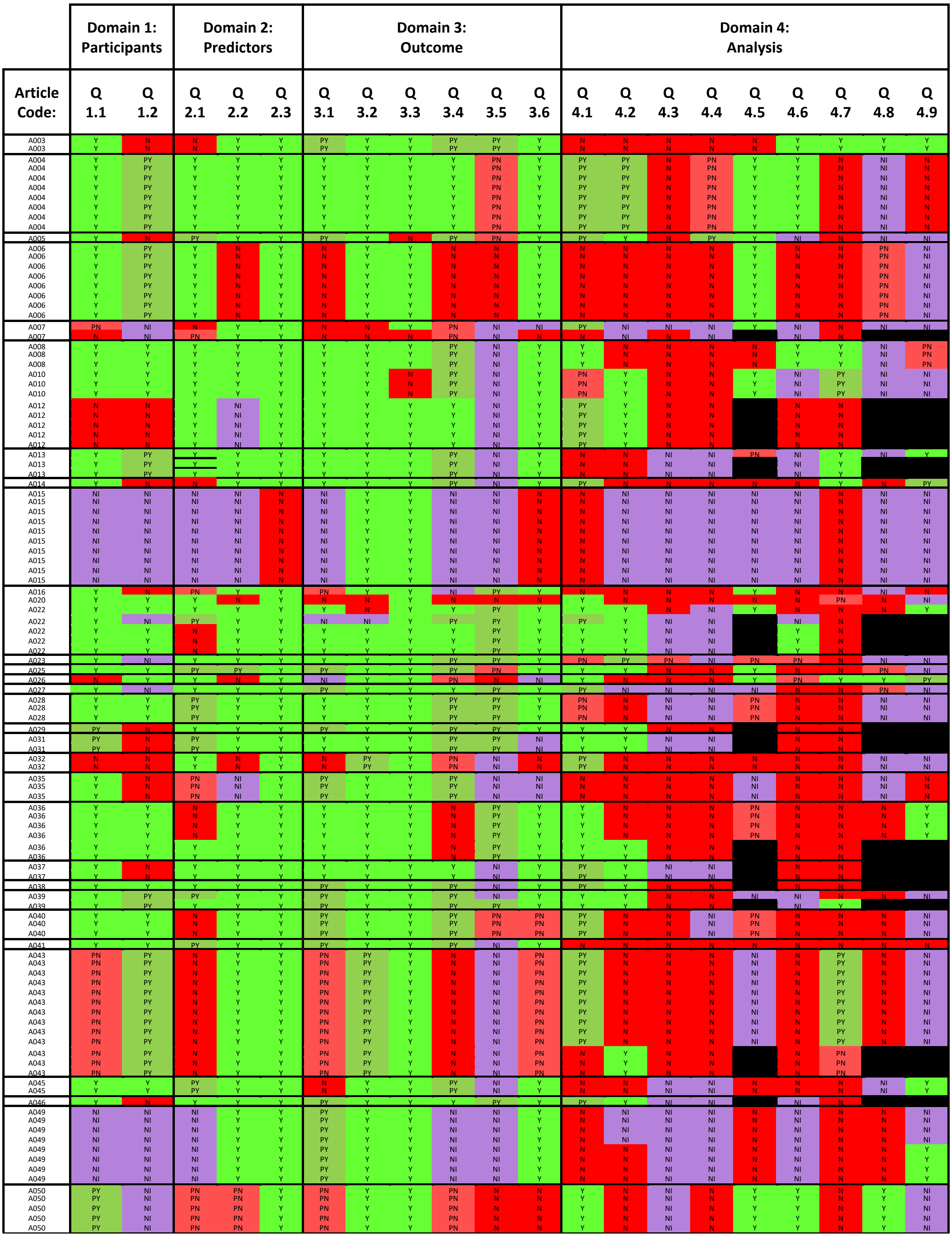
3845

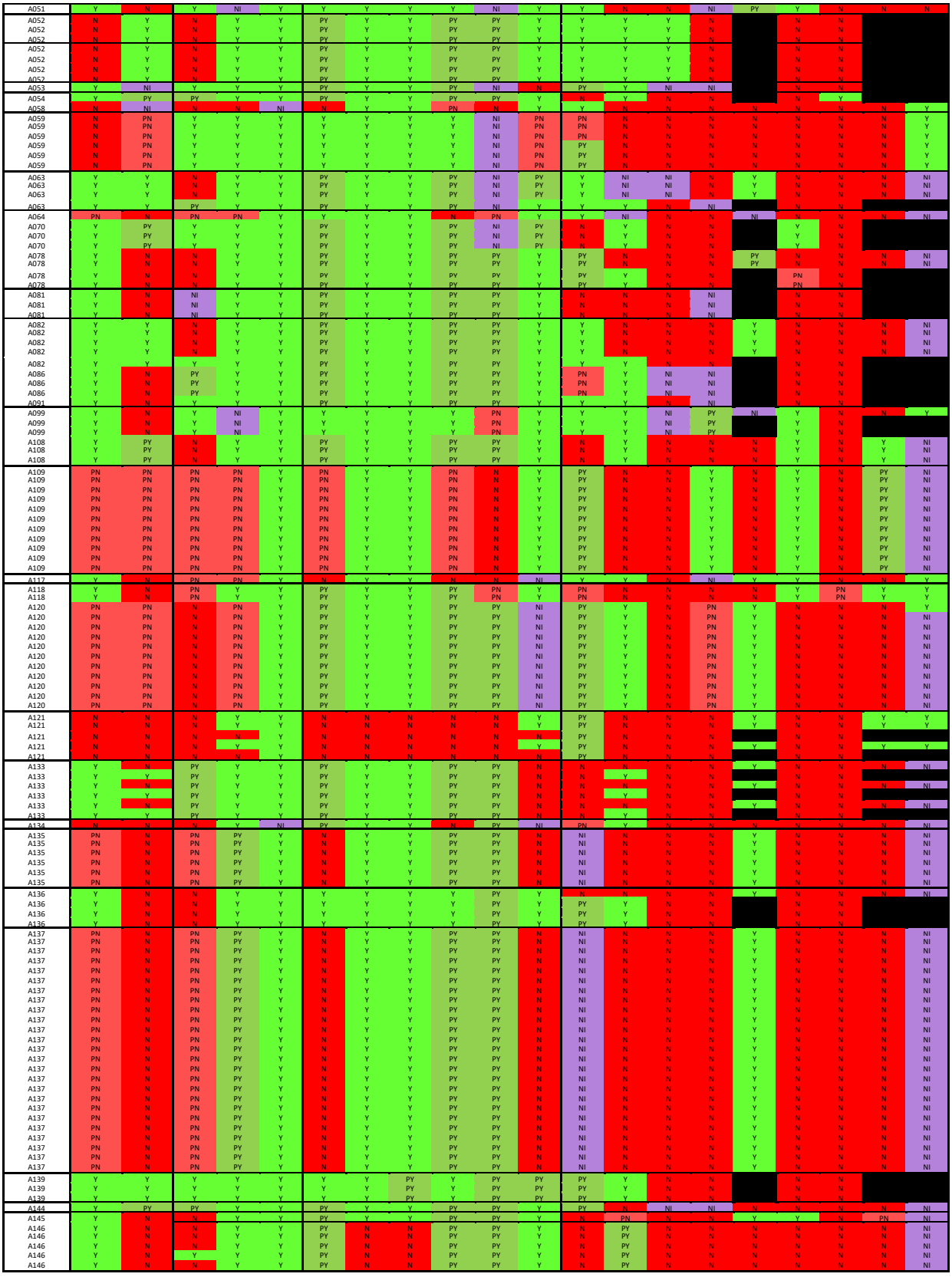
Figure S2.4 – Heat map of individual screening question ratings

Each model on separate row. All data collected about each study and model is available via the following link (tinyurl.com/442f4bdz).

Key: Colour of question responses

No:	Probably no:	No information	Probably Yes:	Yes:	Not applicable:
N	PN	NI	PY	Y	





3849

3850 Table S2.8 – Individual question risk of bias ratings by type of methodology used

3851 Grouped as positive, negative or no information and reported for developed models by type of methodology used

	Title of domain / Screening Question	Regression methods (n=97)			Flexible machine learning methods (n=45)			Ensemble machine learning methods (n=35)		
		LOW	HIGH	UNCLEAR	LOW	HIGH	UNCLEAR	LOW	HIGH	UNCLEAR
	Overall risk of bias	n = 0, 0%	n = 97, 100%	n = 0, 0%	n = 0, 0%	n = 45, 100%	n = 0, 0%	n=0, 0%	n = 35, 100%	n = 0, 0%
		LOW	HIGH	UNCLEAR	LOW	HIGH	UNCLEAR	LOW	HIGH	UNCLEAR
Domain 1	**Participants**	n = 31, 32%	n = 49, 51%	n = 17, 17%	n = 7, 16%	n = 28, 62%	n = 10, 22%	n = 7, 20%	n = 20, 57%	n = 8, 23%
Screening questions:		Y or PY	N or PN	NI	Y or PY	N or PN	NI	Y or PY	N or PN	NI
1.1	Were appropriate data sources used, e.g., cohort, randomized controlled trial, or nested case-control study data?	n = 58, 60%	n = 33, 34%	n = 6, 6%	n = 12, 27%	n = 23, 51%	n = 10, 22%	n = 10, 28%	n = 23, 66%	n = 2, 6%
1.2	Were all inclusions and exclusions of participants appropriate?	n = 45, 47%	n = 42, 43%	n = 10, 10%	n = 7, 16%	n = 25, 55%	n = 13, 29%	n = 7, 20%	n = 24, 69%	n = 4, 11%
		LOW	HIGH	UNCLEAR	LOW	HIGH	UNCLEAR	LOW	HIGH	UNCLEAR
Domain 2	**Predictors**	n = 33, 34%	n = 49, 51%	n = 15, 15%	n = 8, 18%	n = 19, 42%	n = 18, 40%	n = 7, 20%	n = 9, 26%	n = 19, 54%
Screening questions:		Y or PY	N or PN	NI	Y or PY	N or PN	NI	Y or PY	N or PN	NI
2.1	Were predictors defined and assessed in a similar way for all participants?	n = 36, 37%	n = 55, 57%	n = 6, 6%	n = 9, 20%	n = 26, 58%	n = 10, 22%	n = 7, 20%	n = 26, 74%	n = 2, 6%
2.2	Were predictor assessments made without knowledge of outcome data?	n = 83, 86%	n = 9, 9%	n = 5, 5%	n = 20, 44%	n = 18, 40%	n = 7, 16%	n = 17, 48%	n = 15, 43%	n = 3, 9%
2.3	Are all predictors available at the time the model is intended to be used?	n = 94, 97%	n = 2, 2%	n = 1, 1%	n = 38, 85%	n = 6, 13%	n = 1, 2%	n = 33, 94%	n = 2, 6%	n = 0, 0%

	Title of domain / Screening Question	Regression methods (n=97)			Flexible machine learning methods (n=45)			Ensemble machine learning methods (n=35)		
Domain 3	Outcome	LOW	HIGH	UNCLEAR	LOW	HIGH	UNCLEAR	LOW	HIGH	UNCLEAR
		n = 45, 46%	n = 48, 50%	n = 4, 4%	n = 12, 27%	n = 32, 71%	n = 1, 2%	n = 11, 31%	n = 24, 69%	n = 0, 0%
Screening questions:		**Y or PY**	**N or PN**	**NI**	**Y or PY**	**N or PN**	**NI**	**Y or PY**	**N or PN**	**NI**
3.1	Was the outcome determined appropriately?	n = 64, 66%	n = 30, 31%	n = 3, 3%	n = 14, 31%	n = 25, 56%	n = 6, 13%	n = 11, 31%	n = 22, 63%	n = 2, 6%
3.2	Was a prespecified or standard outcome definition used?	n = 88, 91%	n = 9, 9%	n = 0, 0%	n = 43, 96%	n = 2, 4%	n = 0, 0%	n = 35, 100%	n = 0, 0%	n = 0, 0%
3.3	Were predictors excluded from the outcome definition?	n = 86, 89%	n = 11, 11%	n = 0, 0%	n = 44, 98%	n = 1, 2%	n = 0, 0%	n = 35, 100%	n = 0, 0%	n = 0, 0%
Screening questions:		**Y or PY**	**N or PN**	**NI**	**Y or PY**	**N or PN**	**NI**	**Y or PY**	**N or PN**	**NI**
3.4	Was the outcome defined and determined in a similar way for all participants?	n = 65, 67%	n = 25, 26%	n = 7, 7%	n = 19, 42%	n = 16, 36%	n = 10, 22%	n = 24, 68%	n = 9, 26%	n = 2, 6%
3.5	Was the outcome determined without knowledge of predictor information?	n = 40, 41%	n = 20, 21%	n = 37, 38%	n = 17, 38%	n = 15, 33%	n = 13, 29%	n = 19, 54%	n = 13, 37%	n = 3, 9%
3.6	Was the time interval between predictor assessment and outcome determination appropriate?	n = 59, 61%	n = 34, 35%	n = 4, 4%	n = 18, 40%	n = 20, 44%	n = 7, 16%	n = 11, 31%	n = 17, 49%	n = 7, 20%

	Title of domain / Screening Question	Regression methods (n=97)			Flexible machine learning methods (n=45)			Ensemble machine learning methods (n=35)		
Domain 4	**Analysis**	**LOW**	**HIGH**	**UNCLEAR**	**LOW**	**HIGH**	**UNCLEAR**	**LOW**	**HIGH**	**UNCLEAR**
		n = 0, 0%	n = 97, 100%	n = 0, 0%	n = 0, 0%	n = 45, 100%	n = 0, 0%	n = 0, 0%	n = 35, 100%	n = 0, 0%
Screening questions:		**Y or PY**	**N or PN**	**NI**	**Y or PY**	**N or PN**	**NI**	**Y or PY**	**N or PN**	**NI**
4.1	Were there a reasonable number of participants with the outcome?	n = 50, 52%	n = 39, 40%	n = 8, 8%	n = 18, 40%	n = 17, 38%	n = 10, 22%	n = 16, 46%	n = 6, 17%	n = 13, 37%
4.2	Were continuous and categorical predictors handled appropriately?	n = 21, 22%	n = 72, 74%	n = 4, 4%	n = 7, 16%	n = 27, 60%	n = 11, 24%	n = 10, 28%	n = 23, 66%	n = 2, 6%
4.3	Were all enrolled participants included in the analysis?	n = 0, 0%	n = 82, 85%	n = 15, 15%	n = 0, 0%	n = 31, 69%	n = 14, 31%	n = 0, 0%	n = 31, 89%	n = 4, 11%
4.4	Were participants with missing data handled appropriately?	n = 2, 2%	n = 75, 77%	n = 20, 21%	n = 7, 15%	n = 26, 58%	n = 12, 27%	n = 4, 11%	n = 29, 83%	n = 2, 6%
4.5	Was selection of predictors based on univariable analysis avoided?	n = 36, 37%	n = 44, 45%	n = 17, 18%	n = 24, 53%	n = 14, 31%	n = 7, 16%	n = 28, 80%	n = 4, 11%	n = 3, 9%
4.7	Were relevant model performance measures evaluated appropriately?	n = 21, 22%	n = 76, 78%	n = 0, 0%	n = 0, 0%	n = 45, 100%	n = 0, 0%	n = 0, 0%	n = 35, 100%	n = 0, 0%
4.8	Were model overfitting and optimism in model performance accounted for?	n = 13, 13%	n = 63, 65%	n = 21, 22%	n = 8, 18%	n = 26, 58%	n = 11, 24%	n = 6, 17%	n = 22, 63%	n = 7, 20%
4.9	Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?	n = 31, 32%	n = 10, 10%	n = 56, 58%	n = 0, 0%	n = 2, 4%	n = 43, 96%	n = 0, 0%	n = 5, 14%	n = 30, 86%
Sensitivity analysis:		**LOW**	**HIGH**	**UNCLEAR**	**LOW**	**HIGH**	**UNCLEAR**	**LOW**	**HIGH**	**UNCLEAR**
Risk of bias rating not including analysis domain:		*n = 13, 13%*	*n = 80, 83%*	*n = 4, 4%*	*n = 2, 4%*	*n = 43, 96%*	*n = 0, 0%*	*n = 4, 11%*	*n = 31, 89%*	*n = 0, 0%*

3852 Please note: 3 models assessed did not report method, and are thus not included in this table

3853 Table S2.9 – Individual screening question ratings for the top ranking

3854 developed and validated model from all articles

3855 This was conducted as a sensitivity analysis to assess whether articles with a large number of developed

3856 models had a disproportionate effect on the results.

		Developed models (n=50) n, %					Validated models (n=27) n, %				
	Question	**Y**	**PY**	**PN**	**N**	**NI**	**Y**	**PY**	**PN**	**N**	**NI**
1.1	Were appropriate data sources used, e.g., cohort, randomized controlled trial, or nested case control study data?	n = 34, 68%	n = 1, 2%	n = 7, 14%	n = 6, 12%	n = 2, 4%	n = 20, 74%	n = 2, 7%	n = 1, 4%	n = 4, 15%	n = 0, 0%
1.2	Were all inclusions and exclusions of participants appropriate?	n = 13, 26%	n = 7, 14%	n = 3, 6%	n = 20, 40%	n = 7, 14%	n = 7, 26%	n = 5, 19%	n = 0, 0%	n = 12, 44%	n = 3, 11%
2.1	Were predictors defined and assessed in a similar way for all participants?	n = 14, 28%	n = 8, 16%	n = 9, 18%	n = 17, 34%	n = 2, 4%	n = 14, 52%	n = 6, 22%	n = 1, 4%	n = 5, 18%	n = 1, 4%
2.2	Were predictor assessments made without knowledge of outcome data?	n = 33, 66%	n = 3, 6%	n = 5, 10%	n = 5, 10%	n = 4, 8%	n = 24, 89%	n = 0, 0%	n = 0, 0%	n = 1, 4%	n = 2, 7%
2.3	Are all predictors available at the time the model is intended to be used?	n = 47, 94%	n = 0, 0%	n = 0, 0%	n = 1, 2%	n = 2, 4%	n = 27, 100%	n = 0, 0%	n = 0, 0%	n = 0, 0%	n = 0, 0%
3.1	Was the outcome determined appropriately?	n = 14, 28%	n = 20, 40%	n = 4, 8%	n = 10, 20%	n = 2, 4%	n = 8, 30%	n = 15, 55%	n = 1, 4%	n = 2, 7%	n = 1, 4%
3.2	Was a prespecified or standard outcome definition used?	n = 43, 86%	n = 2, 4%	n = 0, 0%	n = 5, 10%	n = 0, 0%	n = 23, 85%	n = 1, 4%	n = 0, 0%	n = 2, 7%	n = 1, 4%
3.3	Were predictors excluded from the outcome definition?	n = 46, 92%	n = 0, 0%	n = 0, 0%	n = 4, 8%	n = 0, 0%	n = 24, 89%	n = 1, 4%	n = 0, 0%	n = 2, 7%	n = 0, 0%
3.4	Was the outcome defined and determined in a similar way for all participants?	n = 7, 14%	n = 26, 52%	n = 6, 12%	n = 8, 16%	n = 3, 6%	n = 6, 22%	n = 17, 63%	n = 1, 4%	n = 3, 11%	n = 0, 0%
3.5	Was the outcome determined without knowledge of predictor information?	n = 0, 0%	n = 19, 38%	n = 7, 14%	n = 8, 16%	n = 16, 32%	n = 0, 0%	n = 16, 59%	n = 1, 4%	n = 1, 4%	n = 9, 33%
3.6	Was the time interval between predictor assessment and outcome determination appropriate?	n = 33, 66%	n = 1, 2%	n = 3, 6%	n = 7, 14%	n = 6, 12%	n = 19, 70%	n = 2, 7%	n = 1, 4%	n = 4, 15%	n = 1, 4%

	Question	Developed models (n=50) n, %					Validated models (n=27) n, %				
		Y	PY	PN	N	NI	Y	PY	PN	N	NI
4.1	Were there a reasonable number of participants with the outcome?	n = 15, 30%	n = 13, 26%	n = 6, 12%	n = 14, 28%	n = 2, 4%	n = 9, 33%	n = 10, 37%	n = 1, 4%	n = 7, 26%	n = 0, 0%
4.2	Were continuous and categorical predictors handled appropriately?	n = 10, 20%	n = 3, 6%	n = 1, 2%	n = 30, 60%	n = 6, 12%	n = 23, 85%	n = 0, 0%	n = 0, 0%	n = 3, 11%	n = 1, 4%
4.3	Were all enrolled participants included in the analysis?	n = 0, 0%	n = 0, 0%	n = 1, 2%	n = 38, 76%	n = 11, 22%	n = 1, 4%	n = 0, 0%	n = 0, 0%	n = 18, 66%	n = 8, 30%
4.4	Were participants with missing data handled appropriately?	n = 1, 2%	n = 2, 4%	n = 2, 4%	n = 32, 64%	n = 13, 26%	n = 0, 0%	n = 1, 4%	n = 0, 0%	n = 15, 55%	n = 11, 41%
4.5	Was selection of predictors based on univariable analysis avoided?	n = 20, 40%	n = 2, 4%	n = 5, 10%	n = 16, 32%	n = 7, 14%					
4.6	Were complexities in the data (e.g. censoring, competing risks, sampling of control participants) accounted for appropriately?	n = 11, 22%	n = 0, 0%	n = 2, 4%	n = 30, 60%	n = 7, 14%	n = 2, 7%	n = 0, 0%	n = 1, 4%	n = 19, 70%	n = 5, 19%
4.7	Were relevant model performance measures evaluated appropriately?	n = 5, 10%	n = 2, 4%	n = 2, 4%	n = 41, 82%	n = 0, 0%	n = 3, 11%	n = 0, 0%	n = 1, 4%	n = 23, 85%	n = 0, 0%
4.8	Were model overfitting and optimism in model performance accounted for?	n = 6, 12%	n = 1, 2%	n = 4, 8%	n = 26, 52%	n = 13, 26%					
4.9	Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?	n = 12, 24%	n = 2, 4%	n = 1, 2%	n = 5, 10%	n = 30, 60%					

3857

Table S2.10 – Individual screening questions for top ranked models in articles

Grouped ratings for the top ranking developed and validated model from all articles, conducted as a sensitivity analysis to assess whether articles with a large number of developed models had a disproportionate effect on the results.

	Title of domain / screening question	Developed models (n=50) n, %			Validated models (n=27) n, %		
		LOW	**HIGH**	**UNCLEAR**	**LOW**	**HIGH**	**UNCLEAR**
Overall risk of bias		n = 0, 0%	n = 50, 100%	n = 0, 0%	n = 0, 0%	n = 27, 100%	n = 0, 0%
		LOW	**HIGH**	**UNCLEAR**	**LOW**	**HIGH**	**UNCLEAR**
Domain 1	**Participants**	n = 17, 34%	n = 28, 56%	n = 5, 10%	n = 10, 37%	n = 15, 56%	n = 2, 7%
Screening questions:		**Yes / probably yes**	**No / probably no**	**No information**	**Yes / probably yes**	**No / probably no**	**No information**
1.1	Were appropriate data sources used, e.g., cohort, randomized controlled trial, or nested case-control study data?	n = 35, 70%	n = 13, 26%	n = 2, 4%	n = 22, 81%	n = 5, 19%	n = 0, 0%
1.2	Were all inclusions and exclusions of participants appropriate?	n = 20, 40%	n = 23, 46%	n = 7, 14%	n = 12, 44%	n = 12, 44%	n = 3, 12%
		LOW	**HIGH**	**UNCLEAR**	**LOW**	**HIGH**	**UNCLEAR**
Domain 2	**Predictors**	n = 20, 40%	n = 22, 44%	n = 8, 16%	n = 21, 78%	n = 5, 18%	n = 1, 4%
Screening questions:		**Yes / probably yes**	**No / probably no**	**No information**	**Yes / probably yes**	**No / probably no**	**No information**
2.1	Were predictors defined and assessed in a similar way for all participants?	n = 22, 44%	n = 26, 52%	n = 2, 4%	n = 20, 74%	n = 6, 22%	n = 1, 4%
2.2	Were predictor assessments made without knowledge of outcome data?	n = 36, 72%	n = 10, 20%	n = 4, 8%	n = 24, 89%	n = 1, 4%	n = 2, 7%
2.3	Are all predictors available at the time the model is intended to be used?	n = 47, 94%	n = 1, 2%	n = 2, 4%	n = 27, 100%	n = 0, 0%	n = 0, 0%
		LOW	**HIGH**	**UNCLEAR**	**LOW**	**HIGH**	**UNCLEAR**
Domain 3	**Outcome**	n = 27, 54%	n = 20, 40%	n = 3, 6%	n = 19, 70%	n = 5, 19%	n = 3, 11%
Screening questions:		**Yes / probably yes**	**No / probably no**	**No information**	**Yes / probably yes**	**No / probably no**	**No information**
3.1	Was the outcome determined appropriately?	n = 34, 68%	n = 14, 28%	n = 2, 4%	n = 23, 85%	n = 3, 11%	n = 1, 4%
3.2	Was a prespecified or standard outcome definition used?	n = 45, 90%	n = 5, 10%	n = 0, 0%	n = 24, 89%	n = 2, 7%	n = 1, 4%
3.3	Were predictors excluded from the outcome definition?	n = 46, 92%	n = 4, 8%	n = 0, 0%	n = 25, 93%	n = 2, 7%	n = 0, 0%

	Title of domain / screening question	Developed models (n=50) n, %			Validated models (n=27) n, %		
Domain 3	**Outcome**	**Yes / probably yes**	**No / probably no**	**No information**	**Yes / probably yes**	**No / probably no**	**No information**
3.4	Was the outcome defined and determined in a similar way for all participants?	n = 33, 66%	n = 14, 28%	n = 3, 6%	n = 23, 85%	n = 4, 15%	n = 0, 0%
3.5	Was the outcome determined without knowledge of predictor information?	n = 19, 38%	n = 15, 30%	n = 16, 32%	n = 16, 59%	n = 2, 8%	n = 9, 33%
3.6	Was the time interval between predictor assessment and outcome determination appropriate?	n = 34, 68%	n = 10, 20%	n = 6, 12%	n = 21, 78%	n = 5, 18%	n = 1, 4%
		LOW	**HIGH**	**UNCLEAR**	**LOW**	**HIGH**	**UNCLEAR**
Domain 4	**Analysis**	n = 0, 0%	n = 50, 100%	n = 0, 0%	n = 0, 0%	n = 27, 100%	n = 0, 0%
Screening questions:		**Yes / probably yes**	**No / probably no**	**No information**	**Yes / probably yes**	**No / probably no**	**No information**
4.1	Were there a reasonable number of participants with the outcome?	n = 28, 56%	n = 20, 40%	n = 2, 4%	n = 19, 70%	n = 8, 30%	n = 0, 0%
4.2	Were continuous and categorical predictors handled appropriately?	n = 13, 26%	n = 31, 62%	n = 6, 12%	n = 23, 85%	n = 3, 11%	n = 1, 4%
4.3	Were all enrolled participants included in the analysis?	n = 0, 0%	n = 39, 78%	n = 11, 22%	n = 1, 4%	n = 18, 67%	n = 8, 29%
4.4	Were participants with missing data handled appropriately?	n = 3, 6%	n = 34, 68%	n = 13, 26%	n = 1, 4%	n = 15, 55%	n = 11, 41%
4.5	Was selection of predictors based on univariable analysis avoided?	n = 22, 44%	n = 21, 42%	n = 7, 14%			
4.6	Were complexities in the data (e.g. censoring, competing risks, sampling of control participants) accounted for appropriately?	n = 11, 22%	n = 32, 64%	n = 7, 14%	n = 2, 7%	n = 20, 74%	n = 5, 19%
4.7	Were relevant model performance measures evaluated appropriately?	n = 7, 14%	n = 43, 86%	n = 0, 0%	n = 3, 11%	n = 24, 89%	n = 0, 0%
4.8	Were model overfitting and optimism in model performance accounted for?	n = 7, 14%	n = 30, 60%	n = 13, 26%			
4.9	Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?	n = 14, 28%	n = 6, 12%	n = 30, 60%			
Sensitivity analysis:		**LOW**	**HIGH**	**UNCLEAR**	**LOW**	**HIGH**	**UNCLEAR**
Risk of bias rating not including analysis domain:		n = 8, 16%	n = 40, 80%	n = 2, 4%	n = 8, 29%	n = 1, 4%	n = 18, 67%

3862

Supplementary Checklists 2.1

Checklist 1: TRIPOD-SRMA checklist for reporting systematic reviews of prediction model studies

Section and topic	Item No	Checklist item	Page of Thesis
Title			
Title	1	Identify the report as a systematic review or meta-analysis (or both) of diagnostic or prognostic model studies. Specify the target population and outcome(s) predicted as relevant to the review question.	Page 24
Abstract			
Abstract	2	See the TRIPOD-SRMA checklist for abstracts	Page 25 (checklist below)
Introduction			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Page 27
Objectives	4	Provide an explicit statement of the objective(s) being addressed with reference to: target population, index and comparator models (as relevant), outcome(s), time (prediction horizon and intended moment of using the model), and setting.	Page 27
Methods			
Study eligibility criteria	5	Specify study characteristics used as eligibility criteria, including any prediction models of specific interest, and whether development or validation studies (or both) were eligible.	Page 28
Information sources	6	Specify all databases, registers, websites, organisations, reference lists, and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Page 28
Search strategy	7	Present the full search strategies for all databases, registers, and websites, including any filters and limits used.	Page 28 / Table S2.1
Study selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they	Page 28

Section and topic	Item No	Checklist item	Page of Thesis
		worked independently, and if applicable, details of automation tools used in the process.	
Data collection process	9	Specify the methods used to collect data from study reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Page 29
Data Items	10a	List and define all items for which data were sought from each study.	Page 29/Table S2.2/ Table S2.3
	10b	State the model performance measures that were sought (eg, measures of calibration, discrimination, overall model fit, clinical utility).	Page 28
	10c	Describe how any desired but unreported data items (items 10a, 10b) were handled (eg, contacted authors, calculated from other reported information).	Page 28
Risk of bias and applicability assessment	11	Specify the methods used to assess risk of bias in the included studies and their applicability to the review question. This should be done separately for each model development and validation. Include details of any tool(s) used, how many reviewers assessed each study and whether they worked independently.	Page 27 to 28
Synthesis methods	12a	Describe any methods for synthesising estimates of performance measures for each model. If meta-analysis was carried out, describe the methods used, including any transformations of data before pooling, how any heterogeneity in model performance was quantified and handled, and software package(s) used.	Page 28
	12b	Describe any methods used to explore possible causes of heterogeneity in model performance (eg, subgroup analysis, meta-regression), including whether or not they were planned.	Page 27
	12c	Describe any sensitivity analyses conducted to assess robustness of the synthesised results.	Page 31
Certainty assessment	13	Describe any methods used to assess certainty (or confidence) in the body of evidence for a prediction model.	Page 27

Section and topic	Item No	Checklist item	Page of Thesis
Results			
Study selection	14	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies and models included in the review, ideally using a flow diagram.	Page 29/Figure 2.1/ Table S2.4
Study and model characteristics	15	Present study characteristics and model details extracted (as per item 10a), and cite the study reports.	Page 29 / Figure 2.1
Risk of bias and applicability	16	Present results of risk of bias and applicability assessment. This should be done separately for each model development and validation in each included study.	Table 2.2 / Figure 2.2 / Table S2.6 / Table S2.7 / Figure S2.4 / Table S2.8 / Table S2.9 / Table S2.10
Results of model performance in individual studies	17	Present performance estimates and confidence intervals for each model and all evaluations, including whether they relate to the internal or external validation performance. If internal, give details of the method.	Figure S2.4
Results of syntheses	18a	Present the results of any synthesis of model performance, together with details of which study estimates contributed. If meta-analysis was carried out, then for each model and performance measure, present summary results, confidence/credible intervals, and measures of heterogeneity. Forest plots may be useful.	Table S2.7
	18b	For each model, present results of all investigations of possible causes of heterogeneity in model performance.	Table S2.7
	18c	Present results of all sensitivity analyses conducted to assess the robustness of the synthesised results.	Page 31 / Table 2.2
Certainty of evidence	19	Present any assessments of certainty (or confidence) in the body of evidence for each prediction model of interest.	Page 30
Discussion			
Summary of evidence	20	Summarise the main findings including the strengths and limitations of the evidence.	Page 32 / Page 33
Limitations	21	Discuss the strengths and limitations of the review process.	Page 33

Section and topic	Item No	Checklist item	Page of Thesis
Implications	22	Discuss implications of the results in the context of other evidence and for practice, policy, and future research.	Page 32 / Page 33
Other information			
Registration and protocol	23a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Page 28
	23b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Page 28
	23c	Describe and explain any amendments to information provided at registration or in the protocol.	N/A Reference 24 – protocol not changed
Support	24	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Page 24
Competing interests	25	Declare any competing interests of review authors.	Page 24
Availability of data, code, and other materials	26	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Page 37

3866

3867 Checklist 2: TRIPOD-SRMA checklist for abstracts

Section and topic	Item No	Checklist item	Present in Chapter 2 abstract
Abstract Title			
Title	1	Identify the report as a systematic review or meta-analysis (or both) of diagnostic or prognostic model studies. Specify the target population and outcome(s) predicted as relevant to the review question.	✓
Abstract Background			
Objectives	2	Provide an explicit statement of the main objective(s) being addressed with reference to: target population, index and comparator models (as relevant), outcome(s), time (prediction horizon and intended moment of using the model), and setting.	✓
Abstract Methods			
Study eligibility criteria	3	Specify study characteristics used as eligibility criteria, including any prediction models of specific interest, and whether development or validation studies (or both) were eligible.	✓
Information sources	4	Specify the information sources (eg, databases, registers) used to identify studies and the date when each was last searched.	✓
Risk of bias and applicability	5	Specify the methods used to assess risk of bias and applicability in the included studies.	✓
Synthesis methods	6	Specify the methods used to synthesise performance measures for each model of interest.	✓

Section and topic	Item No	Checklist item	Present in Chapter 2 abstract
Abstract Results			
Included studies	7	Give the total number of included studies and models, and summarise relevant study characteristics and model details.	✓
Results of syntheses	8	Present results for each of the main models of interest. If meta-analysis was used to synthesise study estimates of model performance, report the summary result and confidence/credible interval for each performance measure, together with the number of study estimates contributing.	✓
Abstract Discussion			
Limitations of evidence	9	Provide a brief summary of the limitations of the evidence included in the review.	✓
Interpretation	10	Provide a general interpretation of the results and important implications for research and practice.	✓

3868

A3 Supplementary material for Chapter 3

Figure S3.1 – Satisfaction with services and frequency of hospital assessments, alongside clinician access to multidisciplinary resources

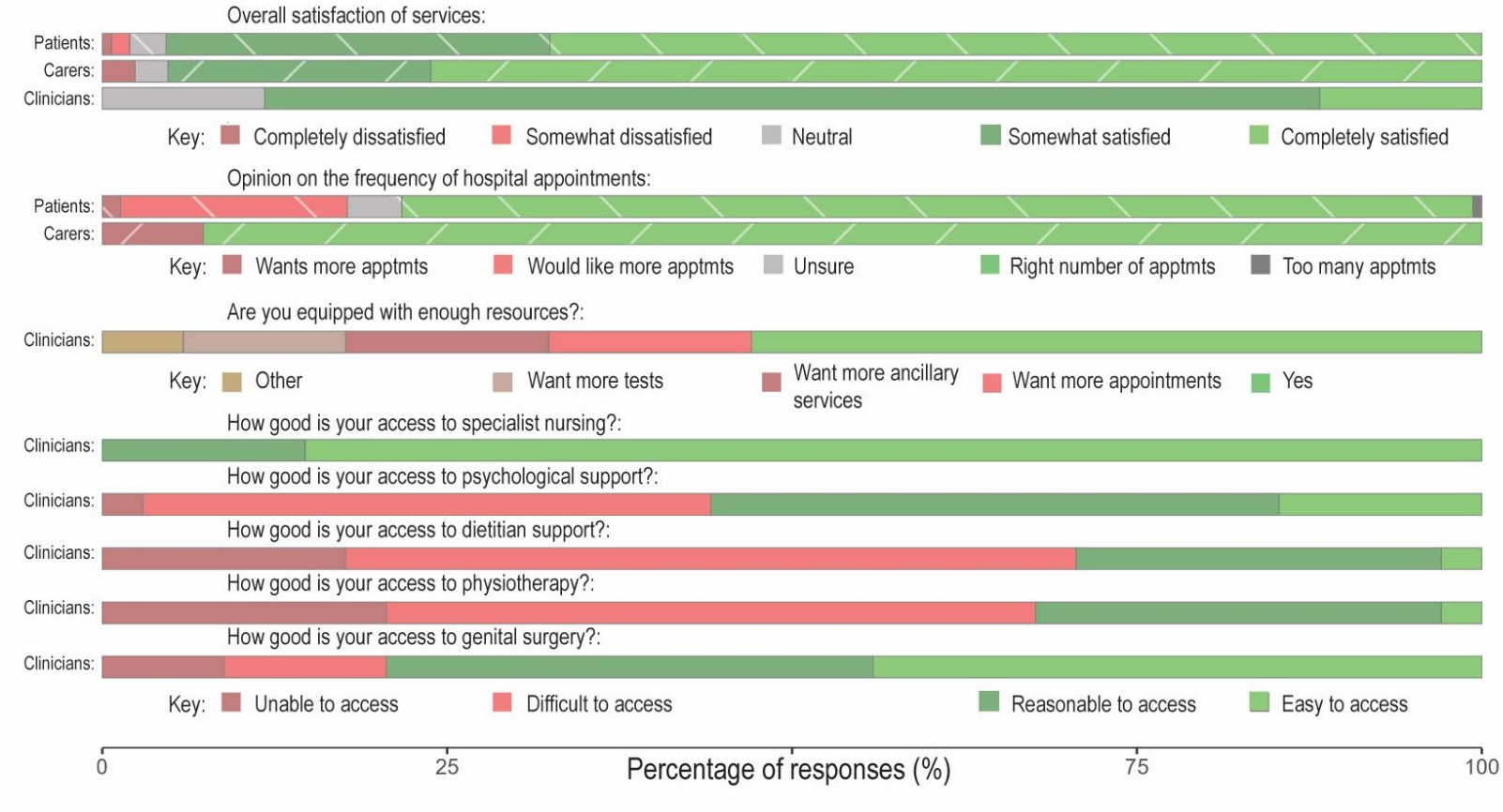


Figure S3.2 – Report of attendance at formal education, and desire to access a formal education course about CAH

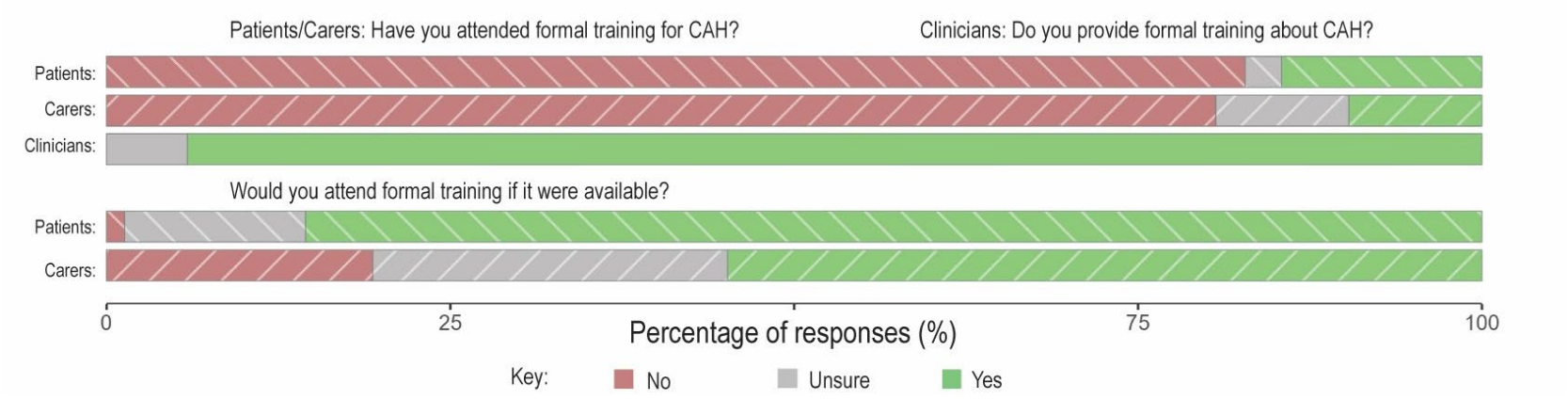


Table S3.1 – All statistical software packages used within *R: A language and environment for statistical computing.*

Our thanks to all authors, who without these packages this work would have not been possible.

Name of software package	**Citation**
Base R	https://www.R-project.org/
dplyr	https://github.com/tidyverse/dplyr
ggplot2	https://ggplot2.tidyverse.org
ggpattern	https://coolbutuseless.github.io/package/ggpattern/
DescTools	https://cran.r-project.org/package=DescTools
summarytools	https://github.com/dcomtois/summarytools

Supplementary weighting methods 3.1:

To account for the distribution of the number of responses per centre being unequal, a sensitivity analysis was conducted to weight responses in relation to the number of responses that would be expected for each hospital as a proportion of its catchment population. The trust catchment population was taken from the UK Office for Health Improvement & Disparities (https://tinyurl.com/ar7saj5e), and from the Scottish Health and Social Care Open Data website (https://www.opendata.nhs.scot/dataset/population-estimates). The underlying assumption was that data missing from hospitals that did not take part in the service evaluation, or did not receive any responses, was that this data was missing completely at random. The target of weighting was to prevent any hospitals with a disproportionately large number of responses from having an undue influence on any particular category of response.

Patient and carer response weighting

The weighting for each question response was calculated for each question, to account for a variable number of missing responses across different questions. The total catchment population surveyed was calculated as the sum of the catchment areas of all hospitals that featured in response to the question. The proportion of the catchment population of that hospital was multiplied by the inverse proportion of the responses from that hospital to create an individual weighting for responses pertaining to that hospital. Responses for the small number of patients or carers that did not enter a centre (1 and 2 respectively) were not included in weighted response percentages.

Example: If a hospital served 5% of the catchment population of all of the hospitals that had an answer to that particular question, and the total number of answers received pertaining to that hospital was 5% of the total number of answers, the weight for each of those answers would be (5/100 x 100/5) = 1, so all answers for that hospital would be multiplied by 1 because the proportion of answers was appropriate.

If instead the total number of answers pertaining to a hospital was 20% of the total number of answers, but the hospital only served 5% of the catchment population pertaining to the answers to that question, the weight for each answer would be (5/100 x 100/20) = 0.25. All answers from that hospital would be multiple by 0.25, because they effectively have four times as many answers as they would do if the proportion of answers matched the appropriate proportion of catchment population.

Clinician weighting

Clinician responses were weighted to equilibrate responses to the equivalent of 1 clinician per centre.

Example: If 3 clinicians from a centre submitted responses to the survey, their responses multiplied by 1/3.

After responses to each question were weighted, percentage of weighted responses was calculated and is reported in supplementary table 2 below, alongside the raw percentage and raw proportion of responses for each question for comparison.

3907 Table S3.2 – Responses to all questions across questionnaires, alongside response percentages adjusted for weighting.

3908
3909

	Patients **Percent of all responses (n)** **[*weighted percentage of responses*]**	**Parents** **Percent of all responses (n)** **[*weighted percentage of responses*]**	**Clinicians** **Percent of all responses (n)** **[*weighted percentage of responses*]**
Demographics			
Number of respondents	43	152	34
Number of centres	27 (1 unknown)	34 (2 unknown)	18 (0 unknown)
How old is the patient who attended the most recent appointment for CAH?	<12 months: 0% (0/43) [*0%*] 1-5 years: 0% (0/43) [*0%*] 6-9 years: 27.9% (12/43) [*26.7%*] 10-14 years: 39.5% (17/43) [*38.7%*] 15-19 years: 32.6% (14/43) [*34.5%*]	<12 months: 9.2% (14/152) [*8.5%*] 1-5 years: 29.6% (45/152) [*29.5%*] 6-9 years: 25.6% (39/152) [*21.5%*] 10-14 years: 26.3% (40/152) [*23.0%*] 15-19 years: 9.2% (14/152) [*17.5%*]	-
How old are you? (Clinicians: What is your job role?)	-	<20 years: 0.7% (1/151) [*0.5%*] 20-29 years: 10.6% (16/151) [*7.2%*] 30-39 years: 39.1% (59/151) [*33.0%*] 40-49 years: 39.7% (60/151) [*50.4%*] >50 years: 9.9% (15/151) [*8.9%*]	Consultant Endocrinologist: 70.6% (24/34) [*74.1%*] Consultant SPIN: 11.8% (4/34) [*13.0%*] Registrar: 5.9% (2/34) [*6.5%*] Nurse: 11.8% (4/34) [*6.5%*]
How would you describe the sex of the patient?	Male: 44.2% (19/43) [*41.3%*] Female: 55.8% (24/43) [*58.7%*] Prefer not to say 0.0% (0/43) [*0.0%*]	Male: 46.7% (71/152) [*42.0%*] Female: 52.6% (80/152) [*57.8%*] Prefer not to say 0.7% (1/152) [*0.2%*]	-
How would you describe the ethnic background of the patient?	Asian: 25.8% (8/31) [*29.3%*] Black: 3.2% (1/31) [*5.9%*] Mixed: 9.7% (3/31) [*12.3%*] Other: 3.2% (1/31) [*1.6%*] White: 58.1% (18/31) [*51.0%*]	Asian: 13.2% (20/152) [*11.7%*] Black: 1.3% (2/152) [*0.9%*] Mixed: 5.3% (8/152) [*5.4%*] Other: 2.0% (3/152) [*4.7%*] White: 78.3% (119/152) [*77.2%*]	-

Time treated / practicing			
How long have you been treated at this hospital for CAH? (Clinicians: How long have you treating children with CAH?)	1 visit: 3.2% (1/31) [*3.6%*] < 12 months: 0.0% (0/31) [*0.0%*] 1-2 years: 6.5% (2/31) [*7.4%*] 3-5 years: 12.9% (4/31) [*14.8%*] 6-10 years: 12.9% (4/31) [*15.0%*] >10 years: 64.5% (20/31) [*59.2%*]	1 visit: 2.0% (3/152) [*2.7%*] < 12 months: 11.2% (17/152) [*12.3%*] 1-2 years: 15.1% (23/152) [*12.5%*] 3-5 years: 24.3% (37/152) [*22.3%*] 6-10 years: 27.0% (41/152) [*23.9%*] >10 years: 20.4% (31/152) [*26.2%*]	1-2 years: 5.9% (2/34) [*3.7%*] 3-5 years: 14.7% (5/34) [*11.1%*] 6-10 years: 23.5% (8/34) [*14.8%*] >10 years: 55.9% (19/34) [*70.4%*]
In the last 12 months, how many appointments have you had at this hospital for CAH? (Clinicians: How many patients with CAH have you treated in the last 12 months?)	1: 16.1% (5/31) [*15.6%*] 2: 38.7% (12/31) [*33.7%*] 3: 25.8% (8/31) [*32.0%*] 4: 6.5% (2/31) [*7.8%*] >4: 12.9% (4/31) [*10.9%*]	1: 12.6% (19/152) [*15.0%*] 2: 27.8% (42/152) [*31.5%*] 3: 24.5% (37/152) [*22.8%*] 4: 18.5% (28/152) [*14.6%*] >4: 16.6% (25/152) [*16.0%*]	1: 2.9% (1/33) [*5.6%*] 2-5: 17.6% (6/33) [*8.3%*] 6-10: 29.4% (10/33) [*31.5%*] 11-20: 26.5% (9/33) [*24.1%*] 21-50: 11.8% (4/33) [*15.7%*] >50: 11.8% (4/33) [*14.8%*]
Satisfaction with services			
Overall, are you satisfied with the service provided to care for your child with CAH? (Clinicians: Are you satisfied with the service you provide?)	Completely dissatisfied: 2.4% (1/42) [*1.2%*] Somewhat dissatisfied: 0.0% (0/42) [*0.0%*] Neutral: 2.4% (1/42) [*2.3%*] Somewhat satisfied: 19.0% (8/42) [*19.1%*] Completely satisfied: 76.2% (32/42) [*77.4%*]	Completely dissatisfied: 0.7% (1/151) [*0.0%*] Somewhat dissatisfied: 1.3% (2/151) [*0.8%*] Neutral: 2.6% (4/151) [*4.6%*] Somewhat satisfied: 27.8% (42/151) [*35.4%*] Completely satisfied: 67.5% (102/151) [*59.3%*]	Completely dissatisfied: 0.0% (0/34) [*0.0%*] Somewhat dissatisfied: 0.0% (0/34) [*0.0%*] Neutral: 11.8% (4/34) [*22.2%*] Somewhat satisfied: 76.5% (26/34) [*69.4%*] Completely satisfied: 11.8% (4/34) [*8.3%*]
Discussions in clinic			
At these appointments in the last 12 months, have you had the opportunity to discuss:			(Clinicians: How often do you discuss the following topics:)
Ideas and goals about management	Not sure: 3.2% (1/31) [*4.7%*] No, I didn't want to: 0.0% (0/31) [*0.0%*] No, I would have liked to: 9.7% (3/31) [*12.3%*] Yes, but not enough: 6.5% (2/31) [*3.1%*] Yes, thoroughly: 80.6% (25/31) [*79.8%*]	Not sure: 4.1% (6/147) [*2.0%*] No, I didn't want to: 2.0% (3/147) [*0.7%*] No, I would have liked to: 3.4% (5/147) [*1.2%*] Yes, but not enough: 19.0% (28/147) [*26.3%*] Yes, thoroughly: 71.4% (105/147) [*69.8%*]	Never: 2.9% (1/18) [*0.9%*] Sometimes: 20.6% (7/18) [*13.9%*] Often: 41.2% (14/18) [*44.4%*] Always: 35.3% (12/18) [*40.7%*]
How to increase steroid doses when unwell	Not sure: 12.5% (3/24) [*6.4%*] No, I didn't want to: 8.3% (2/24) [*12.3%*] No, I would have liked to: 0.0% (0/24) [*0.0%*] Yes, but not enough: 4.2% (1/24) [*4.2%*] Yes, thoroughly: 75.0% (18/24) [*77.0%*]	Not sure: 0.7% (1/147) [*0.6%*] No, I didn't want to: 4.1% (6/147) [*5.5%*] No, I would have liked to: 0.7% (1/147) [*0.3%*] Yes, but not enough: 6.8% (10/147) [*7.3%*] Yes, thoroughly: 87.8% (129/147) [*86.3%*]	Never: 0.0% (0/18) [*0.0%*] Sometimes: 0.0% (0/18) [*0.0%*] Often: 11.8% (4/18) [*11.1%*] Always: 88.2% (30/18) [*88.9%*]
Psychological effects	Not sure: 16.7% (2/12) [*19.5%*] No, I didn't want to: 8.3% (1/12) [*8.4%*] No, I would have liked to: 16.7% (2/12) [*16.1%*] Yes, but not enough: 41.7% (5/12) [*38.5%*] Yes, thoroughly: 16.7% (2/12) [*17.6%*]	Not sure: 5.5% (8/146) [*2.8%*] No, I didn't want to: 15.8% (23/146) [*16.5%*] No, I would have liked to: 27.4% (40/146) [*29.1%*] Yes, but not enough: 20.5% (30/146) [*30.4%*] Yes, thoroughly: 30.8% (45/146) [*21.2%*]	Never: 2.9% (1/34) [*0.9%*] Sometimes: 70.6% (24/34) [*71.3%*] Often: 26.5% (9/34) [*27.8%*] Always: 0.0% (0/34) [*0.0%*]
Levels of physical activity	Not sure: 8.3% (1/12) [*11.1%*] No, I didn't want to: 8.3% (1/12) [*8.4%*] No, I would have liked to: 0.0% (0/12) [*0.0%*] Yes, but not enough: 16.7% (2/12) [*8.4%*] Yes, thoroughly: 66.7% (8/12) [*72.2%*]	Not sure: 6.1% (9/147) [*7.0%*] No, I didn't want to: 10.2% (15/147) [*11.4%*] No, I would have liked to: 20.4% (30/147) [*19.6%*] Yes, but not enough: 16.3% (24/147) [*17.0%*] Yes, thoroughly: 46.9% (69/147) [*45.0%*]	Never: 5.9% (2/34) [*6.5%*] Sometimes: 64.7% (22/34) [*62.0%*] Often: 17.6% (6/34) [*13.9%*] Always: 11.8% (4/34) [*17.6%*]

General wellbeing (Clinicians: Mental health)	Not sure: 8.3% (1/12) [*11.1%*] No, I didn't want to: 0.0% (0/12) [*0.0%*] No, I would have liked to: 0.0% (0/12) [*0.0%*] Yes, but not enough: 16.7% (2/12) [*8.4%*] Yes, thoroughly: 75.0% (9/12) [*80.5%*]	Not sure: 0.0% (0/150) [*0.0%*] No, I didn't want to: 1.3% (2/150) [*0.5%*] No, I would have liked to: 7.3% (11/150) [*4.3%*] Yes, but not enough: 8.7% (13/150) [*14.0%*] Yes, thoroughly: 82.7% (124/150) [*81.2%*]	Never: 6.1% (2/33) [*2.8%*] Rarely: 48.5% (16/33) [*42.1%*] Sometimes: 42.4% (14/33) [*49.5%*] Always: 3.0% (1/33) [*5.6%*]
Fertility	-	Not sure: 2.8% (4/145) [*2.7%*] No, I did not want to: 37.2% (54/145) [*48.2%*] No, I would have liked to: 21.4% (31/145) [*19.7%*] Yes, but not enough: 14.5% (21/145) [*11.0%*] Yes, thoroughly: 24.1% (35/145) [*18.4%*]	Never: 0.0% (0/33) [*0.0%*] Rarely: 27.3% (9/33) [*21.5%*] Sometimes: 66.7% (22/33) [*70.1%*] Always: 6.1% (2/33) [*8.4%*]
Heart disease	-	Not sure: 11.0% (16/145) [*17.0%*] No, I did not want to: 37.9% (55/145) [*38.6%*] No, I would have liked to: 33.8% (49/145) [*32.2%*] Yes, but not enough: 6.2% (9/145) [*4.0%*] Yes, thoroughly: 11.0% (16/145) [*8.2%*]	Never: 3.0% (1/33) [*0.9%*] Rarely: 45.5% (15/33) [*41.1%*] Sometimes: 42.4% (14/33) [*49.5%*] Always: 9.1% (3/33) [*8.4%*]
Genital surgery	-	Not sure: 6.4% (9/140) [*11.6%*] No, I didn't want to: 45.0% (63/140) [*42.4%*] No, I would have liked to: 15.7% (22/140) [*15.1%*] Yes, but not enough: 13.6% (19/140) [*13.1%*] Yes, thoroughly: 19.3% (27/140) [*17.8%*]	-
Genital surgery – carers of girls	-	Not sure: 1.3% (1/76) [*11.1%*] No, I didn't want to: 32.9% (25/76) [*25.7%*] No, I would have liked to: 19.7% (15/76) [*19.5%*] Yes, but not enough: 18.4% (14/76) [*18.3%*] Yes thoroughly: 27.6% (21/76) [*25.5%*]	-
Genital surgery – carers of girls < 5 (29)	-	Not sure: 0.0% (0/30) [*0.0%*] No, I didn't want to: 13.3% (4/30) [*18.4%*] No, I would have liked to: 20.0% (6/30) [*13.4%*] Yes, but not enough: 33.3% (10/30) [*31.2%*] Yes thoroughly: 33.3% (10/30) [*36.9%*]	-
Reduced adult height	-	-	Never: 0.0% (0/33) [*0.0%*] Rarely: 6.1% (2/33) [*1.9%*] Sometimes: 63.6% (21/33) [*72.9%*] Always: 30.3% (10/33) [*25.2%*]
Bone disease	-	-	Never: 9.1% (3/33) [*4.7%*] Rarely: 48.5% (16/33) [*57.9%*] Sometimes: 39.4% (13/33) [*31.8%*] Always: 3.0% (1/33) [*5.6%*]
Mental health	-	-	Never: 6.1% (2/33) [*2.8%*] Rarely: 48.5% (16/33) [*42.1%*] Sometimes: 42.4% (14/33) [*49.5%*] Always: 3.0% (1/33) [*5.6%*]
Tumours	-	-	Never: 6.1% (2/33) [*1.9%*] Rarely: 51.5% (17/33) [*61.7%*] Sometimes: 39.4% (13/33) [*33.6%*] Always: 3.0% (1/33) [*2.8%*]
During your last appointment for your child's CAH, did you discuss and agree on a plan of how to manage the CAH until your next appointment?	Not sure: 9.7% (3/31) [*10.9%*] No: 0.0% (0/31) [*0.0%*] Yes: 90.3% (28/31) [*89.1%*]	Not sure: 0.7% (1/150) [*8.2%*] No: 3.3% (5/150) [*4.0%*] Yes: 96.0% (144/150) [*87.8%*]	-

During your last appointment for your child's CAH, were you offered a printed or electronic copy of you care plan? (Clinician: Do you provide a care plan?)	Clinic Letter only: 26.3% (5/19) [*30.7%*] Not sure: 0.0% (0/19) [*0.0%*] No, I would have liked one: 36.8% (7/19) [*34.2%*] No, I didn't want one: 15.8% (3/19) [*13.7%*] Yes: 21.1% (4/19) [*21.3%*]	Clinic letter only: 42.8% (65/152) [*45.0%*] Not sure: 2.6% (4/152) [*6.1%*] No, I would have liked one: 22.4% (34/152) [*15.6%*] No, I didn't want one: 6.6% (10/152) [*5.1%*] Yes: 25.7% (39/152) [*28.2%*]	Clinic Letter only: 32.4% (11/34) [*25.0%*] Sometimes: 5.9% (2/34) [*2.8%*] Often: 20.6% (7/34) [*16.7%*] Always: 41.2% (14/34) [*55.6%*]
Advice, education, and confidence			
Do you feel confident managing your child's CAH Day to day? Day to day	No: 0.0% (0/31) [*0.0%*] Somewhat confident: 29.0% (9/31) [*30.9%*] Very confident: 71.0% (22/31) [*69.1%*]	No: 0.0% (0/151) [*0.0%*] Somewhat confident: 14.6% (22/151) [*17.4%*] Very confident: 85.4% (129/151) [*82.6%*]	-
Do you feel confident managing your child's CAH when they have another illness?	-	No: 1.5% (2/135) [*4.3%*] Somewhat confident: 42.2% (57/135) [*39.9%*] Very confident: 56.3% (76/135) [*55.7%*]	-
Do you feel confident managing your child's CAH in an emergency?	-	No: 10.1% (14/135) [*12.8%*] Somewhat confident: 49.6% (69/135) [*47.6%*] Very confident: 40.3% (56/135) [*39.6%*]	-
Do you think patients and families are confident managing CAH?	-	-	No: 0.0% (0/34) [*0.0%*] To some extent: 88.2% (30/34) [*95.4%*] Yes, definitely: 11.8% (4/34) [*4.6%*]
In some diseases like diabetes, patients can attend a formal education course to help them manage the disease. Have you attended anything like this for CAH? (Clinicians: Do you provide formal education?)	No: 80.6% (25/31) [*78.2%*] Unsure: 9.7% (3/31) [*9.3%*] Yes: 9.7% (3/31) [*12.5%*]	No: 82.8% (125/151) [*90.1%*] Unsure: 3.3% (5/151) [*1.8%*] Yes: 13.9% (21/151) [*8.1%*]	No: 0.0% (0/34) [*0.0%*] Not sure: 5.9% (2/34) [*4.6%*] Yes: 94.1% (32/34) [*95.4%*]
If a formal educational course teaching about CAH was available, would you attend?	Yes: 54.8% (17/31) [*59.8%*] No: 19.4% (6/31) [*13.7%*] Unsure: 25.8% (8/31) [*26.5%*]	Yes: 85.5% (130/152) [*80.4%*] No: 1.3% (2/152) [*1.5%*] Unsure: 13.2% (20/152) [*18.1%*]	-
Clinicians: What professionals provide the education: (Note: multiple selections allowed)	-	-	Doctor: 71.9% (23/32) [*64.1%*] Nurse: 100.0% (32/32) [*100.0%*] Youth worker: 3.1% (1/32) [*2.9%*] Pharmacist: 3.1% (1/32) [*5.8%*] Psychologist: 15.6% (5/32) [*14.6%*] Dietitian: 3.1% (1/32) [*5.8%*]
Clinicians: How long does formal education take:	-	-	<1 hour: 32.3% (10/31) [*40.2%*] 1-4 hours: 67.7% (21/31) [*59.8%*]

Clinicians: written guidance at education	-	-	No: 0.0% (0/30) [*0.0%*] Not sure: 10.0% (3/30) [*13.8%*] Yes: 90.0% (27/30) [*86.2%*]
Clinicians: do you liaise with school?	-	-	No: 5.9% (2/18) [*2.8%*] Not sure: 8.8% (3/18) [*4.6%*] Sometimes: 35.3% (12/18) [*31.5%*] Often: 2.9% (1/18) [*0.9%*] Yes: 47.1% (16/18) [*60.2%*]
How easy to contact endocrine in working hours	-	-	No: 3.0% (1/33) [*0.9%*] Not sure: 30.3% (10/33) [*26.2%*] Yes: 66.7% (22/33) [*72.9%*]
How easy outside of working hours:	-	-	Unable: 15.2% (5/33) [*23.4%*] Not easily: 30.3% (10/33) [*17.8%*] Not sure: 12.1% (4/33) [*13.1%*] Reasonably easy: 33.3% (11/33) [*38.3%*] Very easy: 9.1% (3/33) [*7.5%*]
How easy in an emergency:	-	-	Unable: 9.1% (3/33) [*16.8%*] Not easily: 21.2% (7/33) [*15.0%*] Not sure: 9.1% (3/33) [*7.5%*] Reasonably easy: 45.5% (15/33) [*45.8%*] Very easy: 15.2% (5/33) [*15.0%*]
In the last 12 months have you experienced conflicting advice from different health professionals in relation to CAH? (Clinicians: Do families get conflicting advice within the NHS?)	Yes, often: 2.4% (1/41) [*1.2%*] Yes, sometimes: 22.0% (9/41) [*20.7%*] No: 68.3% (28/41) [*73.4%*] Unsure: 7.3% (3/41) [*4.8%*]	Yes, often: 7.3% (11/151) [*8.0%*] Yes, sometimes: 19.9% (30/151) [*26.8%*] No: 72.8% (110/151) [*65.2%*] Unsure: 0.0% (0/151) [*0.0%*]	Never: 0.0% (0/18) [*0.0%*] Sometimes: 41.2% (14/18) [*35.2%*] Often: 52.9% (18/18) [*58.3%*] Always: 5.9% (2/18) [*6.5%*]
Have people you have spoken to from other families ever made you think that the advice get is different from the advice they have received? (Clinicians: Would patients be managed differently at a different hospital?)	Not spoken to anyone else: 75.6% (31/41) [*74.9%*] No, the same advice: 22.0% (9/41) [*23.3%*] Yes, somewhat different: 0.0% (0/41) [*0.0%*] Yes, very different advice: 2.4% (1/41) [*1.8%*]	Not spoken to anyone else: 54.3% (82/151) [*47.8%*] No, the same advice: 11.9% (18/151) [*11.1%*] Yes, somewhat different: 29.8% (45/151) [*35.9%*] Yes, very different advice: 4.0% (6/151) [*5.2%*]	Unlikely: 30.3% (10/33) [*26.7%*] Possible: 66.7% (22/33) [*67.6%*] Very likely: 3.0% (1/33) [*5.7%*]
Do you think you know enough about the medications that your child is prescribed to manage their CAH?	I don't take medication: 0.0% (0/42) [*0.0%*] No, I want to know more: 7.1% (3/42) [*6.2%*] Unsure: 11.9% (5/42) [*9.5%*] Yes, but want to know more: 23.8% (10/42) [*20.7%*] Yes: 57.1% (24/42) [*63.5%*]	Doesn't take medication: 0.7% (1/152) [*0.6%*] No, I want to know more: 4.6% (7/152) [*3.4%*] Unsure: 0.0% (0/152) [*0.0%*] Yes, but want to know more: 42.8% (65/152) [*47.2%*] Yes: 52.0% (79/152) [*48.8%*]	-

Do you think that a doctor at a different hospital might treat your child's CAH differently? (Clinicians: Would patients be managed differently at a different hospital?)	-	No: 22.4% (34/152) [*15.2%*] Unsure: 53.9% (82/152) [*60.4%*] Yes: 23.7% (36/152) [*24.4%*]	Unlikely: 30.3% (10/33) [*26.7%*] Possible: 66.7% (22/33) [*67.6%*] Very likely: 3.0% (1/33) [*5.7%*]
Metrics assessed in clinic			
Has your child had the following things checked in the last 12 months in relation to their CAH: (Clinicians: How frequently do you check the following?)			
Weight	-	Yes: 97.4% (147/151) [*96.8%*] No: 2.0% (3/151) [*2.9%*] Unsure: 0.7% (1/151) [*0.3%*]	Never: 0.0% (0/34) [*0.0%*] Less than once a year: 0.0% (0/34) [*0.0%*] At least once a year: 0.0% (0/34) [*0.0%*] Frequently: 0.0% (0/34) [*0.0%*] Every visit: 100.0% (34/34) [*100%*]
Height	-	Yes: 97.4% (147/151) [*96.8%*] No: 2.0% (3/151) [*2.9%*] Unsure: 0.7% (1/151) [*0.3%*]	Never: 0.0% (0/34) [*0.0%*] Less than once a year: 0.0% (0/34) [*0.0%*] At least once a year: 0.0% (0/34) [*0.0%*] Frequently: 0.0% (0/34) [*0.0%*] Every visit: 100.0% (34/34) [*100%*]
Blood pressure	-	Yes: 88.6% (132/149) [*87.7%*] No: 8.7% (13/149) [*9.3%*] Unsure: 2.7% (4/149) [*2.9%*]	Never: 0.0% (0/34) [*0.0%*] Less than once a year: 2.9% (1/34) [*0.9%*] At least once a year: 8.8% (3/34) [*2.8%*] Frequently: 17.6% (6/34) [*17.6%*] Every visit: 70.6% (24/34) [*78.7%*]
Examined without clothes (Clinician: pubertal staging)	-	Yes: 64.6% (95/147) [*57.5%*] No: 32.7% (48/147) [*41.1%*] Unsure: 2.7% (4/147) [*1.4%*]	Never: 0.0% (0/34) [*0.0%*] Less than once a year: 5.9% (2/34) [*11.1%*] At least once a year: 17.6% (6/34) [*14.8%*] Frequently: 55.9% (19/34) [*51.9%*] Every visit: 20.6% (7/34) [*22.2%*]
X-ray to assess bone age	-	Yes: 55.6% (84/151) [*47.5%*] No: 41.7% (63/151) [*50.2%*] Unsure: 2.6% (4/151) [*2.3%*]	Never: 0.0% (0/33) [*0.0%*] Less than once a year: 42.4% (14/33) [*53.3%*] At least once a year: 54.5% (18/33) [*45.7%*] Frequently: 3.0% (1/33) [*1.0%*] Every visit: 0.0% (0/33) [*0.0%*]
Lab 17OHP	-	Yes: 79.9% (119/149) [*80.1%*] No: 8.7% (13/149) [*10.8%*] Unsure: 11.4% (17/149) [*9.1%*]	Never: 6.1% (2/33) [*2.0%*] Less than once a year: 9.1% (3/33) [*2.9%*] At least once a year: 48.5% (16/33) [*65.7%*] Frequently: 15.2% (5/33) [*16.7%*] Every visit: 21.2% (7/33) [*12.7%*]

Lab Androstenedione hormone	-	Yes: 72.8% (110/151) [*69.8%*] No: 6.6% (10/151) [*9.2%*] Unsure: 20.5% (31/151) [*21.0%*]	Never: 5.9% (2/34) [*1.9%*] Less than once a year: 8.8% (3/34) [*2.8%*] At least once a year: 52.9% (18/34) [*68.5%*] Frequently: 11.8% (4/34) [*14.8%*] Every visit: 20.6% (7/34) [*12.0%*]
Lab Renin	-	Yes: 78.4% (116/148) [*74.8%*] No: 8.8% (13/148) [*10.0%*] Unsure: 12.8% (19/148) [*15.2%*]	Never: 0.0% (0/34) [*0.0%*] Less than once a year: 2.9% (1/34) [*0.9%*] At least once a year: 61.8% (21/34) [*66.7%*] Frequently: 17.6% (6/34) [*21.3%*] Every visit: 17.6% (6/34) [*11.1%*]
Dried blood spot 17OHP	-	Yes: 49.3% (73/148) [*45.3%*] No: 45.9% (68/148) [*49.9%*] Unsure: 4.7% (7/148) [*4.8%*]	Never: 33.3% (11/33) [*40.2%*] Less than once a year: 0.0% (0/33) [*0.0%*] At least once a year: 15.2% (5/33) [*17.6%*] Frequently: 21.2% (7/33) [*12.7%*] Every visit: 30.3% (10/33) [*29.4%*]
A urine test to check for hormones	-	Yes: 8.8% (13/148) [*7.4%*] No: 79.7% (118/148) [*86.7%*] Unsure: 11.5% (17/148) [*5.8%*]	Never: 41.2% (14/34) [*48.1%*] Less than once a year: 47.1% (16/34) [*40.7%*] At least once a year: 11.8% (4/34) [*11.1%*] Frequently: 0.0% (0/34) [*0.0%*] Every visit: 0.0% (0/34) [*0.0%*]
Frequency of assessments and access to services			
What do you think about how closely your CAH is monitored? (Clinicians: Are you equipped with enough resources to manage CAH?)	Too many appointments: 0.0% (0/41) [*0.0%*] Right no. of appointments: 92.7% (38/41) [*90.8%*] Want more appointments: 7.3% (3/41) [*9.2%*] Unsure: 0.0% (0/41) [*0.0%*]	Too many appointments: 0.7% (1/152) [*1.3%*] Right no. of appointments: 77.6% (118/152) [*77.6%*] Want more appointments: 17.7% (27/152) [*18.4%*] Unsure: 3.9% (6/152) [*2.6%*]	Want more appointments: 14.7% (5/34) [*13.9%*] Want more ancillary services: 20.6% (7/34) [*26.9%*] Want to be able to do more tests: 11.8% (4/34) [*10.2%*] Yes: 52.9% (18/34) [*49.1%*]
Clinicians: How good is your access to specialist nursing?	-	-	Unable: 0.0% (0/34) [*0.0%*] Difficult: 0.0% (0/34) [*0.0%*] Reasonable: 14.7% (5/34) [*17.6%*] Easy: 85.3% (29/34) [*82.4%*]
Clinicians: How good is your access to psychological support?	-	-	Unable: 2.9% (1/34) [*2.8%*] Difficult: 41.2% (14/34) [*51.9%*] Reasonable: 41.2% (14/34) [*31.5%*] Easy: 14.7% (5/34) [*13.9%*]
Clinicians: How good is your access to dietitian support?	-	-	Unable: 17.6% (6/34) [*15.7%*] Difficult: 52.9% (18/34) [*45.4%*] Reasonable: 26.5% (9/34) [*33.3%*] Easy: 2.9% (1/34) [*5.6%*]
Clinicians: How good is your access to physiotherapy support?	-	-	Unable: 20.6% (7/34) [*18.5%*] Difficult: 47.1% (16/34) [*51.9%*] Reasonable: 29.4% (10/34) [*24.1%*] Easy: 2.9% (1/34) [*5.6%*]
Clinicians: How good is your access to genital surgery?	-	-	Unable: 8.8% (3/34) [*8.3%*] Difficult: 1.8% (4/34) [*10.2%*] Reasonable: 35.3% (12/34) [*38.9%*] Easy: 44.1% (15/34) [*42.6%*]
Number of free text comments	5/43	56/152	25/34

Table S3.3 – Sensitivity analysis to assess timing of responses

Method:

The questionnnaire remained open to responses for a period of 18 months. To assess whether the time period in which results were obtained had an effect on results, a sensitivity analysis was performed by dividing the responses into three time periods and comparing the percentage of responses to each question.

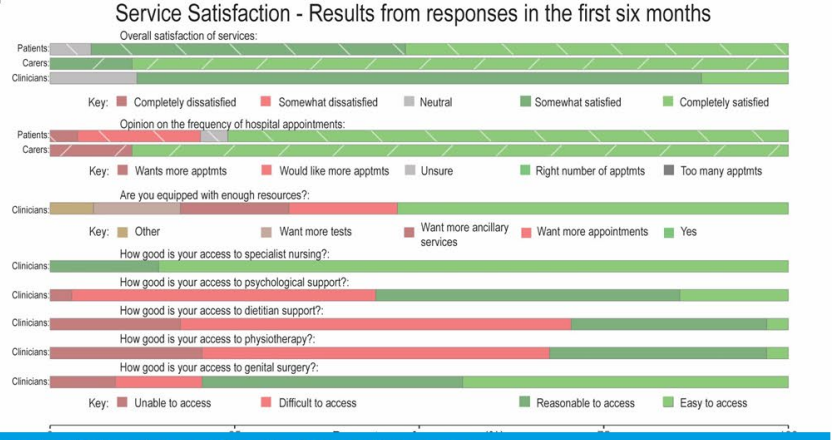
Results:

Responses were divided into questionnaires completed within each of 6 month periods:

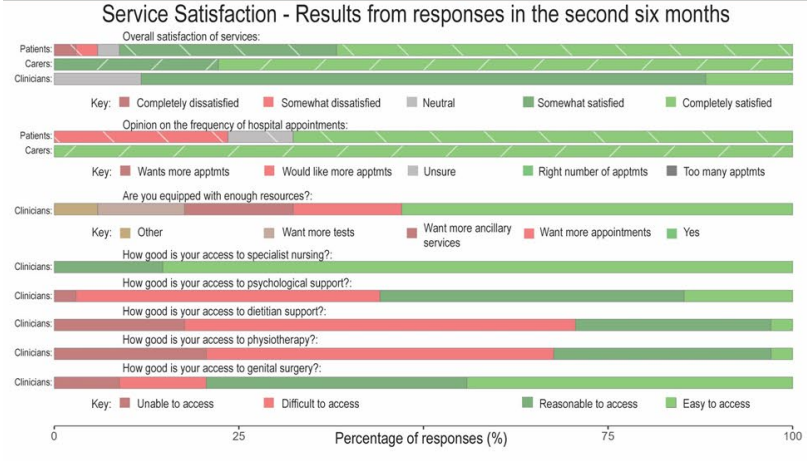
Time Period	Dates of questionnaire completion	Number of responses
Period 1	10/12/2021 to 30/06/2022	Patients=9, carers=34, clinicians=0
Period 2	01/07/2022 to 31/12/2022	Patients=9, carers=54, clinicians=0
Period 3	01/01/2023 to 30/06/2023	Patients=25, carers=64, clinicians=34

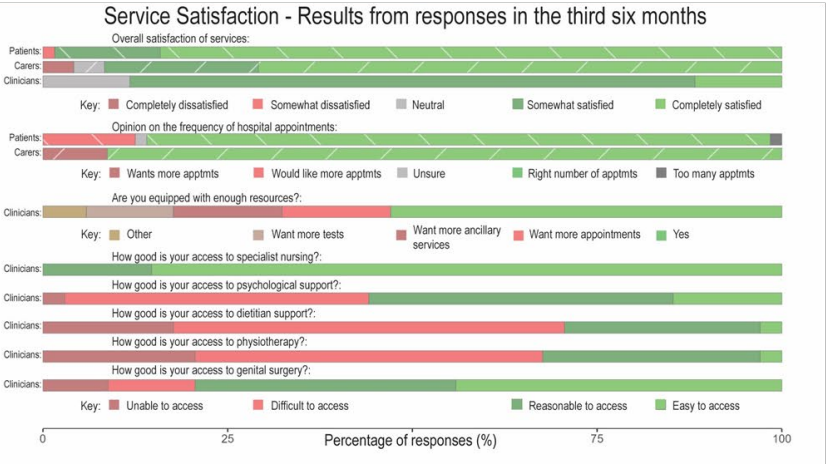
These three time periods are best compared by visual assessment of the responses to questions, that is contained within the attached slideshow found here: (https://github.com/neilxlawrence/CAH_service_evaluation).

Comparison graphs are also produced below. Please note – clinician responses reproduced in each graph as all clinicians responded within period 3.

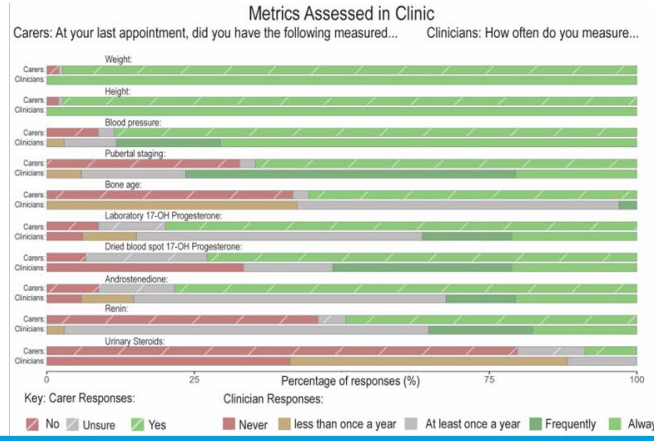


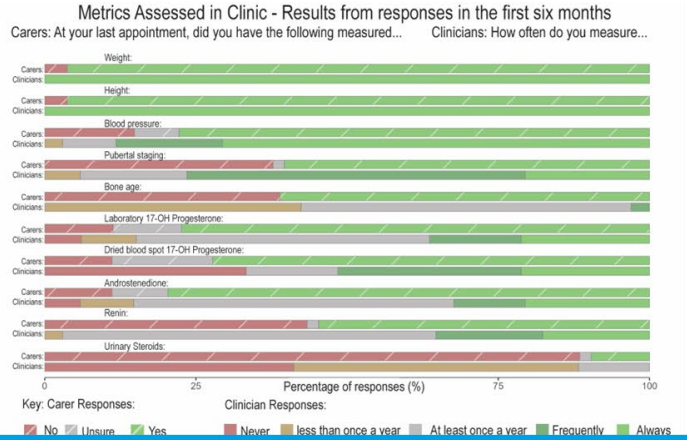
Similar responses to the total frequencies in each of the six month windows of responses



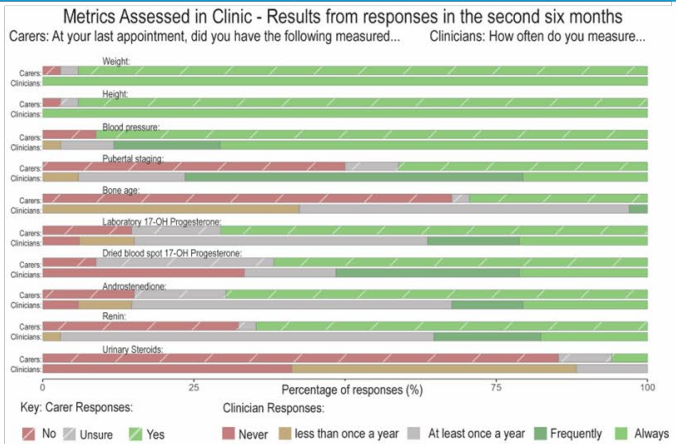


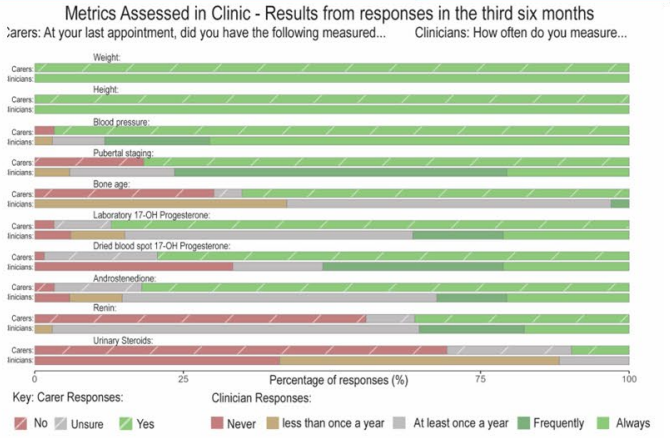
3922



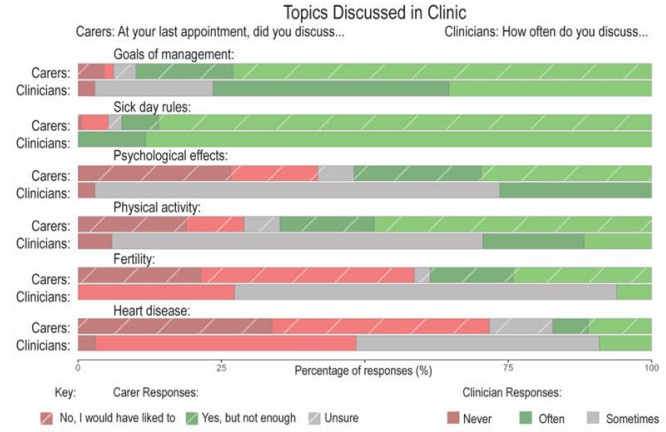


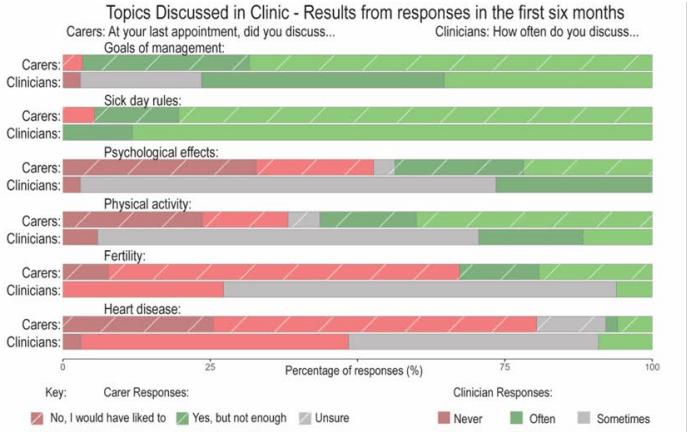
Similar responses to the total frequencies in each of the six month windows of responses



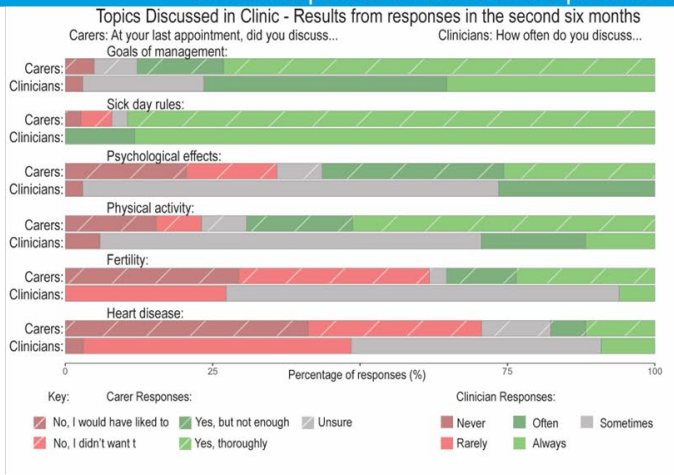


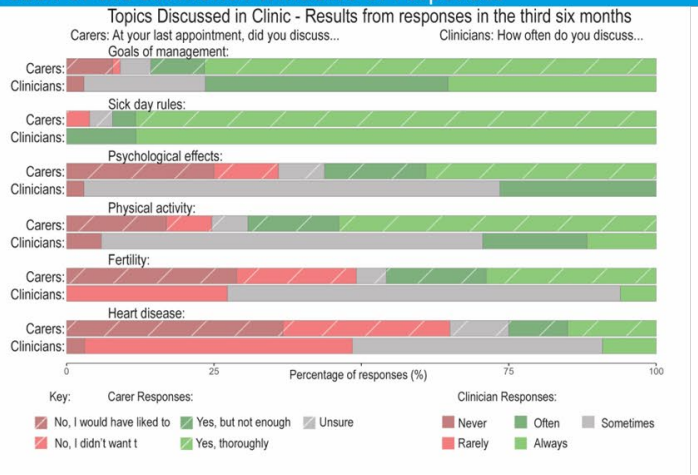
3923



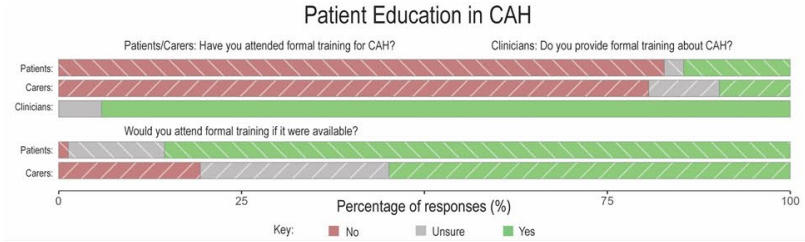


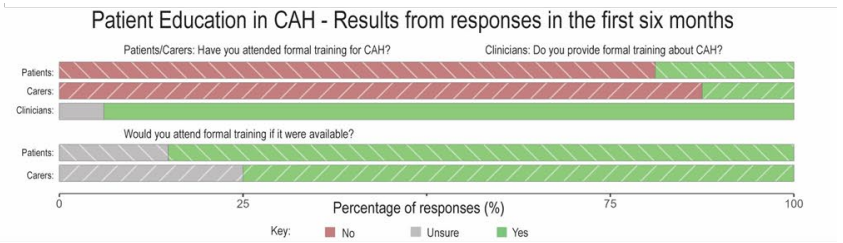
Similar responses to the total frequencies in each of the six month windows of responses



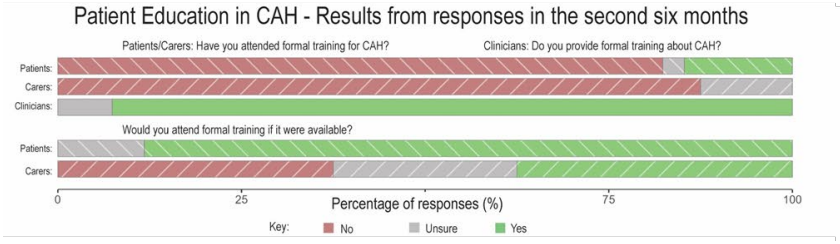


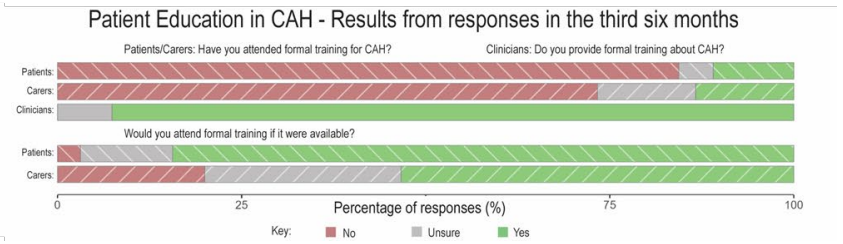
3924





Similar responses to the total frequencies in each of the six month windows of responses





Conclusion to temporal sensitivity analysis

These comparison graphs show remarkable similarity in proportions of answers between the three time periods, and show that there was no significant difference in the responses obtained from patients and carers at different times throughout the response window.

3929 Table S3.4 – Thematic analysis of free text comments

Theme of comments	Number of comments within theme	Quote from comment
Comments from patients under 20 years of age		
Transition	1/5	as a child they give you a 'doom and gloom' outlook on life
Standard of care	2/5	The management of my condition transformed from terrible to very good as soon as I got moved
Standard of care	2/5	Both doctors made considerable effort to ensure both parent and child have a full understanding of CAH
Surgery	1/5	Why can't we still get our surgery when we were born
Education	1/5	I would like more information
Comments from carers of patients under 20		
Medication	2/56	I wish that a slow release hydrocortisone was available
Medication	2/56	I do worry with the links of mental health/depression
Peer support	2/56	would be nice to be put in touch with other families of children with CAH
Peer support	2/56	would like to meet some families who [have the] same problem
Support	6/56	Would like to be able to get in touch with the CAH more easily
Support	6/56	to discuss our management and psychologically
Support	6/56	I find the staff nurses to be very helpful and informative
Support	6/56	Feel supported enough
Support	6/56	We would like ... to speak to a psychologist
Support	6/56	Parents should be offered more support
Education and information	11/56	no-one has to the time to discuss it the long term implications of cah
Education and information	11/56	There should be more programs to discuss this condition
Education and information	11/56	Training of injectable steroids [is needed for] teachers in school

Education and information	11/56	What discussions do we need to have about heart disease?
Education and information	11/56	better explanations to children [about] CAH
Education and information	11/56	The main difficulty has been trying to connect medication to his nursery
Education and information	11/56	I would be thankful for some information … on DSD as we approach puberty
Education and information	11/56	Would like to understand why things change during puberty.
Education and information	11/56	I would like to know a bit more about what different of stages may be like
Education and information	11/56	I would like to see focused mental health care and age appropriate education
Education and information	11/56	We worry about what to do when our children become sick
Transition	4/56	transition is being taken seriously
Transition	4/56	no transition to adult services available
Transition	4/56	it has now started to impact her adult life
Transition	4/56	management is completely different [in adult services]
Standard of care	28/56	standard of care is shocking
Standard of care	28/56	We are so lucky
Standard of care	28/56	[outreach clinics] appear to result in less monitoring
Standard of care	28/56	I am so grateful for all the medical advice and support
Standard of care	28/56	we have to wait 8 weeks for results
Standard of care	28/56	the NHS is over run
Standard of care	28/56	care is amazing
Standard of care	28/56	I am very happy with how they manage her condition
Standard of care	28/56	I am very happy with the care we receive
Standard of care	28/56	having been under three different hospitals, the difference in approach is marked

Standard of care	28/56	prefer to have access to 24hr profiles
Standard of care	28/56	Switching to local hospital from tertiary centre ... was big improvement
Standard of care	28/56	I would like to have androstenedione and renin checked
Standard of care	28/56	24 hour profiles should be routinely offered to ensure correct dosing
Standard of care	28/56	The Endocrine team at ... have been fantastic
Standard of care	28/56	He was actually missed
Standard of care	28/56	We feel that results of blood tests are not always discussed openly
Standard of care	28/56	I'm very happy with the care
Standard of care	28/56	We have been unable to get blood tests for a few years as my child has got issues with needles
Standard of care	28/56	24 hour blood profile is not offered
Standard of care	28/56	I feel that overall my daughter has been treated very well
Standard of care	28/56	Always satisfied
Standard of care	28/56	Our team … do a wonderful job
Standard of care	28/56	there have been no problems whatsoever
Standard of care	28/56	they have been amazing
Standard of care	28/56	A 24 [hour] stay appointment needs to have its place
Standard of care	28/56	levels of care have dropped off since the pandemic
Standard of care	28/56	Can we not have a test in place before the child's birth like they have in america
Advice	2/56	The advice differs so much from one doctor to another
Advice	2/56	The advice maybe slightly different
Surgery	1/56	I think parents should be given the option to [pursue] early surgery
Comments from clinicians caring for those with CAH		
Standard of care	19/25	would like to get more psycholohgy support
Standard of care	19/25	Ideally we would hold a dedicated clinic just for young people with CAH
Standard of care	19/25	We use saliva samples

Standard of care	19/25	We use saliva samples
Standard of care	19/25	We use saliva samples
Standard of care	19/25	we have aspirations for care [that we cannot fulfill]
Standard of care	19/25	involvement with counsellors is vital
Standard of care	19/25	no specialised CAH clinics
Standard of care	19/25	We have a 24/7 consultant led on-call service
Standard of care	19/25	Dietician and psychology support can be improved.
Standard of care	19/25	can always be better
Standard of care	19/25	I would like to see a full time [clinical nurse specialist]
Standard of care	19/25	Variable… dosing of hydrocortisone and transition
Standard of care	19/25	linking with ... primary and secondary care challenging
Standard of care	19/25	Services are variable
Standard of care	19/25	[Large] Variation in practice
Standard of care	19/25	all children with CAH [should be seen] in tertiary care
Standard of care	19/25	shared care with a DGH tend to have less good care
Standard of care	19/25	psychology … is not adequate anywhere
Surgery	3/25	We have access to surgery but don't do it
Surgery	3/25	We are pro the child deciding for themselves when they are 18 years old
Surgery	3/25	Controversy about virilised girls and surgical option is difficult
Education and information	1/25	a remote course that can be attended by parents and children may be quite useful
Research	2/25	Has improved since CHASE [research]
Research	2/25	I think the [special interest group] should have more clinical nurse specialists

3930

A4 Supplementary material for Chapter 4

Table S4.1 – Statistical packages employed within R for data analysis

Name of software package	**Citation**
Base R	https://www.R-project.org/
dplyr	https://github.com/tidyverse/dplyr
piecewiseSEM	http://dx.doi.org/10.1111/2041-210X.12512
missForest	https://cran.r-project.org/web/packages/missForest/
ggplot2	https://ggplot2.tidyverse.org
Ggally	https://github.com/ggobi/ggally
gridExtra	https://cran.r-project.org/web/packages/gridExtra/
DescTools	https://cran.r-project.org/package=DescTools
stats	https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/smooth.spline
mcp	https://cran.r-project.org/web/packages/mcp/index.html

Supplementary Analysis 4.1 – Cosinor modelling of biomarkers

Method

Multilevel cosinor modelling was undertaken to assess average peak of biomarker levels within profiles over 24 hours. To assess consistent rhythmicity within the profiles, a straight line model was fit across the whole data set assessing each biomarker as the dependent variable, and time in radians as the independent variable with a random intercept and random slope added for each individual visit. This was used as the null model. A similar model adding a sine term for time, then a cosine term for time and finally both a sine and cosine term for time was then fit. Each model was then compared to the null model using a likelihood ratio test, and the process repeated for each serum 17-hydroxyprogesterone (17OHP) and serum androstenedione profile within patients taking Chronocort, patients taking standard glucocorticoid replacement therapy and healthy participants. Bootstrapping of the modelling using replicated datasets of the same size as the parent dataset created by sampling with replacement was employed to develop 95% confidence intervals for each estimate.

Results

Models containing both sine and cosinor terms were superior to the null linear model (likelihood ratio test $p<0.05$). Supplementary table 2 shows the regression model fits and corresponding time at peak for each group.

Table S4.2 – Parameters of cosinor models of biomarkers

Independent variable = ln 17OHP (ln(nmol▪day/l)						
	CAH Patients taking standard glucocorticoid therapy		**CAH Patients taking Chronocort**		**Healthy participants**	
Model parameters:	Null model	Cosinor model	Null model	Cosinor model	Null model	Cosinor model
Intercept (SEM)	1.813 (0.096)	2.187 (0.095)	1.550 (0.116)	1.412 (0.116)	0.729 (0.202)	0.998 (0.167)
Time coefficient (SEM)	0.105 (0.008)	*Not in model*	-0.045 (0.009)	*Not in model*	0.067 (0.028)	*Not in model*
Sin(Time) coefficient (SEM)	*Not in model*	-0.636 (0.021)	*Not in model*	0.157 (0.022)	*Not in model*	0.008 (0.027)
Cos(Time) coefficient (SEM)	*Not in model*	0.147 (0.018)	*Not in model*	0.119 (0.018)	*Not in model*	0.153 (0.018)
Conditional R^2	0.757	0.806	0.860	0.866	0.725	0.739
Time of peak	*N/A (Linear model)*	0754hrs	*N/A (Linear model)*	1830hrs	*N/A (Linear model)*	0712hrs
Amplitude of ln model on absolute scale (nmol/l)*	*N/A (Linear model)*	7.73 (95% CI: 7.67 to 7.79)	*N/A (Linear model)*	0.83 (95% CI: 0.77 to 0.90)	*N/A (Linear model)*	0.44 (95% CI: 0.33 to 0.54)
Independent variable = ln androstenedione (ln(nmol▪day/l)						
Model parameters:	Null model	Cosinor model	Null model	Cosinor model	Null model	Cosinor model
Intercept (SEM)	0.428 (0.079)	0.531 (0.079)	0.182 (0.094)	0.152 (0.094)	1.119 (0.112)	0.982 (0.114)
Time coefficient (SEM)	0.034 (0.004)	*Not in model*	-0.010 (0.005)	*Not in model*	0.012 (0.011)	*Not in model*
Sin(Time) coefficient (SEM)	*Not in model*	-0.221 (0.010)	*Not in model*	0.018 (0.010)	*Not in model*	0.121 (0.017)
Cos(Time) coefficient (SEM)	*Not in model*	0.026 (0.010)	*Not in model*	0.006 (0.010)	*Not in model*	0.226 (0.012)
Conditional R^2	0.901	0.908	0.936	0.931	0.684	0.777
Time of peak	*N/A (Linear model)*	0724hrs	*N/A (Linear model)*	1718hrs	*N/A (Linear model)*	0854hrs
Amplitude of ln model on absolute scale (nmol/l)*	*N/A (Linear model)*	0.50 (95% CI: 0.47 to 0.54)	*N/A (Linear model)*	0.03 (95% CI: 0.01 to 0.05)	*N/A (Linear model)*	0.71 (95% CI: 0.64 to 0.78)

CAH=Congenital adrenal hyperplasia, SEM=Standard error of the mean; CI=Confidence Interval

Supplementary Analysis 4.2 - Change point analysis

Method

Bayesian multiple change point analysis was employed to assess any change in the relationship between 17OHP and androstenedione dependent upon the magnitude of androstenedione as the independent variable (x axis). Assessment was carried out on Ln transformed values due to the positive skew of both biomarkers across the population. The value of the change point was then converted back to an absolute value by taking the exponent of the derived estimate.

Bayesian priors were set to assess the possibility of two relationships. Change point model A assessed the likelihood of a horizontal relationship before a change point that turned to a log linear relationship. Change point model B assessed a log linear relationship before the change point with an altered log linear relationship after the change point. Change point model C assessed the same relationship exclusively in male patients, change point model D exclusively in female patients, change point model E exclusively in patients taking Chronocort and change point model F exclusively in patients taking hydrocortisone replacement. Change point model G replicated the most stable model in healthy participants.

Results

The comprehensive results of this modelling are contained in supplementary tables 3 and 4. The most stable model was that of a slope to slope relationship (model B), which translated to a change point occurring at an absolute value of 17OHP of 4.5nmol/l (95% confidence interval 4.0 to 5.0). The models run exclusively in male and female patients were less stable ($\hat{R}$>1.1), but both estimated a change point that had a confidence interval overlapping the change point of the overall model. The models run exclusively in those taking Chronocort and those taking hydrocortisone also showed change points with confidence intervals overlapping that of model B. There was therefore no indication that there was a difference in this change point between sexes, or between those taking Chronocort compared to standard glucocorticoid replacement.

Healthy participants also exhibited a clear change point in the relationship that was at a similar level of 17OHP as the CAH patients, but occurring at a higher level of androstenedione. This is consistent with the clustering of points from healthy participants in the lower right quadrant of Figure 4.4 within the main chapter.

3981 Table S4.3 – Change point analysis in the main cohort

3982

Independent variable (x axis) = ln androstenedione (ln(nmol/l)) Dependent variable (y axis) = ln 17OH-progesterone (ln(nmol/l))				
Multiple change point model estimate	**CAH Patients**			
Change point	A: Both sexes change point model: flat to slope	**B: Both sexes change point model: slope to slope**	C: Male change point model: slope to slope	D: Female change point model: slope to slope
Change point (ln androstenedione (ln nmol/l)) (95% CI)	-1.72 (-1.82 to -1.62) [$\hat{R}$=1.011]	**0.41** **(0.33 to 0.49)** **[$\hat{R}$=1.007]**	-0.10 (-0.93 to 0.35) [$\hat{R}$=7.459]	-0.32 (-2.18 to 0.69) [$\hat{R}$=31.942]
Intercept of first segment (ln androstenedione (ln nmol/l)) (95% CI)	-0.70 (-0.80 to -0.61)) [$\hat{R}$=1.010]	**1.19** **(1.15 to 1.23)** **[$\hat{R}$=1.003)**	0.71 (-0.87 to 1.19) [$\hat{R}$=4.641]	0.53 (-1.09 to 1.30) [$\hat{R}$=17.400]
Slope of first segment (95% CI)	0 (prior)	**0.73** **(0.70 to 0.77)** **[$\hat{R}$=1.004]**	0.44 (-0.94 to 1.03) [$\hat{R}$=4.484]	0.48 (-0.13 to 0.79) [$\hat{R}$=15.404]
Slope of second segment (95% CI)	-0.70 (-0.80 to -0.61)	**1.49** **(1.44 to 1.55)**	1.41 (1.34 to 1.47) [$\hat{R}$=3.254]	1.35 (1.01 to 1.57) [$\hat{R}$=11.219]
Calculated value of absolute 17OHP at change point (95% CI) (nmol/L)	0.50 (0.45 to 0.54)	**4.46** **(4.00 to 4.99)**	1.94 (1.01 to 4.72	1.46 (0.45 to 6.34)

3983 CAH=Congenital Adrenal Hyperplasia; CI=Confidence Interval

3984 Table S4.4 - Change point analysis by preparation and in healthy participants

Independent variable (x axis) = ln androstenedione (ln(nmol/l)) Dependent variable (y axis) = ln 17OH-progesterone (ln(nmol/l))			
Multiple change point model estimate	**CAH Patients**		**Healthy Participants**
Change point	E: Both sexes change point model: Slope to slope (Chronocort)	F: Both sexes change point model: Slope to slope (Hydrocortisone replacement)	G: Both sexes change point model: slope to slope
Change point (ln androstenedione (ln nmol/l)) (95% CI)	0.70 (0.46 to 0.98) [$\hat{R}$=1.029]	0.50 (0.34 to 0.67) [$\hat{R}$<1.001]	1.21 (0.92 to 1.54) [$\hat{R}$=1.020]
Intercept of first segment (ln androstenedione (ln nmol/l)) (95% CI)	1.04 (0.96 to 1.13) [$\hat{R}$=1.019]	1.29 (1.21 to 1.38) [$\hat{R}$<1.001)	0.28 (0.17 to 0.40) [$\hat{R}$<1.001]
Slope of first segment (95% CI)	0.73 (0.66 to 0.80) [$\hat{R}$=1.017]	0.69 (0.63 to 0.74) [$\hat{R}$<1.001]	0.50 (0.31 to 0.65) [$\hat{R}$=1.017]
Slope of second segment (95% CI)	1.74 (1.58 to 1.94) [$\hat{R}$=1.024]	1.45 (1.39 to 1.52) [$\hat{R}$=1.001]	1.21 (0.97 to 1.46) [$\hat{R}$=1.003]
Calculated value of absolute 17OHP at change point (95% CI) (nmol/L)	4.72 (3.54 to 6.76)	5.14 (4.15 to 6.52)	2.42 (1.57 to 4.09)

3985 CAH=Congenital Adrenal Hyperplasia; CI=Confidence Interval

Supplementary Analysis 4.3 – Relationship between serum 17OHP, androstenedione and testosterone

Method

To assess the relationship between testosterone and serum 17OHP and androstenedione, a simple linear regression model was applied with testosterone (ln(nmol/l)) as the dependent variable and 17OHP (nmol/l) or androstenedione (nmol/l) as the independent variable.

Results

In females, the level of serum testosterone correlated with 17OHP in both CAH patients (R^2_{adj} = 0.51) and healthy participants (R^2_{adj}=0.77), as did androstenedione (R^2_{adj}=0.77 and R^2_{adj}=0.56 respectively). In male patients, there was a very weak negative correlation between 17OHP, A4 and testosterone (17OHP: R^2_{adj}=0.02, androstenedione: R^2_{adj}=0.04), whereas a positive correlation existed within healthy participants (17OHP: R^2_{adj} =0.37, androstenedione: R^2_{adj}=0.31) (supplementary table 5).

4000 Table S4.5 – Regression results of testosterone on 17OHP and

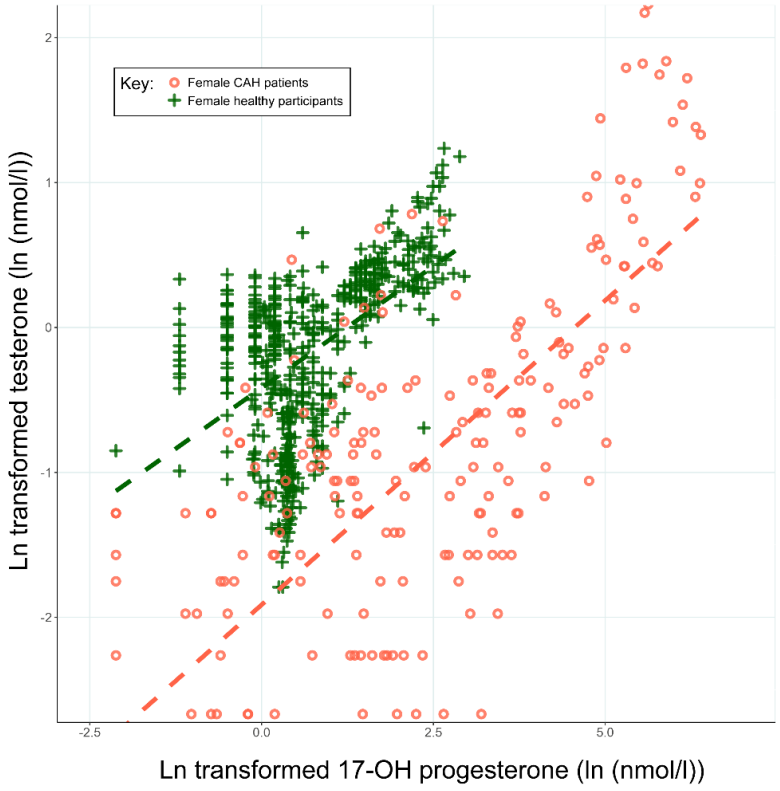
4001 androstenedione

Independent variable ln transformed testosterone				
CAH patients:				
	Male		**Female**	
Dependent variable:	Intercept	Slope	Intercept	Slope
ln 17OHP (ln nmol/l)	2.86 (0.098)	-0.04 (0.025)	-1.91 (0.088)	0.42 (0.028)
	R^2_{adj} = 0.016		R^2_{adj} = 0.505	
ln androstenedione (ln nmol/l)	2.86 (0.075)	-0.10 (0.039)	-1.23 (0.043)	0.76 (0.028)
	R^2_{adj} = 0.043		R^2_{adj} = 0.773	
Healthy participants:				
Dependent variable:	Intercept	Slope	Intercept	Slope
ln 17OHP (ln nmol/l)	2.44 (0.019)	0.29 (0.013)	-0.42 (0.028)	0.34 (0.024)
	R^2_{adj} = 0.367		R^2_{adj} = 0.274	
ln androstenedione (ln nmol/l)	2.25 (0.030)	0.44 (0.024)	-0.88 (0.033)	0.70 (0.028)
	R^2_{adj} = 0.309		R^2_{adj} = 0.558	

4002

Figure S4.1 – Regression of testosterone on serum 17OHP and androstenedione

17OHP=17OH-Progesterone



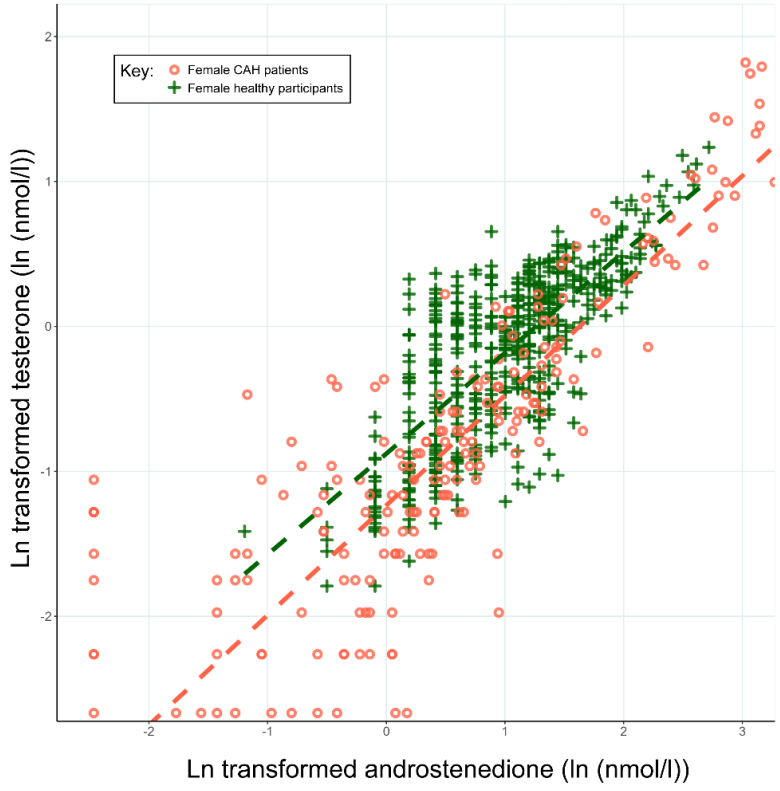


Table S4.6

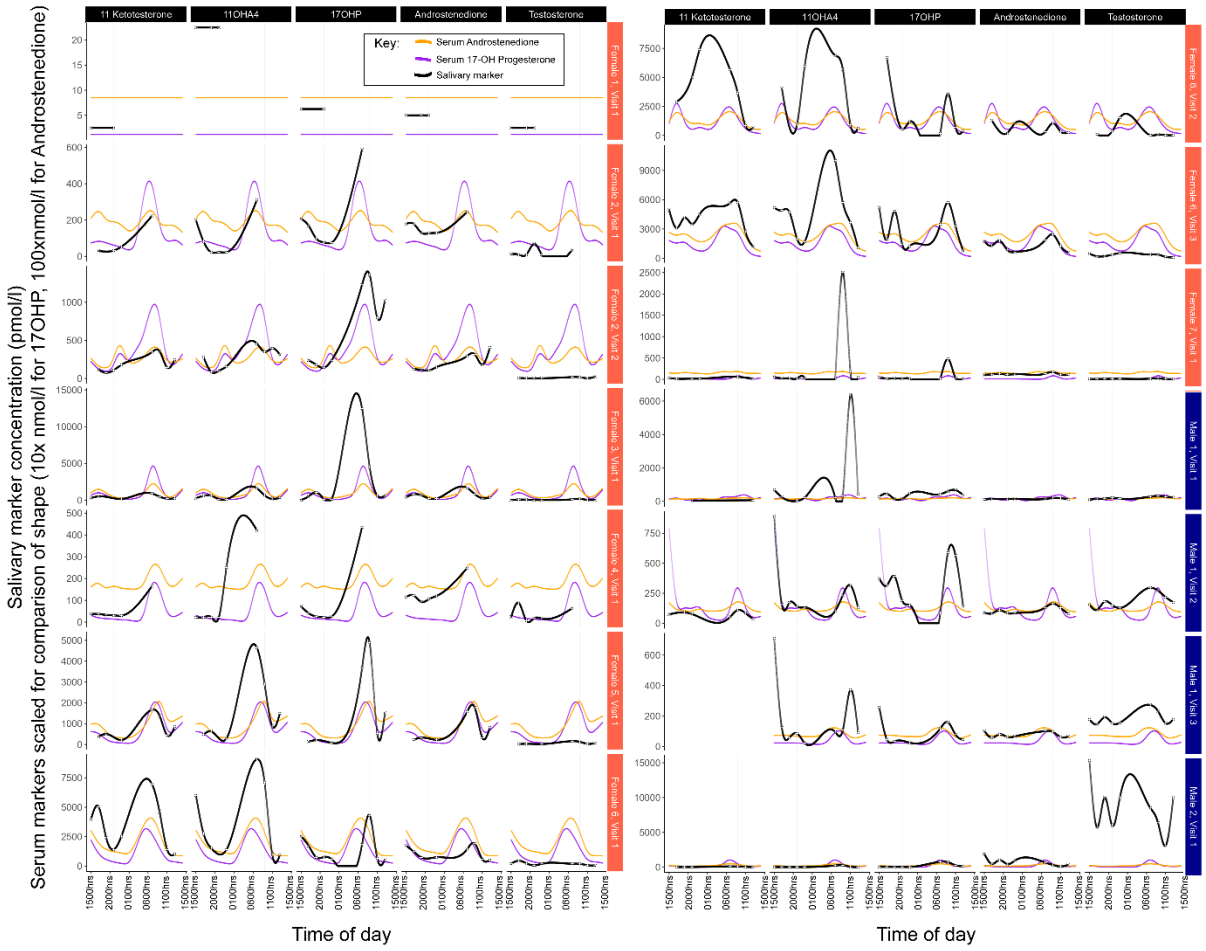
Regression of salivary steroids on serum steroids

Correlation of salivary markers with serum steroids, all measurements natural log transformed. Numbers show mean of estimate (standard error of estimate). Correlation between serum steroids reported to facilitate comparison

Independent variable: serum 17OHP (ln (nmol/l))						
	All CAH patients (n=9, 14 profiles)		**Male CAH patients (n=3, 4 profiles)**		**Female CAH patients (n=6, 10 profiles)**	
Dependent variable:	**Intercept**	**Slope**	**Intercept**	**Slope**	**Intercept**	**Slope**
Salivary 11KT (ln (pmol/l))	2.48 (0.225)	0.98 (0.062)	2.89 (0.303)	0.41 (0.107)	2.78 (0.215)	0.97 (0.056)
	R^2_{adj} = 0.76		R^2_{adj} = 0.48		R^2_{adj} = 0.82	
Salivary 11OHA4 (ln (pmol/l))	3.37 (0.193)	0.80 (0.057)	3.70 (0.486)	0.50 (0.190)	3.52 (0.205)	0.81 (0.056)
	R^2_{adj} = 0.65		R^2_{adj} = 0.15		R^2_{adj} = 0.75	
Salivary 17OHP (ln (pmol/l))	3.12 (0.134)	0.87 (0.040)	3.30 (0.291)	0.77 (0.113)	3.12 (0.163)	0.88 (0.044)
	R^2_{adj} = 0.82		R^2_{adj} = 0.57		R^2_{adj} = 0.85	
Salivary androstenedione (ln (pmol/l))	3.99 (0.124)	0.54 (0.037)	4.42 (0.345)	0.27 (0.135)	4.03 (0.121)	0.56 (0.033)
	R^2_{adj} = 0.67		R^2_{adj} = 0.08		R^2_{adj} = 0.80	
Salivary testosterone (ln (pmol/l))	3.43 (0.329)	0.369 (0.098)	5.41 (0.669)	0.32 (0.263)	2.02 (0.230)	0.53 (0.063)
	R^2_{adj} = 0.11		R^2_{adj} = 0.01		R^2_{adj} = 0.49	
Serum androstenedione (ln (nmol/l))	-0.77 (0.015)	0.62 (0.006)	-0.45 (0.021)	0.59 (0.006)	-0.85 (0.018)	0.59 (0.008)
	R^2_{adj} = 0.67		R^2_{adj} = 0.79		R^2_{adj} = 0.57	
Serum testosterone (ln (nmol/l))	-1.02 (0.126)	0.51 (0.037)	2.87 (0.072)	-0.04 (0.018)	-1.91 (0.072)	0.42 (0.024)
	R^2_{adj} = 0.28		R^2_{adj} = 0.02		R^2_{adj} = 0.49	
Independent variable: serum androstenedione (ln (nmol/l))						
Dependent variable:	**Intercept**	**Slope**	**Intercept**	**Slope**	**Intercept**	**Slope**
Salivary 11KT (ln (pmol/l))	3.33 (0.125)	1.48 (0.063)	3.71 (0.198)	0.57 (0.296)	3.33 (0.157)	1.49 (0.071)
	R^2_{adj} = 0.87		R^2_{adj} = 0.15		R^2_{adj} = 0.87	
Salivary 11OHA4 (ln (pmol/l))	4.32 (0.150)	1.10 (0.085)	4.77 (0.260)	0.34 (0.446)	4.03 (0.187)	1.22 (0.088)
	R^2_{adj} = 0.61		R^2_{adj} = -0.01		R^2_{adj} = 0.73	
Salivary 17OHP (ln (pmol/l))	4.30 (0.144)	1.06 (0.082)	4.74 (0.185)	1.18 (0.319)	3.78 (0.203)	1.24 (0.096)
	R^2_{adj} = 0.61		R^2_{adj} = 0.27		R^2_{adj} = 0.70	
Salivary androstenedione (ln (pmol/l))	4.51 (0.070)	0.84 (0.040)	4.70 (0.137)	1.13 (0.235)	4.26 (0.065)	0.91 (0.031)
	R^2_{adj} = 0.81		R^2_{adj} = 0.40		R^2_{adj} = 0.93	
Salivary testosterone (ln (pmol/l))	4.07 (0.326)	0.326 (0.143)	5.56 (0.273)	1.93 (0.474)	2.14 (0.168	0.91 (0.080)
	R^2_{adj} = 0.04		R^2_{adj} = 0.31		R^2_{adj} = 0.65	
Serum 17OHP (ln (nmol/l))	1.45 (0.014)	1.08 (0.010)	1.17 (0.024)	1.34 (0.015)	1.44 (0.018	0.97 (0.013)
	R^2_{adj} = 0.67		R^2_{adj} = 0.79		R^2_{adj} = 0.57	
Serum testosterone (ln (nmol/l))	-0.35 (0.079)	0.93 (0.048)	2.84 (0.055)	-0.07 (0.028)	-1.26 (0.035	0.78 (0.024)
	R^2_{adj} = 0.44		R^2_{adj} = 0.03		R^2_{adj} = 0.78	

Figure S4.2 – Salivary biomarker spline plots

Salivary markers (pmol/l) on time of day in CAH patients plotted over scaled serum 17OHP and androstenedione for comparison of profile shape. Serum 17OHP in 10x nmol/l (purple) is overlaid on serum androstenedione in 100x nmol/l (orange) to allow comparison to the profile of each salivary marker in pmol/l (black). CAH=congenital adrenal hyperplasia; 17OHP=17OH-Progesterone



Supplementary Analysis 4.4 - Assessing biomarkers of disease control with dose

Method

To assess the consistency of disease control within patients maintained on the same dose, visits within patients where dose or preparation was not changed were paired. The AUC of 17OHP and androstenedione was regressed between each paired visit to assess consistency between visits with a calibration plot.

To assess a consistent change in disease control with total daily dose across all patients, a simple linear regression model was applied with ln area under the curve of 17OHP (ln(nmol▪day/l)) as the dependent variable and total daily hydrocortisone equivalent (mg) as the independent variable. Conversion factors for hydrocortisone equivalent applied were 5x for prednis(ol)one and 80x for dexamethasone.

Results

Across all visits there were 93 paired visits within 39 men maintained on the same dose and preparation, and 218 paired visits within 71 women maintained on the same dose and preparation. Within patient comparisons were more consistent within men (17OHP R^2_{adj}=0.79, androstenedione R^2_{adj} =0.90) than within women (17OHP R^2_{adj} =0.44, androstenedione R^2_{adj} =0.55) (Appendix Table S4.6, Appendix Figure S4.3). There was no relationship between ln AUC 17OHP and total daily hydrocortisone equivalent at visit (p=0.34, R^2_{adj} =0.002) (Appendix Table S4.7, Appendix Figure S4.4).

4037 Table S4.6 – AUC of biomarkers within patients regressed between visits

4038 maintained on the same dose and preparation

4039 Numbers show mean of estimate (standard error of estimate)

Independent variable: ln AUC 17OHP at latter visit in same patient maintained on same preparation and same dose (ln(nmol▪day/l))				
	Male CAH patients		**Female CAH patients**	
Dependent variable:	**Intercept**	**Slope**	**Intercept**	**Slope**
Ln AUC 17OHP at former visit within patient (ln(nmol▪day/l))	0.25 (0.287)	0.96 (0.052)	1.30 (0.261)	0.70 (0.054)
	R^2_{adj} = 0.79		R^2_{adj} = 0.44	
Ln AUC androstenedione at former visit within patient (ln(nmol▪day/l))	-0.12 (0.139)	1.02 (0.036)	0.60 (0.158)	0.79 (0.049)
	R^2_{adj} = 0.90		R^2_{adj} = 0.55	

4040

4041 Table S4.7 – Change in disease control with total daily dose across the sample

4042 population

4043 Numbers show mean of estimate (standard error of estimate)

Independent variable: ln Area under curve 17OHP at visit (ln(nmol▪day/l))		
Dependent variable:	**Intercept**	**Slope**
Total daily Hydrocortisone equivalent (mg)	5.60 (0.24)	-0.007 (0.007)
	R^2 = 0.002	

4044

Figure S4.3 – Variability in biomarkers of disease control within patients between visits maintained on the same dose and preparation

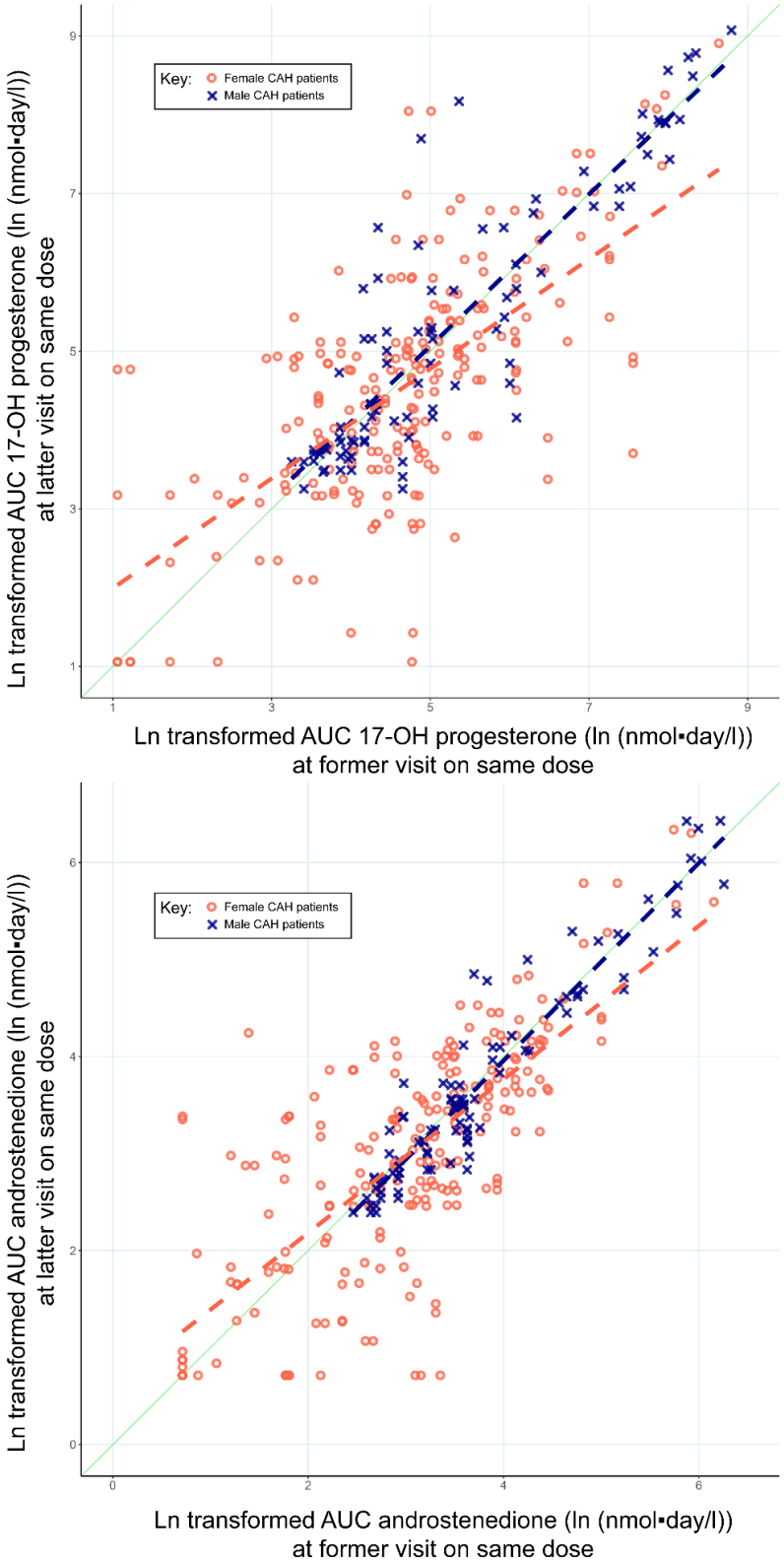
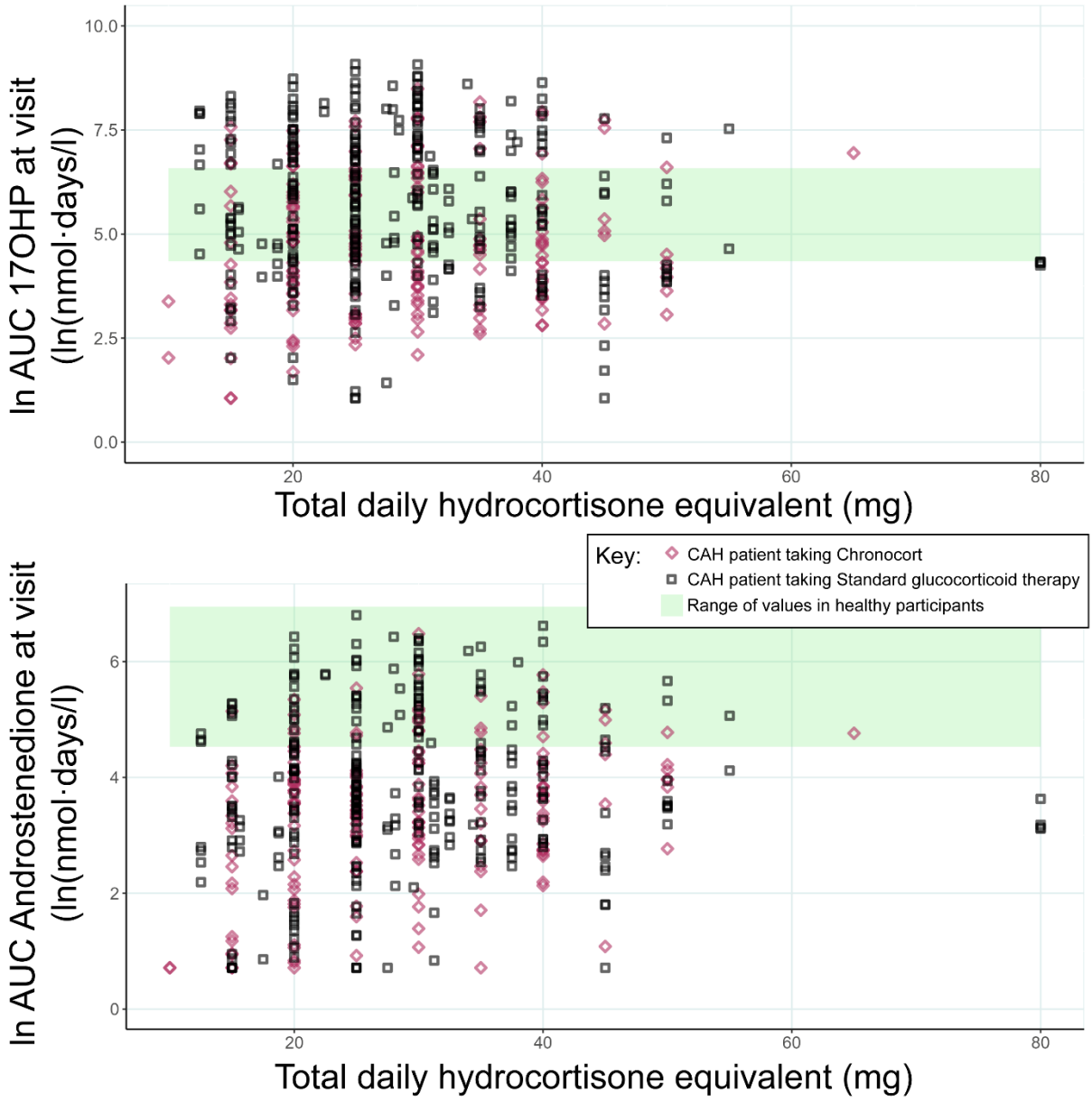


Figure S4.4 – Ln AUC 17OHP and androstenedione on daily hydrocortisone equivalent dose

No relationship between dose and overall control. AUC=area under the curve; 17OHP=17OH-Progesterone



Supplementary Analysis Code 4.1

The code used to model data within R to carry out this research can be found:

https://github.com/neilxlawrence/chronocort_modelling

A5 Supplementary Material for Chapter 5

Supplementary methods 5.1 – Full details of data preparation

Data was prepared for analysis by removal of duplicated entries, and manual assessment of outlying values using data visualisation, followed by contact with individual centres for clarification about possible data entry errors and units of measurement.

Interpolation of missing height and weight data

Due to varying number of visits within patients, a hierarchical approach was employed to interpolation of missing height and weight. In patients with 11 or more visits in which height and weight were available, a generalised additive model was estimated for height on age with weight as a covariate. Pearson scaled residuals were calculated for each data point within this fit to assess for outliers. Individual model fits were examined visually and an absolute Pearson scaled residual of 3 or more was selected to declare a datapoint as an implausible outlier allowing deletion of that point and imputation. If a datapoint was declared an outlier, the modelling was reiterated without the outlying point before establishing the final GAM model fit that was then used to interpolate missing or outlying data points. In patients with 11 or more visits in which height was available but the covariate weight was not available, the process was repeated to fit a GAM model without the covariate of weight.

In patients with between 4 and 10 known data points for height, a cubic spline with 4 degrees of freedom was fit through the data, without covariate adjustment. Residuals were calculated manually from the spline fit by calculating the difference between the model fit and each data point, and then standardising by dividing them by the standard deviation of the height measurements within that patient. A threshold of greater than 0.5 was used for this standardised residual to declare a data point as an outlier following visual inspection of the model fits. In patients with between 2 and 4 available data points for height, a linear interpolation method was employed. Points were not extrapolated beyond known data points.

A similar process was applied for measurements of weight, with different outlier threshold parameters due to the greater variability in this metric with true values able to decrease as well as increase. The residual threshold for GAM models was thus set at a pearson residual of 4, and for spline models a standardised residual threshold of 1.2.

Imputation of missing dose data

Missing data patterns were reviewed and correlated with other data fields of relevance. The field ‘has dose changed from previous visit’, evidenced that doses patients were prescribed at that visit going forward were entered into the registry, with a preference for dose changes to be entered into the registry: Patients who were maintained on the same dose were more likely to have their precise dose

details missing, with dose changes more likely to include precise dosing details. A pragmatic approach of data imputation using last observation carried forward was therefore employed, with limits on how far data could be carried according to age, restricted to carrying for 6 months under age 1, 12 months under age 3, 24 months under age 5 and 36 months over age 5. If data was not available to carry forward, then next observation carried backward was employed with the same age limits. The dose a patient was on prior to clinic (as defined by the previous visit) was therefore used for modelling to assess the effect of that dose on the patient parameters measured in clinic. Hydrocortisone equivalent dosing was converted using the conversion factors in table S2.

Joint modelling multiple imputation of BP and biomarkers

The remaining variables of interest exhibited more random variation within patients, and were thus more suited to multiple imputation, as carrying values or interpolating values would result in less plausible random variation around the trend. These metrics were imputed using a joint modelling multiple imputation model with complete age, sex, interpolated height and interpolated weight used as covariates. Biomarkers were converted to consistent units (table S3) and natural log transformed prior to imputation to achieve a closer to normal shaped distribution. Imputation was carried out 10 times to create 10 different imputed datasets. Following imputation, the standardised lower and upper limits of detection (table S3) were applied to maintain consistency across the datasets and prevent the imputation of excessively large or small values that would have undue leverage on multivariable modelling. Biomarkers were censored to consistent lower and upper levels of detection to prevent undue leverage of high outlying values or extreme values created by log transforming values equal to or close to zero using the values in table S3.

This hybrid approach to missing data imputation was complicated, and tested by visualising the data and by rerunning multivariable models on complete case data without data imputation. Appendix Figure S5.2 shows a similar trend with age across all the variables, and the sensitivity analyses rerunning models on complete cases showed no change in the direction, and very little change in the magnitude of any of the estimates of interest (Table S8).

4115 # Table S5.1 – R software packages employed during analysis

Name of software package	Citation
Base R	https://www.R-project.org/
tibble	https://tibble.tidyverse.org/
rlang	https://cran.r-project.org/web/packages/rlang/index.html
dplyr	https://github.com/tidyverse/dplyr
doRNG	https://cran.r-project.org/web/packages/doRNG/index.html
piecewiseSEM	http://dx.doi.org/10.1111/2041-210X.12512
lme4	https://cran.r-project.org/web/packages/lme4/index.html
mice	https://cran.r-project.org/web/packages/mice/index.html
miceRanger	https://cran.r-project.org/web/packages/miceRanger/index.html
mgcv	https://cran.r-project.org/web/packages/mgcv/index.html
sjstats	https://cran.r-project.org/web/packages/sjstats/index.html
rlist	https://renkun-ken.github.io/rlist/
pedsbp	https://cran.r-project.org/web/packages/pedbp/readme/README.html
ggplot2	https://ggplot2.tidyverse.org
ggpubr	https://rpkgs.datanovia.com/ggpubr/
gridExtra	https://cran.r-project.org/web/packages/gridExtra/
DescTools	https://cran.r-project.org/package=DescTools
summarytools	https://cran.r-project.org/web/packages/summarytools/vignettes/introduction.html
Hmisc	https://cran.r-project.org/web/packages/Hmisc/index.html
rms	https://cran.r-project.org/web/packages/rms/index.html
mfp	https://cran.r-project.org/package=mfp
glmnet	https://cran.r-project.org/web/packages/glmnet/index.html
brms	https://cran.r-project.org/web/packages/brms/index.html
Shiny	http://www.rstudio.com/shiny/

4116

Table S5.2 – Conversion factors for glucocorticoid equivalents

Conversion factors to create hydrocortisone equivalent doses

Preparation of glucocorticoid	Multiplication factor to create hydrocortisone equivalent
Hydrocortisone	1
Prednisolone	4
Dexamethasone	80
Cortisone acetate	0.8
Methylprednisolone	5

4120 Table S5.3 – Conversion factors for biomarkers

Marker	Original unit	Unit to convert to	Multiplication factor	Standardised lower limit of detection	Standardised upper limit of detection
17OHP	ng/ml	nmol/l	0.030261	0.05nmol/l	1000nmol/l
Androstenedione	ng/dl	nmol/l	0.034916	0.05nmol/l	100nmol/l
Plasma Renin Activity (PRA)	ng/ml/hr	nmol/l/hr	0.77154	*Converted to renin first*	*Converted to renin first*
PRA to renin	nmol/l/hr	µIU/ml	0.158	*Converted to renin first*	*Converted to renin first*
Renin	ng/l	µIU/ml	1.67	0.05µIU/ml	1000µIU/ml

4121

4122 Table S5.4 – Summary statistics with imputed values

Sex assigned at birth:	Male	Female	Not assigned	Total sample
Number of countries	17	17	4	18
Number of centres	31	31	4	35
Number of patients	257	291	6	554
Number of visits	3018	3361	57	6436
Number of visits per patient Median (Q1 to Q3)	9 (7 to 16)	9 (5 to 16)	8 (2 to 11)	9 (6 to 16)
Number of years visits spanned within patients Median (Q1 to Q3)	3.2 (2.7 to 6.9)	3.2 (2.1 to 8.0)	2.1 (0.2 to 3.1)	3.2 (2.5 to 7.3)
Age of patients at youngest visit (years) Median (Q1 to Q3)	0.13 (0.04 to 0.99)	0.21 (0.04 to 3.03)	0.08 (0.03 to 0.26)	0.16 (0.04 to 2.17)
Age of patients at oldest visit (years) Median (Q1 to Q3)	5.70 (3.09 to 11.47)	6.01 (3.07 to 14.35)	2.25 (0.46 to 3.17)	5.81 (3.07 to 12.87)
Systolic BP at visit (mmHg) Median, (n) [Q1 to Q3]‡	107 (n=1556) [97 to 118] *105 (n=2936) [96 to 116]*	105 (n=1652) [96 to 116] *104 (n=3257) [95 to 114]*	99 (n=21) [85 to 107] *96 (n=53) [85 to 105]*	106 (n=3229) [97 to 117] *105 (n=6246) [95 to 115]*
Systolic BP SDS at visit‡ Median (n) [Q1 to Q3]	1.5 (n=1492) [0.5 to 2.4] *1.5 (n=2792) [0.5 to 2.5]*	1.1 (n=1564) [0.3 to 2.1] *1.2 (n=3080) [0.3 to 2.2]*	-	1.3 (n=3056) [0.4 to 2.2] *1.4 (n=5872) [0.4 to 2.3]*
Diastolic BP at visit (mmHg) Median, (n) [Q1 to Q3]‡	64 (n=1542) [57 to 71] *64 (n=2936) [57 to 71]*	64 (n=1645) [57 to 70] *64 (n=3256) [57 to 71]*	60 (n=20) [50 to 71.25] *59 (n=53) [50 to 69]*	64 (n=3207) [57 to 70] *63.97 (n=6245) [57 to 71]*
Diastolic BP SDS at visit‡ Median (n) [Q1 to Q3]	1.1 (n=1478) [0.4 to 2] *1.3 (n=2792) [0.5 to 2.2]*	0.9 (n=1559) [0.3 to 1.6] *1.1 (n=3080) [0.3 to 1.9]*	-	1.0 (n=3037) [0.3 to 1.8] *1.2 (n=5872) [0.4 to 2.1]*
Visits prescribed hydrocortisone* n (%) [missing n, % missing]‡	2200 (72.9%) [610, 20.2%] *2670 (88.5%) [133 (4.4%)]*	2214 (65.9%) [648, 19.3%] *2668 (79.4%) [134 (4.0%)]*	35 (61.4%) [22, 38.6%] *51 (89.5%) [6 (10.5%)]*	4449 (69.1%) [1280, 19.9%] *5389 (83.7%) [273 (4.2%)]*
Total Hydrocortisone equivalent at visit per BSA (mg/m²)†‡ Median (n) [Q1 to Q3]	14.1 (n= 2237) [9.8 to 15.5] *13.8 (n= 2774) (9.7 to 15.2)*	14.5 (n= 2438) [9.9 to 15.7] *14.1 (n= 3099) (9.8 to 15.6)*	18.4 (n= 25) [13.5 to 20.3] 17.2 (n= 48) (13.5 to 19.9)	14.3 (n= 4700) [9.9 to 15.6] *14.0 (n= 5921) (9.8 to 15.5)*
Visits prescribed fludrocortisone n (%) [missing n, % missing] ‡	2577 (85.4%) [180, 6.0%] *2618 (86.8%) [125, 4.1%]*	2754 (81.9%) [140, 4.2%] *2804 (83.4%) [77, 2.3%]*	51 (89.5%) [6, 10.5%] *55 (96.5%) [2, 3.5%]*	5382 (83.6%) [326, 5.1%] *5477 (85.1%) [204, 3.2%]*
Total Fludrocortisone at visit per body surface area (when prescribed) (µg/m²) ‡ Median (n) [Q1 to Q3]	318 (n= 2442) [103 to 396] *322 (n= 2536) [105 to 400]*	292 (n= 2527) [99 to 321] *296 (n= 2693) [99 to 323]*	555 (n= 38) [207 to 484] *561 (n= 52) [203 to 505]*	307 (n= 5007) [102 to 356] *311 (n= 5281) [103 to 359]*
Visits prescribed salt n (%), [missing n, % missing] ‡	400 (13.3%) [290, 9.6%] *400 (13.3%) [265, 8.8%]*	486 (14.5%) [204, 6.1%] *486 (14.5%) [175, 5.2%]*	35 (61.4%) [5, 8.8%] *35 (61.4%) [4, 7.0%]*	921 (14.3%) [499, 7.8%] *921 (14.3%) [444, 6.9%]*
Renin (µIU/ml) Median (n) [Q1 to Q3] ‡	5.0 (n=1059) [0.3 to 69.1] *9.2 (n=2945) [0.9 to 75.2]*	4.0 (n=1041) [0.4 to 54.0] *6.8 (n=3261) [0.8 to 56.0]*	0.4 (n=16) [0.1 to 3.2] *2.6 (n=53) [0.2 to 38.2]*	4.5 (n=2116) [0.4 to 61.8] *7.7 (n=6259) [0.8 to 64.4]*
17-OH Progesterone (nmol/ml) Median (n) [Q1 to Q3] ‡	18.2 (n=1380) [2.0 to 100.0] *17.0 (n=2965) [2.8 to 92.1]*	26.9 (n=1325) [3.0 to 140.0] *20.93 (n=3279) [2.96 to 117.61]*	12.1 (n=31) [4.0 to 62.8] *15.94 (n=54) [3.81 to 65.54]*	21.18 (n=2736) [2.72 to 115.0] *18.87 (n=6298) [2.95 to 102.84]*
Androstenedione (nmol/ml) Median (n) [Q1 to Q3] ‡	0.1 (n=1285) [0.1 to 3.0] *0.5 (n=2960) [0.1 to 2.6]*	1.0 (n=1211) [0.1 to 6.0] *1.0 (n=3270) [0.1 to 5.0]*	0.1 (n=29) [0.1 to 1.0] *0.2 (n=53) [0.1 to 1.2]*	0.7 (n=2525) [0.1 to 3.5] *0.8 (n=6283) [0.1 to 3.6]*

4123 n=Number, BSA=Body surface area; Q1=Quartile 1; Q3=Quartile 3; SDS=Standard deviation score
4124 ‡Normal font, black text = original data; Italics, grey text = imputed data (mean of median and quartiles calculated on each imputed data set)
4125 *remaining visits patients prescribed either cortisone acetate (n=352), dexamethasone (134), prednisone (37), prednisolone (105), methylprednisolone (1), mixed dosing (59) or no glucocorticoid (86) (Table S5)
4126 †Hydrocortisone equivalent calculated by multiplying preparations by the following factors: prednisolone/prednisone x4; dexamethasone x80; cortisone acetate x0.8; methylprednisolone x5 (Table S2 for full frequency tables of preparations)

4127 Table S5.5 – Full frequency table of preparations of glucocorticoid prescribed

4128 at visit

Preparation	All visits	Male visits	Female visits	Visits sex not assigned
Hydrocortisone	5389 (83.7%)	2670 (88.5%)	2668 (79.4%)	51 (89.5%)
Prednisolone	105 (1.6%)	12 (0.4%)	93 (2.8%)	0 (0.0%)
Prednisone	37 (0.6%)	12 (0.4%)	25 (0.7%)	0 (0.0%)
Dexamethasone	134 (2.1%)	36 (1.2%)	98 (2.9%)	0 (0.0%)
Methylprednisolone	1 (0.1%)	1 (0.1%)	0 (0.0%)	0 (0.0%)
Cortisone Acetate	352 (5.5%)	112 (3.7%)	240 (7.2%)	0 (0.0%)
Mixed dosing	59 (0.9%)	4 (0.1%)	55 (1.6%)	0 (0.0%)
None	86 (1.3%)	38 (1.2%)	48 (1.4%)	0 (0.0%)
Unknown	273 (4.2%)	133 (4.4%)	134 (4.0%)	6 (10.5%)
Total	6436 (100%)	3018 (100%)	3361 (100%)	57 (100%)

4129

Table S5.6 – Bayesian multiple change point analysis

	Male systolic	**Male Diastolic**	**Female Systolic**	**Female diastolic**
Intercept (mmHg) (95% CI) [R̂]	21.1 (18.2 to 23.9) [1.033]	33.4 (31.5 to 35.4) [1.007]	18.4 (17.7 to 19.1) [1.023]	25.1 (22.5 to 27.7) [1.24]
Slope (mmHg/year) (95% CI) [R̂]	-1.1 (-1.7 to -0.6) [1.042]	-4.4 (-5 to -3.8) [1.006]	-0.9 (-1 to -0.8) [1.031]	-2.8 (-3.6 to -2.1) [1.296]
Change point age (years) (95% CI) [R̂]	11.5 (7.8 to 17.6) [1.039]	5.9 (5.5 to 6.4) [1.006]	13.1 (12.2 to 14.1) [1.021]	7.0 (5.9 to 8.2) [1.274]
Plateau (mmHg) (95% CI)	9.2 (6.7 to 10.7) [-]	7.3 (6.8 to 7.9) [-]	6.4 (5.8 to 6.9) [-]	5.6 (4.8 to 6.5) [-]

R̂ is the Gelman-Rubin convergence statistic assessing the strength of Markov chain Monte Carlo model fits, values closer to 1.0 showing a stronger model estimate. Plateau is calculated from estimated changepoint, slope and intercept, and therefore does not have a calculated R̂.

Table S5.7 – Comprehensive Bayesian Joint Modelling parameters

Model 1: Estimation of drug dose effects on BP

Number of patients in model: 452 **Number of visits in model:** 4187

Parameter	Estimate	Confidence Interval	Significance
Outcome: Systolic BP (Absolute values)			
Intercept	102.047	96.551 to 107.542	*
Sex (Male)	1.440	-0.124 to 3.005	NS
Sex (Not assigned)	-7.643	-16.714 to 1.427	NS
Age (years)	0.084	-0.428 to 0.595	NS
Height (cm)	-0.083	-0.156 to -0.009	*
Weight (kg)	0.370	0.220 to 0.521	*
Daily fludrocortisone (µg)	0.012	0.005 to 0.020	*
Daily hydrocortisone equivalent (mg)	0.052	-0.057 to 0.162	NS
Daily salt (g)	-0.686	-2.138 to 0.766	NS
SD of random slope | random intercept : 0.80 | 4.97		**R^2 systolic BP (95% CI) :** 0.27 (0.14 to 0.40)	
Outcome: Diastolic BP (Absolute values)			
Intercept	64.564	60.633 to 68.495	*
Sex (Male)	0.713	-0.394 to 1.820	NS
Sex (Not assigned)	-3.061	-9.261 to 3.138	NS
Age (years)	0.069	-0.288 to 0.425	NS
Height (cm)	-0.069	-0.118 to -0.019	*
Weight (kg)	0.184	0.107 to 0.262	*
Daily fludrocortisone (µg)	0.008	0.003 to 0.014	*
Daily hydrocortisone equivalent (mg)	0.031	-0.045 to 0.107	NS
Daily salt (g)	-0.562	-1.540 to 0.416	NS
SD of random slope | random intercept : 0.23 | 2.33		**R^2 diastolic BP (95% CI) :** 0.17 (0.06 to 0.28)	

BP = Blood pressure; SD = Standard deviation; CI = Confidence interval; NS = Not significant; * = CI does not span zero. CI calculated by iterating Bayesian Model using 400 iterations each through 10 bootstrap replications (sampling all data available from one patient with replacement) for each of the 10 imputed datasets.

Model 1: Sensitivity analysis A: Complete case analysis: Estimation of drug dose effect on BP

Number of patients in model: 370 **Number of visits in model:** 2204

Parameter	Estimate	Confidence Interval	Significance
Outcome: Systolic BP (Absolute values)			
Intercept	98.682	90.895 to 106.468	*
Sex (Male)	0.998	-0.544 to 2.541	NS
Sex (Not assigned)	-9.694	-13.366 to -6.021	*
Age (years)	0.068	-0.473 to 0.609	NS
Height (cm)	-0.080	-0.158 to -0.002	*
Weight (kg)	0.410	0.222 to 0.599	*
Daily fludrocortisone (µg)	0.022	0.013 to 0.031	*
Daily hydrocortisone equivalent (mg)	0.079	-0.056 to 0.215	NS
Daily salt (g)	-1.655	-4.258 to 0.949	NS
SD of random slope | random intercept : 1.12 | 7.40		**R^2 systolic BP (95% CI) :** 0.27 (0.21 to 0.32)	
Outcome: Diastolic BP (Absolute values)			
Intercept	62.648	56.310 to 68.985	*
Sex (Male)	0.688	-0.452 to 1.828	NS
Sex (Not assigned)	-2.778	-6.878 to 1.322	NS
Age (years)	0.069	-0.376 to 0.514	NS
Height (cm)	-0.059	-0.132 to 0.013	NS
Weight (kg)	0.179	0.123 to 0.234	*
Daily fludrocortisone (µg)	0.014	0.006 to 0.021	*
Daily hydrocortisone equivalent (mg)	0.061	-0.025 to 0.147	NS
Daily salt (g)	-1.389	-2.888 to 0.110	NS
SD of random_slope | random_intercept : 0.35 | 3.44		**R^2 diastolic BP (95% CI) :** 0.17 (0.10 to 0.24)	

BP = Blood pressure; SD = Standard deviation; CI = Confidence interval; NS = Not significant; * = CI does not span zero. CI calculated by iterating Bayesian Model using 400 iterations each through 10 bootstrap replications on complete data.

Conclusion of sensitivity analysis: imputation strategy is robust and leads to estimates of similar direction and magnitude as complete case analysis.

Model 1: Sensitivity analysis B: Estimation of drug dose by BSA effect on BP SDS

Number of patients in model: 422 **Number of visits in model:** 3223

Parameter	Estimate	Confidence Interval	Significance
Outcome: Systolic BP (SD Score)			
Intercept	1.428	1.162 to 1.694	*
Sex (Male)	0.163	-0.007 to 0.334	NS
Age (years)	-0.068	-0.093 to -0.044	*
Daily fludrocortisone per BSA ($\mu g/m^2$)	0.001	0.001 to 0.002	*
Daily hydrocortisone equivalent per BSA (mg/m^2)	0.003	-0.010 to 0.016	NS
Daily salt per BSA (g/m^2)	-0.021	-0.116 to 0.073	NS
BMI SD Score	0.081	0.021 to 0.141	*
SD of random slope | random intercept : 0.08 | 0.54		**R^2 systolic BP (95% CI) :** 0.31 (0.18 to 0.44)	
Outcome: Diastolic BP (SD Score)			
Intercept	1.755	1.545 to 1.965	*
Sex (Male)	0.163	0.019 to 0.307	*
Age (years)	-0.140	-0.161 to -0.119	*
Daily fludrocortisone per BSA ($\mu g/m^2$)	0.001	0.001 to 0.002	*
Daily hydrocortisone equivalent per BSA (mg/m^2)	0.005	-0.005 to 0.015	NS
Daily salt per BSA (g/m^2)	0.023	-0.045 to 0.092	NS
BMI SD Score	-0.007	-0.052 to 0.037	NS
SD of random slope | random intercept : 0.08 | 0.38		**R^2 diastolic BP (95% CI) :** 0.42 (0.33 to 0.51)	

BP = Blood pressure; SD = Standard deviation; BSA = Body surface area (calculated by the mosteller formula); BMI = body mass index; CI = Confidence interval; NS = Not significant; * = CI does not span zero. CI calculated by iterating Bayesian Model using 400 iterations each through 10 bootstrap replications (sampling all data available from one patient with replacement) for each of the 10 imputed datasets. *National Heart, Lung and Blood Institute data for calculation of SD scores of BP only available for those over 1 year, and not calculated for those patients declared sex 'not assigned' at birth, thus less data available for modelling.

Conclusion of sensitivity analysis: modelling of drugs by surface area and SD scores instead of absolute values does not make a significant difference to the direction or magnitude of the effect size of drugs on BP

Model 1: Sensitivity analysis C: Estimation of drug dose effect on BP restricted to patients under 5 years of age

Number of patients in model: 381 **Number of visits in model:** 2895

Parameter	Estimate	Confidence Interval	Significance
Outcome: Systolic BP (Absolute values)			
Intercept	85.87	74.96 to 96.78	*
Sex (Male)	1.80	-0.36 to 3.96	NS
Sex (Not assigned)	-7.97	-20.62 to 4.68	NS
Age (years)	-1.66	-3.10 to -0.22	*
Height (cm)	0.25	0.04 to 0.45	*
Weight (kg)	-0.27	-0.88 to 0.35	NS
Daily fludrocortisone (µg)	0.01	0.01 to 0.02	*
Daily hydrocortisone equivalent (mg)	0.02	-0.13 to 0.17	NS
Daily salt (g)	-0.71	-2.18 to 0.77	NS
SD of random slope | random intercept : 0.63 | 4.47		**R^2 systolic BP (95% CI) :** 0.20 (0.05 to 0.35)	
Outcome: Diastolic BP (Absolute values)			
Intercept	46.85	38.96 to 54.74	*
Sex (Male)	0.50	-1.05 to 2.05	NS
Sex (Not assigned)	-2.79	-13.03 to 7.44	NS
Age (years)	-1.90	-3.09 to -0.72	*
Height (cm)	0.29	0.15 to 0.44	*
Weight (kg)	-0.46	-0.91 to -0.00	*
Daily fludrocortisone (µg)	0.01	0.00 to 0.02	*
Daily hydrocortisone equivalent (mg)	-0.01	-0.13 to 0.12	NS
Daily salt (g)	-0.50	-1.45 to 0.45	NS
SD of random_slope | random_intercept : 0.45 | 2.28		**R^2 diastolic BP (95% CI) :** 0.14 (0.02 to 0.26)	

BP = Blood pressure; SD = Standard deviation; CI = Confidence interval; NS = Not significant; * = CI does not span zero. CI calculated by iterating Bayesian Model using 400 iterations each through 10 bootstrap replications on complete data.

Conclusion of sensitivity analysis: Despite salt treatment being administered to only 15% of the sample and concentrated in those under 5, there is still no statistically significant effect on blood pressure of salt replacement when restricting analysis to patients under 5 years of age only.

Model 2: Extent to which renin predicts BP

Number of patients in model: 544 **Number of visits in model:** 6130

Parameter	Estimate	Confidence Interval	Significance
Outcome: Systolic BP (Absolute values)			
Intercept	108.85	103.06 to 114.63	*
Sex (Male)	1.88	0.51 to 3.25	*
Sex (Not assigned)	-8.33	-15.64 to -1.01	*
Age (years)	0.33	-0.07 to 0.73	NS
Height (cm)	-0.15	-0.22 to -0.07	*
Weight (kg)	0.43	0.29 to 0.56	*
ln renin (ln(μIU/ml))	-1.00	-1.47 to -0.53	*
SD of random slope | random intercept : 0.40 | 4.55		**R^2 systolic BP (95% CI) :** 0.32 (0.19 to 0.45)	
Outcome: Diastolic BP (Absolute values)			
Intercept	68.7	64.20 to 73.20	*
Sex (Male)	1.07	0.21 to 1.93	*
Sex (Not assigned)	-3.38	-9.30 to 2.54	NS
Age (years)	0.29	-0.07 to 0.66	NS
Height (cm)	-0.10	-0.16 to -0.05	*
Weight (kg)	0.21	0.15 to 0.28	*
ln renin (ln(μIU/ml))	-0.71	-1.13 to -0.29	*
SD of random slope | random intercept : 0.18 | 2.60		**R^2 diastolic BP (95% CI) :** 0.21 (0.09 to 0.32)	

BP = Blood pressure; SD = Standard deviation; CI = Confidence interval; NS = Not significant; * = CI does not span zero. CI calculated by iterating Bayesian Model using 400 iterations each through 10 bootstrap replications (sampling all data available from one patient with replacement) for each of the 10 imputed datasets.

Model 2: Sensitivity analysis A: Complete Case Analysis: Extent to which renin predicts BP

Number of patients in model: 256 **Number of visits in model:** 1027

Parameter	Estimate	Confidence Interval	Significance
Outcome: Systolic BP (Absolute values)			
Intercept	99.74	88.17 to 111.30	*
Sex (Male)	1.40	-0.44 to 3.24	NS
Sex (Not assigned)	-7.35	-10.05 to -4.65	*
Age (years)	0.17	-0.66 to 1.01	NS
Height (cm)	-0.05	-0.20 to 0.09	NS
Weight (kg)	0.35	0.11 to 0.58	*
ln renin (ln(μIU/ml))	-1.13	-1.71 to -0.54	*
SD of random slope | random intercept : 0.95 | 8.27		**R^2 systolic BP (95% CI) :** 0.32 (0.26 to 0.39)	
Outcome: Diastolic BP (Absolute values)			
Intercept	61.28	54.33 to 68.24	*
Sex (Male)	0.63	-1.00 to 2.26	NS
Sex (Not assigned)	-2.54	-4.51 to -0.57	*
Age (years)	-0.02	-0.66 to 0.61	NS
Height (cm)	-0.02	-0.10 to 0.07	NS
Weight (kg)	0.16	0.05 to 0.26	*
ln renin (ln(μIU/ml))	-0.80	-1.20 to -0.41	*
SD of random slope | random intercept : 0.54 | 4.73		**R^2 diastolic BP (95% CI) :** 0.24 (0.17 to 0.30)	

BP = Blood pressure; SD = Standard deviation; CI = Confidence interval; NS = Not significant; * = CI does not span zero. CI calculated by iterating Bayesian Model using 400 iterations each through 10 bootstrap replications (sampling all data available from one patient with replacement) on complete data.

Conclusion of sensitivity analysis: imputation strategy is robust and leads to estimates of similar direction and magnitude as complete case analysis.

Model 2: Sensitivity analysis B: Extent to which renin predicts BP SDS

Number of patients: 499 Number of visits: 4564

Parameter	Estimate	Confidence Interval	Significance
Outcome: Systolic BP (SD Score*)			
Intercept	1.89	1.75 to 2.03	*
Sex (Male)	0.16	-0.01 to 0.32	NS
Age (years)	-0.09	-0.11 to -0.07	*
ln renin ln(μIU/ml)	-0.08	-0.12 to -0.04	*
BMI SD Score	0.08	0.01 to 0.15	*
SD of random slope | random intercept : 0.08 | 0.50		**R^2 systolic BP (95% CI) :** 0.33 (0.20 to 0.45)	
Outcome: Diastolic BP (SD Score*)			
Intercept	2.28	2.15 to 2.42	*
Sex (Male)	0.18	0.05 to 0.30	*
Age (years)	-0.17	-0.20 to -0.15	*
ln renin ln(μIU/ml)	-0.05	-0.08 to -0.01	*
BMI SD Score	-0.02	-0.07 to 0.02	NS
SD of random slope | random intercept : 0.09 | 0.38		**R^2 diastolic BP (95% CI) :** 0.42 (0.34 to 0.51)	

BP = Blood pressure; SD = Standard deviation; BMI = body mass index calculated using world health organisation reference data; CI = Confidence interval; NS = Not significant; * = CI does not span zero. CI calculated by iterating Bayesian Model using 400 iterations each through 10 bootstrap replications (sampling all data available from one patient with replacement) for each of the 10 imputed datasets. *National Heart, Lung and Blood Institute data for calculation of SD scores of BP only available for those over 1 year, and not calculated for those patients declared sex 'not assigned' at birth, thus less data available for modelling.

Conclusion of sensitivity analysis: modelling of SD scores instead of absolute values does not make a significant difference to the direction or magnitude of the prediction of renin on BP

Model 3: Extent to which 17OH-progesterone predicts BP

Number of patients in model: 544 **Number of visits in model:** 6130

Parameter	Estimate	Confidence Interval	Significance
Outcome: Systolic BP (Absolute values)			
Intercept	106.64	101.58 to 111.69	*
Sex (Male)	1.47	0.07 to 2.87	*
Sex (Not assigned)	-8.12	-15.63 to -0.62	*
Age (years)	0.18	-0.20 to 0.55	NS
Height (cm)	-0.12	-0.18 to -0.05	*
Weight (kg)	0.42	0.28 to 0.55	*
ln 17OHP (ln(nmol/l))	-0.64	-1.00 to -0.27	*
SD of random slope | random intercept : 0.41 | 4.35		**R^2 systolic BP (95% CI) :** 0.31 (0.17 to 0.45)	
Outcome: Diastolic BP (Absolute values)			
Intercept	67.37	63.72 to 71.02	*
Sex (Male)	0.76	-0.17 to 1.68	NS
Sex (Not assigned)	-3.11	-9.54 to 3.33	NS
Age (years)	0.18	-0.16 to 0.51	NS
Height (cm)	-0.08	-0.13 to -0.03	*
Weight (kg)	0.21	0.14 to 0.28	*
ln 17OHP (ln(nmol/l))	-0.50	-0.78 to -0.22	*
SD of random slope | random intercept : 0.19 | 2.40		**R^2 diastolic BP (95% CI) :** 0.20 (0.07 to 0.32)	

BP = Blood pressure; SD = Standard deviation; 17OHP = 17OH-progesterone; CI = Confidence interval; NS = Not significant; * = CI does not span zero. CI calculated by iterating Bayesian Model using 400 iterations each through 10 bootstrap replications (sampling all data available from one patient with replacement) for each of the 10 imputed datasets.

Model 3: Sensitivity analysis A: Complete Case Analysis: Extent to which 17OH-progesterone predicts BP

Number of patients in model: 288 **Number of visits in model:** 1215

Parameter	Estimate	Confidence Interval	Significance
Outcome: Systolic BP (Absolute values)			
Intercept	94.52	84.07 to 104.97	*
Sex (Male)	0.79	-1.65 to 3.24	NS
Sex (Not assigned)	-5.99	-10.13 to -1.86	*
Age (years)	0.07	-0.63 to 0.77	NS
Height (cm)	0.02	-0.12 to 0.15	NS
Weight (kg)	0.30	0.10 to 0.50	*
ln 17OHP (ln(nmol/l))	-0.86	-1.26 to -0.46	*
SD of random slope | random intercept : 0.81 | 6.94		**R^2 systolic BP (95% CI) :** 0.30 (0.24 to 0.37)	
Outcome: Diastolic BP (Absolute values)			
Intercept	60.14	54.60 to 65.69	*
Sex (Male)	0.75	-0.33 to 1.82	NS
Sex (Not assigned)	-1.61	-5.33 to 2.11	NS
Age (years)	0.14	-0.40 to 0.68	NS
Height (cm)	-0.01	-0.07 to 0.06	NS
Weight (kg)	0.15	0.09 to 0.21	*
ln 17OHP (ln(nmol/l))	-0.64	-0.88 to -0.40	*
SD of random slope | random intercept : 0.62 | 4.47		**R^2 diastolic BP (95% CI) :** 0.22 (0.17 to 0.28)	

BP = Blood pressure; SD = Standard deviation; CI = Confidence interval; NS = Not significant; * = CI does not span zero. CI calculated by iterating Bayesian Model using 400 iterations each through 10 bootstrap replications (sampling all data available from one patient with replacement) on complete data.

Conclusion of sensitivity analysis: imputation strategy is robust and leads to estimates of similar direction and magnitude as complete case analysis.

Model 3: Sensitivity analysis B: Extent to which 17OH-progesterone predicts BP SDS

Number of patients in model: 499 **Number of visits in model:** 4564

Parameter	Estimate	Confidence Interval	Significance
Outcome: Systolic BP (SD Score*)			
Intercept	1.89	1.75 to 2.03	*
Sex (Male)	0.13	-0.03 to 0.29	NS
Age (years)	-0.09	-0.11 to -0.07	*
ln 17OHP (ln (nmol/l))	-0.05	-0.08 to -0.02	*
BMI SD Score	0.08	0.01 to 0.15	*
SD of random slope | random intercept : 0.07 | 0.49		**R^2 systolic BP (95% CI) :** 0.32 (0.19 to 0.45)	
Outcome: Diastolic BP (SD Score*)			
Intercept	2.3	2.17 to 2.43	*
Sex (Male)	0.16	0.03 to 0.29	*
Age (years)	-0.17	-0.20 to -0.15	*
ln 17OHP (ln (nmol/l))	-0.04	-0.07 to -0.01	*
BMI SD Score	-0.02	-0.07 to 0.02	NS
SD of random slope | random intercept : 0.09 | 0.37		**R^2 diastolic BP (95% CI) :** 0.42 (0.33 to 0.51)	

BP = Blood pressure; SD = Standard deviation; BMI = body mass index calculated using world health organisation reference data; CI = Confidence interval; NS = Not significant; * = CI does not span zero. CI calculated by iterating Bayesian Model using 400 iterations each through 10 bootstrap replications (sampling all data available from one patient with replacement) for each of the 10 imputed datasets. *National Heart, Lung and Blood Institute data for calculation of SD scores of BP only available for those over 1 year, and not calculated for those patients declared sex 'not assigned' at birth, thus less data available for modelling.

Conclusion of sensitivity analysis: modelling of SD scores instead of absolute values does not make a significant difference to the direction or magnitude of the prediction of 17OHP on BP

Model 4: Extent to which Androstenedione predicts BP

Number of patients in model: 543 **Number of visits in model:** 6187

Parameter	Estimate	Confidence Interval	Significance
Outcome: Systolic BP (Absolute values)			
Intercept	105.60	101.92 to 109.29	*
Sex (Male)	1.30	-0.35 to 2.95	NS
Sex (Not assigned)	-9.42	-19.72 to 0.87	NS
Age (years)	0.30	-0.06 to 0.67	NS
Height (cm)	-0.14	-0.19 to -0.08	*
Weight (kg)	0.44	0.35 to 0.54	*
ln androstenedione (ln(nmol/l))	-0.73	-1.18 to -0.28	*
SD of random slope | random intercept : 0.41 | 4.38		**R^2 systolic BP (95% CI) :** 0.31 (0.17 to 0.45)	
Outcome: Diastolic BP (Absolute values)			
Intercept	66.67	64.00 to 69.33	*
Sex (Male)	0.76	-0.57 to 2.09	NS
Sex (Not assigned)	-4.20	-12.85 to 4.46	NS
Age (years)	0.29	0.03 to 0.56	*
Height (cm)	-0.10	-0.13 to -0.06	*
Weight (kg)	0.22	0.16 to 0.28	*
ln androstenedione (ln(nmol/l))	-0.61	-0.96 to -0.25	*
SD of random slope | random intercept : 0.20 | 2.39		**R^2 diastolic BP (95% CI) :** 0.20 (0.08 to 0.32)	

BP = Blood pressure; SD = Standard deviation; 17OHP = 17OH-progesterone; CI = Confidence interval; NS = Not significant; * = CI does not span zero. CI calculated by iterating Bayesian Model using 400 iterations each through 10 bootstrap replications (sampling all data available from one patient with replacement) for each of the 10 imputed datasets.

Model 4: Sensitivity analysis A: Complete Case Analysis: Extent to which androstenedione predicts BP

Number of patients in model: 282 **Number of visits in model:** 1229

Parameter	Estimate	Confidence Interval	Significance
Outcome: Systolic BP (Absolute values)			
Intercept	92.95	80.68 to 105.23	*
Sex (Male)	0.87	-0.60 to 2.34	NS
Sex (Not assigned)	-10.83	-24.93 to 3.27	NS
Age (years)	0.15	-0.63 to 0.92	NS
Height (cm)	-0.02	-0.18 to 0.14	NS
Weight (kg)	0.38	0.27 to 0.48	*
ln androstenedione (ln(nmol/l))	-0.86	-1.43 to -0.30	*
SD of random slope | random intercept : 0.91 | 7.73		**R^2 systolic BP (95% CI) :** 0.32 (0.26 to 0.38)	
Outcome: Diastolic BP (Absolute values)			
Intercept	61.74	55.96 to 67.53	*
Sex (Male)	0.78	-1.00 to 2.56	NS
Sex (Not assigned)	-3.42	-9.81 to 2.96	NS
Age (years)	0.12	-0.24 to 0.47	NS
Height (cm)	-0.05	-0.13 to 0.03	NS
Weight (kg)	0.21	0.12 to 0.30	*
ln androstenedione (ln(nmol/l))	-0.72	-0.95 to -0.49	*
SD of random slope | random intercept : 0.57 | 5.20		**R^2 diastolic BP (95% CI) :** 0.25 (0.20 to 0.30)	

BP = Blood pressure; SD = Standard deviation; CI = Confidence interval; NS = Not significant; * = CI does not span zero. CI calculated by iterating Bayesian Model using 400 iterations each through 10 bootstrap replications (sampling all data available from one patient with replacement) on complete data.

Conclusion of sensitivity analysis: imputation strategy is robust and leads to estimates of similar direction and magnitude as complete case analysis.

Model 4: Sensitivity analysis B: Extent to which androstenedione predicts BP SDS

Number of patients in model: 499 **Number of visits in model:** 4564

Parameter	Estimate	Confidence Interval	Significance
Outcome: Systolic BP (SD Score*)			
Intercept	1.78	0.73 to 2.84	*
Sex (Male)	0.12	-0.04 to 0.28	NS
Age (years)	-0.10	-0.28 to 0.08	NS
ln androstenedione (ln(nmol/l))	-0.06	-0.25 to 0.13	NS
BMI SD Score	0.08	0.01 to 0.16	*
SD of random slope | random intercept : 0.07 | 0.49		**R^2 systolic BP (95% CI) :** 0.32 (0.18 to 0.45)	
Outcome: Diastolic BP (SD Score*)			
Intercept	2.21	1.68 to 2.74	*
Sex (Male)	0.15	0.02 to 0.28	*
Age (years)	-0.18	-0.27 to -0.08	*
ln androstenedione (ln(nmol/l))	-0.04	-0.19 to 0.10	NS
BMI SD Score	-0.02	-0.07 to 0.02	NS
SD of random slope | random intercept : 0.09 | 0.37		**R^2 diastolic BP (95% CI) :** 0.42 (0.33 to 0.51)	

BP = Blood pressure; SD = Standard deviation; BMI = body mass index calculated using world health organisation reference data; CI = Confidence interval; NS = Not significant; * = CI does not span zero. CI calculated by iterating Bayesian Model using 400 iterations each through 10 bootstrap replications (sampling all data available from one patient with replacement) for each of the 10 imputed datasets. *National Heart, Lung and Blood Institute data for calculation of SD scores of BP only available for those over 1 year, and not calculated for those patients declared sex 'not assigned' at birth, thus less data available for modelling.

Conclusion of sensitivity analysis: modelling of SD scores instead of absolute values does not make a significant difference to the direction or magnitude of the prediction of androstenedione on BP

4228 Table S5.8 – Sensitivity analysis results: impact of imputation strategy on

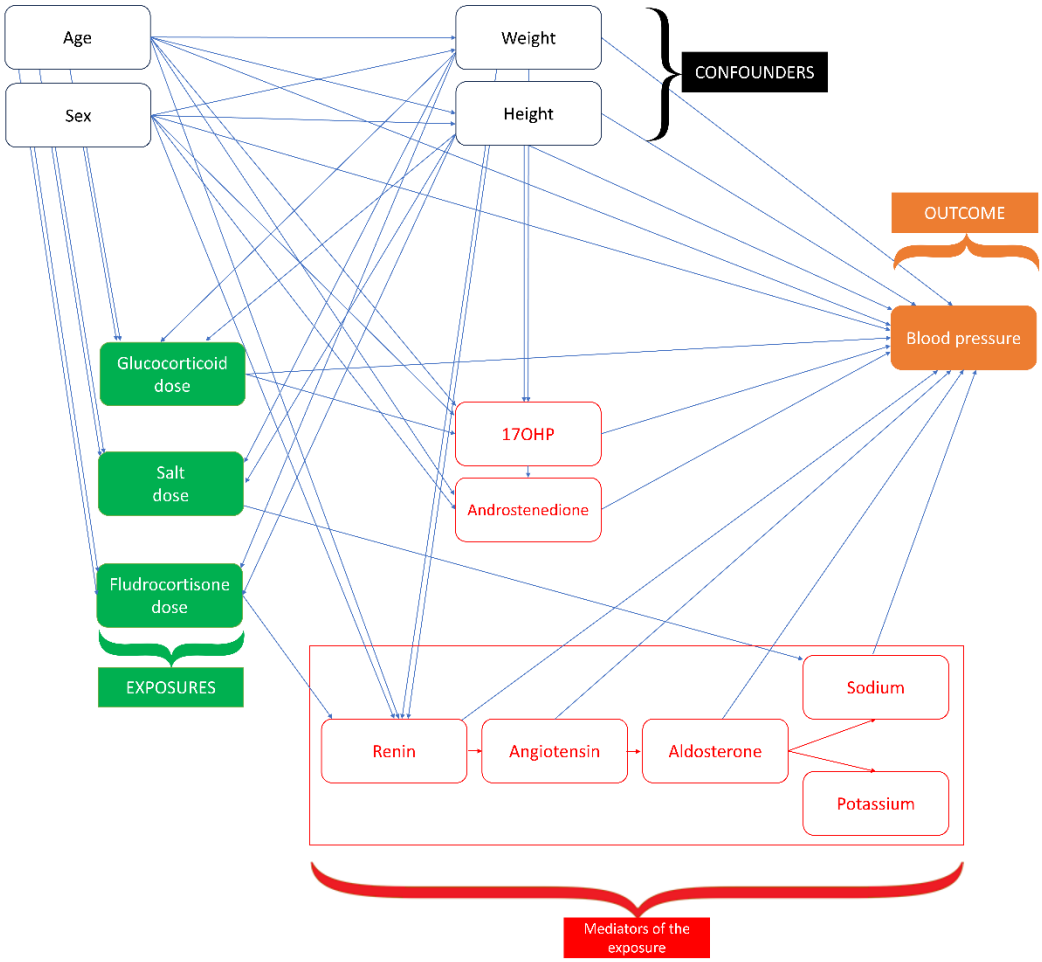
4229 sample size and estimates

Model number: Target of estimation	Sample size with imputation (Patients| Visits)	Sample size without imputation (Patients| Visits)	Estimate with imputation (systolic)	Estimate without imputed data (systolic	Estimate with imputation (diastolic)	Estimate without imputed data (diastolic)
1: Effect of fludrocortisone on BP	452 | 4187	370 | 2204	0.012 (0.005 to 0.020)	0.022 (0.013 to 0.031)	0.008 (0.003 to 0.014)	0.014 (0.006 to 0.021)
2: Degree to which renin predicts BP	544 | 6130	256 | 1027	-1.00 (-1.47 to -0.53)	-1.13 (-1.71 to -0.54)	-0.71 (-1.13 to -0.29)	-0.80 (-1.20 to -0.41)
3: Degree to which 17OHP predicts BP	544 | 6130	288 | 1215	-0.64 (-1.00 to -0.27)	-0.86 (-1.26 to -0.46)	-0.50 (-0.78 to -0.22)	-0.64 (-0.88 to -0.40)
4: degree to which androstenedione predicts BP	543 | 6187	282 | 1229	-0.73 (-1.18 to -0.28)	-0.86 (-1.43 to -0.30)	-0.61 (-0.96 to -0.25)	-0.72 (-0.95 to -0.49)

4230 BP=Blood Pressure. The sign and magnitude of all the estimates of interest are similar, and show a robust data imputation strategy and a

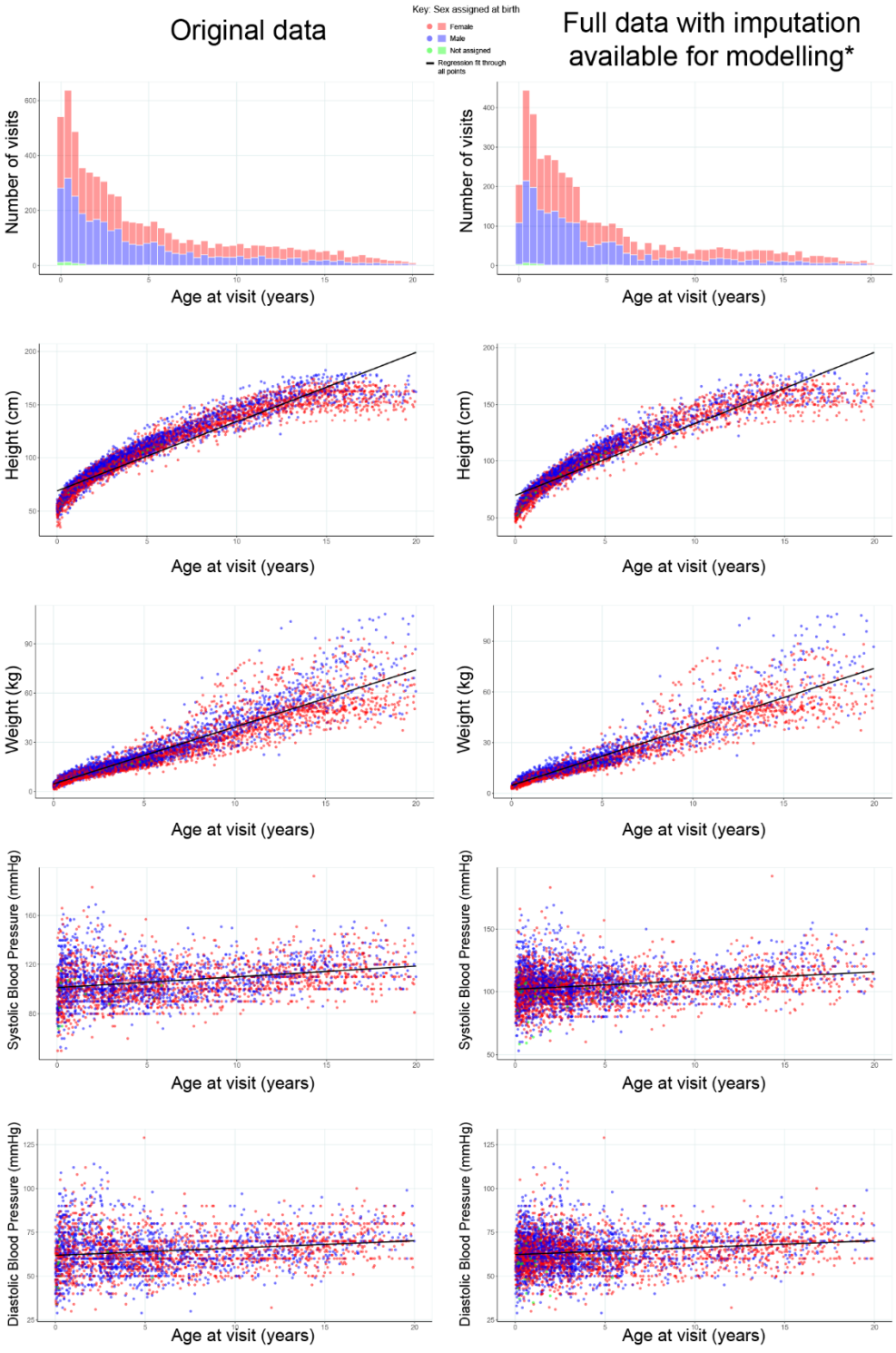
4231 high level of confidence in the effect sizes.

Figure S5.1 – Directed acyclic graph

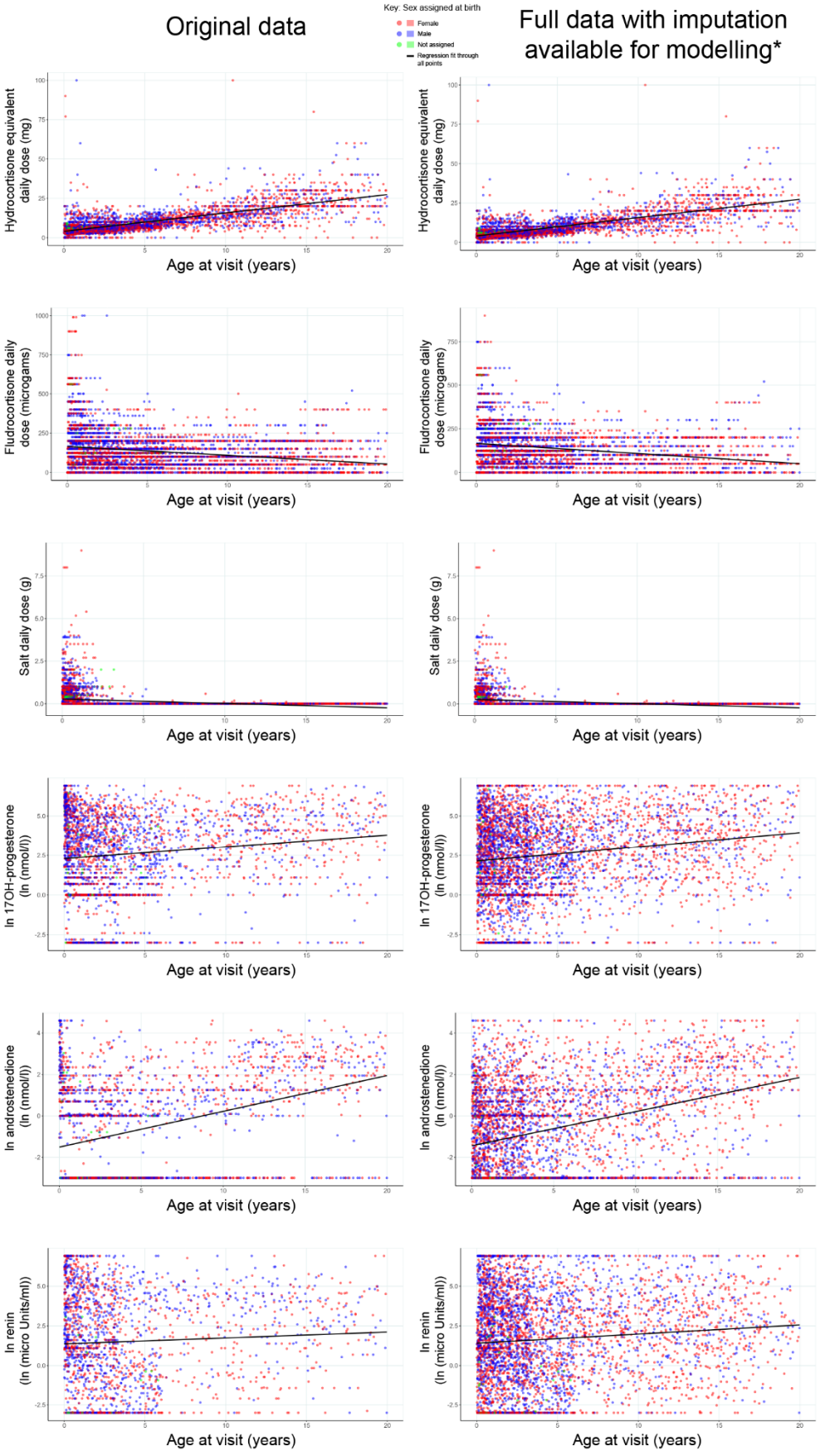


Relationship between variables of interest agreed through expert consensus. Arrow depicts direction of effect, but does not define the size or sign of the effect. Constructed to inform which confounding variables should be controlled for in analysis, whilst avoiding mediating variables or ancestors of variables that would introduce bias when interpreting regression coefficients. e.g. age is a covariate affecting BP and the dose of medication a patient is prescribed, and should thus be controlled for when estimating the effect of dose on BP. However, the effect of salt treatment will be mediated by increasing a patients' sodium, and thus sodium should not be controlled for when assessing the effect of salt on BP. Appropriate covariate adjustment set for each target of estimation reported in table 3.

Figure S5.2 – Histograms and regression of variables within models

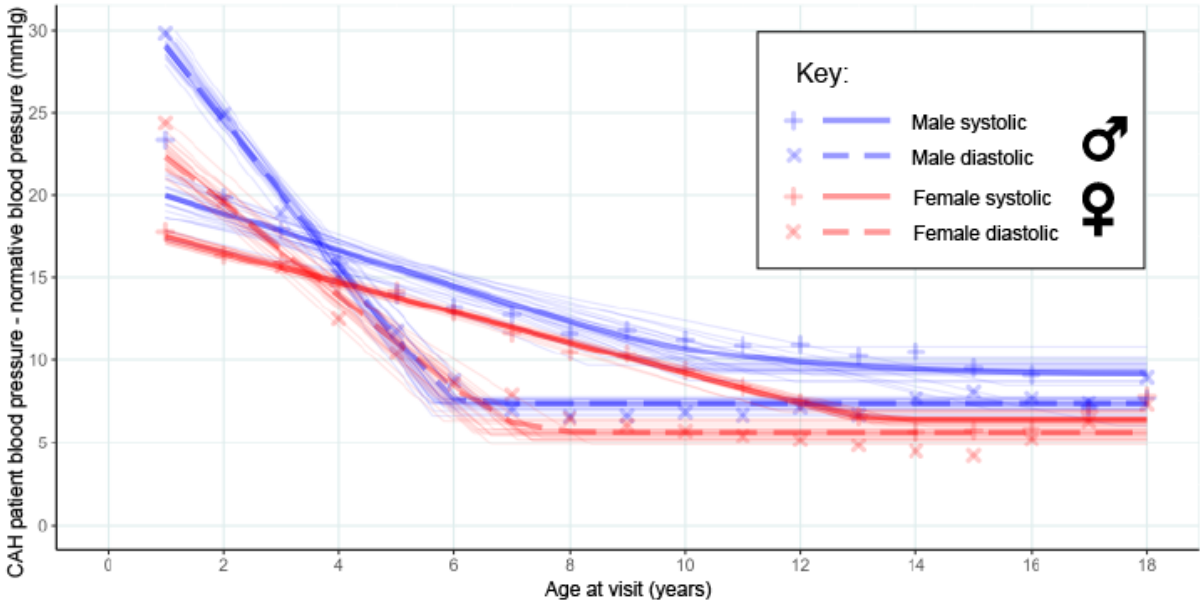


*Registry data contains dose prescribed from that clinic visit. To model the biometrics measured at a clinic visit, the dose the patient was prescribed at the previous visit is used. Height and weight are interpolated between points, but not extrapolated prior to the first data point available or beyond the last data point available. Dose data was imputed using last observation carried forward or next observation carried backward, but restricted to carrying for 6 months under age 1, 12 months under age 3, 24 months under age 5 and 36 months over age 5. Joint modelling multiple imputation of biomarkers and blood pressure only employed to visits with known or interpolated height and weight. Imputation therefore increases the amount available for modelling, but does not facilitate modelling across the entire dataset because of the pragmatic restrictions imposed to ensure the imputation strategy produced plausible data. Right hand column shows 1 imputation set as an example, but please note imputation was carried out 10 times with estimates calculated for each imputation and then collated using Rubin's rules.


Original data
Key: Sex assigned at birth
Female
Male
Not assigned
Regression fit through all points
Full data with imputation available for modelling*
Hydrocortisone equivalent daily dose (mg)
Fludrocortisone daily dose (micrograms)
Salt daily dose (g)
ln 17OH-progesterone (ln (nmol/l))
ln androstenedione (ln (nmol/l))
ln renin (ln (micro Units/ml))
Age at visit (years)

4252

Figure S5.3 – Bayesian multiple change point analysis of difference between CAH patient blood pressure and normative values with age



Difference between median normative healthy child blood pressure values subtracted from median LMS modelled blood pressure of CAH (Congenital Adrenal Hyperplasia) patients with age. CAH patients have a larger difference in blood pressure at earlier ages. Lines show Bayesian multiple change point analysis that estimates the age where this difference stops decreasing to a plateau. Analysis shows the difference in diastolic blood pressure reduces at a faster rate than the difference in systolic blood pressure for both sexes. indicating ages at which the difference between CAH patients and normative values stops decreasing. Plateau of each line shows the amount CAH patients have greater blood pressure than normative values for each reading.

A6 Supplementary Material for Chapter 6

Supplementary methods 6.01

Data preparation

Data was prepared for analysis by removal of duplicated entries, and manual assessment of outlying values using data visualisation, followed by contact with individual centres for clarification about possible data entry errors and units of measurement.

Interpolation of missing height and weight data

Due to varying number of visits within patients, a hierarchical approach was employed to interpolation of missing height and weight. In patients with 11 or more visits in which height and weight were available, a generalised additive model was estimated for height on age with weight as a covariate. Pearson scaled residuals were calculated for each data point within this fit to assess for outliers. Individual model fits were examined visually and an absolute Pearson scaled residual of 3 or more was selected to declare a datapoint as an implausible outlier allowing deletion of that point and imputation. If a datapoint was declared an outlier, the modelling was reiterated without the outlying point before establishing the final GAM model fit that was then used to interpolate missing or outlying data points. In patients with 11 or more visits in which height was available but the covariate weight was not available, the process was repeated to fit a GAM model without the covariate of weight.

In patients with between 4 and 10 known data points for height, a cubic spline with 4 degrees of freedom was fit through the data, without covariate adjustment. Residuals were calculated manually from the spline fit by calculating the difference between the model fit and each data point, and then standardising by dividing them by the standard deviation of the height measurements within that patient. A threshold of greater than 0.5 was used for this standardised residual to declare a data point as an outlier following visual inspection of the model fits. In patients with between 2 and 4 available data points for height, a linear interpolation method was employed. Points were not extrapolated beyond known data points.

A similar process was applied for measurements of weight, with different outlier threshold parameters due to the greater variability in this metric with true values able to decrease as well as increase. The residual threshold for GAM models was thus set at a Pearson residual of 4, and for spline models a standardised residual threshold of 1.2.

Measurements for bone age were interpolated between points without any covariate adjustment, and thus GAM modelling was not used, but strictly spline interpolation when four or more data points were available, and linear interpolation when two or three data points were available.

Carrying of missing dose data

Missing data patterns were reviewed and correlated with other data fields of relevance. The field 'has dose changed from previous visit', evidenced that doses patients were prescribed at that visit going forward were entered into the registry, with a preference for dose changes to be entered into the registry: Patients who were maintained on the same dose were more likely to have their precise dose details missing, with dose changes more likely to include precise dosing details. A pragmatic approach of data imputation using last observation carried forward was therefore employed, with limits on how far data could be carried according to age, restricted to carrying for 6 months under age 1, 12 months under age 3, 24 months under age 5 and 36 months over age 5. If data was not available to carry forward, then next observation carried backward was employed with the same age limits. The dose a patient was on prior to clinic (as defined by the previous visit) was therefore used for modelling to assess the effect of that dose on the patient parameters measured in clinic. Hydrocortisone equivalent dosing was converted using the conversion factors in table S1.

Bootstrapping and uncertainty estimation

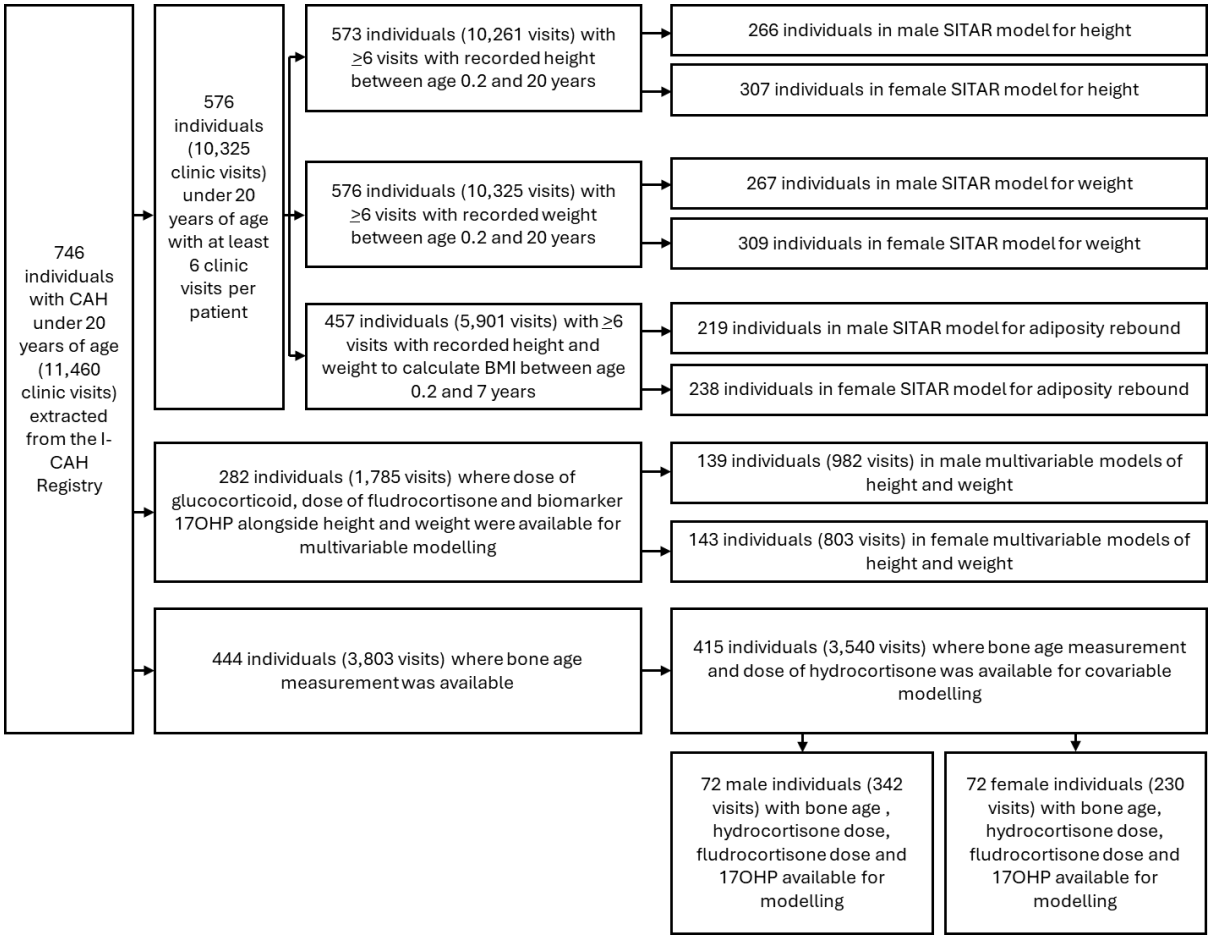
As regression coefficients are difficult to interpret directly from multivariable spline models, clinically meaningful values of dose and disease control were inserted into the models defined by expert consensus. Uncertainty around the estimates provided by these models was defined by using bootstrapping, creating 500 replications of the dataset using sampling with replacement of patient data and calculating the difference between patients with clinically meaningful values for each replication. The 95% confidence interval of each estimate was calculated using the 2.5^{th} and 97.5^{th} centile of estimates across all bootstrapped replications. Variations on the clinically meaningful parameters are presented in table S7.

4320 Table S6.01 – Conversion factors

Conversion factors for glucocorticoid equivalent dosing					
Preparation of glucocorticoid	**Multiplication factor to create hydrocortisone equivalent**				
Hydrocortisone	1				
Prednisolone	4				
Dexamethasone	27				
Cortisone acetate	0.8				
Methylprednisolone	5				
Conversion factors for biomarker unit equivalence					
Marker	**Original unit**	**Unit to convert to**	**Multiplication factor**	**Standardised lower limit of detection**	**Standardised upper limit of detection**
17OHP	ng/ml	nmol/l	0.0303	0.05nmol/l	1000nmol/l
Androstenedione	ng/dl	nmol/l	0.0349	0.05nmol/l	100nmol/l

4321

Figure S6.01 – Data flow diagram



Flow diagram indicating proportion of registry data extracted that had sufficient variables for each model assessment. CAH=Congenital adrenal hyperplasia, SITAR=SuperImposition by Translation And Rotation, 17OHP=17-hydroxyprogesterone

Table S6.02 – Full frequency table of preparations of glucocorticoid prescribed at visit

Preparation	All Visits n (%)	Male visits n (%)	Female visits n (%)
Hydrocortisone	9731 (84.9%)	4996 (87.7%)	4735 (82.1%)
Prednisolone	50 (0.4%)	1 (0.1%)	49 (0.9%)
Prednisone	50 (0.4%)	9 (0.2%)	41 (0.7%)
Dexamethasone	48 (0.4%)	17 (0.3%)	31 (0.5%)
Cortisone Acetate	108 (1.0%)	49 (0.9%)	59 (1.0%)
Mixed dosing	144 (1.3%)	62 (1.1%)	82 (1.4%)
None	135 (1.2%)	81 (1.4%)	54 (0.9%)
Unknown	1194 (10.4%)	475 (8.3%)	719 (12.5%)
Total	11460	5690	5770

Table S6.03 – SITAR model estimates and basis function parameters

	Male patients			Female patients		
SITAR outcome metric:	**Height (cm)**	**Weight (kg)**	**BMI (kg/m^2)**	**Height (cm)**	**Weight (kg)**	**BMI (kg/m^2)**
Number of patients used to fit SITAR	266	267	219	307	309	238
Number of data points used to fit SITAR	5196	5224	3030	5065	5101	2871
SITAR degrees of freedom*	5	6	3	5	6	3
SITAR internal knots (ln (years))	-0.735, -0.098, 0.342, 0.717	-0.896, -0.284, 0.136, 0.472, 0.793	-0.241, 0.392	-0.761, -0.111, 0.356, 0.748	-0.905, -0.313, 0.143, 0.477, 0.817	-0.256, 0.383
SITAR boundary knots (ln (years))	-1.888, 1.449	-1.928, 1.467	-1.356, 0.929	-1.900, 1.426	-1.894, 1.432	-1.345, 0.950
SITAR fixed spline coefficient 1	48.921	10.324	16.492	37.613	11.457	-1.651
SITAR fixed spline coefficient 2	70.033	18.284	68.933	53.792	24.025	5.607
SITAR fixed spline coefficient 3	111.617	31.194	14.959	77.711	57.395	1.51
SITAR fixed spline coefficient 4	137.554	51.196	-	95.257	76.904	-
SITAR fixed spline coefficient 5	107.238	69.068	-	73.475	83.846	-
SITAR fixed spline coefficient 6		74.172	-		80.107	-
SITAR fixed spline coefficient a ('Size')	52.934	10.091	-14.575	75.979	5.206	15.314
SITAR fixed spline coefficient b ('Timing')	-0.005	0.342	-0.485	0.289	0.399	-0.030
SITAR fixed spline coefficient c ('Intensity')	-0.158	0.313	-0.841	0.384	-0.241	-0.161
SITAR random effect a ('Size') standard deviation	18.202	4.705	1.486	8.883	4.912	1.516
SITAR random effect b ('Timing') standard deviation	0.432	0.305	0.396	0.212	0.310	0.275
SITAR random effect c ('Intensity') standard deviation	0.263	0.351	0.262	0.113	0.438	0.405

SITAR=SuperImposition by Translation and Rotation. Models are fit on the natural log of age and then results converted back onto the absolute scale for purposes of plotting. All code available at https://github.com/neilxlawrence/I-CAH_Adiposity_rebound

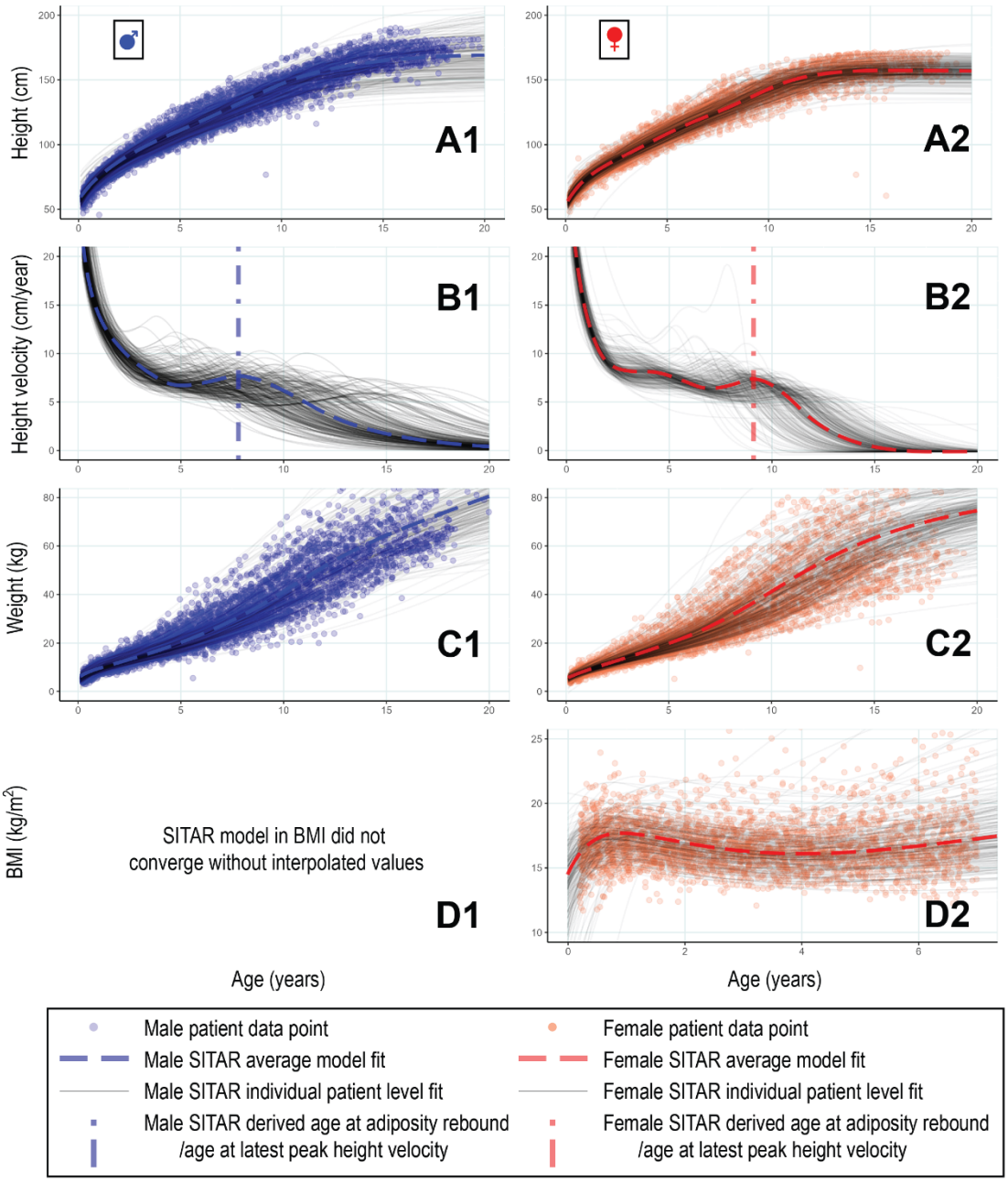
Table S6.04 – SITAR Model derived metrics of growth interpolated values excluded from modelling

	Male patients			Female patients		
Age at latest peak height velocity (years) Mean (SD)	8.1 (2.8)			9.1 (1.5)		
Latest peak height velocity (cm/year) Mean (SD)	7.8 (1.4)			7.5 (1.4)		
Age at adiposity rebound Mean (SD)	*did not converge*			4.1 (1.3)		
BMI at adiposity rebound Mean (SD)	*did not converge*			16.0 (1.4)		
SITAR outcome variable	**Height (cm)**	**Weight (kg)**	**BMI (kg/m^2)**	**Height (cm)**	**Weight (kg)**	**BMI (kg/m^2)**
Number of patients used to fit SITAR model	264	265	*did not converge*	303	306	230
Number of data points used to fit SITAR model	5006	5082	*did not converge*	4859	4956	2720
SITAR predicted value at age 18 Mean (SD)	166.0 (13.6)	74.0 (13.8)	*did not converge*	156.8 (6.7)	69.0 (9.8)	28.0 (4.5)
WHO z-score of SITAR predicted value at age 18 Mean (SD)*	-1.35 (1.81)	-	*did not converge*	-1.0 (1.0)	-	1.6 (1.0)
SITAR degrees of freedom	5	6	3	5	6	3
Restricted age range of patients to fit SITAR (years)	0.2 to 20	0.2 to 20	0.2 to 7	0.2 to 20	0.2 to 20	0.2 to 7
Root Mean Square Error	1.79	1.53	*did not converge*	2.66	2.14	0.894
Marginal R^2	0.75	0.94	*did not converge*	0.91	0.91	0.06
Conditional R^2	0.99	0.99	*did not converge*	0.99	0.99	0.80

This table reports the metrics calculated within table 2 of the main manuscript, excluding any interpolated values in modelling. SD=Standard deviation, variability defined by between patient random effects. SITAR: SuperImposition by Translation And Rotation. Optimum degrees of freedom defined by lowest Bayesian Information Criterion. Patients' included in models if they had six or more measures of the necessary biometric.

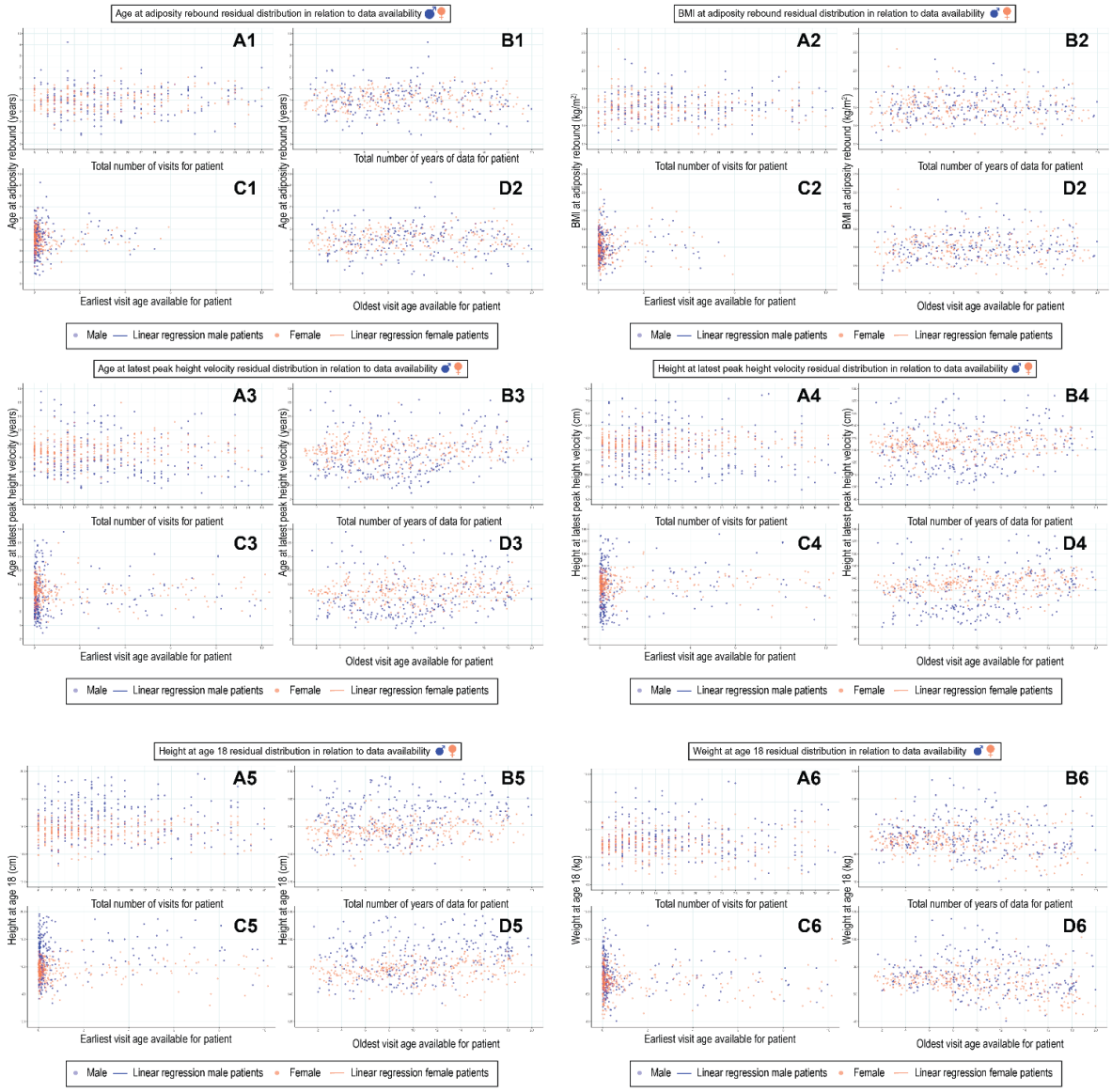
*WHO do not provide reference standards for weight for age beyond 10 years

Figure S6.02 – SITAR model results estimated without interpolated values



All models estimated without the use of any interpolated values. Prefix A = Height; B = Height Velocity; C = Weight; D = BMI. Suffix 1=Male; 2=Female. SITAR: SuperImposition by Translation And Rotation. SITAR model of BMI in males did not converge without interpolated values.

Figure S6.03 – Residual plots of model derived metrics



Patients contributed data to the study within the I-CAH Registry at variable points in their childhood. These plots assess the SITAR model derived metrics related to BMI, height, and weight to ensure no heteroskedasticity dependent upon the age or amount of data a patient contributes to the model.

Prefix letter refers to metric of data availability for each patient: A=Total number of visits for patient, B=Total number of years of data for patient from first to last visit, C=Age at earliest visit of patient, D=Age at oldest visit of patient

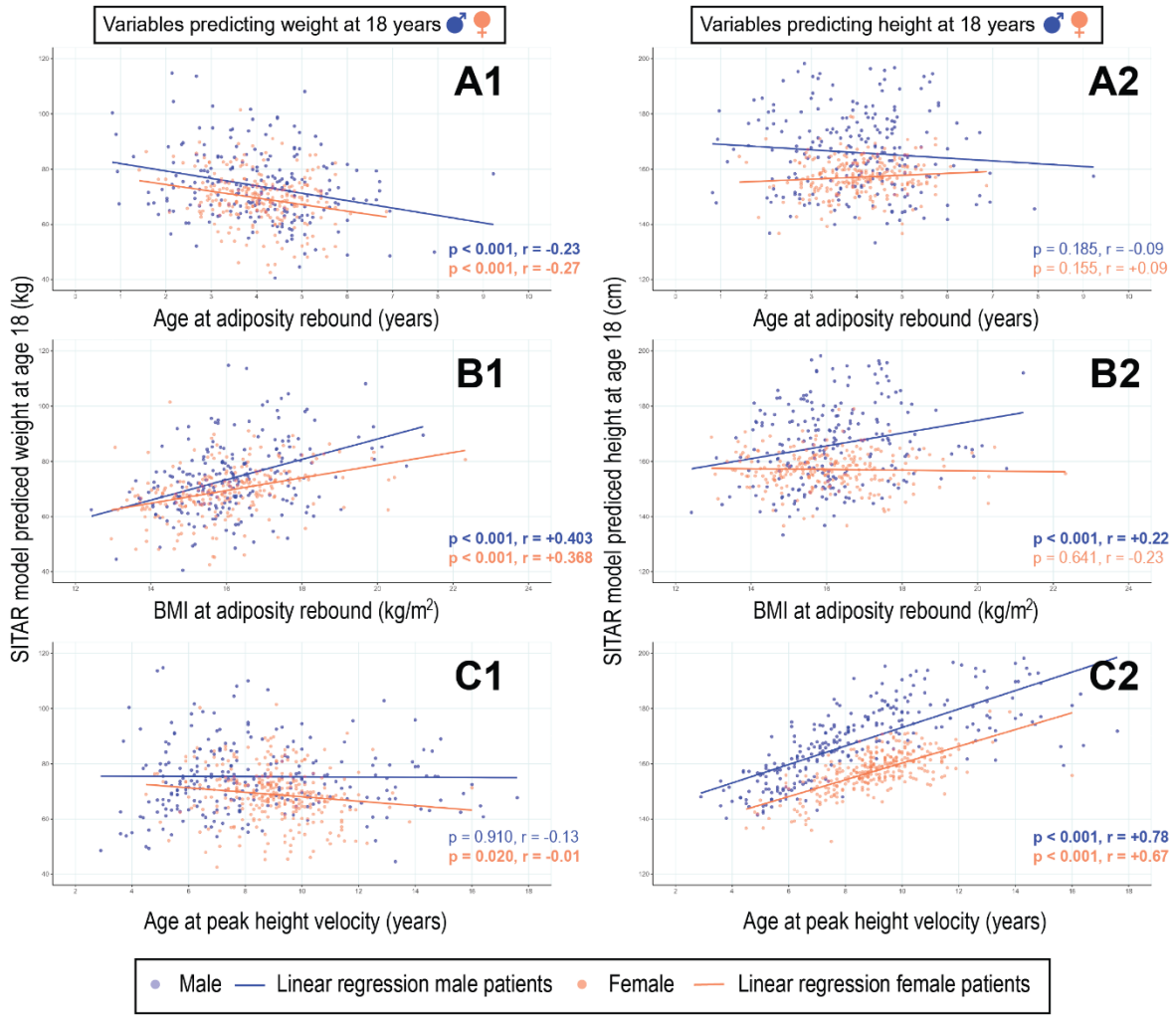
Suffix number refers to metric derived for each patient: 1=Age at adiposity rebound, 2=BMI at adiposity rebound, 3=Age at latest peak height velocity, 4=Height at latest peak height velocity, 5=SITAR model predicted height at 18 years of age, 6=SITAR model predicted weight at 18 years of age

Table S6.05 – Association between metrics of adiposity rebound and latest peak height velocity to predict height and weight at 18 years

Models predicting SITAR estimated height at 18 years of age (cm)										
	Male model estimates (95% CI) [*p-value*]					Female model estimates (95% CI) [*p-value*]				
Predictor variable:	n	Intercept	Slope	R^2	Model *p*	n	Intercept	Slope	R^2	Model *p*
Age at adiposity rebound (years)	219	170.0 (163.7 to 176.2) [p<0.001]	-1.0 (-2.5 to 0.5) [p=0.185]	0.01	0.185	238	154.3 (150.4 to 158.3) [p<0.001]	0.7 (-0.3 to 1.6) [p=0.156]	0.01	0.155
BMI at adiposity rebound (mg/m^2)	219	128.6 (107.4 to 150.0) [p<0.001]	**2.3 (1.0 to 3.6)** [p<0.001]	0.05	<0.001	238	159.3 (150.0 to 168.6) [p<0.001]	-0.1 (-0.7 to 0.4) [p=0.640]	0.01	0.641
Age at peak height velocity (years)	258	139.7 (136.3 to 143.1) [p<0.001]	**3.3 (3.0 to 3.7)** [p<0.001]	0.53	<0.001	307	130.0 (126.6 to 133.4) [p<0.001]	**3.0 (2.7 to 3.4)** [p<0.001]	0.45	<0.001
Models predicting SITAR estimated weight at 18 years of age (kg)										
	Male model estimates (95% CI) [*p-value*]					Female model estimates (95% CI) [*p-value*]				
Predictor variable:	n	Intercept	Slope	R^2	Model p	n	Intercept	Slope	R^2	Model p
Age at adiposity rebound (years)	219	84.7 (79.2 to 90.1) [p<0.001]	**-2.7 (-4.0 to -1.4)** [p<0.001]	0.07	<0.001	238	79.2 (73.9 to 84.6) [p<0.001]	**-2.4 (-3.7 to -1.1)** [p<0.001]	0.05	<0.001
BMI at adiposity rebound (mg/m^2)	219	14.5 (-3.6 to 32.5) [p=0.118]	**3.7 (2.6 to 4.8)** [p<0.001]	0.16	<0.001	238	32.7 (20.8 to 44.7) [p<0.001]	**2.3 (1.6 to 3.0)** [p<0.001]	0.14	<0.001
Age at peak height velocity (years)	258	75.7 (69.5 to 81.8) [p<0.001]	-0.04 (-0.731 to 0.651) [p=0.910]	0.01	0.910	307	76.1 (70.0 to 82.3) [p<0.001]	**-0.8 (-1.5 to -0.1)** [p=0.020]	0.02	0.020

CI=Confidence Interval. Statistically significant slopes highlighted in bold type

Figure S6.04 – Association between adiposity rebound and timing of pubertal growth spurt to predict height and weight at 18 years



Prefix (predictors): A=Age at adiposity rebound as predictor; B=BMI at adiposity rebound; C=Age at peak height velocity. Suffix (outcome): 1=SITAR model predicted weight at age 18; 2=SITAR model predicted height at age 18. SITAR=SuperImposition by Translation And Rotation

4371 Table S6.06 – Full multivariable model coefficients to predict Height covarying

4372 by weight, dose of medications and disease control

Dependent variable = Height	Male model estimate (95% CI)	Female model estimate (95% CI)
Number of patients	139	143
Number of visits	982	803
Intercept	44.7 (40.3 to 48.5)	43.7 (40.2 to 47.0)
ln age spline basis 1	52.7 (45.5 to 60.8)	50.2 (39.6 to 59.7)
ln age spline basis 2	106.9 (94.3 to 119.8)	112.0 (107.0 to 132.3)
ln age spline basis 3	100.6 (87.5 to 114.2)	105.8 (90.3 to 121.4)
Weight	2.0 (1.4 to 2.7)	1.94 (1.26 to 2.63)
Weight : ln age spline basis 1	-1.3 (-1.8 to -0.7)	-0.81 (-1.54 to -0.07)
Weight : ln age spline basis 2	-2.2 (-3.3 to -1.2)	-2.943 (-4.31 to -1.62)
Weight : ln age spline basis 3	-1.5 (-2.0 to -1.0)	-1.13 (-1.75 to -0.54)
ln 17ohp	0.05 (-0.26 to 0.42)	-0.08 (-0.39 to 0.19)
ln 17OHP : ln age spline basis 1	0.47 (0.02 to 0.89)	0.51 (-0.18 to 1.09)
ln 17OHP : ln age spline basis 2	-0.16 (-1.28 to 0.85)	0.42 (-0.31 to 1.19)
ln 17OHP : ln age spline basis 3	-0.31 (-1.35 to 0.68)	0.09 (-0.49 to 0.62)
hydrocortisone equivalent	-0.079 (-0.29 to 0.16)	-0.04 (-0.19 to 0.24)
hydrocortisone equivalent : ln age spline basis 1	-0.032 (-0.37 to 0.23)	-0.33 (-0.73 to 0.02)
hydrocortisone equivalent : ln age spline basis 2	0.45 (-0.05 to 0.99)	0.43 (-0.30 to 0.90)
hydrocortisone equivalent : ln age spline basis 3	0.13 (-0.19 to 0.42)	-0.10 (-0.42 to 0.24)
fludrocortisone	-0.003 (-0.014 to 0.007)	-0.009 (-0.019 to -0.001)
fludrocortisone : ln age spline basis 1	-0.023 (-0.049 to 0)	-0.001 (-0.037 to 0.036)
fludrocortisone : ln age spline basis 2	0.001 (-0.031 to 0.031)	0.023 (-0.009 to 0.052)
fludrocortisone : ln age spline basis 3	0.005 (-0.021 to 0.033)	-0.001 (-0.024 to 0.022)
marginal R^2	0.982 (0.977 to 0.987)	0.979 (0.973 to 0.985)
conditional R^2	0.995 (0.994 to 0.996)	0.995 (0.994 to 0.996)
SD of random intercept for centre	0.83	0.76
SD of random intercept for id	3.14	3.59

4373 17OHP=17-Hydroxyprogesterone. Units employed within models: Height=cm, Weight=kg, 17OHP=nmol/l, hydrocortisone

4374 equivalent=mg, fludrocortisone=micrograms. Age modelled as a spline basis function on the natural log scale with 3

4375 degrees of freedom, interacted with each variable of height, dose of medications and disease control

4376 Table S6.07 – Full multivariate model coefficients to predict Weight, covarying

4377 by height, dose of medications and disease control

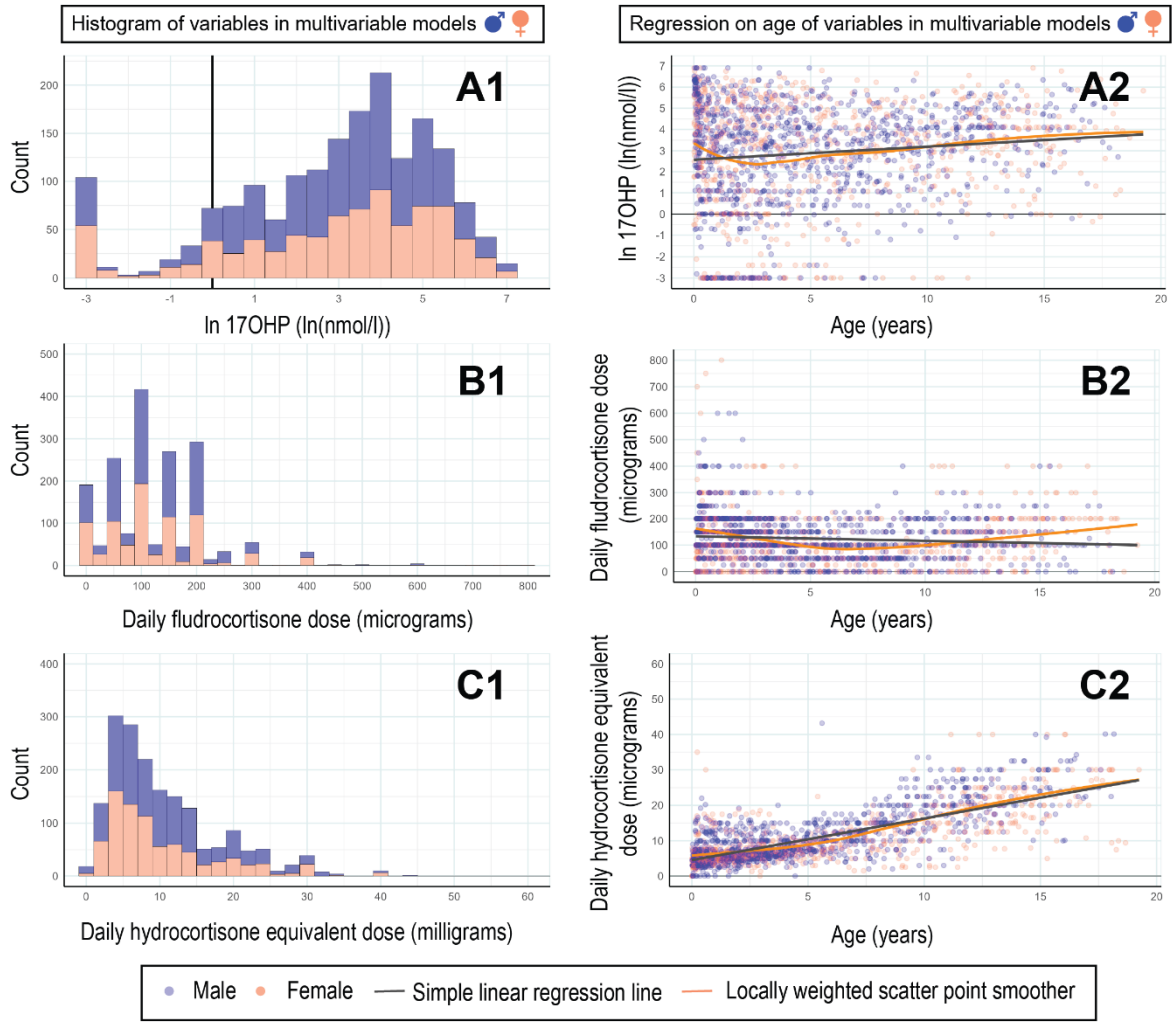
Dependent variable = Weight	Male model estimate (95% CI)	Female model estimate (95% CI)
Number of patients	139	143
Number of visits	982	803
Intercept	4.8 (-10.2 to 17.4)	-1.2 (-12.6 to 11.8)
ln age spline basis 1	-53.9 (-77.1 to -28.8)	-20.2 (-56.3 to 13.5)
ln age spline basis 2	-88.4 (-118.9 to -54.2)	-33.9 (-75.4 to 8.2)
ln age spline basis 3	-92.4 (-143.2 to -33.1)	-96.7 (-142.9 to -46.4)
Height	0.02 (-0.20 to 0.29)	0.02 (-0.23 to 0.26)
Height : ln age spline basis 1	0.55 (0.20 to 0.84)	0.36 (0.05 to 0.71)
Height : ln age spline basis 2	0.97 (0.60 to 1.32)	0.52 (0.09 to 1.01)
Height : ln age spline basis 3	0.76 (0.33 to 1.14)	0.90 (0.52 to 1.27)
ln 17ohp	-0.46 (-0.83 to -0.11)	0.19 (-0.22 to 0.73)
ln 17OHP : ln age spline basis 1	-0.12 (-0.64 to 0.37)	-0.57 (-1.85 to 0.50)
ln 17OHP : ln age spline basis 2	1.26 (-0.02 to 2.61)	0.37 (-1.50 to 2.17)
ln 17OHP : ln age spline basis 3	0.93 (-0.15 to 2.17)	0.68 (-1.13 to 2.71)
hydrocortisone equivalent	0.02 (-0.20 to 0.26)	0.07 (-0.12 to 0.36)
hydrocortisone equivalent : ln age spline basis 1	0.20 (-0.19 to 0.68)	0.50 (-0.09 to 1.07)
hydrocortisone equivalent : ln age spline basis 2	-0.10 (-1.01 to 0.68)	-0.36 (-1.26 to 0.58)
hydrocortisone equivalent : ln age spline basis 3	0.00 (-0.65 to 0.68)	0.56 (0.03 to 1.16)
fludrocortisone	-0.001 (-0.011 to 0.007)	0.019 (0.001 to 0.035)
fludrocortisone : ln age spline basis 1	0.025 (0 to 0.053)	0.023 (-0.036 to 0.079)
fludrocortisone : ln age spline basis 2	0.007 (-0.023 to 0.043)	-0.039 (-0.079 to 0.009)
fludrocortisone : ln age spline basis 3	0.009 (-0.027 to 0.048)	-0.02 (-0.052 to 0.024)
marginal R^2	0.938 (0.922 to 0.954)	0.875 (0.834 to 0.916)
conditional R^2	0.973 (0.964 to 0.981)	0.964 (0.951 to 0.979)
SD of random intercept for centre	0.61	0.89
SD of random intercept for id	2.58	4.30

4378 17OHP=17-Hydroxyprogesterone. Units employed within models: Height=cm, Weight=kg, 17OHP=nmol/l, hydrocortisone

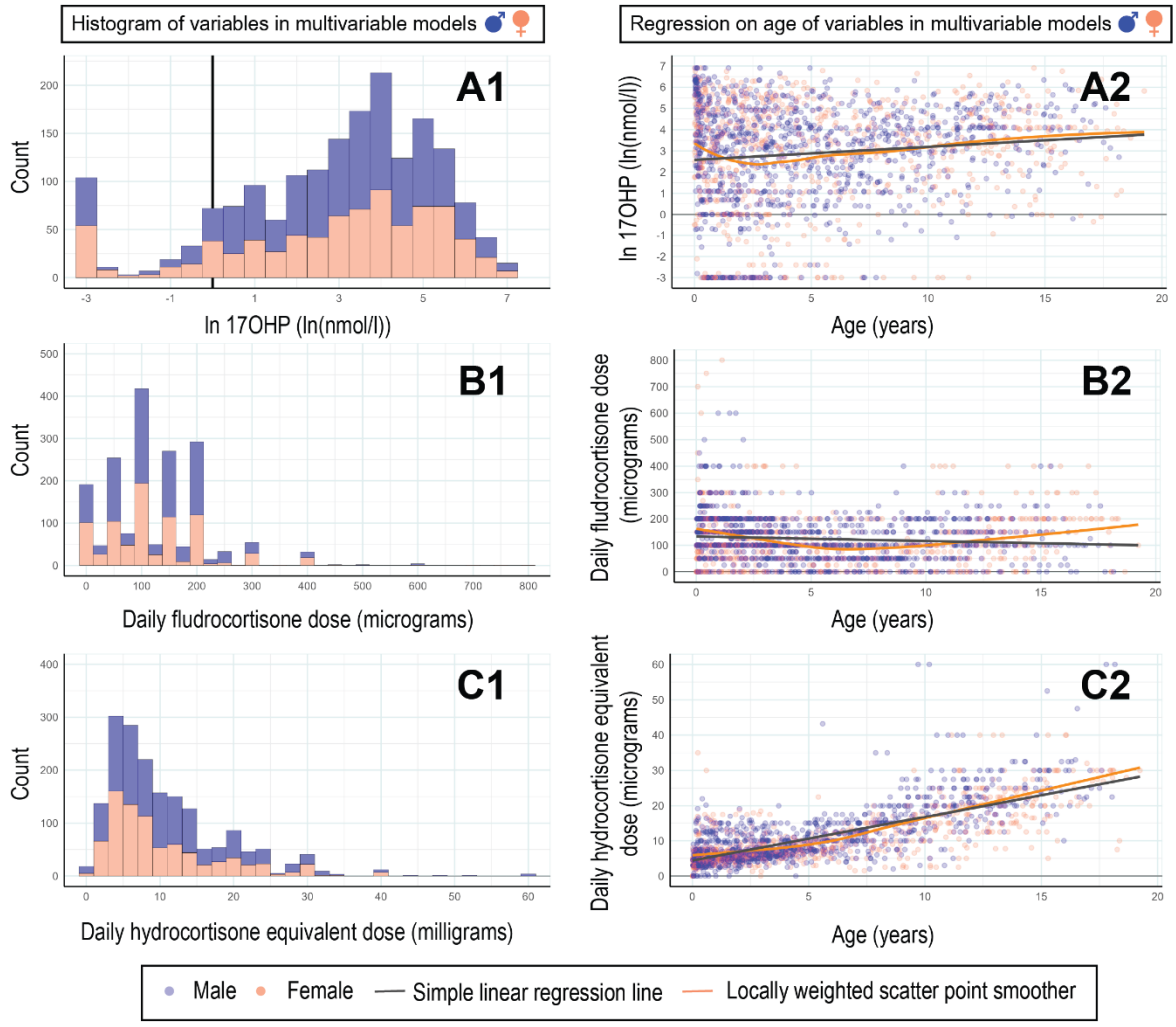
4379 equivalent=mg, fludrocortisone=micrograms. Age modelled as a spline basis function on the natural log scale with 3

4380 degrees of freedom, interacted with each variable of height, dose of medications and disease control

4381 Figure S6.05 – Histograms and regression of variables within multivariate mode



4382



4383

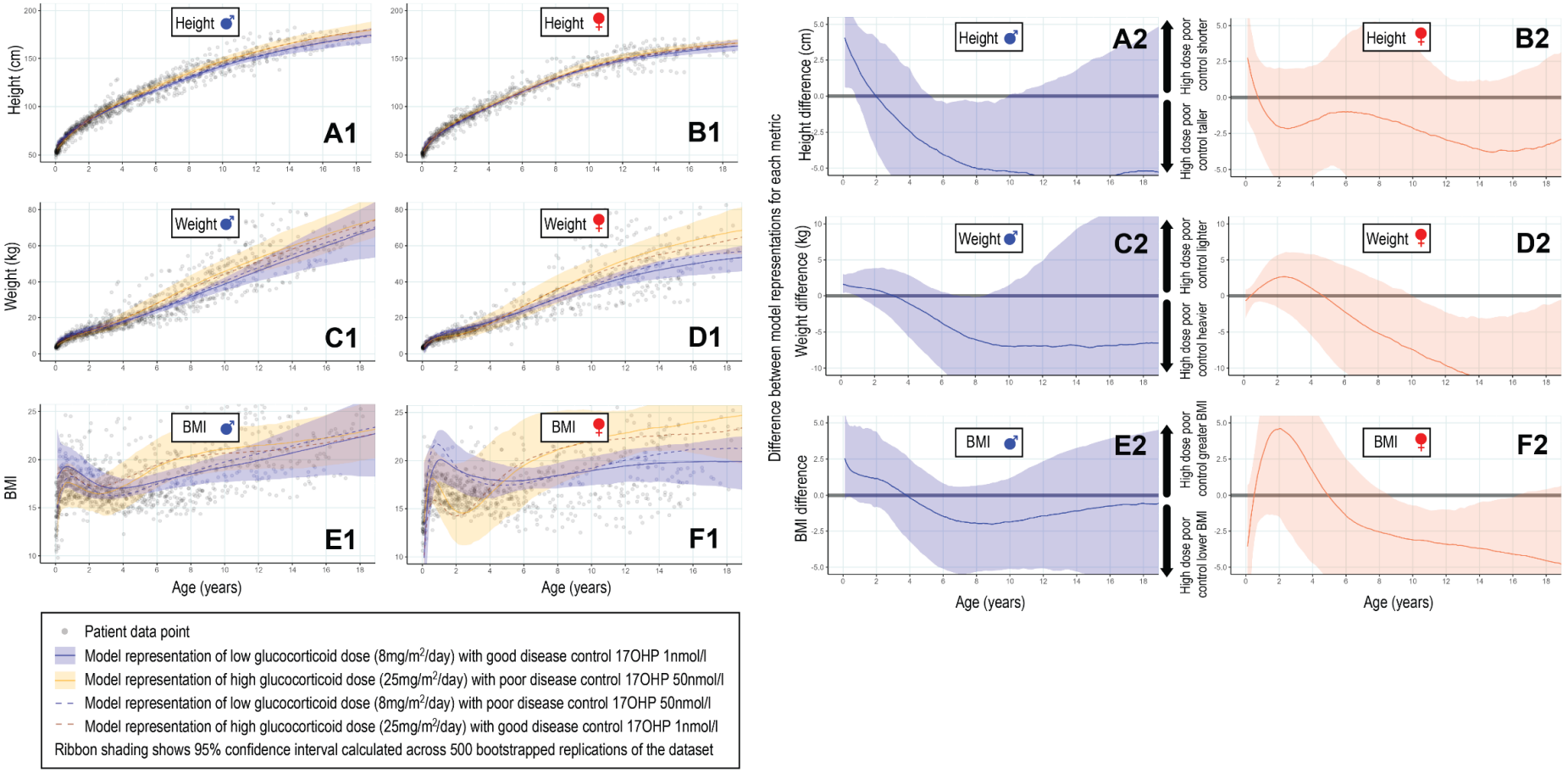
4384 17OHP=17-Hydroxyprogesterone

Table S6.08 – Covariate adjusted multivariate model metrics assessing disease associations with growth with alternative example values

Sex of patients in underlying model	Male					Female			
Data points with covariates available to train model	**982 Visits | 139 patients**					**803 Visits | 143 patients**			
Target of estimation	**Low dose good control**	**Low dose poor control**	**High dose good control**	**High dose poor control**		**Low dose good control**	**Low dose poor control**	**High dose good control**	**High dose poor control**
Metrics inserted to demonstrate covariate model fit									
Hydrocortisone equivalent (mg/m^2/day)	**8**	**8**	**25**	**25**		**8**	**8**	**25**	**25**
Fludrocortisone dose (micrograms/day)	**100**	**100**	**100**	**100**		**100**	**100**	**100**	**100**
17-Hydroxyprogesterone (nmol/l)	**1**	**50**	**1**	**50**		**1**	**50**	**1**	**50**
Model derived estimates: (* indicates statistically significant difference between those on low dose with good control versus those with high dose and poor control)									
Height at age 8 (95% CI)	130.2 (127.8 to 132.4)	131.6 (129.4 to 133.4)	133.8 (130.9 to 136.8)	135.2 (131.9 to 138.5)		129.0 (126.9 to 131.3)	129.6 (127.9 to 131.2)	129.7 (125.3 to 134.0)	130.3 (125.4 to 134.8)
Height SDS at age 8 (95% CI)	0.5 (0.1 to 0.9)	0.8 (0.4 to 1.1)	1.2 (0.7 to 1.7)	1.4 (0.8 to 2.0)		0.4 (0.1 to 0.8)	0.5 (0.2 to 0.8)	0.5 (-0.2 to 1.3)	0.6 (-0.2 to 1.4)
Weight at age 8 (95% CI)	31.4 (28.8 to 33.5)	32.4 (29.7 to 34.8)	36.4 (32.0 to 42.2)	37.5 (32.3 to 43.8)		30.7 (28.5 to 33.0)	31.0 (28.7 to 33.1)	35.8 (27.7 to 43.5)	35.9 (27.9 to 43.9)
Weight SDS at age 8 (95% CI)	1.4 (0.8 to 1.8)	1.6 (1.0 to 2.0)	2.3 (1.5 to 3.1)	2.4 (1.6 to 3.3)		1.2 (0.8 to 1.6)	1.2 (0.8 to 1.6)	2.0 (0.6 to 2.9)	2.0 (0.7 to 3.0)
BMI at age 8 (95% CI)	18.5 (17.3 to 19.5)	18.7 (17.6 to 19.8)	20.3 (18.2 to 23.1)	20.4 (18.3 to 23.2)		18.5 (17.1 to 19.8)	18.4 (17.3 to 19.6)	21.2 (17.1 to 25.6)	21.2 (17.5 to 25.1)
BMI SDS at age 8 (95% CI)	1.5 (0.9 to 1.9)	1.6 (1.1 to 2.1)	2.2 (1.4 to 3.1)	2.3 (1.4 to 3.1)		1.3 (0.7 to 1.8)	1.3 (0.8 to 1.7)	2.2 (0.7 to 3.2)	2.2 (0.9 to 3.1)
Height at age 12 (95% CI)	152.0 (148.1 to 155.3)	152.5 (149.7 to 155.1)	156.9 (152.2 to 162.0)	157.4 (153.0 to 162.9)		147.7 (145.6 to 150.1)	149.0 (147.2 to 150.9)	149.4 (144.6 to 154.1)	150.6 (146.4 to 155.2)
Height SDS at age 12 (95% CI)	0.4 (-0.1 to 0.9)	0.5 (0.1 to 0.9)	1.1 (0.4 to 1.8)	1.2 (0.6 to 2.0)		-0.5 (-0.8 to -0.2)	-0.3 (-0.6 to -0.1)	-0.3 (-1.0 to 0.4)	-0.1 (-0.7 to 0.6)
Weight at age 12 (95% CI)	46.1 (40.5 to 51.7)	47.9 (42.2 to 53.4)	51.2 (44.0 to 58.7)	53.1 (45.5 to 60.9)		42.5 (38.4 to 46.6)	44.5 (41.1 to 48.1)	49.9 (40.1 to 58.6)	52.2 (44.5 to 59.4)
BMI at age 12 (95% CI)	19.9 (17.9 to 21.8)	20.6 (18.5 to 22.3)	20.9 (18.3 to 23.1)	21.5 (19.0 to 23.5)		19.5 (17.8 to 21.2)	20.1 (18.8 to 21.5)	22.3 (19.0 to 25.2)	22.9 (20.4 to 25.2)
BMI SDS at age 12 (95% CI)	1.0 (0.2 to 1.6)	1.2 (0.4 to 1.7)	1.3 (0.4 to 1.9)	1.5 (0.7 to 2.0)		0.6 (-0.1 to 1.1)	0.8 (0.3 to 1.2)	1.4 (0.4 to 2.1)	1.6 (0.9 to 2.0)
Age at adiposity rebound (95% CI)	3.7 (2.3 to 5.6)	3.5 (2.4 to 4.5)	2.5 (1.3 to 3.9)	2.6 (0.5 to 4.6)		4.8 (2.1 to 8.8)	4.7 (2.6 to 6.6)	2.0 (1.6 to 8.2)	2.3 (1.6 to 3.2)
BMI at adiposity rebound (95% CI)	16.9 (15.8 to 17.9)	16.2 (15.2 to 17.1)	17.0 (14.9 to 18.6)	16.2 (13.9 to 18.3)		17.7 (16.0 to 19.2)	17.3 (16.0 to 18.4)	14.2 (10.6 to 21.0)	14.4 (11.2 to 19.5)

BMI z-score at adiposity rebound (95% CI)	1.1 (0.1 to 1.8)	0.6 (-0.3 to 1.2)	0.9 (-0.7 to 1.94)	0.28 (-1.8 to 1.4)		1.4 (0.4 to 2.2)	1.2 (0.5 to 1.8)	-1.2 (-4.7 to 2.7)	-0.9 (-4.1 to 2.5)

4387 Figure S6.06 – Visualisations of covariate adjusted models with alternative example values



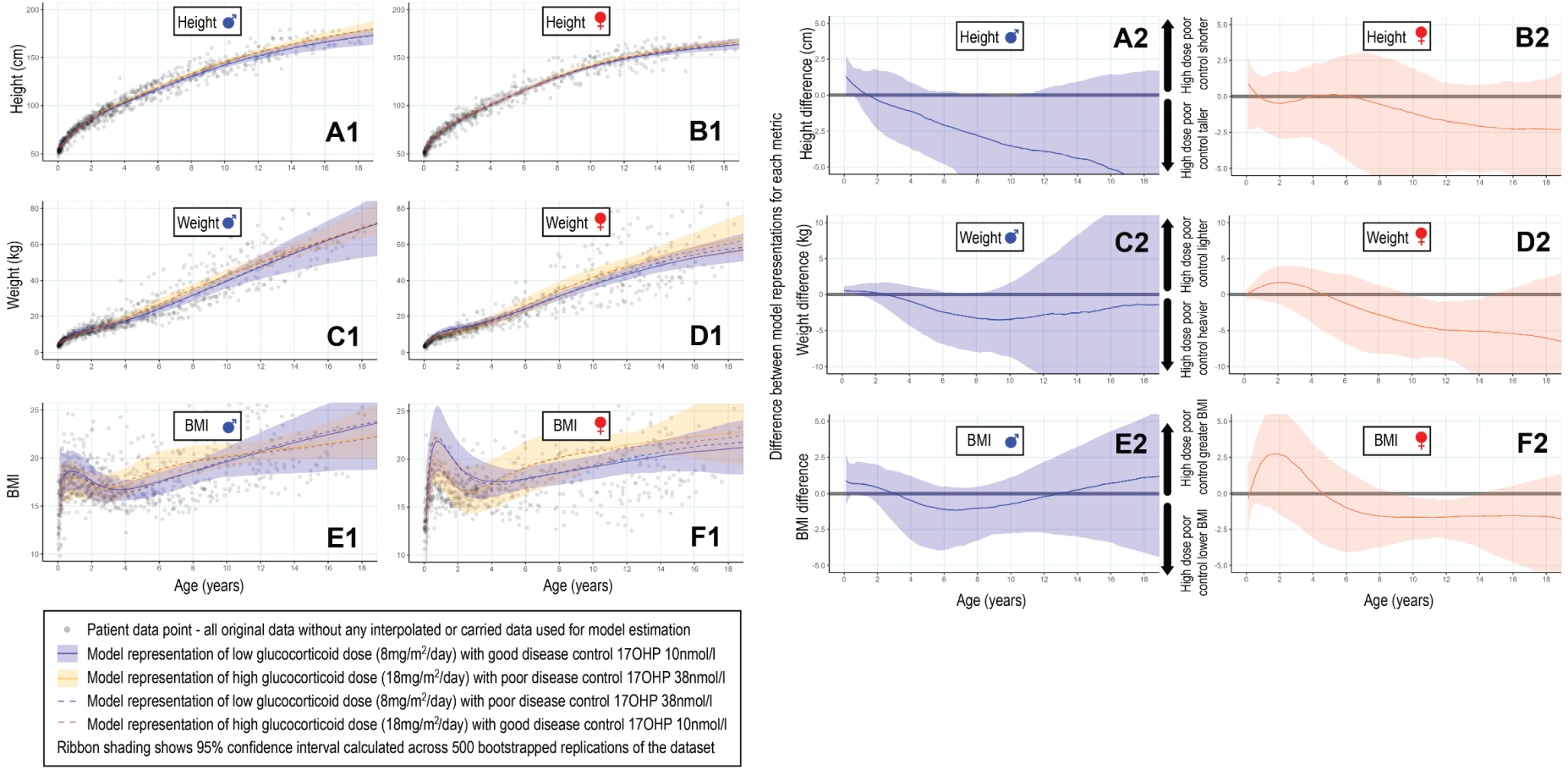
4388

4389 Prefix figure key: A = Male Height; B = Female Height; C = Male weight; D = Female weight. E = Male BMI; F = Female BMI.
4390 Suffix figure key: Figures ending '1' show trajectories of growth for modelled values contained within Appendix Table S6.7 as an alternative set of example comparisons to those presented in the main manuscript.

Table S6.09 – Covariate adjusted multivariate model metrics assessing disease associations with growth estimated without any interpolated or carried data

Sex of patients in underlying model	Male					Female			
Data points with covariates available to train model	745 visits | 130 patients (no interpolations or carried data)					618 visits |134 patients (no interpolations or carried data)			
Target of estimation	Low dose good control	Low dose poor control	High dose good control	High dose poor control		Low dose good control	Low dose poor control	High dose good control	High dose poor control
Metrics inserted to demonstrate covariate model fit									
Hydrocortisone equivalent (mg/m2/day)	8	8	18	18		8	8	18	18
Fludrocortisone dose (micrograms/day)	100	100	100	100		100	100	100	100
17-Hydroxyprogesterone (nmol/l)	10	38	10	38		10	38	10	38
Model derived estimates: (indicates statistically significant difference between those on low dose with good control versus those with high dose and poor control)*									
Height at age 8 (95% CI)	130.4 (127.7 to 133.0)	130.5 (127.9 to 133.1)	133.2 (131.0 to 135.4)	133.4 (130.9 to 135.5)		129.4 (127.1 to 131.5)	129.5 (127.0 to 131.5)	129.9 (127.3 to 133.5)	130.0 (127.0 to 133.6)
Height SDS at age 8 (95% CI)	0.6 (0.1 to 1.0)	0.6 (0.1 to 1.0)	1.1 (0.7 to 1.4)	1.1 (0.7 to 1.5)		0.5 (0.1 to 0.9)	0.5 (0.1 to 0.9)	0.6 (0.1 to 1.2)	0.6 (0.1 to 1.2)
Weight at age 8 (95% CI)	31.7 (28.5 to 34.2)	31.8 (28.4 to 34.5)	34.9 (32.1 to 38.2)	35.0 (31.9 to 38.2)		31.1 (28.8 to 33.3)	31.2 (28.7 to 33.5)	34.2 (28.9 to 39.4)	34.2 (29.0 to 39.4)
Weight SDS at age 8 (95% CI)	1.4 (0.8 to 1.9)	1.5 (0.8 to 2.0)	2.0 (1.5 to 2.5)	2.1 (1.5 to 2.6)		1.3 (0.8 to 1.6)	1.3 (0.8 to 1.7)	1.8 (0.9 to 2.5)	1.8 (0.9 to 2.5)
BMI at age 8 (95% CI)	18.6 (17.2 to 19.7)	18.6 (17.1 to 19.7)	19.6 (18.3 to 21.2)	19.7 (18.2 to 21.3)		18.6 (17.3 to 19.8)	18.6 (17.3 to 19.7)	20.2 (17.4 to 22.7)	20.2 (17.7 to 22.7)
BMI SDS at age 8 (95% CI)	1.6 (0.9 to 2.0)	1.6 (0.8 to 2.0)	2.0 (1.4 to 2.6)	2.0 (1.4 to 2.6)		1.3 (0.8 to 1.8)	1.4 (0.8 to 1.7)	1.9 (0.8 to 2.6)	1.9 (1.0 to 2.6)
Height at age 12 (95% CI)	151.8 (147.7 to 155.3)	151.7 (148.2 to 154.8)	155.8 (152.6 to 158.9)	155.6 (152.8 to 158.7)		148.3 (146.3 to 150.4)	148.9 (146.8 to 151.1)	149.5 (146.9 to 152.8)	150.1 (147.6 to 153.0)
Height SDS at age 12 (95% CI)	0.4 (-0.2 to 0.9)	0.4 (-0.1 to 0.8)	1.0 (0.5 to 1.4)	0.9 (0.5 to 1.4)		-0.4 (-0.7 to -0.1)	-0.3 (-0.7 to -0.0)	-0.3 (-0.6 to 0.2)	-0.2 (-0.5 to 0.3)
Weight at age 12 (95% CI)	47.5 (40.8 to 53.8)	47.9 (41.0 to 53.8)	49.8 (46.4 to 53.1)	50.2 (46.7 to 53.6)		43.4 (40.3 to 47.1)	44.4 (40.9 to 48.8)	47.3 (43.2 to 52.0)	48.4 (44.8 to 52.6)
BMI at age 12 (95% CI)	20.6 (18.3 to 22.7)	20.8 (18.4 to 22.8)	20.6 (19.4 to 21.7)	20.7 (19.7 to 21.8)		19.8 (18.5 to 21.2)	20.1 (18.6 to 21.7)	21.2 (19.6 to 22.7)	21.5 (20.2 to 22.8)
BMI SDS at age 12 (95% CI)	1.2 (0.4 to 1.8)	1.3 (0.4 to 1.8)	1.2 (0.8 to 1.5)	1.3 (0.9 to 1.6)		0.7 (0.2 to 1.1)	0.8 (0.2 to 1.3)	1.1 (0.6 to 1.5)	1.2 (0.8 to 1.5)
Age at adiposity rebound (95% CI)	3.7 (2.5 to 5.1)	3.6 (2.6 to 5.0)	2.7 (0.7 to 4.7)	2.8 (0.4 to 4.5)		4.4 (2.9 to 7.4)	4.2 (2.8 to 6.5)	2.6 (1.9 to 4.4)	2.7 (2.0 to 4.1)
BMI at adiposity rebound (95% CI)	16.6 (15.3 to 17.6)	16.3 (15.0 to 17.4)	16.9 (15.6 to 18.2)	16.6 (15.2 to 18.0)		17.49 (15.8 to 18.8)	17.3 (15.5 to 18.6)	16.6 (14.6 to 19.3)	16.4 (14.3 to 19.1)
BMI z-score at adiposity rebound (95% CI)	0.9 (-0.3 to 1.6)	0.7 (-0.4 to 1.4)	0.9 (-0.6 to 1.9)	0.7 (-1.0 to 1.8)		1.4 (0.3 to 2.1)	1.3 (0.1 to 1.9)	0.8 (-0.7 to 2.5)	0.7 (-1.0 to 2.3)

Figure S6.07 – Visualisations of covariate adjusted models estimated without any interpolated or carried data



Prefix figure key: A = Male Height; B = Female Height; C = Male weight; D = Female weight. E = Male BMI; F = Female BMI.

Suffix figure key: Figures ending '1' show trajectories of growth for modelled values contained within table SEXTRA## as a sensitivity analysis to assess estimates in data without any interpolated or carried forward points..

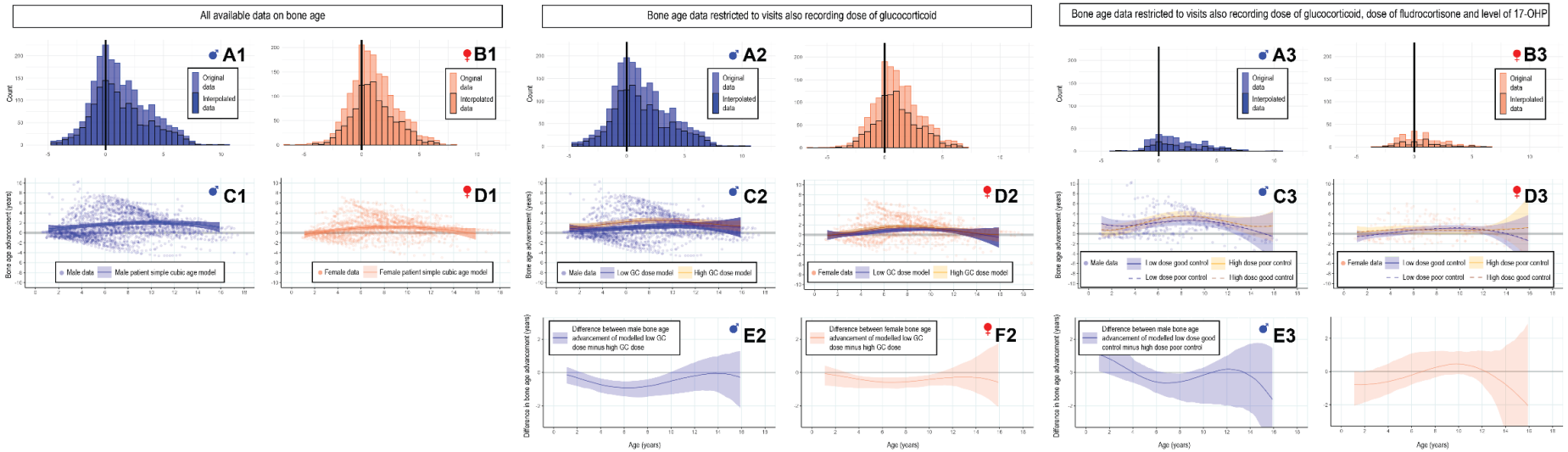
4401 Table S6.10 – Bone age advancement modeling estimates

Dependent variable: Bone age advancement	Male model estimate (95% CI) No covariates	Female model estimate (95% CI) No covariates	Male model estimate (95% CI) Covaried GC dose only	Female model estimate (95% CI) Covaried GC dose only	Male model estimate (95% CI) All covariates	Female model estimate (95% CI) All covariates
Number of patients	204	240	1.51 (-0.26 to 3.27)	-0.19 (-2.07 to 1.69)	3.02 (-0.73 to 6.77)	-0.29 (-4.35 to 3.78)
Number of visits	2116	1687	-0.64 (-1.35 to 0.05)	-0.15 (-0.92 to 0.62)	-0.46 (-2.08 to 1.15)	-0.38 (-2.70 to 1.94)
Intercept	0.64 (-0.68 to 1.82)	-0.84 (-1.98 to 0.23)	0.097 (0.001 to 0.193)	0.057 (-0.040 to 0.155)	0.09 (-0.16 to 0.34)	0.11 (-0.25 to 0.48)
Age	0.15 (-0.32 to 0.68)	0.46 (-0.03 to 0.96)	-0.004 (-0.008 to 0.000)	-0.003 (-0.007 to 0.001)	-0.006 (-0.018 to 0.007)	-0.006 (-0.023 to 0.011)
Age^2	0.01 (-0.05 to 0.07)	-0.03 (-0.08 to 0.04)	-0.040 (-0.211 to 0.131)	-0.026 (-0.246 to 0.193)	-0.477 (-0.875 to -0.080)	0.256 (-0.149 to 0.662)
Age^3	-0.001 (-0.003 to 0.001)	0.000 (-0.002 to 0.002)	0.067 (0.009 to 0.124)	0.040 (-0.028 to 0.109)	0.191 (0.055 to 0.326)	-0.057 (-0.223 to 0.109)
Hydrocortisone equivalent	-	-	-0.008 (-0.015 to -0.002)	-0.006 (-0.012 to 0.001)	-0.021 (-0.036 to -0.006)	0.003 (-0.019 to 0.024)
Hydrocortisone equivalent : Age	-	-	0.000 (0.000 to 0.001)	0.000 (0.000 to 0.000)	0.001 (0.000 to 0.001)	0.000 (-0.001 to 0.001)
Hydrocortisone equivalent : Age^2	-	-	-	-	0.007 (-0.013 to 0.028)	-0.010 (-0.034 to 0.014)
Hydrocortisone equivalent : Age^3	-	-	-	-	-0.004 (-0.013 to 0.006)	0.005 (-0.008 to 0.018)
Fludrocortisone	-	-	-	-	0.000 (-0.001 to 0.002)	0.000 (-0.002 to 0.001)
Fludrocortisone : Age	-	-	-	-	0.000 (0.000 to 0.000)	0.000 (0.000 to 0.000)
Fludrocortisone : Age^2	-	-	-	-	0.083 (-0.296 to 0.461)	-0.026 (-0.507 to 0.455)
Fludrocortisone : Age^3	-	-	-	-	-0.097 (-0.305 to 0.111)	-0.028 (-0.286 to 0.231)
ln 17OHP	-	-	-	-	0.016 (-0.018 to 0.049)	0.008 (-0.033 to 0.048)
ln 17OHP : Age	-	-	-	-	-0.001 (-0.002 to 0.001)	0.000 (-0.002 to 0.001)
ln 17OHP : Age^2	-	-	0.06 (0.03 to 0.11)	0.07 (0.03 to 0.12)	0.09 (0.04 to 0.26)	0.08 (0.04 to 0.16)

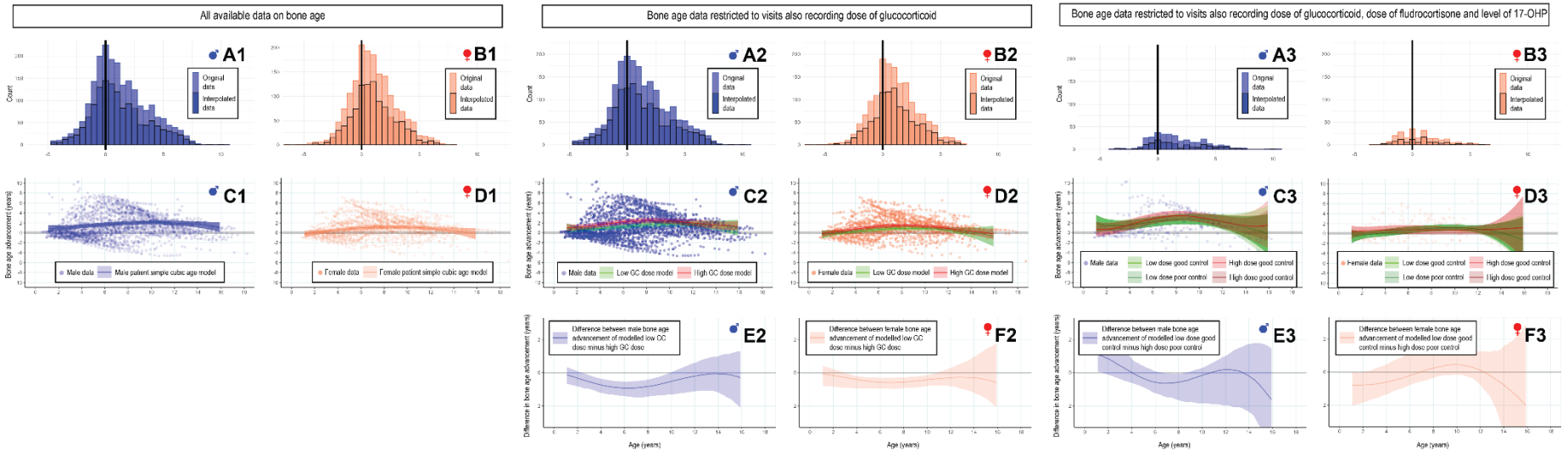
Dependent variable: Bone age advancement	Male model estimate (95% CI) No covariates	Female model estimate (95% CI) No covariates	Male model estimate (95% CI) Covaried GC dose only	Female model estimate (95% CI) Covaried GC dose only	Male model estimate (95% CI) All covariates	Female model estimate (95% CI) All covariates
ln 17OHP : Age^3	-	-	0.84 (0.77 to 0.89)	0.84 (0.80 to 0.88)	0.92 (0.82 to 0.96)	0.92 (0.85 to 0.96)
Marginal R^2	0.02 (0.01 to 0.06)	0.03 (0.01 to 0.07)	1.41	0.55	2.26	0.73
Conditional R^2	0.81 (0.76 to 0.86)	0.85 (0.80 to 0.88)	0.06	0.07	0.05	0.06
SD of random intercept for centre	1.03	0.53	1.51 (-0.26 to 3.27)	-0.19 (-2.07 to 1.69)	3.02 (-0.73 to 6.77)	-0.29 (-4.35 to 3.78)
SD of random intercept for id	0.06	0.07	-0.64 (-1.35 to 0.05)	-0.15 (-0.92 to 0.62)	-0.46 (-2.08 to 1.15)	-0.38 (-2.70 to 1.94)

4402 Cubic model estimates for models predicting bone age advancement with varying sets of covariates. 17OHP=17-Hydroxyprogesterone (nmol/l). Hydrocortisone equivalent (mg), fludrocortisone (mg)

4403 Figure S6.08 – Bone age advancement covariate modelling sensitivity analysis



4404



4405

Bone age advancement modelling with varying levels of available covariates.
Prefix letter: A=Male histogram of bone age data available, B=Female histogram of bone age data available, C=Male bone age advancement model, D=Female bone age advancement model, E=male difference between clinically relevant example patient values dependent upon covariates of relevant model, F=female difference between clinically relevant example patient values dependent upon covariates of relevant model
Suffix number: 1=No covariates in models, maximum bone age data employed; 2=Dose of hydrocortisone incorporated as a covariate with each age term in model; 3=Dose of hydrocortisone, dose of fludrocortisone, and the natural log of the level of 17-Hydroxyprogesterone incorporated as covariates with each age term in model.
This figure shows the significant number of visits that are lost when attempting to model the bone age advancement with all three covariates (figures with suffix 3). Figures with suffix 2 are the more parsimonious model only covarying with dose of glucocorticoid that are presented in the main manuscript.
17OHP=17-HydroxyProgesterone. GC=Glucocorticoid (hydrocortisone equivalent); Model coefficients contained within table S8

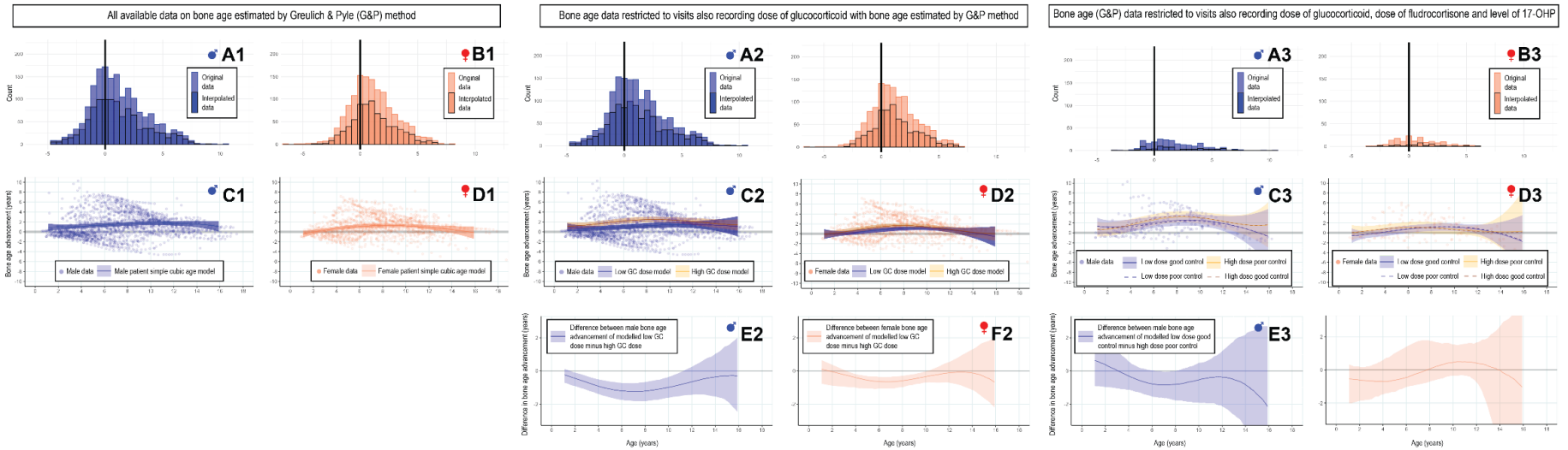
Table S6.11 – Bone age advancement modeling estimates restricted only to Greulich & Pyle estimates

Dependent variable: Bone age advancement	Male model estimate (95% CI) No covariates	Female model estimate (95% CI) No covariates	Male model estimate (95% CI) Covaried GC dose only	Female model estimate (95% CI) Covaried GC dose only	Male model estimate (95% CI) All covariates	Female model estimate (95% CI) All covariates
Number of patients	181	216	168	200	59	62
Number of visits	1726	1356	1623	1246	239	161
Intercept	0.90 (-0.20 to 1.99)	-1.08 (-2.16 to -0.00)	1.33 (-0.16 to 2.83)	0.22 (-1.94 to 2.37)	1.31 (-3.66 to 6.29)	-0.07 (-5.35 to 5.21)
Age	0.00 (-0.47 to 0.48)	0.59 (0.10 to 1.09)	-0.64 (-1.38 to 0.09)	-0.25 (-1.14 to 0.65)	-0.19 (-2.41 to 2.02)	-0.51 (-3.44 to 2.43)
Age^2	0.02 (-0.04 to 0.08)	-0.04 (-0.10 to 0.02)	0.09 (-0.02 to 0.20)	0.07 (-0.05 to 0.18)	0.09 (-0.24 to 0.41)	0.13 (-0.31 to 0.58)
Age^3	-0.001 (-0.004 to 0.001)	0.001 (-0.002 to 0.003)	-0.003 (-0.008 to 0.002)	-0.004 (-0.008 to 0.001)	-0.006 (-0.021 to 0.009)	-0.008 (-0.028 to 0.013)
Hydrocortisone equivalent	-	-	0.006 (-0.146 to 0.157)	-0.082 (-0.340 to 0.176)	-0.289 (-0.815 to 0.237)	0.164 (-0.314 to 0.643)
Hydrocortisone equivalent : Age	-	-	0.054 (-0.004 to 0.112)	0.061 (-0.022 to 0.144)	0.140 (-0.031 to 0.311)	-0.010 (-0.217 to 0.197)
Hydrocortisone equivalent : Age^2	-	-	-0.007 (-0.014 to 0.000)	-0.008 (-0.017 to 0.001)	-0.017 (-0.035 to 0.002)	-0.003 (-0.030 to 0.024)
Hydrocortisone equivalent : Age^3	-	-	0.000 (0.000 to 0.000)	0.000 (0.000 to 0.001)	0.001 (0.000 to 0.001)	0.000 (-0.001 to 0.001)
Fludrocortisone	-	-	-	-	0.009 (-0.018 to 0.035)	-0.009 (-0.038 to 0.020)
Fludrocortisone : Age	-	-	-	-	-0.003 (-0.017 to 0.010)	0.005 (-0.011 to 0.020)
Fludrocortisone : Age^2	-	-	-	-	0.000 (-0.002 to 0.002)	0.000 (0.000 to 0.000)
Fludrocortisone : Age^3	-	-	-	-	0.000 (0.000 to 0.000)	0.000 (-0.003 to 0.002)
ln 17OHP	-	-	-	-	0.19 (-0.30 to 0.68)	-0.05 (-0.75 to 0.65)
ln 17OHP : Age	-	-	-	-	-0.16 (-0.43 to 0.10)	0.01 (-0.35 to 0.37)
ln 17OHP : Age^2	-	-	-	-	0.021 (-0.020 to 0.062)	-0.001 (-0.055 to 0.054)
ln 17OHP : Age^3	-	-	-	-	-0.001 (-0.003 to 0.001)	0.000 (-0.002 to 0.002)
Marginal R^2	0.02 (0.01 to 0.05)	0.04 (0.02 to 0.08)	0.06 (0.03 to 0.11)	0.07 (0.03 to 0.13)	0.09 (0.04 to 0.33)	0.12 (0.05 to 0.27)
Conditional R^2	0.81 (0.75 to 0.87)	0.84 (0.80 to 0.89)	0.83 (0.75 to 0.88)	0.85 (0.80 to 0.89)	0.93 (0.80 to 0.97)	0.89 (0.79 to 0.96)
SD of random intercept for centre	1.05	0.46	1.31	0.53	2.51	0.70
SD of random intercept for id	0.06	0.08	0.07	0.07	0.05	0.06

Cubic model estimates for models predicting bone age advancement with varying sets of covariates. 17OHP=17-Hydroxyprogesterone (nmol/l). Hydrocortisone equivalent (mg), fludrocortisone (mg)

4418

Figure S6.09 – Bone age advancement covariate modelling restricted to only Greulich & Pyle



Bone age advancement modelling with varying levels of available covariates, all data in these models restricted to that measured by Greulich & Pyle bone age dating methodology.
Prefix letter: A=Male histogram of bone age data available, B=Female histogram of bone age data available, C=Male bone age advancement model, D=Female bone age advancement model, E=male difference between clinically relevant example patient values dependent upon covariates of relevant model, F=female difference between clinically relevant example patient values dependent upon covariates of relevant model
Suffix number: 1=No covariates in models, maximum bone age data employed; 2=Dose of hydrocortisone incorporated as a covariate with each age term in model; 3=Dose of hydrocortisone, dose of fludrocortisone, and the natural log of the level of 17-OH Progesterone incorporated as covariates with each age term in model.
This figure shows the significant number of visits that are lost when attempting to model the bone age advancement with all three covariates (figures with suffix 3). Figures with suffix 2 are the more parsimonious model only covarying with dose of glucocorticoid that are presented in the main manuscript.
17OHP=17-Hydroxyprogesterone. GC=Glucocrticoid (hydrocortisone equivalent); Model coefficients contained within table S8

End